UNDERSTANDING THE FUNCTION OF CTCF RECRUITMENT TO ONCOGENIC HUMAN PAPILLOMAVIRUS GENOMES DURING THE VIRAL LIFE CYCLE

by

IEISHA PENTLAND

A thesis submitted to the University of Birmingham for the degree of DOCTOR OF PHILOSOPHY

UNIVERSITY OF BIRMINGHAM

Institute of Cancer and Genomic Sciences
College of Medical and Dental Sciences
University of Birmingham
September 2017

Abstract

The CCCTC-binding factor (CTCF) is a DNA binding protein essential for genome-wide organization of chromatin. A conserved CTCF binding site was identified in the E2 open reading frame of high-risk HPV types, but was absent in low-risk HPV types. Abrogation of CTCF binding at the E2 site in HPV18 genome containing primary human foreskin keratinocytes causes significant upregulation of transcripts encoding the early viral oncoproteins E6 and E7, resulting in epithelial hyperproliferation. Notably, abrogation of CTCF binding results in a more open conformation within the viral long control region (LCR), which is positioned 3 kilobases upstream of the E2-CTCF binding site. In addition, there is a loss of recruitment of the transcriptional repressor protein YY1 and of the polycomb repressive complexes. Chromatin conformation capture was used to demonstrate DNA looping between the E2-CTCF binding site and the YY1-bound viral LCR. The formation of this chromatin loop is dynamic and reduced upon epithelial differentiation. Together, these data show that high-risk HPV genomes recruit CTCF to the E2 ORF to form a repressive chromatin loop with the YY1-bound LCR to control viral oncoprotein expression. Ultimately this strategy will allow the virus to coordinate life cycle events to maintain a persistent infection.

Acknowledgements

I would firstly like to thank my supervisor Dr Jo Parish for her incredible support, encouragement and guidance over the past 4 years. Jo has always provided insight and knowledge during experimental discussions and has enabled me to fulfil my full potential. I would like to thank my second supervisor Dr Sally Roberts for her guidance and support. Her knowledge and input has been invaluable for my PhD project. Thank you for providing me with endless supplies of Waitrose cake. I would like to acknowledge Karen Campos Leon for being the best Post-doc anyone could have by your side. From the moment I started Karen has been with me on my incredible journey, through both the good times and the can't cope times. She has always provided me with a plentiful supply of hot chocolate and treats, and her lovely and kind persona is always a boost on a difficult day. I would like to acknowledge my partner in crime Paul McCormack for making my time in Birmingham full of fun and laughter. I will miss walking in the office amongst a mist of Paco Rabanne and being greeted by "Alright Princess". Thank you to the other great students and staff that have made my time in Birmingham a great place to live and work. Sarah Leonard would also put a smile on my face with her cheery ways and Tom Clarke would also be great fun and a good friend. I will always remember our great nights in the Village. Thank you to Cancer Research UK for funding my PhD and providing the resources to enable us to carry out such a fantastic and forward thinking

project. I wouldn't have been able to cope without the endless love and support from Matthew. He has encouraged and supported me throughout my PhD and I can't thank him enough. Thank you to Jane for being my Birmingham mum and providing me with lovely food and cups of tea. Lastly, I would like to thank my family for their endless support. Rochelle and Elliott have been the most supportive siblings. My mum has always been my biggest inspiration and would like to thank her for guiding me to where I am today.

Table of contents

List of tables

List of figures

Abbreviations

°C- Degrees Celsius

µCi- Microcurie

µg- Microgram

µL- Microlitre

µm- Micromolar

x g- Times gravity

3C- Chromosome conformation capture

3D- Three-dimensional

ΔCTCF- HPV18 genome with a 3bp mutation in the CTCF binding in the E2 ORF

J2-3T3- Murine fibroblasts

Ab- Antibody

AC- Adenocarcinoma

AP-1- Activator protein 1

ALTER- Agitated low temperature antigen retrieval

APC- Annual percentage change

APS- Ammoniumpersulfate

ASR- Age standardised incident rate

ATAQ-Seq- Assay for transposase accessible chromatin with high-throughput sequencing

ATM- Ataxia-telangiectasia mutated

ATP- Adenosine triphosphate

ATR- Ataxia telangiectasia and Rad3 related

AUG- Start codon

***Bgl*II**- Restriction endonuclease (*Bacillus amyloli*)

BET- Bromodomain and extra terminal domain

BME- β-Mercaptoethanol

bp- Base pair

BPV- Bovine papillomavirus

BrdU- Bromodeoxyuridine

BRCA1- Breast cancer type 1 susceptibility protein

Ab- Antibody

BRD2- Bromodomain-containing protein 2

BRD4- Bromodomain-containing protein 4

BrdU- Bromodeoxyuridine

BSA- Bovine serum albumin

CDK- Cyclin dependent kinase

CEM- Complete E media

cDNA- Complementary DNA

ChIA-PET- Chromatin interaction analysis by paired-end tag sequencing

ChIP- Chromatin immunoprecipitation

ChIP-Seq- Chromatin immunoprecipitation sequencing

CHK2- Checkpoint kinase 2

ChlR1- ATP-dependent DNA helicase DDX11

CIN612- HPV type 31b genome containing cell line derived from a low-grade cervical lesion

CIN- Cervical intra-epithelial neoplasia

cm- Centimetre

CMV- Cytomegalovirus

CO_2- Carbon dioxide

Cp- EBV C promoter

CPS- Counts per second

Ct- Cycle threshold

CTCF- CCCTC binding factor

CTCFBSDS- CTCF binding site database

CTD- C-terminal repeat domain

ddH$_2$0- Double distilled water

DMEM- Dulbeccos modified eagles medium

DMSO- Dimethyl sulfoxide

DNA- Deoxyribonucleic acid

DEPC- Diethyl pyrocarbonate

dNTPS- Deoxynucleotide Triphosphates

D*pn*1- Methylation specific nuclease

DTT- 1,4-dithiothreitol

E1- Papillomavirus E1 protein

E2- Papillomavirus E2 protein

E2F- The E2 factor

E4- Papillomavirus E4 protein

E5- Papillomavirus E5 protein

E6- Papillomavirus E6 protein

E6AP- Ubiquitin-protein ligase E3A

E7- Papillomavirus E7 protein

EBNA1- Epstein-Barr nuclear antigen 1

EBNA2- Epstein-Barr nuclear antigen 2

EBV- Epstein Barr virus

***Eco*RI**- Restriction endonuclease (*Escherichia coli*)

EDTA- Ethylenediaminetetraacetic acid

EED- Embryonic Ectoderm Development

EGF- Epidermal growth factor

EGFR- Epidermal growth factor receptor

EMSA- Electrophoretic mobility shift assay

Enh- Enhancer

Ep- HPV early promoter

ESE- Exonic splicing enhancer

ESS- Exonic splicing silencer

EZH2- Enhancer of zeste homolog 2

FAIRE- Formaldehyde assisted isolation of regulatory elements

FBS- Fetal bovine serum

FCS- Fetal calf serum

FLAG- Polypeptide protein tag

FOX1A- Forkhead box protein A1

Fw- Forward

g- Gram

G- Gauge

G418- Geneticin

GAPDH- Glyceraldehyde 3-phosphate dehydrogenase

GMEM- Glasgow minimum essential medium

GTFs- General transcription factors

Gy- Gray

H&E- Haematoxylin and eosin

H_20- Water

HAT- Histone acetyltransferase

HBV- Hepatitis B

HCV- Hepatitis C

HCl- Hydrochloric acid

HDAC- Histone deacetylase

HEPES- 4-(2-Hydroxyethyl)piperazine-1-ethanesulfonic acid

HINGS- Heat inactivated goat serum

HIF-1- Hypoxia-inducible factor 1

HFKs- Human foreskin keratinocytes

hnRNP- Heterogeneous nuclear ribonucleoproteins

HNSCC- Head and neck squamous cell carcinoma

HPV- Human papillomavirus

HSV-1- Herpes simplex virus

HVS- Herpesvirus saimiri

ICR- Imprinting control region

IF- Immunofluorescence

IFN- Interferon

Igf2/H19- Imprinting gene locus

IgG- Immunoglobulin G

IMS- Industrial methylated spirits

ISE- Intronic splicing enhancer

ISS- Intronic splicing enhancer

JAK3- Tyrosine-protein kinase JAK3

Kb- Kilobases

kDa- Kilo Dalton

KDM6A- Lysine-specific demethylase 6A

KDM6B- Lysine-specific demethylase 6B

KSHV- Kaposi's sarcoma associated virus

L1- Papillomavirus major capsid protein L1

L2- Papillomavirus minor capsid protein L2

L- Linearised

LANA- latency-associated nuclear antigen

LAT- Latency- associated transcript

LCR- Long control region

LMP1- Latent membrane protein 1

LMP2A- Latent membrane protein 2

M- Molar

MIE- Major immediate early

mES- Murine embryonic cells

mg- Milligram

ml- Millilitre

mM- Millimolar

mRNA- Messenger ribonucleic acid

MYC- Proto-oncogene c-Myc

NaOH- Sodium hydroxide

$NaHCO_3$- Sodium bicarbonate

NCBI- National Center for Biotechnology Information

NES- Nuclear export signal

NF-1- Nuclear factor 1

ng- Nanogram

***Nla*III-** Restriction endonuclease (Ne*isseria lactamica*)

NLS- Nuclear localisation signal

Nm- Nanometre

NTC- No template control

OriP- Origin of replication

ORF- Open reading frame

PBS- Phosphate buffered saline

pGEM2- Vector containing HPV18 genome

OC- Open circle

Oct-1- Octamer transcription factor 1

P^{32}- Phosphorus-32

p53- Tumour Protein P53

P_{97}- HPV16 early promoter

P_{105}- HPV18 early promoter

p107- Retinoblastoma-like protein 1

p130- Retinoblastoma-like protein 2

P_{670}- HPV16 late promoter

P_{811}- HPV18 late promoter

PARP- Poly(ADP-ribose) polymerase

PARPs- Poly(ADP-ribose) chains

PBS- Phosphate buffered saline

PcG- Polycomb-group proteins

PCR- Polymerase chain reaction

PDGFRA- Platelet-derived growth factor receptor A

PenStrep- Penicillin-Streptomycin

pH- potential of hydrogen

PHD- Plant homeodomain

PHO- Polycomb protein

PIC- Protease inhibitor cocktail

pmol- Picomole

PMSF- Phenylmethanesulfonylfluoride

pRb- Retinoblastoma-associated protein

PRC1- Polycomb Repressive Complex 1

PRC2- Polycomb Repressive Complex 2

PVs- Papillomaviruses

PVDF- Polyvinylidene difluoride

PWM- Position weight matrix

O/N- Overnight

Qp- EBV Q promoter

qPCR- Quantitative polymerase chain reaction

RACE- Rapid amplification of cDNA ends

RAD21- Double-strand-break repair protein rad21 homolog

RCB- Reconstitution buffer

RING1B- E3 ubiquitin-protein ligase

RNA- Ribonucleic acid

RNA-seq- RNA sequencing

RPLPO- Large ribosomal protein

Rpm- Revolutions per minute

RRP- Respiratory papillomatosis

RT- Room temperature

RT-PCR- Reverse transcriptase polymerase chain reaction

Rv- Reverse

s- second

SA- Splice acceptor

SC- Supercoiled

SCC- Squamous cell carcinoma

Sd- Standard deviation

SD- Splice donor

SDS- Sodium dodecyl sulphate

SEM- Standard error of the mean

SET- Histone-lysine N-methyltransferase

SFM- Serum free media

shRNA- Short hairpin RNA

SiRNA- Small interfering RNA

SMC1- Structural maintenance of chromosomes protein 1

SMC3- Structural maintenance of chromosomes protein 3

SP1- Specificity Protein 1

SRSF- Serine/arginine-rich splicing factor

SSC- Saline sodium citrate

SUZ12- Suppressor of zeste 12 protein homolog

TAD- Topologically associating domain

TBE- Tris-borate-EDTA buffer

TBST- Tris-buffered saline with Tween 20

TE- Tris-EDTA buffer solution

TEF-1- Transcriptional enhancer factor

TEMED- N,N,N,N-Tetramethylethylenediamine

TFIID- Transcription factor II D

Tn5- Transposase

TSS- Transcriptional start site

U- Units

ULB- Urea lysis buffer

UTR- Untranslated region

UV- Ultraviolet

V- Volt

v/v- %Volume per volume

VLP- Virus like particle

W12- HPV-16 genome containing cell line derived from a low-grade cervical lesion

WHO- World Health Organisation

Wp- EBV W promoter

WT- Wild type

YY1- Ying-yang

Chapter 1: Introduction

1 Chapter 1: Introduction

1.1 Human Papillomavirus

1.1.1 Introduction

The human papillomavirus (HPV) belongs to the *Papillomaviradae* family of small epitheliotropic DNA viruses, which are known to be present in most mammals as well as other diverse hosts including avian and reptile species. HPV-related diseases within the host can range from benign warts and papillomas to invasive cancer, most commonly cervical cancer.

Historically, it was first postulated that the development of cervical cancer was associated with the sexual transmission of herpes simplex-2 virus (HSV-2) (zur Hausen, 2009). However earlier research in domestic and cotton-tail rabbits had already demonstrated the carcinogenic potential of the rabbit papillomaviruses to induce papillomas (Rous and Beard, 1935), (Syverton, 1952). The same researchers also postulated that persistence of a papilloma correlated with the risk of malignant progression, as sites with lesion regression were no longer at risk of carcinoma development. However, it wasn't until 1975 when zur Hausen first hypothesized that HPV infection and persistence may be the causative agent of cervical cancer in humans (zur Hausen et al., 1975). Since then, the HPV life cycle and the mechanisms this virus employs to induce malignant transformation have been further

investigated, leading to a greater understanding of HPV-related cancer progression and also the development of the HPV prophylactic vaccines.

1.1.2 Types and classification

The papillomavirus family is highly diverse and there are currently more than 200 identified papillomaviruses and over 150 sequenced HPV types (Doorbar et al., 2012, Doorbar et al., 2015). Historically the papillomavirus family was combined with the polyomaviruses to form the Papo-vaviridae family, based on their double-strand DNA genomes and non-enveloped capsid structures. They were later divided into separate virus groups because they have dissimilar genome organization and a lack of sequence homology (de Villiers et al., 2004). Subtypes exist within the papillomavirus family, which are categorized based on their L1 gene sequence being at least 10% dissimilar to other PV types (Bernard, 2005). Overall, the L1 gene forms the basis of PV classification and construction of the phylogenetic tree.

The human papillomaviruses are clustered among five genera; alpha, beta, gamma, nu, and mu (Fig. 1). HPV types across the five genera can differ in their tissue tropism and disease associations (reviewed by (Doorbar et al., 2012). The alpha genus, which is comprised of over 60 types, includes both high- and low-risk mucosal types associated with benign and malignant lesions, as well as low-risk cutaneous types associated with benign lesions

and skin warts (Bernard, 2005). The alpha genus is the most extensively studied due to the presence of the most pathogenic high-risk types that reside in this group. The alpha-7 species contains the genotypes HPV18, 39, 45, 59, 68, 70, 85 and 97 and the alpha-9 species contains HPV16, 31, 33, 35, 52, 58 and 67 (Chen et al., 2013b). The beta species are known to infect cutaneous epithelium. The remaining Gamma, Nu, and Mu papillomaviruses are all associated with benign cutaneous lesions (Bernard, 2005). It is apparent that these PVs can complete their viral life cycle and either cause asymptomatic infections in individuals, so that no apparent disease is detected, or infection can cause visible papilloma formation.

HPV is transmitted via direct skin-to-skin contact but is more often spread via sexual intercourse or sexual contact; over time 40 HPV types have been identified that can infect the genital tract (de Villiers et al., 2004). Initial HPV infections are acquired during the first years of sexual activity, with increased risk associated with the number of sexual partners (Winer et al., 2003). HPV infection is often transient and the infection is cleared by the immune system within 1-2 years of initial infection (Woodman et al., 2007). However, the virus can persist in a subset of individuals who fail to clear the infection, which increases the probability of HPV-mediated malignant progression (Woodman et al., 2007).

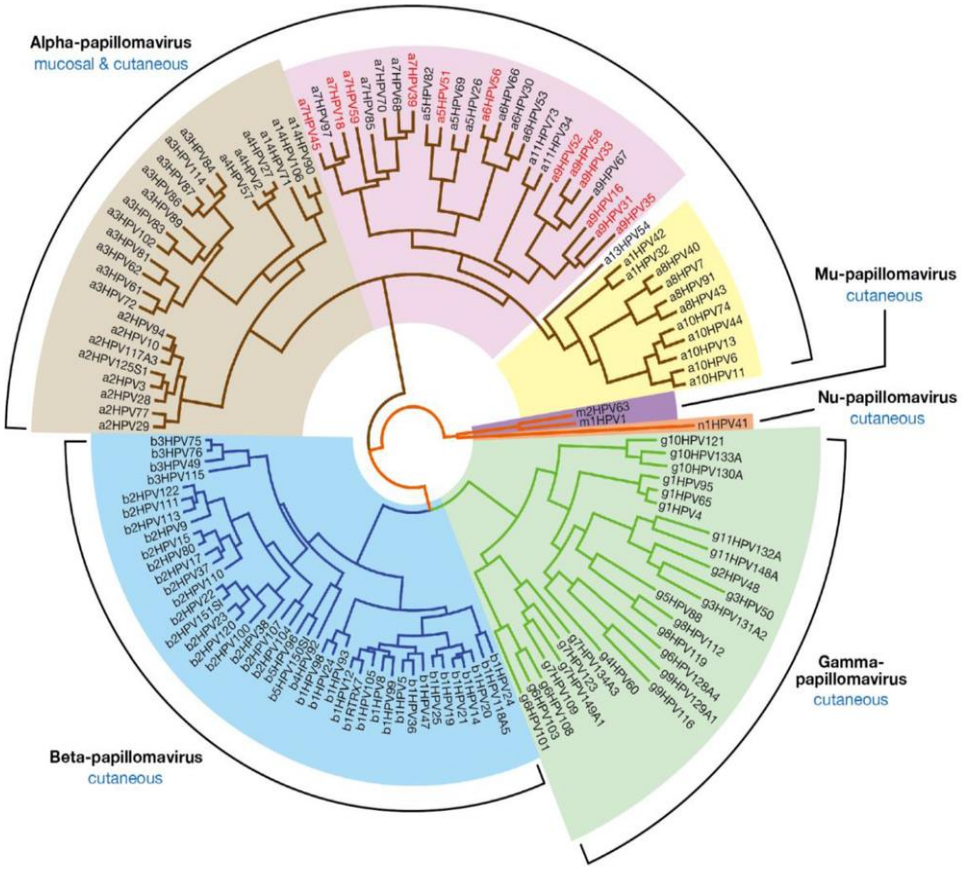


Figure 1. Evolutionary map of the *Papillomaviradae* family. The *Papillomaviradae* family consists of five genera: Alpha, Beta, Gamma, Nu and Mu. Tissue tropisms are indicated below each genera. Within each genus there are several different species, and within each species multiple genotypes exist. The known high-risk HPV types are highlighted in red. Image taken from (Doorbar et al., 2012).

1.1.3 High and low-risk types

The virus is typically sub-categorized into high-, intermediate-, or low-risk types. HPV types found in cervical, head and neck, or anogenital cancers are categorized as high-risk, whereas those identified in genital warts and non-malignant lesions are normally low-risk types. According to the World Health

Organization (WHO) there have been 13 identified HPV types classified as high-risk cancer causing types: -16, -18, -31, -33, -35, -39, -45, -51, -52, -56, -58, -59 and 68. HPV type 73 has been detected in invasive cervical cancer and has been classified as possibly cancer-causing (Cogliano et al., 2005, Humans, 2007). The low-risk HPV types include HPV6, -11, -40, -42, -43, -44, -53, -54, -61, -72, and -81 (Steben and Duarte-Franco, 2007). It is estimated that 99.7% of all cervical squamous cell cancers worldwide contain HPV, which is the main attributable risk factor (Walboomers et al., 1999). The high-risk alpha types are strongly associated with the development of squamous cell carcinoma (SCC) and adenocarcinoma (AC). Overall, the HPV 16 and 18 high-risk types are the most prevalent in the human population, as they are attributable to 70% of reported cervical cancer cases (Clifford et al., 2003). HPV18 is predominantly linked to AC of the cervix, accounting for 32% of AC cases, whereas HPV 16 is causative of 62% of all SCCs (Bulk et al., 2006). These high-risk types have the ability to persist in the host for long periods of time and possess the potential for inducing tumour progression (Doorbar et al., 2012). HPV6 and 11 are predominately the main types that cause condyloma acuminatum (genital warts), accounting for 90% of all known cases (Gissmann et al., 1983). Low-risk Alpha types HPV6 and HPV11 have also been associated with the development of respiratory papillomatosis (RRP), of which ~5% of cases can lead to cancer (Doorbar et al., 2012), (Derkay, 1995).

HPV infection is also associated with the malignant progression at other epithelial sites, including the head and neck region, as well as anogenital sites of the anus, penis, vulva and vagina (Doorbar et al., 2012). It is estimated that HPV is present in 23-35% of head and neck squamous cell carcinomas (HNSCC), but most predominantly in the oropharynx where HPV is detected in 45-90% of cases. Similar to cervical cancer, HPV16 is found in 68-87% of these HPV positive cases (Kreimer et al., 2005).

Furthermore, the beta HPV types have been associated with the development of keratinocyte carcinoma. A wide range of beta HPV types have been identified in SCC of patients with epidermodysplasia verruciformis (EV), especially HPV5 but also HPV8, 14, 17, 20, or 47 (Orth, 1986, Orth, 2005). However the role of beta types in the development of skin cancer is complex and may involve additional factors including immunosuppression and UV exposure (Wieland et al., 2014, Jackson et al., 2000, Underbrink et al., 2008, Viarisio et al., 2011).

1.1.4 Epidemiology

HPV is the leading cause of cervical cancer worldwide, with an estimated 500,000 cases per year and approximately 270,000 deaths amongst women (Parkin et al., 2005). Globally, cervical cancer is the seventh most common cancer but the fourth most common in women, after breast, bowel and lung

cancer (Ferlay et al., 2010, Ferlay et al., 2012). Of the 12.7 million estimated global cancer cases approximately 2 million are attributable to infectious agents, including: Epstein Barr Virus (EBV), Hepatitis B & C (HBV and HCV), *Helicobacter Pylori* and HPV. It is estimated that HPV infection is attributable to around 30% of these two million cancer cases and subsequently 4.8% of the total cancer burden (de Martel et al., 2012). HPV is attributable to 1.6% of all cancer cases in the UK, predominantly cervical cancer but also cancers in the head and neck and anogenital regions (Parkin, 2011).

Cervical cancer is disproportionately higher in developing countries, where an estimated 85% of global cases occur (Fig. 2). It is the most prevalent form of cancer among women in developing countries, accounting for 15% of cancer cases (Ferlay et al., 2010). Conversely, it accounts for only 3.6% of new cancers in the female population of the developed world (Parkin and Bray, 2006). The global age standardized rate (ASR) of cervical cancer is depicted in figure 3; darker regions represent higher rates of cervical cancer and lighter regions depict lower rates.

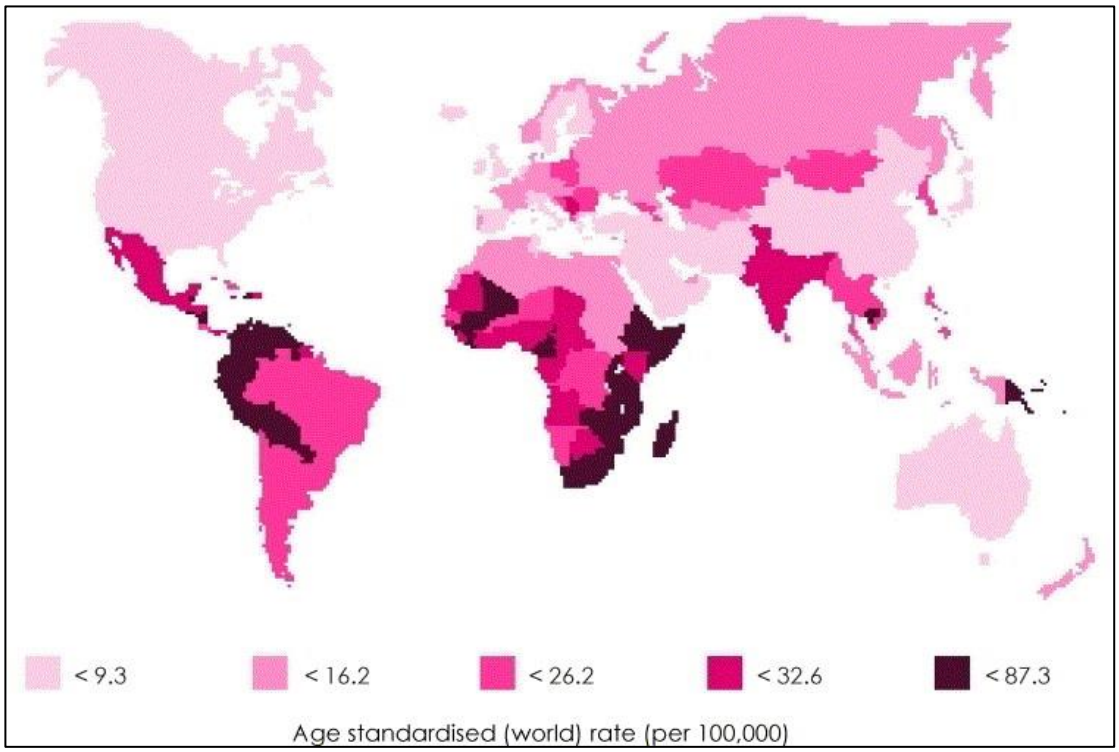


Figure 2. Global estimated incidence risk of cervical cancer. Shown is the age-standardized rate (per 100,00) of cervical cancer worldwide, with the highest rates illustrated in darker regions and the lowest rates in lighter regions. Image taken from (Parkin and Bray, 2006).

The lowest incidence rates can be found in the UK, Western Europe, North America, and Australia with annual rates below 9.3 per 100,000 individuals. Conversely, higher incidence rates are found in the developing countries of Latin American, the Caribbean, sub-Saharan Africa, and South-central Asia, with more than 30 individuals per 100,000 affected each year (Parkin and Bray, 2006). It is estimated that 88% of all cervical cancer deaths occur in the developing world. Of the 1.9% of women who develop cervical cancer, around 1.1% will die of the disease before the age of 75 years (Arbyn et al., 2011). Despite the high incidence in some areas the overall global ratio of

mortality to incidence is 52% (Ferlay et al., 2010). However it is predicted there will be 700,000 cases of cervical cancer in 2020, with potentially 90% of the cases occurring in the developing world (Parkin and Bray, 2006).

HPV infection is strongly associated with HNSCC and other anogenital cancers. The proportion of head and neck cancers caused by HPV vary across regions due to the contribution of tobacco and alcohol risk factors, but it is estimated that 40–80% of oropharyngeal cancers in the US are caused by HPV (Marur et al., 2010). Whilst the number of tobacco associated head and neck cancers has been declining, the prevalence of HPV positive head and neck cancers has been increasing over the past four decades in the USA (Shiboski et al., 2005, Chaturvedi et al., 2008, Frisch et al., 2000, Ryerson et al., 2008). Chaturvedi *et al* concluded that the incidence rates of oropharnygeal squamous cell carcinoma (OSCC) have increased significantly between 1973 and 2004 (annual percentage change [APC]= 0.8; $P<0.001$), and this is occurring in a younger cohort of patients, when compared to non-HPV related OSCC (Chaturvedi et al., 2008). Interestingly, patients with HPV positive HNSCC respond better to chemotherapy and chemoradiaton treatment compared to those with HPV negative HNSCC, and display a better survival outcome (2 year overall survival; 95% [95% CI, 87%-100%] versus 62% [95% CI, 49%-74%]) (Fakhry et al., 2008). Given the current trends, the number of HPV related oropharyngeal cancer cases is expected to surpass the

number of cervical cancers in US by the year 2020 (Zandberg et al., 2013, Chaturvedi et al., 2008).

Cancers of the external genitalia are generally rare, with cancer of the penis attributing to 0.5% of all male cancers. Similarly, vulva cancer has an age standardized incident rate (ASR) of between 0.5 and 1.5 per 100,000. There is a lower rate at 0.3-0.7 ASR for HPV-associated vaginal cancer (Parkin and Bray, 2006). HPV infection is also associated with anal cancer; approximately 85% of anal carcinomas contain HPV, with slightly higher rates in homosexual males (Grulich et al., 2010). Overall, anal cancer accounts for 2 in every 100,000 cancer cases but the incidence has been on the increase over the past 30-50 years (Franceschi and De Vuyst, 2009).

1.1.5 Screening

Since the introduction of cervical screening programs, there has been a significant reduction in the incidence and mortality of women from invasive cervical cancer over the past 40 years, with a decrease in mortality in excess of 50% in some developed nations, such as the USA (Cuzick et al., 2008). However, the same decline in incidence and mortality is not apparent in developing countries, most likely due to the lack of screening and vaccination programs available. Of those countries where screening programs are

implemented, low population coverage or poor quality cytology can be attributable to high incidence of the disease (Parkin and Bray, 2006).

Abnormalities in the cervical epithelium can be detected using the papanicolaou staining smear test. These changes are termed as cervical intraepithelial neoplasia (CIN) and are graded according to the depth of abnormal cells present (Fig. 3) (Braaten and Laufer, 2008). CIN1 is characterized by mild dysplasia of only one third of cervical epithelium. In a large proportion of cases these lesions will regress spontaneously and only 1% of cases will develop in to invasive cervical cancer (Braaten and Laufer, 2008). CIN2 cases exhibit dysplasia of up to two-thirds of the epithelium and in cases of CIN3 the majority of the epithelium is abnormal. It is estimated that cervical cancer will only develop in 12% of CIN3 cases if left untreated (Braaten and Laufer, 2008). Women over the age of 30 with persistent infection are at greatest risk of developing cervical cancer, so screening is hugely beneficial for this cohort of women. However, HPV screening programs are less effective for the younger age bracket, as women under 30 years have a higher incidence of HPV infection, the majority of which will spontaneously regress and never develop in to invasive cervical cancer (Arney and Bennett, 2010). Therefore it is more effective to implement vaccination programs in females of school age, to in order to generate neutralising antibodies prior to the initial infection of high-risk HPV.

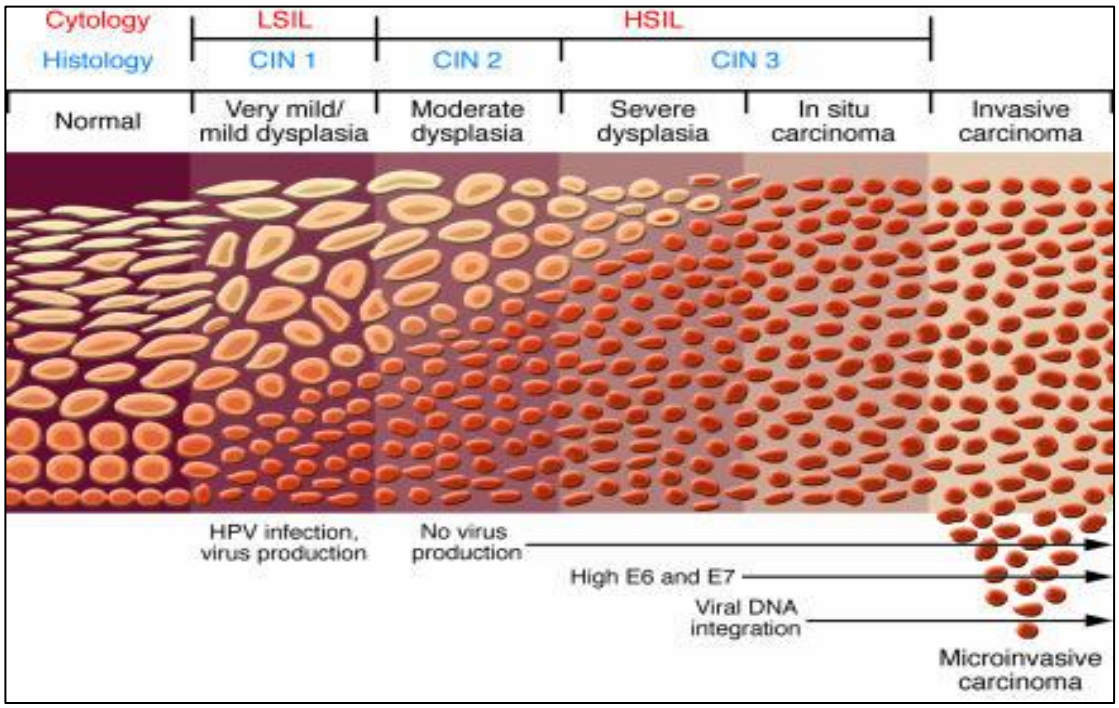


Figure 3. Progression of dysplasia to invasive carcinoma in cervical epithelium. High-risk HPV infection may cause mild dysplasia of the epithelium that can either regress or continue further to moderate dysplasia**.** Mild dysplasia of the epithelium is referred to as CIN1 and moderate dysplasia is known as CIN2. Further dysplasia and carcinoma *in situ* are known as CIN3. Expression of the HPV E6 and E7 oncoproteins and integration of HPV DNA, along with further accumulation of genomic instabilities can result in carcinoma *in situ* and invasive carcinoma**.** Image taken from (Lowy and Schiller, 2006).

Alternative screening methods have been investigated to maximize the detection of abnormal lesions in low resource regions. In a randomized control study in rural India, a single round of screening with a high-risk HPV assay was shown to significantly reduce the incidence of advanced cervical cancer and associated deaths (Sankaranarayanan et al., 2009). This type of testing also requires less advanced training and resources than cytological screening methods. Another group demonstrated that the Qiagen *care*HPV

kit was a predictive tool for ≥CIN 2 lesions (Ying et al., 2014). From the results obtained in both trials it is expected that HPV testing will be available in low resource regions in the near future, as a primary screening tool for women over 30 years of age.

1.1.6 Vaccine

There are currently three commercially available recombinant HPV vaccines. Cervarix (GlaxoSmithKline) is a bivalent vaccine, which protects against both high-risk HPV 16 and 18, and Gardasil (Merck) is a quadrivalent vaccine, which offers vaccination against four HPV types: 6, 11, 16, and 18 (Schiffman et al., 2007). Both vaccines offer protection against the development of premalignant lesions and cervical cancer, but Gardasil provides additional protection against genital warts caused by HPV 6 and 11 (Crum et al., 2006) . Most recently, Gardasil 9 has been developed to prevent the development of cancers associated with HPV types 16, 18, 31, 33, 45, 52 and 58 and genital warts caused by HPV types 6 and 11. The vaccines are comprised of HPV L1 virus-like particles (VLPs) that successfully induce high concentrations of neutralizing antibodies to the HPV L1 protein (Crum et al., 2006).

Each of the HPV vaccines was manufactured using recombinant DNA technologies. The viral L1 gene was sub-cloned into yeast for the quadrivalent vaccine and baculovirus for the bivalent type (Mariani and

Venuti, 2010, Schwarz and Leo, 2008). The recombinant DNA overexpresses L1 proteins that spontaneously self-assemble to form virus-like particles (VLPs), which are morphologically and antigenically similar to the structure of wild type virions, and are capable of inducing high levels of virus neutralizing antibodies in the host (Giuliano et al., 2011, Palefsky et al., 2011).

Koutsky *et al* conducted the first phase II clinical trial using VLPs for a HPV16 vaccine in women aged 16-23 years, and this was shown to reduce the incidence of HPV16 infection and cervical intraepithelial neoplasia (Koutsky et al., 2002). In previous clinical trials both the quadrivalent and bivalent vaccines have elicited a sustained and potent immune response, much higher than a natural infection. Peak antibody titres were reached after the third vaccination but gradually declined thereafter (GlaxoSmithKline Vaccine et al., 2009, Villa et al., 2006). A randomized placebo-controlled phase III clinical study was conducted to assess the effectiveness of the quadrivalent vaccine in preventing genital warts, cervical, vulvar or vaginal intraepithelial neoplasia and cervical cancer in women between 16 and 24 years. Overall, the vaccine efficacy was 100% for each of the end points thus reduced HPV associated anogenital disease in this cohort of women (Garland et al., 2007). Randomized clinical trials using the Gardasil vaccine have shown protection against genital warts and premalignant anal neoplasia in male subjects (Giuliano et al., 2011) (Palefsky et al., 2011). There is also evidence supporting a 2 dose

regimen of Cervarix vaccine instead of the previous regimen of 3 doses. Several trials have shown that a 2 dose regimen in 9-14 year old girls provided an immunogenic response that was non-inferior to 3 doses administered to a cohort 15-25 year olds (Romanowski et al., 2011, Puthanakit et al., 2016).

Evidence from a recent trial has demonstrated that the HPV 16/18 AS04-adjuvent vaccine shows cross protective efficacy against other oncogenic HPV types: 31, 33, 45 and 51 (Wheeler et al., 2012). Recently the new Merck V503 vaccine has been evaluated in phase III clinical trials to assess its effectiveness in targeting 7 high-risk and 2 low-risk HPV types. This study demonstrated that the vaccine prevented 97% of high-grade, precancerous vaginal, vulvar, and cervical disease caused by HPV types 31, 33, 45, 52, and 58-which are not targeted by the Gardasil or Cervix vaccines-and was also shown to generate immune responses to HPV types 6, 11, 16, and 18 (2014, Vesikari et al., 2015, Merck, 2014).

A major drawback of existing vaccines is the expensive production of VLPs in the eukaryotic model systems. The high production costs restrict the use of vaccines in the developing world, where there is already a lack of cytological screening programs. A plausible alternative and more cost-effective method is the use of plants for the production of oral vaccines. Previously HPV11 L1

protein has been expressed in transgenic potato plants, demonstrating self-assembly into immunogenic VLPs (Warzecha et al., 2003). HPV16 L1 protein has also been expressed in transgenic potato and tobacco plants, but the overall yield obtained still remains a challenge to overcome (Maclean et al., 2007).

Unfortunately these vaccines are not curative for pre-existing HPV infections or precancerous lesions. Therefore high coverage of the vaccine in school age children before the onset of sexual activity is the most effective strategy for the reduction in HPV associated cancers (Hildesheim and Herrero, 2007, Hildesheim et al., 2007). However, if the female vaccination coverage falls below 50% then the herd immunity effect is ineffective for protection of males who do not receive the vaccine (Drolet et al., 2015). This raises the question as to whether males should also receive the HPV vaccine, but the overall costs versus benefit considerations are being evaluated.

1.1.7 Treatment

Early stages of pre-cancer have a relatively successful response to treatment, however locally advanced cervical cancer cases are associated with poor prognosis and high recurrence rates. CIN1 does not require initial treatment as the lesions will often spontaneously regress, but follow up screening is used to monitor potential changes in the epithelium. However, treatment is

routinely given to patients who present with CIN2 or CIN3. Treatment options for these precancerous lesions include laser therapy or cryotherapy to destroy abnormal tissue, or conisation and large loop excision to remove cervical stroma (Nuovo et al., 2000). The standard treatment of locally advanced cancer is usually radiotherapy combined with cisplatin chemotherapy, yet the 5-year patient survival rate remains below 50%. In late stage or recurrent cancer cisplatin and toptecan combination therapy only support a median survival of less than one year (Haedicke and Iftner, 2013). These findings highlight the importance of regular Pap smear testing for early detection and treatment.

1.1.8 Structure, genome and gene expression

HPVs have circular double-stranded DNA genomes that are approximately 8kb in size and packaged in to icosahedral capsids. The viral genome is comprised of 8 major open reading frames (ORFs) that are usually divided into early and late regions (Fig. 4). The early region encodes six proteins (E1, E2, E1^E4, E5, E6 and E7) with regulatory functions throughout the HPV life cycle; however, E1^E4 is a known late protein as it is expressed in the mid to upper layers of the epithelium. The late region encodes the L1 and L2 structural capsid proteins, which are exclusively expressed in the later stages of differentiation in the granular epithelium (Zheng and Baker, 2006).

Transcription occurs on only one DNA strand and is initiated from more than one promoter region, thus producing multiple mRNAs with numerous ORFs (Graham, 2010). Using 5' RACE the HPV18 early viral transcriptional start sites (TSS) have been mapped at nucleotides 52 and 102 and the late transcripts commonly start at nucleotide 811, although several other sites have been identified (Wang et al., 2011). Other transcription maps have identified the early and late TSS at nucleotides 97 and 670 in HPV16 (Grassmann et al., 1996, Smotkin and Wettstein, 1986) and nucleotides 99 and 742 in HPV31 respectively (Hummel et al., 1992). The viral genome also contains a non-coding long control region (LCR), which despite occupying only 850bp of the genome contains the origin of replication and important host and viral transcription factor binding sites (Zheng and Baker, 2006). Once the host cell is infected the early promoter in the LCR is constitutively active throughout the virus life cycle (Graham, 2010).

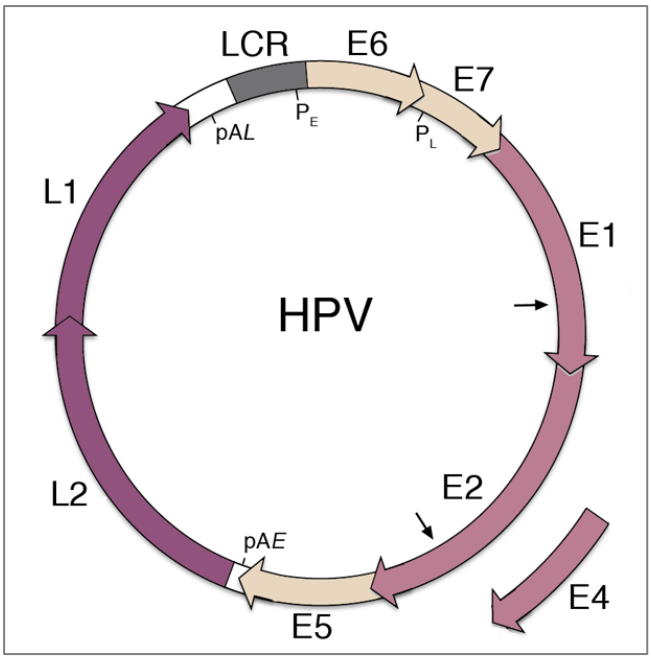


Figure 4. Organisation of the HPV genome. The genome contains a long control region (LCR) and eight ORFs, whose products are alternatively spliced. The LCR contains transcription factor, E2 and E1 binding sites for viral replication and gene expression. The early promoter (P_E) and late promotor (P_L) control early and late gene expression during epithelial differentiation. Gene expression is also regulated by the early and late polyadenylation sites $_P$AE and $_P$AL. Image taken from Roberts, 2015.

Although HPV16 and HPV18 share a similar library of transcription factors that bind to their LCR regions, the nucleotide positions and abundance of binding sites vary between the two high-risk HPV types (Fig. 5). The central portion of the LCR occupies around 400bp and is referred to as the enhancer region. This region in both HPV16 and 18 is flanked by E2 binding sites #3 and #4, and it also contains binding sites for several transcription factors, including NF1, AP1, Oct, transcriptional enhancer factor 1 (TEF-1) and Yin Yang 1 (YY1). The abundance of transcription factor binding sites, together

with the differential binding of these factors throughout the viral life cycle is thought to modulate HPV gene expression. The 5′ segment of the LCR occupies around 300bp and is flanked by the late polyadenylation site and the E2 binding site #4. The 3′ segment is the smallest region occupying around 150bp and contains an E1 binding site that is indicative of the origin of replication and the early promoter. The early promoter region is controlled by binding of the viral E2 protein and transcription factors in a differentiation-dependent manner (Carson and Khan, 2006). More specifically, this region contains E2 binding sites #1 and #2, a binding site for the transcription factor SP1 and TATA box for TFIID binding, which collectively are all very important for early E6/E7 promoter activity. It is believed that initial binding of TFII is required for the recruitment of the pre-initiation complex, then binding of SP1 is responsible for activating the E6/E7 promoter. Importantly, due to overlapping sites in the early promoter region, SP1 also blocks the binding of the E2 protein. However, when levels of E2 protein are elevated, the E2 protein can displace SP1 and lead to initiation of E1 dependent viral replication and repression of early gene expression. Furthermore, E2 binding at position #4 can displace TFII, which further represses the early promoter activity. HPV DNA integration events are a common occurrence in most cervical cancers. During these integration events there is often loss of the intact E2 ORF, thus loss of the repressive activity and a concomitant increase

in E6 and E7 expression, giving rise to a cell growth advantage (Dowhanick et al., 1995, Jeon et al., 1995).

Additional transcription factors that bind to the viral LCR they have varying roles in modulating gene expression (reviewed by O'Connor et al., 1995). Most PV types contain binding sites for the TEF-1, which has previously been shown to contribute to enhancer activity in HPV16 (Ishiji et al., 1992b). Interestingly the TEF-1 binding motif overlaps the YY1 binding sequence, so there may be competitive binding between these two factors (O'Connor et al., 1995). The high-risk HPV types 16, 18 and 31 contain the highest number of YY1 binding sites across the LCR, with 10, 7 and 8 identified sites respectively. Binding of YY1 in the viral LCR is known to have repressive effects on early gene expression in HPV18 and HPV16 (Bauknecht et al., 1992, Lace et al., 2009, O'Connor et al., 1996, O'Connor et al., 1998). Naturally occurring point mutations or deletions in YY1 binding sites have been identified in viral episomes obtained from cervical carcinomas and interestingly these correlated with increased enhancer activity (May et al., 1994, Dong et al., 1994). YY1 has also been shown to repress gene expression in other DNA viruses such as adenovirus and EBV, which are also associated with persistent infection (Montalvo et al., 1995, Shi et al., 1991). Activator protein 1 (AP1) consensus sites are conserved amongst PV types, with either one or two sites existing within the central enhancer region observed across

the mucosal types. Binding of AP1 is thought to contribute to epithelial specific activation of viral genes (Thierry et al., 1992), and may be needed to modulate viral genes in response to epidermal growth factor (EGF) and protein kinase C (PKC) signaling (Chan et al., 1990). Neurofibromin 1 (NF1) binding sites are predominantly located in the central segment of the viral LCR, and have been shown to be important for enhancer function in HPV16 (Baldwin et al., 2007). In HPV16, binding of Oct-1 to the 3' region of the LCR is next to a NF1 binding site. O'connor *et al* demonstrated that binding of Oct-1 stabilises NFI, which itself causes increased levels of enhancer activity (O'Connor et al., 1995). Most recently, several additional transcription factors have been identified using the TRANSFAC database; FOXA1 and MYC were confirmed to bind to the LCR regions of HPV16 and HPV18 by ChIP assays (Sichero et al., 2012). The emergence of additional transcription factors that are capable of binding to the viral LCR indicate that the level of transcriptional regulation is more complicated than originally considered.

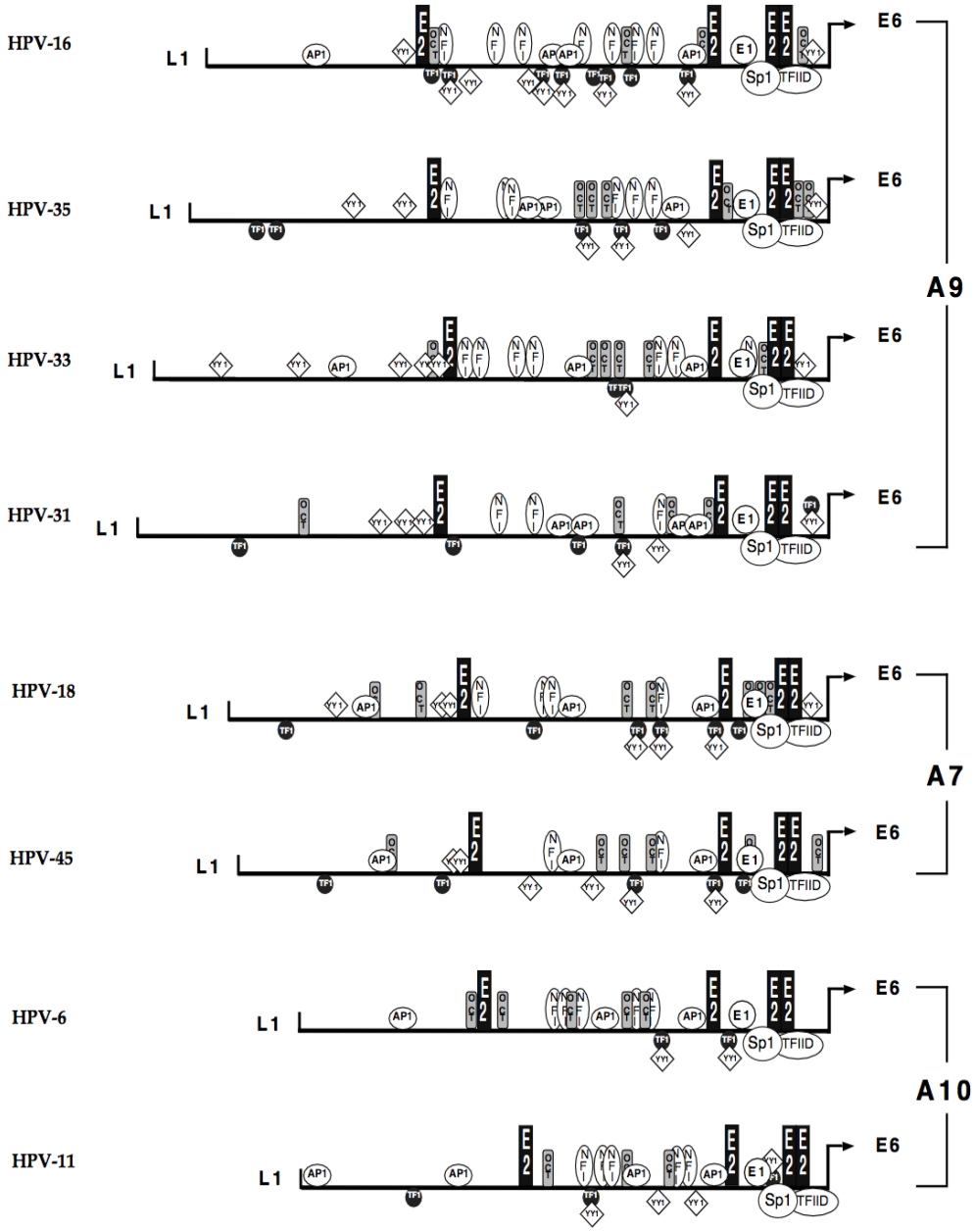


Figure 5. Viral LCR regions of high- and low-risk HPV types. The alpha genus contains the Alphapapillomavirus 9 (A9) species, which consists of high-risk HPV16, -31, -33 and -35; A7, which contains high-risk HPV18 and HPV45; and A10, which contains low-risk HPV6 and HPV11. The schematic depicts the predicted and/or confirmed binding profiles of transcription factors and viral proteins to the viral LCR region of the alpha genus. Image taken from O'Connor et al., 1995.

It is well known that alternative promoter usage, alternative splicing and alternative polyadenylation site usage regulate HPV gene expression.

1.1.9 Promoter usage

Depending on the activity of different promoters or polyadenylation sites, transcripts are expressed with different coding potentials. The mRNAs expressed from the early promoter have the potential to encode E1, E2, E1^E4, E5, E6 and E7 and are all polyadenylated at the early polyadenylation site pAE. However some of the mRNAs expressed from the late promoter are also polyadenylated at pAE, and these can encode E1^E4 and E5. The majority of mRNAs expressed from the late promoter are polyadenylated at the late polyadenylation site (pAL) and these can encode E1^E4, E5, L1 and L2. Reviewed by (Johansson and Schwartz, 2013).

1.1.10 Polyadenylation

There are two known polyadenylation sites in the HPV18 genome (Fig. 6); using 3′ RACE the early polyadenylation site (pAE) has been mapped at position 4235 and the late site is at 7278 (Wang et al., 2011). The early polyadenylation site is downstream of E5 and the late polyadenylation site (pAL) is downstream of L1. The early and late regions of the genome are separated by pAE, which effectively blocks L1 and L2 expression during the early stages of the life cycle. This strategy allows the virus to evade immune surveillance and persist in the host cell by preventing the expression of the

most immunogenic proteins until the upper layers of the epithelium (Johansson and Schwartz, 2013). These capsid positive cells are then sloughed off from the cornified layer of epithelium. The presence of upstream regulatory elements are known to stimulate pAE polyadenylation in HPV16, through the binding of the poly(A) polymerase (FIP1) to a short UTR region (Zhao et al., 2005). Regulatory elements located downstream in the L2 region can also regulate pAE activity (Sokolowski et al., 1998, Terhune et al., 1999, Terhune et al., 2001). In HPV16 and HPV31 the cleavage stimulation factor 64kDa subunit (CSTF64) was shown to interact with this regulatory element in the L2 gene and stimulate polyadenylation of mRNAs (Oberg et al., 2005, Terhune et al., 1999). Additionally, the RNA element in HPV16 L2 was also shown to interact with the heterogeneous nuclear ribonucleoprotein H (hnRNP H) to stimulate early polyadenylation at pAE (Oberg et al., 2005). Further work has demonstrated that increased viral E2 protein expression can also repress pAE by blocking formation of the polyadenylation complex, which allows transcriptional read-through, the induction of the late genes and subsequent virion synthesis (Johansson et al., 2012).

1.1.11 Alternative Splicing

Given that the HPV genome is only 8kb, the virus employs alternative splicing to generate polycistronic mRNA, which can encode multiple proteins, thus allowing the virus to increase the repertoire of protein encoding

transcripts (Fig. 6). Alternative splicing is also advantageous because it generates mRNAs with the fewest possible number of AUG codons upstream of each HPV gene, which are known to have a negative effect on translation, reviewed by (Johansson and Schwartz, 2013). HPV18 is known to contain 5 major 5′ splice sites and 6 major 3′ splice sites, which are highly conserved with the splice sites in other papillomavirus genomes (Wang et al., 2011). The retention of the first HPV18 intron between splice donor 233 (SD233) and splice acceptor 416 (SA416) is required for generating the E6 encoding mRNA. However, high levels of hnRNP A1 and A2/B1 stimulate splicing between SD233 and SA416 to generate mRNAs that encode E7 and smaller E6 variants (E6*I, E6*II and E6*III) (Rosenberger et al., 2010). Alternatively spliced protein isoforms can confer different functions, for example, the large HPV16 E6 isoform can bind and augment degradation of procaspase 8, whereas the small E6* isoform can stabilize procaspase 8, thus modulating the apoptotic pathway in the opposite manner (Filippova et al., 2007). Furthermore, splicing between SD1357 and SA3434 generates an mRNA with the potential to encode an alternative E2 protein, E8^E2c, which functions as a transcriptional inhibitor (Wang et al., 2011, Lace et al., 2008).

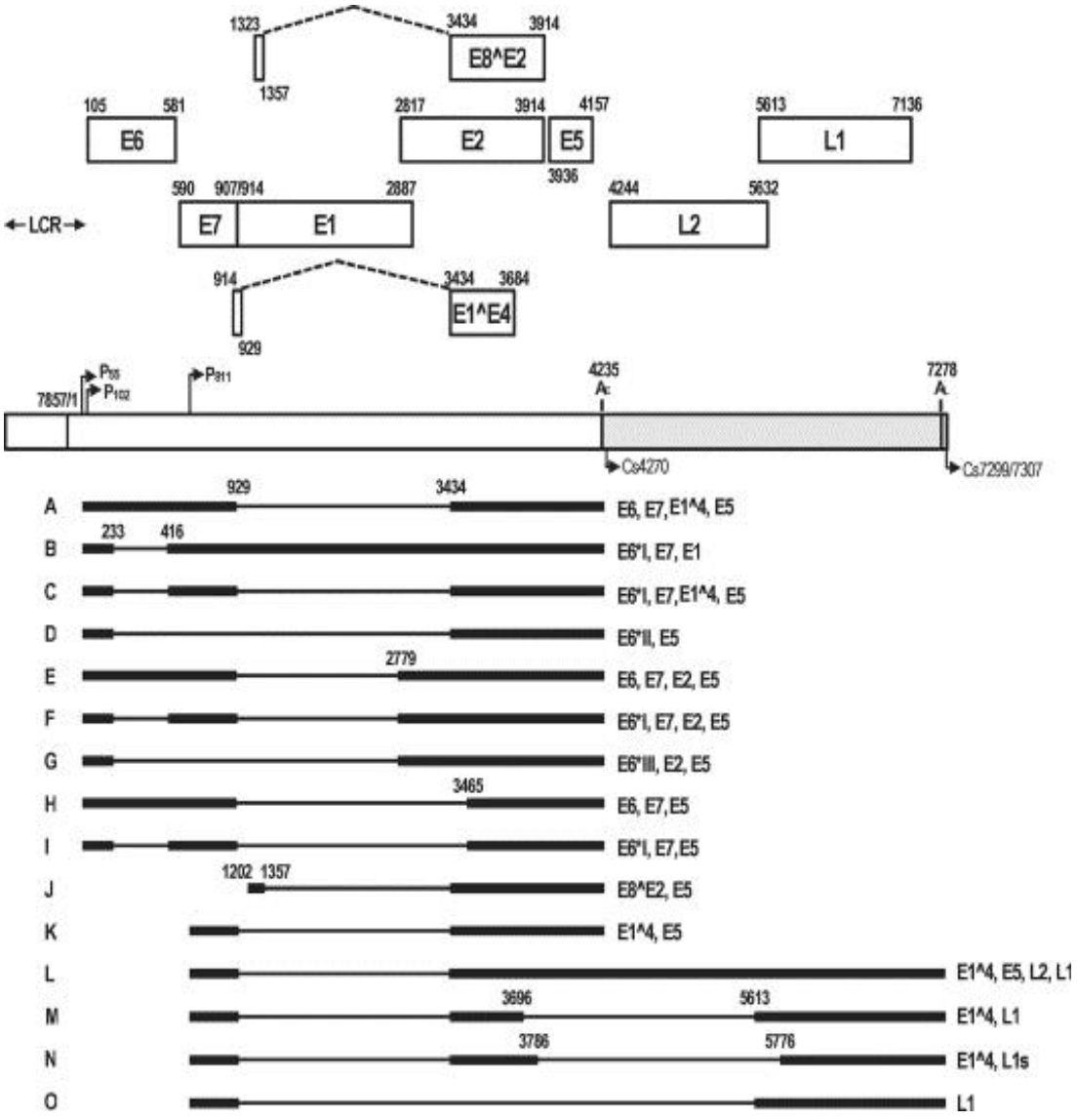


Figure 6. Transcription map of HPV18. The genome is 7851kb in size and encodes a subset of early (E) and late genes (L). The diagram above depicts a linear arrangement of the genome, with boxes representing the ORFs. The numbers at either end of each ORF box indicates the start codon and the last nucleotide. To generate the intact E1^E4 and E8^E2 ORFs splicing is required, which is shown by the dashed line. The early and late polyadenylation sites are represented as A_E and A_L and the long control region is denoted as LCR. RNA species derived from alternate splicing and promoter usage are labelled A-O. The thick black line represents the exons and the thin black lines represent the introns, and the known nucleotide splice sites are labelled above each transcript. The proteins encoded by each RNA species are shown to the right of each transcript. Figure taken from (Wang et al., 2011).

Splicing itself is regulated by spliceosome complexes and expression of intonic/exonic sequence enhancers (ISE/ESE) or silencers (ISS/ESS) (Wang and Burge, 2008). Studies in HPV16 have demonstrated that splicing at the most commonly used 3' splice site (SA3358) is enhanced by the serine-arginine-rich splicing factor 1 (SRSF1), as mutational inactivation of the SRSF1 binding site was shown to cause skipping of SA3358. Conversely, binding of SRSF9 or SRSF3 factors inhibit splicing at SA3358 (Jia et al., 2009, Somberg and Schwartz, 2010). Interestingly, mRNAs that encode E6 and E7 are mainly spliced at SA3358, so it is thought that SRSF1 protein is required to enhance expression of these viral proteins, whilst reducing the splicing of the E2 encoding mRNA. However this splice site is also required for production of mRNAs encoding L1 and L2, so there needs to be a balance in SRSF1 expression to modulate viral gene expression. Overall it has been postulated that in the basal epithelium SRSF1 is expressed at high levels to induce E6 and E7 expression and limit E2 production. However as the epithelium differentiates the levels of SRSF1 are reduced, which subsequently allows the expression of the L1 and L2 transcripts (Johansson and Schwartz, 2013).

1.1.12 HPV life cycle

The HPV life cycle differs from other virus families, as infection requires epidermal or mucosal epithelial cells such as the basal cells of the stratified squamous epithelium, which themselves are mitotically active and capable of

dividing (Fig. 7). The productive HPV life cycle is strictly linked to differentiation of these epithelial cells (reviewed by (Bodily and Laimins, 2011). Infection is thought to reach dividing basal epithelium cells via micro abrasions of the skin or mucosa, with the cervical squamocolumnar (SC) junction particularly vulnerable to infection (Hubert et al., 2014, Herfs et al., 2012). The HPV L1 protein attaches to heparin sulphate proteoglycans on the cell surface (Combita et al., 2001, Giroglou et al., 2001, Joyce et al., 1999). Viral entry may also be mediated by calveolar or clathrin mediated endocytosis, via the α6-integrin receptor (Yoon et al., 2001).

Upon infection the HPV early proteins E1 and E2 are expressed, allowing the initial amplification of viral episomes to 50-100 copies. Although the viral episomes replicate independently from the host cell chromosomal DNA, the virus utilizes the host cell polymerase machinery (Conger et al., 1999). As each basal cell divides, one of the daughter cells will migrate and begin to differentiate, whilst the other will remain as a new basal cell (Bodily and Laimins, 2011). The viral genome is maintained upon cell division by tethering to the host chromosome, which ensures equal segregation of viral episomes in to daughter cells (Graham, 2010, McBride, 2008). This process is mediated by an E2 tethering mechanism involving host cell ChlR1 helicase, which is important for loading E2 to the host cell sister chromatids (Parish et al., 2006a, McBride, 2008). E2 from bovine papillomaviruses (BPV) have also

been shown to mediate viral tethering through binding to the host protein Bromodomain Containing 4 (BRD4) (You et al., 2004).

The early viral genes E5, E6, and E7 are expressed in the infected basal cells, and the co-operative actions of these viral proteins culminates in enhanced cellular proliferation and lateral expansion of the epithelium (zur Hausen, 2002). The actions of E6 and E7, which disrupt the p53 and retinoblastoma pathways, respectively, cause normally differentiating epithelial cells to remain cycling, along with abrogation of cellular differentiation and a reduction in apoptosis (Hamid et al., 2009, Graham, 2010). The E5 protein is thought to bind to EGFR in order to induce mitogenic signals for proliferation (Genther Williams et al., 2005). Together E5, E6 and E7 maintain an environment favorable for unscheduled HPV replication.

The expression of the viral E2 protein is important for the regulation of the E6 and E7 oncoproteins. The increase in E2 expression in the upper suprabasal layers of the epithelium causes downregulation of the E6 and E7 proteins and late viral proteins are expressed. Upon differentiation to the granular layer the L1 and L2 capsid proteins are expressed and the viral epsiomes are packaged in to capsids and released from the surface epithelium (Graham, 2010). There is also evidence supporting the role of E1^E4 in mediating viral

release; E1^E4 protein can disrupt the keratin network and mediate virion release. (Doorbar et al., 1991, Wang et al., 2004b)

Overall, the completion of the viral life cycle is dependent on the differentiation of the epithelium that it infects; therefore, a controlled balance in the expression of the viral oncoproteins is imperative for efficient viral replication but also the successful completion of virion production and release. Any perturbations in viral oncoprotein expression can deregulate the complete viral life cycle, and lead to the malignant progression observed in HPV-associated cancers. The functions of the viral proteins are discussed in further detail in the following section.

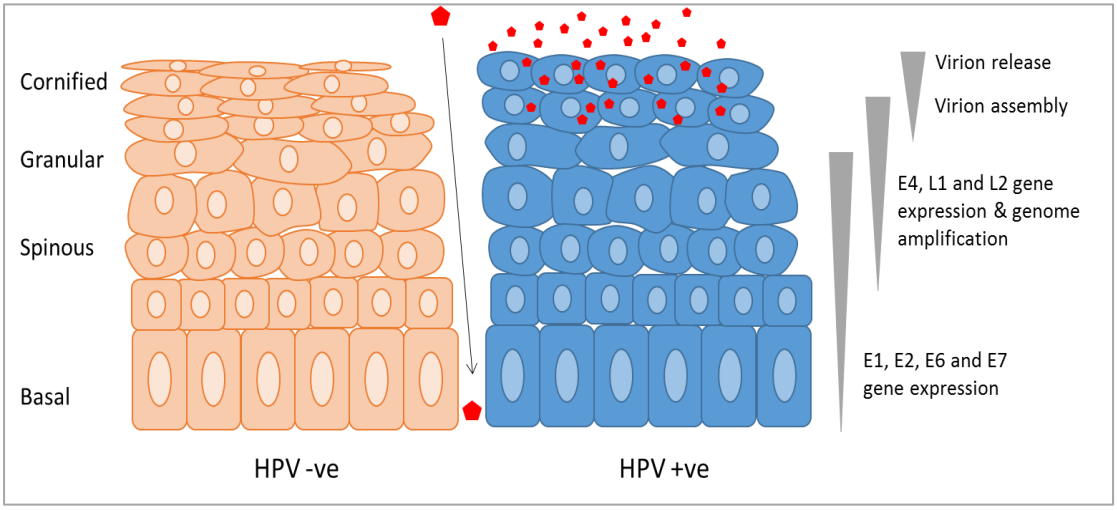


Figure 7. Life cycle of HPV infection in stratified epithelium.

Virus entry via micro abrasions in the surface epithelium and infection of basal cells. Initially the genome is maintained at low copy numbers. In the lower/middle layers, genome amplification occurs and the high risk E6 and E7 proteins cause additional cell cycle re-entry and proliferation. In the upper layers the cells exit the cell cycle and L1 and L2 proteins are synthesized to allow viral encapsidation. Virions are released as infected cells are sloughed from the surface epithelium. Figure adapted from(Bodily and Laimins, 2011).

1.1.13 Early Proteins

1.1.13.1 E1

E1 is an ATP-dependent helicase and primary replication protein. E1 cooperatively binds to the replication origin with E2, where it forms double hexamers to encircle DNA and unwind the viral origin (Mohr et al., 1990) (McBride, 2008, Sanders and Stenlund, 1998). Therefore in co-operation with E2, the E1 protein plays an essential role in the initial replication of viral genomes (Remm et al., 1992, Ustav and Stenlund, 1991). Whilst E1 is necessary for initial replication there is more recent evidence that E1 is not required for the maintenance replication of viral genomes (Egawa et al., 2012). The presence of nuclear localisation signals (NLSs) and nuclear export signals (NES) in the N-terminal region of E1 contribute to the regulation of this protein. Specifically, ERK1/2 kinase phosphorylation of the E1 nuclear localization signal causes nuclear accumulation of E1 (Yu et al., 2007). Accumulation of cyclin A and cyclin E and their cyclin dependent kinases (CDKs) during the S/G2 phases of the cell cycle further contribute to E1 nucleocytoplasmic localization of E1 (Deng et al., 2004, Ma et al., 1999). Overall, the viral E1 protein remains in the nucleus during S phase and G2 of the cell cycle, when CDK1 and CDK2 are active. This nuclear retention of E1 allows the virus to couple HPV DNA replication to the cell cycle, which makes overall viral replication more efficient (Ma et al., 1999, Deng et al.,

2004, Yu et al., 2007). E1 has been shown to contain a caspase cleavage motif that is targeted by caspase-3 and caspase-7, which leads to cleavage of the E1 replication protein (Moody et al., 2007). There is evidence to support that this E1 modification is required to facilitate viral genome amplification (Moody et al., 2007).

1.1.13.2 E2

Papillomavirus E2 is an important regulatory protein implicated in a wide range of events throughout the HPV life cycle, including: the initiation of viral DNA replication, regulation of viral transcription, partitioning the viral genome during mitosis and potentially viral packaging (reviewed by (McBride, 2013). The E2 ORF encodes the full length E2 protein, which is comprised of an N-terminal transactivation domain, C-terminal DNA binding domain, and a linker hinge region (McBride, 2013). During initiation of viral replication, E2 binds via its N-terminus to the E1 helicase protein and recruits it to the origin of replication (Mohr et al., 1990). Once the E1 protein is loaded onto the viral DNA, E1 mediated ATP hydrolysis causes the dissociation of the E1-E2 interaction and displacement of E2 (Lusky et al., 1994, Sanders and Stenlund, 1998, Sanders and Stenlund, 2000).

E2 also has an important role in tethering the viral genome to the host chromosomes through protein-protein interactions during mitosis, in order to

facilitate stable maintenance and partitioning of the viral genome in the host cell (You et al., 2004, Parish et al., 2006a). Overall, this function of E2 essentially permits persistent infection of HPV in the host cell.

Furthermore, E2 is also necessary during the vegetative replication stage as there is a large detectable abundance of E2 during such amplification of the genome (Xue et al., 2010, Johansson and Schwartz, 2013). Formation of viral DNA replication foci is dependent on E2 and within these foci the DNA damage response including the ATR and ATM pathways can be induced, which is believed to recruit cellular repair proteins in order to increase viral genome amplification (Moody and Laimins, 2009b, Sakakibara et al., 2011, Reinson et al., 2013).

It is well established that in HPV positive cell lines E2 can activate or repress the early viral promoter, with the latter causing down-regulation of the E6 and E7 genes (Thierry and Howley, 1991). Papillomaviruses can encode shorter E2 forms (E8^E2) consisting of a short E8 peptide attached to the hinge region and DNA binding domain. These can repress viral transcription and replication, via competition for binding to E2 binding sites, recruitment of repressor complexes, or dimerization with full length E2 (Lambert et al., 1987, Lim et al., 1998, Monini et al., 1993, Barsoum et al., 1992, Ammermann et al., 2008). Disruption of E2 by integration of the viral genome reduces E2

mediated repression, thus allowing higher expression of the E6 and E7 oncoproteins, which further augment progression to malignant transformation (Durst et al., 1987, Bernard et al., 1989).

The majority of E2 binding sites contain at least one CpG site. Methylation of the cytosine residue at these sites is known to disrupt E2 binding and E2 mediated control of transcription. Kim *et al* demonstrated that E2 binding sites were often highly methylated in less differentiated cervical epithelial cells. In contrast, more highly differentiated epithelial cells, which support viral genome amplification, displayed hypomethylation in E2 binding sites in the LCR. (Kim et al., 2003). In HPV16, methylation was more prevalent at the CpGs located within the E2 binding sites proximal to the early viral promoter (Bhattacharjee and Sengupta, 2006). Overall, methylation of E2 binding sites that mediate repression of the early promoter may enhance malignant transformation due to increased E6 and E7 levels. (Kim et al., 2003).

Lastly it is possible that E2 is implicated in the packaging of virions, as earlier studies with BPV demonstrated that E2 enhanced DNA packaging within L1 and L2 pseudovirions (Zhao et al., 2000). There is also evidence to suggest that E2 regulates post transcriptional activities such as viral transcript processing, via depletion of the polyadenylation complex, which allows read

through of the early polyadenylation site (McBride, 2013, Johansson et al., 2012).

1.1.13.3 E5

The E5 protein is highly hydrophobic and associates with internal membrane structures (Krawczyk et al., 2010, Conrad et al., 1993). High-risk E5 proteins exhibit weak transforming activity when expressed alone, but when expressed in tissue culture assays they can augment the transforming activity of E6 and E7, so overall E5 is often classified as an additional oncogene (Moody and Laimins, 2010). E5 is thought to act cooperatively with E6 and E7 to promote hyperproliferation and malignant progression (Venuti et al., 2011). For example, E5 can display transforming activity by inhibiting EGF receptor degradation and inducing the ligand-independent activation of growth factor receptors (McBride, 2008, DiMaio and Mattoon, 2001, Suprynowicz et al., 2010, Straight et al., 1995). Furthermore, when HPV16 E5 was expressed in the basal layer of the stratified squamous epithelium of transgenic mice, epidermal hyperplasia was observed and skin tumour formation occurred at a high frequency (Genther Williams et al., 2005). HPV16 E5 has also been identified as an oligomeric channel forming protein or viroporin, which is thought to mediate the hyperactivation of EGFR and mitogenic signaling (Wetherill et al., 2012).

1.1.13.4 E6 and E7 oncoproteins

E6 and E7 are the primary HPV oncoproteins that target key cellular proteins, thus perturb normal cellular processes in order to generate an environment favorable for replication of viral genomes (Munger et al., 2004). The integration of viral DNA in to the host chromosome is a key event involved in HPV induced carcinogenesis, and is strongly correlated with disease progression (Daniel et al., 1997, Munger et al., 2004). Analysis of cervical biopsy specimens has revealed that 72% of HPV16 positive cervical carcinomas contained viral integration and in those that were HPV18 positive, 100% displayed viral integration (Cullen et al., 1991). Upon integration several viral ORFs are disrupted or lost, with the exception of E6 and E7 expression (Baker et al., 1987, Jeon et al., 1995, el Awady et al., 1987, Shirasawa et al., 1987). Disruption of the E2 ORF, can lead to the loss of repression of E6 and E7, thus causing increased expression of the viral oncoproteins (Baker et al., 1987, Jeon et al., 1995). However, integrated HPV16 DNA is not always present in cervical carcinomas, suggesting that integration is not always required for HPV-induced carcinogenesis (Hudelist et al., 2004). Recent studies have even suggested that the existence of both epsiomal copies and integrated DNA facilitate HPV carcinogenesis, as episomal E1 and E2 proteins may initiate integrated DNA replication and subsequent chromosomal abnormalities (Moody and Laimins, 2010).

Initial studies demonstrated the ability of E6 and E7 from high-risk but not low-risk types to immortalize human foreskin keratinocytes (Hawley-Nelson et al., 1989, Munger et al., 1989). Interestingly E6 alone is not sufficient to immortalize cells but E7 alone is able to cause hyperproliferation, although the cells eventually senesce (Hawley-Nelson et al., 1989). However E6 and E7 alone are not sufficient to completely transform these cells, as additional accumulation of genetic alterations are required for malignant progression, which correlates with the latency period of HPV infection and onset of cancer. An example of this is the ability of high-risk E6 and E7 proteins to induce centrosome defects that lead to chromosomal instability (Duensing et al., 2000). The development of transgenic mice models where E6 and E7 expression was targeted to the basal cells of squamous epithelium demonstrated their ability to induce squamous epithelial neoplasia (Arbeit et al., 1994). Whilst E7 expression was shown to mainly promote tumour formation, E6 expression was predominantly involved in tumour progression, but the combined actions are able to induce transformation (Song et al., 2000).

The E7 protein is comprised of 98 amino acid residues and contains a zinc finger like domain in the C-terminus (CR3) and a flexible N-terminus (CR1 & CR2) and is a well-known phosphoprotein (Smotkin and Wettstein, 1987, McLaughlin-Drubin and Munger, 2009). The main phosphorylation sites located at ser31 and ser32 in the CR2 domain are phosphorylated by casein

kinase II (CKII) (Barbosa et al., 1990), and are important for the transforming ability of E7 (Firzlaff et al., 1991). E7 was first shown to interact with the retinoblastoma tumour suppressor gene products, which consist of consists of pRb, p107 and p130 (Dyson et al., 1989). Under normal conditions the pRb product is known to bind to the activating transcription factor members E2F1-3, whereas p107 and p130 preferentially bind to E2F4-5 proteins (Frolov and Dyson, 2004, Macaluso et al., 2006). In terms of differentiating epithelium the pRb protein is predominately expressed in the basal and proliferating cells, whereas p130 is expressed in the differentiated layers of the epithelium (Macaluso et al., 2006).

It is well known that the hyposphosphorylated from of pRb can bind and inactivate the transcription factor E2F during early G1 phase of the cell cycle. Conversely, activation of cyclin dependent kinases phosphorylate pRb so it no longer binds to E2F, thus allowing E2F dependent expression of S-phase specific genes (Frolov and Dyson, 2004, Macaluso et al., 2006). E7 has been shown to bind to pRb via it CR2 domain and target it for proteosomal degradation, thus leading to E2F mediated cell cycle progression (Fig. 8) (Munger et al., 1989, Barbosa et al., 1990). Moreover, E7 is known to block activity of the cyclin dependent kinase inhibitor p21, thus permitting expression of cyclin-A and cyclin-E associated genes and cell cycle progression (Jones et al., 1997). E7 can also bind to E2F6 and inhibit its role in

mediating transcriptional repression (McLaughlin-Drubin et al., 2008). Overall, deregulation of these pathways leads to S-phase entry in the suprabasal cells that would normally undergo differentiation. However, E7 can also disrupt the E2F/pRb without necessarily causing degradation of pRb (Chellappan et al., 1992), and there is evidence that rafts derived from HPV16 genome containing HFKs actually displayed similar levels of pRb in the suprabsasl layers compared to untransfected rafts (Collins et al., 2005). In addition to pRb, high-risk E7 can also target p107 and p130 for degradation. Interestingly, the degradation of p130 from differentiating epithelium occurs more frequently than pRb, and this action seems to be more important for viral DNA synthesis (Collins et al., 2005). A further role of E7, which is independent of pRb degradation, is to bind to histone deacetylases (HDACs) and increase the level of host cell E2F2-mediated transcription (Longworth et al., 2005).

The E6 protein is comprised of 150 amino acid residues and contains two zinc finger domains. Initial studies demonstrated the association between E6 and the tumour suppressor p53 (Werness et al., 1990). E6 from high-risk HPV16 and 18 is known to bind via its N-terminal to the E6-E6AP ubiquitin ligase complex and then bind to p53 to direct the degradation of the tumour suppressor protein (Fig. 9) (Werness et al., 1990, Scheffner et al., 1990). E6 can also bind directly to p53 to block transcription and DNA binding activity

(Moody and Laimins, 2010, Lechner and Laimins, 1994). This is similar to the SV40 large T antigen and the E1B adenovirus protein, which can also bind and inactivate p53. Crystal structures of E6 bound complexes have been elucidated, and studies have demonstrated that two zinc domains and a linker helix in the E6 protein form a basic-hydrophobic pocket, which binds to helical LxxLL motifs in target proteins (Zanier et al., 2013). Furthermore, the LxxLL motif of E6AP is required for E6-p53 interaction. Specifically, the LxxLL motif of E6AP is thought to structure a p53-binding cleft on E6, enabling E6 to be competent of for p53 binding (Martinez-Zapien et al., 2016). The normal function of p53 is to respond to cellular DNA damage, by arresting cells in G1 to allow DNA repair or if the damage is to great p53 induces apoptosis. Subsequent abrogation of p53 by E6 prevents apoptosis and inhibition of cell growth, but also permits the accumulation of genetic mutations. E6 can also bind to the acetyltransferases p300 and CREB-binding protein, which prevents them acetylating and stabilizing p53 (Patel et al., 1999). Moreover, HPV16 E6 has been shown to activate telomerase in human foreskin keratinocytes via its zinc-binding domain, which allows synthesis of telomere repeat sequences required for immortalization (Klingelhutz et al., 1996). An additional characteristic of high risk E6 proteins is their PDZ-domain binding motif located in the C-terminus, which is able to interact with PDZ targets involved in cell polarity and cell proliferation, most notably Disc

Large I (DlgI) (Kiyono et al., 1997, Lee et al., 1997) and Scribble (Scrib) (Nakagawa and Huibregtse, 2000).

Both E6 and E7 can induce genomic instability through DNA damage and activation of the ATM-ATR DNA repair pathways, which respond to double strand or single strand breaks respectively. Previous work has shown that activation of the ATM pathway, indicated by phosphorylation of downstream CHK2, BRCA1 and NBS1, is important for viral genome amplification and formation of viral replication foci in differentiated cells (Moody and Laimins, 2009a). However, activation of these pathways is not important for stable maintenance of viral episomes in undifferentiated cells (Moody and Laimins, 2009b). The same study demonstrated the induction of caspase cleavage upon keratinocyte differentiation through the action of CHK2, of which is activated via E7 binding to the ATM kinase. Thus E7 binding can induce caspase activity, which is known to cleave E1 for viral genome replication (Moody and Laimins, 2009b).

It is well established that angiogenesis is an early event in the development of HPV-induced malignant lesions and cervical cancer so it has previously been proposed that HPV gene products may facilitate this process (Toussaint-Smith et al., 2004). HPV31 E6 and E7 proteins have been shown to increase HIF-1α levels and induce expression of HIF-1 target genes during hypoxic

conditions, which initiate angiogenesis at the site of proliferating lesions, thus highlighting another role for these oncoproteins in malignant progression (Nakamura et al., 2009).

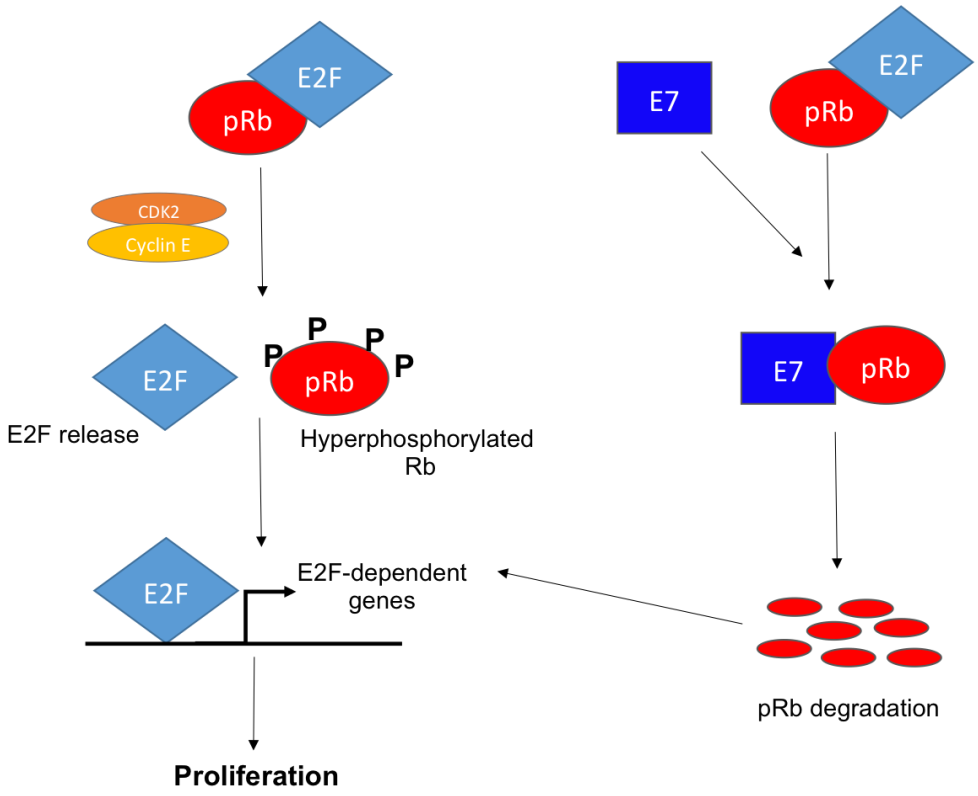


Figure 8. E7-mediated degradation of pRb. In normal cells, CDK2 mediated phosphorylation of the pRb protein causes dissociation with E2F during the late G1 phase of the cell cycle. E2F is released from the complex and binds to target genes for S-phase entry and progression. In HPV infected cells, high-risk E7 protein associates with the E2F-bound pRb protein and targets pRb for proteosomal degradation, thus allowing uncontrolled S-phase entry and proliferation.

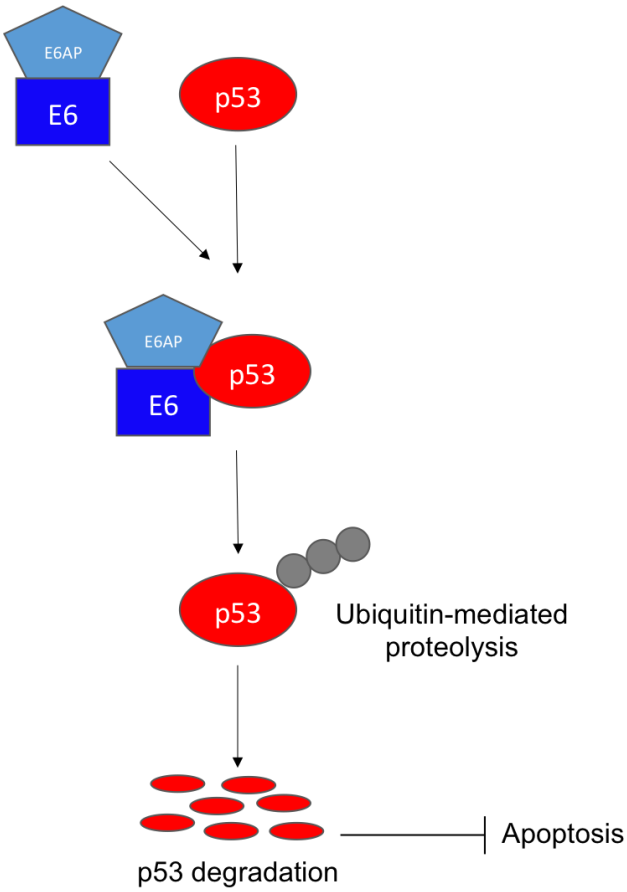


Figure 9. E6-mediated degradation of p53. Aberrant activity of the E2F transcriptional activator proteins by E7 would normally induce expression of p53, to mediate cell cycle arrest or apoptosis. High-risk E6 proteins circumvent this by targeting the E6AP ubiquitin ligase to p53, which transfers ubiquitin peptides to the p53 protein and mediates degradation by the 26S proteosome.

1.1.13.5 E6 and E7 from beta types

Interestingly, the E6 and E7 proteins from the beta HPV types are structurally similar to the alpha genus oncoproteins (Tomaic, 2016). Whilst beta HPV types have been implicated in the development skin carcinogenesis, the E6 and E7 proteins have varying transforming abilities across the different types. An initial study demonstrated that HPV38 is able to reduce pRb expression and immortalize primary human keratinocytes (Caldeira et al., 2003). Furthermore, E6 and E7 from the beta HPV49 have been shown to deregulate both the p53 and pRb pathways and immortalize primary keratinocytes (Cornet et al., 2012). The same study demonstrated that while HPV14 and HPV22 E7 can induce degradation of pRb, these HPV types are not sufficient for immortalization of primary HFKs (Cornet et al., 2012). Meanwhile, HPV24, -36, and -38 E6/E7 expressing HFKs display an increase in ΔNp73α expression, which is a dominant negative inhibitor of the p53/p73 pathways, and is thought to inhibit wild type p53 function (Cornet et al.,2012). Moreover, the E6 proteins from HPV5, -8, -20, -22, -38, -76, -92, and -96 have also been shown to promote degradation of Bak, and prevent UV induced apoptosis (Underbrink et al., 2008, Jackson et al., 2000). Taken this further, in vivo studies demonstrated that HPV38 E6 and E7 expressed from a K14 promoter in the basal and suprabasal epidermis of transgenic mice, enhanced the susceptibility of these mice developing UV-induced carcinogenesis (Viarisio et al., 2011). Overall these findings demonstrate that the E6 and E7

oncoproteins from beta HPV38 cause increased susceptibility to UV-induced carcinogenesis and contribute to SCC development in the skin.

1.1.13.6 Low-risk E6 and E7

Unlike high-risk types, the low-risk HPV types are not associated with the development of cancer; however, they are still causative of benign papilloma formation. This inability of causing malignant transformation is attributable to variations in the functions of the E6 and E7 proteins between high and low-risk types. The E6 proteins of low-risk HPV6 and 11 types have a much lower affinity for p53 than their high-risk HPV16 and 18 counterparts (Lechner and Laimins, 1994). Low-risk E6 can only bind to p53 via its C-terminus so it is unable to target the protein for degradation (Li and Coffino, 1996, Hiller et al., 2006). Furthermore, HPV11 E6 protein has also been shown to sequester p53 protein in the cytoplasm and induce apoptosis (Sun et al., 2008). Furthermore, the PDZ domain that is present in the E6 protein of high-risk HPV types is not found in the low-risk types, suggesting that the PDZ domain interaction is an important pathway in malignant progression. Low-risk E7 has a much lower affinity for pRb, p105, and p107 than its high-risk counterparts (Klingelhutz and Roman, 2012), and is less efficient at disrupting the pRb-E2F interaction. Interestingly, whilst low-risk E7 is able to degrade p130 it is unable to degrade p107 or pRb, whereas the high-risk E7 proteins can target all three pocket proteins for degradation (Zhang 2006(Barrow-Laing et al., 2010).

1.1.14 Late Proteins

1.1.14.1 E1^E4

The E1^E4 protein is expressed from the E1^E4 spliced mRNA product, which contains the first few of amino acids from the E1 ORF (Nasseri et al., 1987). Despite the E1^E4 ORF residing in the early region of the viral genome it is not believed to play an important role in the early stages of the life cycle. Nevertheless, E1^E4 is known to function during the later stages involving genome amplification and virion release (reviewed by (Doorbar, 2013). A major function of E1^E4 is to associate with cytokeratin and disrupt the cytokeratin filament network and formation of the cornified envelope within the host cell (Doorbar et al., 1991, Roberts et al., 1993), which is thought to facilitate virion release from the surface epithelium (Bryan and Brown, 2000). In the upper layers of HPV16 infected cervical epithelium there is accumulation of E1^E4 protein, which is shown to associate with the keratin network (Doorbar et al., 1997). Further research has proposed a model of HPV16 E1^E4-mediated keratin reorganization, whereby E1^E4 targets and disrupts the keratin network, resulting in phosphorylation and ubiquitylation of keratin (McIntosh et al., 2010). E1^E4 function is regulated by post translational modifications such as phosphorylation at specific amino acid residues (Doorbar, 2013). For example, HPV1, -16, and -18 E1^E4 is phosphorylated by SRPK1 (Bell et al., 2007), and sequestering of SRPK1 by

E1^E4 is thought to reduce phosphorylation of host SR proteins (Prescott et al., 2014).

1.1.14.2 L1 and L2

L1 is the major capsid protein that self assembles in to VLPs and L2 is the minor capsid protein, both of which are important for viral genome assembly and packaging. Completely assembled virions contain approximately 360 L1 molecules and only a few L2 molecules, to form an overall icosahedral capsid (Buck et al., 2008). During infection L1 mediates the initial interaction with host cells via heparin sulfate carbohydrates. This interaction causes a conformational change, resulting in exposure of the L2 amino terminus, which is cleaved by furin protease (Richards et al., 2006). Furin cleavage allows the virion to bind to a secondary cell surface receptor and enter the host cell (reviewed by (Buck et al., 2013). The ability of the L1 capsid protein to self-assemble in to VLPs forms the basis of the current prophylactic HPV vaccines (Lowy and Schiller, 2006).

1.1.15 Immune evasion and persistence

HPV must evade and suppress the immune system to allow viral persistence. Throughout the life cycle virion production is limited to the late stages in differentiated cells, in order to evade detection by the immune system. Additionally, virion production does not cause lysis of the host cell or

induction of the inflammatory response, which is another mechanism to circumvent an immune response (reviewed by (Bodily and Laimins, 2011).

Virally infected cells are capable of producing interferon (IFN), which provides a first line of defence against infection. However during HPV infection the IFN response is disrupted and many IFN target genes are suppressed. HPV16 E7 has been shown to inhibit IFN-α activity, via loss of the IFN stimulated gene factor (ISGF) (O'Brien and Saveria Campo, 2002). The E7 protein can also bind to IRF-1 via its Rb binding domain and inhibit the activation of the IFN-β promoter (Park et al., 2000). Similarly E6 protein appears to down regulate IFN-1, IFN- β, and STAT-1 genes (Nees et al., 2001).

The anti-inflammatory IL-10 marker shows increased expression in HPV positive cell lines (Alcocer-Gonzalez et al., 2006), and expression is correlated with disease progression of cervical cancer (Bermudez-Morales et al., 2008). There is also an increase in IL-6 protein expression in HPV positive SCC compared to HPV negative cancers (Tomlins and Storey, 2010). There is evidence that changes in cytokine production disrupt the activity of Langerhans and dendritic cells, which are needed for HPV clearance (Hubert et al., 1999). Further evidence suggests that natural killer (NK) cell activity—an important component of innate immune response—is also disrupted during HPV infection. In one study there were reportedly reduced numbers of NKs during HPV infection (Stentella et al., 1998), and PV infected cell lines

are resistant to NK lysis *in vitro* (O'Brien and Saveria Campo, 2002). It is suggested that E5, E6, and E7 have a role to play in the down regulation of the MHC-I complex, thus compromising the HPV specific T cell immune response (Ashrafi et al., 2006, Georgopoulos et al., 2000, Bodily and Laimins, 2011).

1.1.16 HPV and cancer progression

Cervical cancers predominantly arise in the cervical transformation zone between the squamous epithelium of the ectocervix and the columnar epithelium of the endocervix (Bodily and Laimins, 2011). There are four main steps in the development of cervical cancer; HPV infection of the epithelium at the cervical transformation zone, viral persistence in the host cell, progression of these persistently infected cells to a precancerous state, then cellular invasion through the basement membrane (Schiffman et al., 2007).

The E6 and E7 oncoproteins of high-risk HPV types act co-operatively in maintaining HPV infection and the development of HPV associated cancers. As previously mentioned, E6 and E7 disrupt the p53 and Rb mediated signaling pathways, which promotes epithelial proliferation and maintains an environment permissive for viral replication. However the activity of E6 and E7 alone is not sufficient to completely transform cells, but instead further genomic instability is required for cellular transformation and the

development of cancer. For example, addition of an activated Ras oncogene in tissue culture results in secondary mutations for complete cellular transformation (Munger et al., 2004). Integration of the HPV genome is associated with invasive cancer, as integration often occurs in the E1 and E2 regions, which disrupts the negative feedback control of E6 and E7 oncogenes by the regulatory E2 protein (Woodman et al., 2007). However not all women with invasive cancer have detectable integration, so it may not be direct cause (Vinokurova et al., 2008).

The levels of E6 and E7 expression are generally thought to underlie the neoplastic phenotype of cervical cancer. It is generally considered that CIN1 cases have a lower level of cell proliferation in the basal and suprabasal layers, with insufficient expression of E6 and E7 for cancer progression. In the case of CIN2 there is elevated E6 and E7 expression, which deregulates the normal cell cycle, allowing accumulation of genetic alterations. With regard to CIN2/3+ disease, deregulation of the viral life cycle is thought to facilitate integration into the host chromosome, thus further disrupting the regulation of E6 and E7 (Middleton et al., 2003, Doorbar et al., 2012).

1.2 CTCF

1.2.1 Novel interaction with HPV

Previously the Parish lab has demonstrated an association between CTCF and high-risk HPV genomes. Bioinformatic tools were used to determine CTCF binding sites in the HPV genome, which highlighted a strong binding site around nucleotide 3000 in high-risk HPV and around nucleotide 5400 in both high- and low-risk HPV.

1.2.2 Structure and function

Chromatin organizing CCCTC binding factor (CTCF) is a highly conserved and ubiquitously expressed zinc finger DNA binding protein. The protein is comprised of 727 amino acid residues, and features an N-terminus and C-terminus that can bind to other host proteins, and a central zinc finger domain (Fig. 10) (Zlatanova and Caiafa, 2009). CTCF is implicated in a diverse number of genomic processes, including gene activation/repression, X chromosome inactivation, chromatin insulation, and maintenance of genetic imprinting, which are all mediated through its ability to organize chromatin domains and facilitate long range chromosomal interactions (reviewed by (Phillips and Corces, 2009).

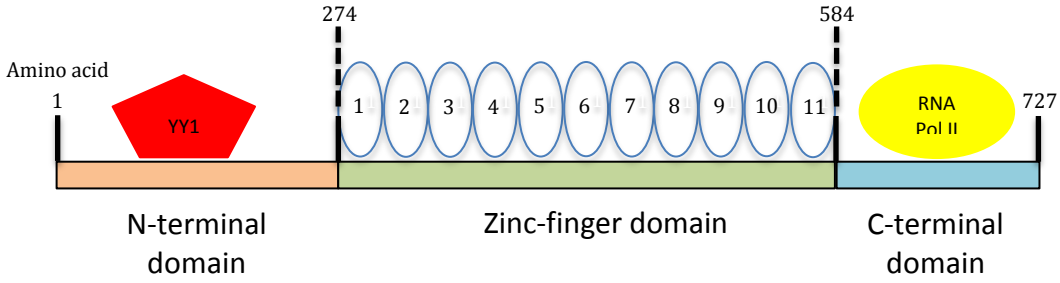


Figure 10. CTCF primary protein structure. CTCF contains 727 amino acid residues. The N-terminus is known to bind to YY1 and the C-terminus interacts with RNA Polymerase II. The central domain consists of 11 zinc fingers. Figure adapted from (Zlatanova and Caiafa, 2009).

CTCF was first discovered as a protein that binds to the chicken c-myc oncogene, where it was shown to mediate transcriptional repression (Lobanenkov et al., 1990). Independently, the NeP1 silencer protein that binds to the chicken lysozyme silencer was also discovered, and was later confirmed as CTCF, and again demonstrated repressor activity (Baniahmad et al., 1990, Burcin et al., 1997). On the other hand CTCF from HeLa cell extracts was shown to bind proximal to the amyloid β-protein precursor promoter where it functioned as a transcriptional activator (Phillips and Corces, 2009), (Vostrov and Quitschke, 1997). Shortly after, CTCF was first shown to mediate insulation in vertebrates at the *Igf2/H19* locus (Bell et al., 1999). Accumulating evidence suggests a wider role of CTCF beyond normal insulator functions, so the protein has been termed an 'architectural' protein

based on its contribution to the 3D structure of chromatin (Ong and Corces, 2014).

CTCF is ubiquitously expressed with variable binding patterns across different cellular genomes. Studies have estimated there to be around 35,000-75,000 sites in mammalian genomes, mainly distributed at linker regions that are surrounded by well-positioned nucleosomes (Wang et al., 2012), (Ong and Corces, 2014, Maurano et al., 2015). A recent study has identified constitutive CTCF binding sites, which display higher occupancy than dynamic ones and constitute 44% of the total number of sites (Beagan et al., 2017). Furthermore, Maurano *et al* profiled 40 cell lines and identified that the majority of CTCF binding sites were dynamic across different cell types; however, there was a subset of CTCF binding sites (20%) that were constitutively bound (Maurano et al., 2015). Around half of CTCF binding sites are located in intergenic regions of the genome, which supports their role as an insulator. Interestingly, only 15% of binding sites are near promoters- much less than other transcription factors- and 40% are within introns and exons, which may account for the CTCF enhancer blocking function (Ong and Corces, 2014, Kim et al., 2007).

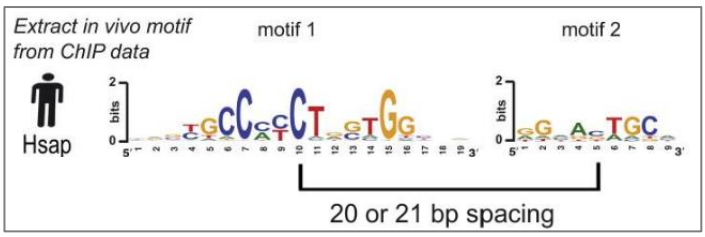


Figure 11. Position weight matrix (PWM) of a CTCF binding site published by Schmidt et al. The PWM was obtained after CTCF ChIP experiments using human liver tissue, followed by next generation sequencing. This PWM consists of a 20bp motif and a second 9bp motif. Image taken from (Schmidt et al., 2012).

The structure of CTCF is highly conserved throughout evolution. Within the CTCF gene, the promoter regions, untranslated regions, and intron-exon organization are identical between mice and humans (Ohlsson et al., 2001). The mammalian CTCF protein contains eleven zinc fingers and one DNA binding motif within exons E2 to E9. Exon 10 contains the stop codon and the 3'UTR. The first 10 zinc fingers contain a pair of cysteine residues separated by 12 amino acids from a pair of histidine residues surrounding a central zinc atom (Ohlsson et al., 2001). The resulting structure forms a DNA recognition alpha helix that inserts in to the major groove of DNA to make specific base contacts. Different combinations of zinc fingers can bind differentially to dissimilar CTCF DNA binding sites, and unlike other multi-ZF proteins, those found in CTCF can bind to both DNA and protein, thus CTCF is often termed a 'multivalent' protein (Ohlsson et al., 2001). Moreover, despite containing 11 zinc fingers Renda *et al* have shown that only 4 of these are essential for

strong binding at the β-globin insulator FII site and the imprinted *Igf2*/*H19* locus, and they recognize a core 12bp binding sequence, common at the majority of CTCF sites (Renda et al., 2007). Further studies have confirmed the presence a 20bp core motif that is present in most previously identified CTCF binding sites, and has shown to be highly evolutionary conserved (Holohan et al., 2007, Kim et al., 2007, Xie et al., 2007, Schmidt et al., 2012). Additionally, a 10bp motif has been identified upstream of the core 20bp motif (Rhee and Pugh, 2011)Schmidt et al., 2012) and this has been shown to bind zinc fingers 9-11 (Xiao et al., 2015). With the advent of CRISPR technology it has been possible to demonstrate that CTCF binding site location and orientation are important for loop establishment and 3D chromosomal architecture (Guo et al., 2015). Interestingly, CTCF binding sites are often mutated in non-coding regions of cancer genomes (Katainen et al., 2015), and abrogation of binding via mutations has recently been shown to activate T-cell acute lymphoblastic leukemia proto-oncogenes (Hnisz et al., 2016).

1.2.3 DNA methylation and CTCF binding sites

Interestingly the CTCF binding site contains CpGs within the DNA consensus sequence, suggesting that DNA methylation has a role to play in regulating CTCF occupancy at specific genes. Wang *et al* have demonstrated CTCF binding is linked to differential DNA methylation across 19 diverse human

cell types (Wang et al., 2012). Comparing bisulphite-sequencing results their data showed 41% of cell specific CTCF binding is linked to differential DNA methylation, concentrated at two specific sites in the CTCF binding sequence. They also show that CTCF binding is disrupted in immortal cell lines, which also correlates with increased methylation (Wang et al., 2012). Methylation at a CpG in the ICR has been shown to reduce binding and modulate expression at the *Igf2*/*H19* locus (Bell and Felsenfeld, 2000, Hark et al., 2000, Kanduri et al., 2000). Most recently, researchers have identified a mutation in the isocitrate dehydrogenase gene in certain gliomas, which causes inhibition of demethylation pathways. Subsequently, loss of CTCF binding and its accompanying insulator function leads to an interaction between a distant enhancer and the glioma-associated oncogene PDGFRA (Flavahan et al., 2016). Whilst these results indicate a strong link between DNA methylation and CTCF occupancy and function, a more recent study has demonstrated that repression of CTCF binding is not significantly controlled by DNA methylation (Maurano et al., 2015).

Somatic missense mutations of amino acids at specific positions in the zinc finger domain have been identified in prostate, breast, and Wilms' tumours. These tumour-derived mutations abrogate CTCF binding to the *Igf2*/*H19* ICR, thus reducing CTCF interaction with these growth regulatory gene promoters (Filippova et al., 2002). These findings suggest that CTCF may represent a

novel tumour suppressor gene, which exhibits a tumour specific 'change of function' characteristic (Filippova et al., 2002).

1.2.4 CTCF as a chromatin insulator protein

Genome wide studies have shown that a significant number of CTCF binding sites localize to boundaries between active and inactive chromatin, suggesting a role for CTCF in the barrier activity of insulators (Cuddapah et al., 2009). These regions of CTCF binding are generally characterized by active marks such as H3K4me3 and H2A.Z. Repressed chromatin is associated with increased H3K27me3 and lamin, whereas the active regions have methylated H3K4 and acetylated H2AK5 (Barski et al., 2007a). The initial discovery of CTCF as an enhancer-blocking insulator came during work with transgene assays and the 5'HS4 insulator element located at the 5' end of the chicken beta-globin locus (Chung et al., 1993).

1.2.5 Insulation via long-range interactions

The formation of loop domains is an important process required for transcriptional control of genes in eukaryotes, as it facilitates or prevents the interaction of enhancers and promoters that are separated on chromosomes. Several groups have demonstrated that CTCF can mediate long-range interactions, and proposed this as a mechanism for its insulating functions (Kurukuti et al., 2006, Ling et al., 2006). A role for CTCF in long-range interactions has previously been demonstrated at the mouse beta-globin locus

in embryonic erythroid progenitor cells and differentiated erythroid cells (reviewed by (Holwerda and de Laat, 2013). Further analysis revealed the presence of three CTCF binding sites upstream of the beta-globin locus and one downstream site (Farrell et al., 2002). Subsequent chromosome conformation capture (3C) experiments showed that in the embryonic erythroid progenitor cells, where the globin gene is not expressed, all of the CTCF binding sites were in contact with each other and were maintaining a compact domain structure (Palstra et al., 2003, Splinter et al., 2006). On the other hand, globin genes were expressed in the differentiated erythroid cells, where distant CTCF control regions clustered to form an active chromatin hub with the LCR and the globin genes. The functional outcomes of such long-range interactions may also depend on the DNA sequences adjacent to specific DNA binding sites but also the involvement additional chromatin proteins (Holwerda and de Laat, 2013). During V(D)J recombination of B cell immunoglobulin and T cell receptor gene segments, looping between distant CTCF binding sites may facilitate the joining of distant gene segments (Ong and Corces, 2014).

The role of CTCF has also been widely studied at the mouse imprinted *Igf*2/*H19* locus, where there are known to be four CTCF binding sites located in the imprinted control region (ICR), which is positioned between the *Igf*2 gene and the downstream *H19* gene and enhancers (reviewed by (Phillips and

Corces, 2009). On the maternally inherited allele CTCF occupies the ICR binding sites, which prevents downstream enhancers from activating *Igf2* expression, but does allows *H19* activation. However, on the paternal allele the ICR is methylated, which prevents CTCF binding. This allows the downstream enhancers to activate the *Igf2* gene, whereas *H19* expression is now repressed due to the promoter methylation in the ICR (Bell and Felsenfeld, 2000, Hark et al., 2000, Kanduri et al., 2000, Szabo et al., 2000). Further work has demonstrated that CTCF functions as more than just an insulator protein and is actively involved in regulating allelic gene expression. Li *et al* have demonstrated that CTCF bound on the ICR of the maternal allele forms an intrachromosomal loop through CTCF dimerization (90kb). The bound CTCF then recruits the PRC2 complex via SUZ12, which mediates H3K27me3 deposition and inactivation of the *Igf2* promoter (Li et al., 2008). Overall, these findings strongly indicate that CTCF-mediated loops around the maternal *Igf2* gene can confer transcriptional silencing and prevent enhancer association with the promoter region.

Recent investigations have demonstrated that CTCF binding can actually promote the interactions between regulatory sequences and promoters, which contradicts the previously observed enhancer blocking function. Using chromosome conformation capture carbon copy (5C) technology one study determined that 79% of long range elements and promoters are not blocked

by CTCF bound sites, with some actually exhibiting active enhancers (Sanyal et al., 2012). It is clear from this study that CTCF has a broader role to play in regulating genome function.

Indeed, the CTCF protein is now referred to as an architectural protein instead of solely an insulator protein, due to its genome wide role in mediating chromosomal interactions between different topological domains in order to regulate genomic processes (Gomez-Diaz and Corces, 2014). The human genome is divided in to approximately 2,000 topically associated domains (TADs) and it is believed that this partitioning and organization is directly linked to the function of the genome (Dixon et al., 2012). One study has demonstrated that 85% of CTCF binding sites are actually located within TADs (Intra), with only 15% positioned at TAD boundaries (Inter) (Dixon et al., 2012). The enrichment of CTCF at TAD boundaries is associated with insulator function, which prevents the aberrant spread of heterochromatin between TADs. Zuin *et al* demonstrated that siRNA-mediated depletion of CTCF reduced intra-TAD interactions, whilst increasing the inter-TAD interactions and altering the overall chromatin topology, thus highlighting that CTCF is indeed necessary to maintain topological domain boundaries (Zuin et al., 2014). The ability of CTCF to bind between TADs or within TADs may explain how it is capable of mediating both enhancer blocking and enhancer promoting functions (reviewed by (Ong and Corces, 2014).

Handoko *et al* used ChIA-PET to identify chromatin loops associated with bound CTCF protein in the murine embryonic stem cells. Here they demonstrated that around 10% of CTCF binding sites were involved in chromatin loop formation. Furthermore, these CTCF mediated loops were separated in to distinct domains each characterized by different histone modifications and gene expression patterns, which revealed both active and silenced gene clusters (Handoko et al., 2011).

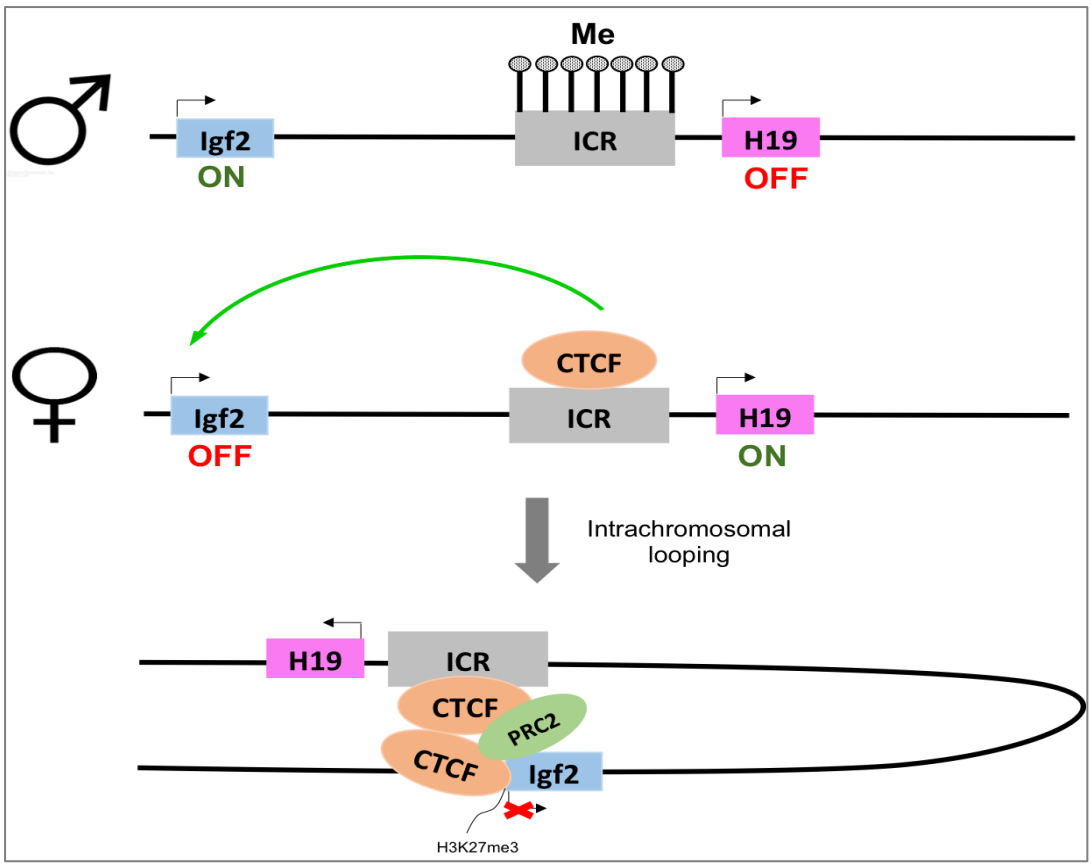


Figure 12. The role of CTCF protein at the imprinted *Igf2/H19* locus. On the maternally inherited allele CTCF binds to the unmethylated ICR and the *Igf2* promoters. CTCF forms a long-range intrachromosomal loop through CTCF dimerization. Bound CTCF recruits the PRC2 complex via SUZ12, which mediates methylation of lysine 27 on histone 3 (H3K27me). On the paternally inherited allele the ICR is methylated, which prevents CTCF binding. This abrogates loop formation and recruitment of the PRC2 complex, leading to activation of the *Igf2* promoters. Figure adapted from (Li et al., 2008)

1.2.6 CTCF binding partners

The diverse biological functions of CTCF may reflect its ability to bind a variety of protein partners. A number of potential CTCF binding partners have been identified, including transcription factors, regulatory proteins, and histones.

1.2.6.1 Poly ADP ribose polymerase (PARP)

Poly(ADP)-ribosylation is another modification of CTCF that is important for insulator function (Yu et al., 2004). The role of Poly(ADP)-ribose polymerases (PARPs) is to catalyse the formation of poly(ADP-ribose) chains (PARs) on to chromatin proteins. The formation of polymer chains introduces a negative charge on proteins, which affects their interactions with other proteins or DNA. Yu *et al* demonstrated that CTCF itself undergoes poly(ADP)-ribosyl)ation on its N-terminus. Further experiments using inhibition of the PARP pathway caused a loss of control of H19 gene imprinting by CTCF, thus illustrating PARylated CTCF is required for imprinting control (Yu et al., 2004). Guastafierro *et al* first demonstrated the role of CTCF in the cross talk between poly(ADP-ribosyl)ation and DNA methylation (Guastafierro et al., 2008). Previous experiments already demonstrated *in vivo* DNA hypermethylation when PARP activity is inhibited. It has been now elucidated that CTCF can bind to both PARP1 and DNAmethyltransferase 1 (DNMT1), resulting in PARP1-mediated poly-(ADP)-ribosylation of DNMT1, and subsequent hypomethylation (Zampieri et al., 2012). This suggests a mechanism whereby PARP1 activation and binding at CTCF target sites preserves their methylation free status (Zampieri et al., 2012).

1.2.6.2 RNA polymerase II

CTCF has been shown to co-localize in the nucleus via its C-terminus with the large subunit of RNA polymerase II, which functions to transcribe all protein-coding genes and non-coding regulatory RNAs (Chernukhin et al., 2007). It is possible that the CTCF and RNA polymerase II interaction is involved in transcriptional activation. Chernukin *et al* demonstrated that a single CTCF binding site was sufficient enough to activate an N-myc luciferase reporter gene. They also showed that CTCF sites are required for RNA polymerase II association at the H19 ICR (Chernukhin et al., 2007). A genome-wide analysis study of CTCF and RNA polymerase II interactions indicates that CTCF may recruit the large subunit to certain subpopulations of CTCF binding sites in the genome, which may confer diverse transcriptional regulation (Chernukhin et al., 2007). Another group has demonstrated the role of CTCF in pre-mRNA splicing. Here, Shukla *et al* demonstrated that CTCF binding to the CD45 exon 5 could promote inclusion of weak upstream exons via CTCF mediated RNA polymerase II pausing. However DNA methylation can inhibit CTCF binding to exon 5, thus preventing inclusion of weak upstream exons (Shukla et al., 2011). This has led to the speculation that methylation-dependent CTCF binding may contribute to tissue specific alternative splicing.

1.2.6.3 Cohesin

The cohesin complex mediates cohesion between replicated sister chromatids and is required for chromosome segregation in dividing cells. CTCF has been shown to co-localize with cohesin in the human genome, with this interaction being the only one required to stabilize most CTCF chromosomal contacts. Using ChIP Wendt *et al* noticed that cohesin was enriched at several sites of CTCF binding, such as the H19 imprinting control region and the β-globin locus (Wendt et al., 2008). The interaction involves binding between the C-terminus of CTCF to the SA2 subunit of cohesin (Xiao et al., 2011). ChIP-qPCR experiments revealed an important role for CTCF in positioning cohesin on DNA, and likewise cohesin may also contribute to CTCF positioning (Wendt et al., 2008). They also found that cohesin was required for the insulator function of CTCF at the H19 ICR, and like CTCF was shown to control H19 and IGF2 transcript levels. Wendt *et al* have proposed that CTCF may in fact function to define binding sites for cohesin, and cohesin itself is the orchestrator of insulator and boundary effects (Wendt et al., 2008). Furthermore, cohesion was also considered to be the protein-binding partner required to mediate DNA looping. Indeed there is accumulating evidence that both CTCF and cohesin co-operate to facilitate DNA looping at different loci (Degner et al., 2009, Kim et al., 2011, Majumder et al., 2008).

1.2.6.4 Ying-yang 1 (YY1)

Integrative modeling has revealed that the transcription factor YY1 is often enriched with CTCF protein at chromatin TADs (Moore et al., 2015). Furthermore, binding of YY1 to the c-fos promoter has been shown to bend DNA (Natesan and Gilman, 1993), so it was postulated that YY1 co-operates with CTCF in DNA loop formation (Xu and Corces, 2016). There is now evidence that YY1 directly interacts with the N-terminus of the CTCF protein (Donohoe et al., 2007, Pena-Hernandez et al., 2015). Clusters of YY1 and CTCF binding sites were identified in the Tsix domain within the X chromosome inactivation centre. Furthermore, YY1 and CTCF were shown to bind via discrete protein-protein interactions at this Tsix domain, forming a necessary complex for X chromosome inactivation (Donohoe et al., 2007). Most recently, CTCF and YY1 have been implicated in co-ordinating 3D organization of chromatin during neural lineage commitment. In the neural progenitor cells YY1 binding mediated enhancer looping interactions, which occurred next to constitutively bound CTCF binding sites (Beagan et al., 2017).

1.2.7 The role of CTCF in other virus life cycles

Previous research groups have demonstrated that CTCF plays an integral role in the life cycles of several other DNA viruses: Epstein Barr virus (EBV), Kaposis sarcoma associated herpes virus (KSHV), Herpesvirus saimiri (HVS) and Herpes simplex virus 1 (HSV-1).

1.2.7.1 Human papillomavirus (HPV)

The role of CTCF and the cohesion protein SMC1 has been investigated in the context of the high-risk HPV31 viral life cycle. Previously, the Laimins group has already demonstrated that HPV31 can induce the ATM DNA damage pathway (Moody and Laimins, 2009b), which activates three further pathways: p53/p21, CHK2/CDC25 and NBS1/SMC1. Mehta *et al* demonstrated that pSMC1 and CTCF co-localise to distinct nuclear foci in HPV31 containing CIN612 cells after calcium induced differentiation; however, these pSMC1/CTCF foci were absent in untransfected HFKs. ChIP experiments revealed significant enrichment of SMC1 to the L2 region of HPV31 containing CIN612 cells in monolayer culture and upon differentiation. Mutation of three CTCF binding motifs in the L2 region of the HPV31 genome resulted in loss of both CTCF and SMC1 binding to the L2 gene region. Mutations of the CTCF binding motifs in the L2 region of HPV31 or shRNA knockdown of CTCF protein caused inhibition of HPV31 genome amplification upon methylcellulose differentiation, and loss of episome maintenance in monolayer culture. Overall these findings suggest that CTCF and pSMC1 recruitment to the L2 region of HPV31 is important for viral genome amplification and viral episome maintenance (Mehta et al., 2015).

1.2.7.2 Epstein Barr Virus (EBV)

EBV is a member of the gamma herpesvirus family of DNA viruses and infects approximately 90% of the human population. The 172kb virus targets B-lymphocytes to establish a long-term latent infection and is maintained as chromatin-associated viral episomes. The viral latency types are associated with different malignancies. Burkitt's lymphoma is associated with type I latency characterized by expression of EBNA1 gene expression (Rowe M 1987). Nasopharyngeal carcinoma is associated with EBV type II latency, with EBNA1 and the additional expression of latent membrane protein 1 (LMP1) and latent membrane protein 2A (LMP2A). Lymphoproliferative diseases are type III associated and express several latent proteins, including Epstein-Barr nuclear antigen 2 (EBNA2) (Young L 1989).

Chau *et al* first indicated that EBV latency type is regulated by CTCF. ChIP experiments revealed that CTCF could bind to the EBV genome between the origin of replication (OriP) enhancer and the C promoter (Cp), which is active in type III but inactive in type I latency. Interestingly, more CTCF was bound to this region in the type I cells than the type III and the overall CTCF protein and mRNA levels were higher in type I cells. Furthermore, upon deletion of the CTCF binding site between OriP and the C promoter there was a 3.5 fold elevation in EBNA2 mRNA levels. The siRNA depletion of CTCF mRNA also caused an increase in EBNA2 mRNA levels in type I cells. Given that the OriP

enhancer activates the C promoter it was hypothesized that binding of CTCF between these two regions in the type I cells may be blocking the enhancer. This provided the first evidence that CTCF was acting as a transcriptional repressor of type III latency and may be critical binding factor in determining the type of EBV latency (Chau et al., 2006).

CTCF binding sites have also been identified at promoter regions upstream of Cp, Wp, EBERS, and Qp (Tempera et al., 2010). During type I latency the Q promoter (Qp) is involved in the maintenance of viral episomes, so it was hypothesized that CTCF binding upstream of this promoter may have a regulatory function. Indeed, when this CTCF binding site was mutated it resulted in a loss of viral episome maintenance. Mutation also caused a reduction in EBNA1 transcripts-which are normally expressed from Qp in type I latency-indicating that loss of CTCF may deregulate promoter usage. Indeed, 16 weeks after transfection, the cells containing the mutated CTCF site exhibited a loss of Qp usage but an increase in Fp usage. Mechanistically, Tempera *et al* were able to demonstrate the boundary function of CTCF in the EBV genome, as binding of CTCF was shown to prevent the spread of CpG methylation; however, the loss of CTCF resulted in the spread of methylation to silence Qp promoter usage.

Tempera *et al* have also demonstrated the role of CTCF in mediating distinct chromatin loops in different EBV latency types (Tempera et al., 2011). Here they demonstrated that in type I latency cells, where Qp is active, there was a chromatin loop formed between Qp and OriP. Furthermore, in type III cells, where Cp is active, there was a chromatin loop formed between Cp and OriP. Overall, these mutually exclusive chromatin loops correlated with the transcriptional activity of the promoters. Combination of 3C and ChIP (3C-ChIP) provided further evidence that these chromatin loops were mediated by CTCF protein binding upstream of either Cp or Qp. Moreover, mutation studies and siRNA depletion of CTCF also confirmed that CTCF was important for mediating these long-range interactions (Tempera et al., 2011).

More recently, a CTCF site has been confirmed in the overlapping 3' region of LMP1 and the first intron of LMP2A, which was also co-occupied by cohesin. Here abrogation of CTCF binding caused a decrease in LMP1 and LMP2A transcript levels concomitant with an increase in the LMP2B transcript. The abrogation in CTCF binding was shown to cause a loss in loop formation between OriP and the LMP1 and LMP2A regions, resulting in repression of these promoter regions, and an increase in LMP2B expression (Chen et al., 2013a).

1.2.7.3 Kaposis Sarcoma virus (KSHV)

KSHV is a gamma herpes virus, which establishes long-term latent infection in B-lymphocytes and is associated with the Kaposi sarcoma, often seen in untreated HIV and immunosuppressed individuals. KSHV persistence depends on the balance between the latent and lytic phase of the life cycle. Latent KSHV exists episomally and encodes latency associated nuclear antigen (LANA/ORF73), viral cyclin (vCyclin/ORF72), vFLIP (ORF71), kaposin (K12), and microRNAs which are needed for viral maintenance (Wen and Damania, 2010).

Several CTCF binding sites were identified within the KSHV genome using ChIP. At one particular site between the ORF73 and K14 within the first intron of the major latency transcript, cohesion was found to co-localize with CTCF binding (Stedman et al., 2008). Deletion of the CTCF binding site caused disruption in cohesin binding, and when Rad21 was depleted by siRNA an induction in lytic genes was observed. These experiments provided strong evidence for the co-operative binding of CTCF and cohesion in regulating viral gene expression, specifically repression of the lytic genes. Another study demonstrated that the binding of CTCF and cohesion to the major latency transcript was cell cycle regulated, with peak binding of cohesion observed during mid S phase (Kang and Lieberman, 2009).

Taking this further, Kang *et al* used 3C to demonstrate a role for CTCF in mediating DNA loop formation in the KSHV genome to regulate viral transcription. Here they identified two separate DNA loops; one loop formed between the CTCF binding site and ORF50 (60kb apart) and another between the CTCF site and K12 3' end. When the three CTCF binding site clusters in the latency control region were mutated there was a 2-3 fold reduction in loop formation between the ORF50 region, as well as a 2-3 fold reduction in ORF50 lytic mRNA and derepression of some latent genes (Kang et al., 2011). These findings reveal that the KSHV genome forms DNA loops, which are mediated by CTCF and cohesion binding and are necessary to co-ordinate latent and lytic gene expression.

Most recently Kang *et al* have proposed that nucleosome positioning and RNA polymerase II recruitment are fundamental for the CTCF mediated regulation of KSHV latency. Their experiments demonstrated that abrogation of CTCF binding resulted in aberrant gene splicing with an observed increase in LANA mRNA. They also showed that CTCF abrogation could alter RNA polymerase II recruitment to the genome, with an observed increase in RNA polymerase II recruitment in the absence of CTCF. Furthermore, CTCF was shown to be involved in nucleosome displacement across the KSHV genome. Overall they proposed that CTCF binding to the KSHV genome can displace

nucleosomes, which can modulate RNA polymerase II activity and thus alter mRNA processing (Kang et al., 2013).

1.2.7.4 Herpes Simplex Virus (HSV-1)

HSV-1 is an alpha herpesvirus that infects sensory neurons where it is establishes a life-long latent infection. Within the sensory neurons HSV-1 is maintained episomally and associates with histones. During latent infection, lytic gene expression is repressed, whereas latency associated transcript (LAT) is fully expressed. Previous work has identified seven CTCF binding clusters in the HSV-1 genome, which interestingly surround the intermediate/early (IE) genes and flank the reactivation region of the LAT (Amelio et al., 2006). During latent infection of murine dorsal root ganglia, ChIP experiments revealed CTCF enrichment at the LAT site. Further work demonstrated that CTCF has transcriptional silencing activity and is capable of blocking the LAT enhancer from acting on adjacent promoters (Amelio et al., 2006). Most recently, during lytic activation, CTCF binding sites have been identified in the HSV-1 genome, which differ to those detected during viral latency. CTCF was shown to support HSV-1 transcription by promoting the binding of RNA polymerase II and preventing the spread of repressive histone marks H3K27me3 and H3K9me3 (Lang et al., 2017).

1.2.7.5 Human Cytomegalovirus (HCMV)

HCMV is a human beta-herpesvirus that expresses immediate early, early and late genes. Martinez *et al* have recently investigated the role of CTCF in the regulation of HCMV gene expression. Depletion of CTCF caused an increase in the major immediate early (MIE) and early gene expression and a subsequent 50-fold increase in virion production. Further work identified a CTCF binding site in the first intron of this MIE gene. They have proposed that CTCF binding in the first intron can block RNA polymerase II elongation during the initial stages of transcription (Martinez et al., 2014).

1.2.7.6 Herpesvirus Saimiri (HVS)

HVS is a gamma-2 herpesvirus that establishes a latent infection in T-lymphocyte cells, where it is maintained as histone associated episomes in the host nuclei. Subgroup C HVS strains are the most oncogenic and are capable of transforming T lymphocyte cells (Zielke et al., 2012). The expression of most viral genes is restricted during latency, with the exception of orf73/LANA, which is involved in suppression of the lytic cascade (Schafer et al., 2003), (Zielke et al., 2012). Recently, two strong and conserved CTCF binding sites were identified in the intergenic region of the orf73/LANA promoter region (Zielke et al., 2012). Latently infected T cells harbouring the CTCF binding site mutations showed impaired proliferation and significantly reduced episomal maintenance, along with reduced orf73/LANA gene

expression. These results suggest that HVS episomal stability is dependent on CTCF (Zielke et al., 2012).

1.2.7.7 Adenovirus

Compared to the herpesviruses the adenovirus has a relatively small genome of around 36kb. Several CTCF binding sites have been identified in the adenovirus genome, with CTCF binding shown to be dependent on viral DNA replication. Knockdown of CTCF protein resulted in suppression of both DNA replication and late gene expression (Komatsu et al., 2013).

1.3 Role of chromatin during transcription

Chromatin is comprised of nucleosome units, which themselves are made up of a protein core of four duplicated histones: H2A, H2B, H3 and H4. These eight histones are wrapped around by 147bp of DNA and a linker DNA region separates each nucleosome unit. Whilst chromatin is essential for DNA compaction and packaging it also has an important role in regulating gene expression.

Each histone contains an N-terminal tail that can undergo a vast number of post-translational modifications, which can include methylation, acetylation, phosphorylation and ubiquitination (Kouzarides, 2007). Depending on the type of modification and the distribution pattern, each can confer different transcriptional outcomes. Active chromatin regions are characterized by the

enrichment of several histone modifications. Of note, acetylation of H3 and H4 is found at active promoters regions as well as the active gene bodies, H3K4me3 is found at active promoters, H3K27ac and H3K4me1 are enriched at active enhancers and H3K36me3 is enriched towards the 3′ end of actively transcribed genes (Barski et al., 2007a). Histone lysine acetylation is highly correlated with active transcription, as it is thought that acetylation of lysine residues can physically alter chromatin structure and disrupt nucleosomal interactions, which leads to the open chromatin conformation associated with transcriptional activation.

H3K4me3 is a hallmark of active chromatin and is enriched at promoter regions and TSS of active genes (Santos-Rosa et al., 2002, Schneider et al., 2004) whereas H3K4me1 is enriched at 3′ ends and H3K4me2 is abundant across the middle of gene bodies. The methylation of H3K4 is catalyzed by the SET1 homologs-SET1A, SET1B, MLL1, MLL2, MLL3 and MLL4. The SET1 and MLL complexes can also recognize and bind to the H3K4me3 and positively reinforce the maintenance of this epigenetic mark (Murton et al., 2010, Shi and Whetstine, 2007). It is proposed that the main role of H3K4me3 deposition is to define the start of transcription, and to initiate the recruitment of the ATP-dependent nucleosome-remodelling complex, NURF (Li et al., 2006, Li et al., 2007) and other chromatin modifying proteins. The PHD finger domain of certain chromatin modifiers, such as HAT complexes can also recognize and

bind to the H3K4me3 mark and promote acetylation of H3 and H4 (Bian et al., 2011). Furthermore, these PHD domains can also bind to HDACs, indicating that the H3K4me3 can dynamically regulate transcription through the recruitment of both of HATs and HDACs (Doyon et al., 2006). Di- and tri-methylation of H3 at lysine 36 is enriched at the 3' region of genes, and only the H3K36me3 mark is positively correlated with active transcription (Barski et al., 2007a, Pokholok et al., 2005). In yeast, H3K36me2/3 can recruit HDACs to actively transcribed regions to initiate deacetylation and this is thought to prevent aberrant transcriptional activation within gene bodies (Carrozza et al., 2005, Keogh et al., 2005, Joshi and Struhl, 2005). Enrichment of the active H3K4me3 and repressive H3K27me3 across genes is not mutually exclusive, as around 15% of gene promoters are enriched in both of these epigenetic modifications, thus possess a bivalent function (Mikkelsen et al., 2007).

The methylation of histone residues is tightly linked to the transcriptional elongation mediated by RNA polymerase II. General transcription factors (GTFs) bind to the initiator or 'TATA' sequences of protein coding genes and then recruit RNA polymerase II to promote transcription. The C-terminus of RNA polymerase II is referred to as the carboxyl-terminal domain (CTD) and contains the hepatapeptide consensus motif Tyr-Ser-Pro-Thr-Ser-Pro-Ser, which is repeated 52 times (Cramer et al., 2001). The differential phosphorylation patterns of the CTD can result in the recruitment of different

effector proteins involved in pre-mRNA capping, splicing or polyadenylation (Fong and Bentley, 2001). Phosphorylation of serine 5 (Ser5) in the CTD is localized to promoter regions, whereas serine 2 phosphorylation (Ser2) is enriched at coding regions and increases towards the 3′ end of genes (Komarnitsky et al., 2000).

In yeast, the SET1 methyltransferase protein has been shown to interact with the polymerase associating factor (PAF) complex as well as RNA polymersae II phosphorylated at serine 5, indicating that the H3K4me3 modification is deposited co-transcriptionally (Ng et al., 2003). Furthermore the SET2 protein, which catalyzes H3K36 methylation, was shown to associate with serine-2 phosphorylated form of RNA Polymerase II, and again this is regulated by the PAF complex (Krogan et al., 2003, Li et al., 2002).

On the other hand, histone modifications such as H3K9me3, H3K27me3 and H2AK199u1 are well known hallmarks of repressive chromatin (Barski et al., 2007b). The Polycomb repressive complexes, PRC1 and PRC2, mediate the deposition of these repressive histone modifications. The polycomb group genes (PcG) were first discovered in *Drosophila melanogaster*, where they were shown to be required for stable repression and maintenance of HOX genes (Paro, 1995, Pirrotta, 1998). Mammalian counterparts to the *Drosophila* PcG genes were subsequently identified.

1.3.1 Polycomb repressive complex

1.3.1.1 Enhancer of Zeste homologue 2

Polycomb repressive complex 2 (PRC2) is a mult-subunit complex with a well-established role in mediating the methylation of lysine 27 on H3, which is a known marker of repressive chromatin and gene silencing. The enhancer of zeste homologue 2 (EZH2) protein is an essential subunit of the PRC2 complex as it is the only known methyltransferase responsible for catalyzing the methylation of H3 at lysine 27 (Cao et al., 2002, Czermin et al., 2002, Kuzmichev et al., 2002, Muller et al., 2002). EZH2 contains a C-terminal SET domain that is responsible for transferring three successive methyl groups to the H3 histone tail to form H3K27me3 (Shi et al., 2006)

Emerging evidence has highlighted the role of EZH2 phosphorylation on modulating its function. Cyclin dependent kinase phosphorylation of EZH2 at T350 was shown to increase EZH2 association with chromatin and thereby augment H3K27me3 deposition (Chen et al., 2010). Conversely, EZH2 phosphorylation at T492 was shown to disrupt binding to the co-factors EED and SUZ12, and thus negatively impact PRC2 function (Wei et al., 2011, Paro, 1995). The Akt signaling pathway has also been shown to phosphorylate EZH2 at serine-21, which subsequently caused a reduction in PRC2 mediated histone methylation (Cha et al., 2005). Furthermore, by using a natural killer/T cell lymphoma cell model, EZH2 was phosphorylated by JAK3 at residue

Y244, which disrupted the association of EZH2 with PRC2 complex. This phosphorylation event demonstrated that EZH2 was switched from mediating gene repression to mediating transcriptional activation and enhancing NK cell proliferation (Yan et al., 2016).

1.3.1.2 Embryonic ectoderm development

The embryonic ectoderm development (EED) protein is another key member of the PRC2 complex. EED has been shown to bind EZH2 via its N-terminus (Han et al., 2007), and a central pocket domain that specifically binds to trimethyalted lysine residues such as H3K27me3, which is thought to propagate this repressive mark within chromatin (Margueron et al., 2009, Xu et al., 2010). The methyltransferase activity of EZH2 is dependent on the interaction between EZH2 and the WD40 domain of EED (Cao et al., 2014). EED has also been shown to directly interact with the PRC1 complex. It is proposed that the PRC2 EED protein recruits PRC1 to areas of H3K27me3 within chromatin, and enhances the PRC1 RING1B-mediated ubiquitination of histone H2A (Cao et al., 2014).

1.3.1.3 Suppressor of zeste 12 protein homologue

Suppressor of zeste 12 protein homolog (SUZ12) binds via its C-terminal domain to EZH2 and this interaction was shown to be important for the enzymatic activity of EZH2 and the silencing function of the PRC2 complex (Pasini et al., 2004). SUZ12 also mediates interactions between the PRC2

complex and co-factor proteins such as Jarid2 (Peng et al., 2009) and retinoblastoma protein (Rb) associated protein 48 (RpAp48).

1.3.2 Polycomb repressive complex 1

The most commonly understood role for the polycomb repressive complex 1 (PRC1) complex is to recognize the H3K27me3 deposited by the PRC2 complex (Cao et al., 2002; Min et al., 2003; Wang et al., 2004b). The PRC1 complex contains a core subset of proteins, including: B lymphoma Mo-MLV insertion region 1 (BMI1), RING1A (also known as RING1) and RING1B (also known as RING2 or RNF2). RING1B is an E3 ligase that has been shown to catalyzes monoubiquitination of H2A on lysine 119 (H2AK119ub), which is associated with compact and silenced chromatin (Wang et al., 2004a). However emerging evidence has revealed that RING1B can occupy target sites in the absence of the repressive mark H3K27me3, indicating that in some scenarios PRC1 recruitment is not always preceded by PRC2 binding (Schoeftner et al., 2006, Tavares et al., 2012).

1.3.3 Polycomb recruitment to chromatin

Studies in Drosophila have demonstrated that the DNA-binding protein, Pho, interacts with specific DNA sequences and recruits polycomb proteins (PcGs) to mediate transcriptional repression (Brown et al., 1998, Fritsch et al., 1999). Further work concluded that the mammalian YY1 protein is a homolog of Pho (Atchison et al., 2003, Srinivasan and Atchison, 2004), and can bind DNA to

recruit polycomb complexes and mediate transcriptional repression (Wilkinson et al., 2006). A more recent study has demonstrated that knockdown of YY1 protein in HeLa cells caused a reduction in polycomb recruitment to DNA (Basu et al., 2014a).

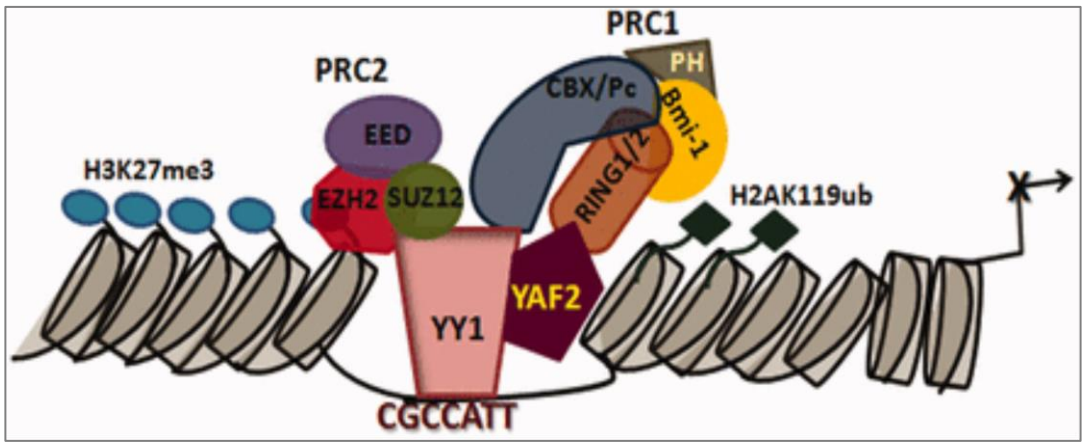


Figure 13. Proposed model for YY1-mediated recruitment of PcG to DNA. YY1 binds to DNA and interacts with EZH2 and SUZ12, thereby recruiting the PRC2 complex to mediate trimethylation of lysine 27 on histone H3. YY1 can also recruit the PRC1 complex, which leads to ubiquitination of H2A on lysine 119. Image taken from (Basu et al., 2014b).

1.3.4 Histone modifications within the HPV genome

As previously mentioned, the virus has a double stranded DNA genome, and it is well known that the viral genome is chromatinized within the host cell and packaged with histones. One group has demonstrated the association of histone modifications with the HPV31 genome in naturally derived HPV31 genome containing CIN612 cells. Here, the acetylated H3 and H4 and dimethylated H3K4 histone marks that are associated with active

transcription were identified across the early and late viral promoters in monolayer culture, and upon differentiation the association of these factors was further enriched (Wooldridge and Laimins, 2008). However the resolution of these studies was limited, as the association of histone marks was not investigated throughout the entire HPV31 genome.

1.4 Hypothesis and Aims

Hypothesis: The HPV18-CTCF interaction is required for the regulatory control of early and late gene expression and the overall completion of the productive viral life cycle.

Aims:

1) Use chromatin immunoprecipitation (ChIP) to confirm if CTCF binds to the E2 ORF in high-risk HPV16 and HPV18 genomes in primary HFKs and naturally infected cervical cells and determine binding dynamics upon cellular differentiation.
2) Determine if CTCF binding is reduced or abrogated in a HFK cell line containing a mutation in the E2 ORF of the HPV18 genome (ΔCTCF).
3) Use the ΔCTCF HFK cell line to determine the importance of CTCF binding to the E2 ORF for early and late viral gene and protein expression and viral genome replication, maintenance and amplification.

4) Determine if CTCF binding to the E2 ORF is required for nucleosome positioning and maintenance of the epigenetic landscape throughout the viral genome.
5) Determine if abrogation of CTCF binding to the E2 ORF affects cellular transcription factor recruitment to the viral LCR region.
6) Determine if CTCF binding to the E2 ORF is required to regulate the 3D chromatin architecture of the viral genome.

Chapter 2: Materials and Methods

2 Chapter 2: Material and Methods

2.1 List of Buffers

Cell culture

HFK E-medium

1200 ml	DMEM (6171)
640 ml	Hams F12 media
20 ml	20 x Cocktail**
40 ml	Pen strep
2 ml	Hydrocortisone (5 mg/mL)
2 ml	Cholera toxin (10 μg/mL)
100 ml	Hyclone FCS

****20 x Cocktail**

20 ml	Adenine (0.18 M)
20 ml	Insulin (5 mg/ml, 0.1 M HCl)
20 ml	Transferrin (5 mg/ml)
20 ml	Tri iodo –lL- thyronine ($2x10^{-8}$ M)
120 ml	PBS

Complete E-medium (keratinocyte culture medium)

500 ml	E-medium*
2 ml	EGF (1 μg/mL)

Organotypic Raft medium (No EGF)

500 ml	E medium*
10 ml	L-glutamine

Methylcellulose (1.5%)

3.75g	Methylcellulose
12.5ml	Hyclone FCS
237.5ml	E medium

Add 125ml E Medium (without EGF). Incubate at 60°C for 20 minutes, swirling every 5 minutes. Cool on ice then add a further 112.5ml E Medium. Stir O/N at 4°C then add 12.5ml Hyclone FCS. Autoclave and store at 4°C for up to 6 weeks.

J2 cryopreservation medium

8 ml	J2 medium
1ml	DMSO
1ml	Bovine Serum

Keratinocyte cryopreservation medium

16 ml	Serum free medium
2 ml	Hyclone FCS
2ml	DMSO

<u>WesternBlot</u>

Urea lysis buffer

54.54 g	Urea
0.876 g	NaCl
5 ml	0.5M Tris
700 µl	14.3M β-Mercaptoethanol

Make up to 100 ml with ddH_2O

6x SDS gel-loading buffer

3ml	1M Tris pH 6.8
1.2 g	SDS pellets
6 ml	Glycerol
0.93 g	DTT
6 mg	Bromophenol blue

Dissolve SDS pellets in Tris at 37°C, cool to RT and add glycerol, then add DTT and bromophenol blue

12% Resolving gel

3.3 ml	H_2O
2.5 ml	Tris(1.5M, pH8.8)
4 ml	Acrylamide-bis (30%)
100 µl	10% SDS
100 µl	10% APS
6 µl	TEMED (Add just before use)

Stacking gel

2.05 ml	H_2O
375 µl	Tris(0.5M, pH6.8)
500 µl	Acrylamide-bis (30%)
30 µl	10% SDS
30 µl	10% APS
3 µl	TEMED (Add just before use)

10x SDS-page running buffer

10 g	SDS pellets
30 g	Tris
144 g	Glycine

Make up to 1 litre with ddH_2O

10x Semi-Dry transfer buffer (pH 8.3)

5.82 g	Tris
2.9 g	Glycine
3.7ml	10% SDS
200ml	Methanol

Make up to 1 litre with ddH_2O

10x TBST

24.2 g	Tris
80 g	NaCl

pH to 7.6, add 10 ml Tween 20 and make up to 1litre with ddH_2O

Ponceau S Stain

1 g	Ponceau S
50 ml	Acetic acid

Make up to 1 L with ddH_2O

<u>Southern Blot</u>

Hybridisation buffer

10 ml	50 X Denhardts[$]
20 ml	20 X SSC Buffer[$$]
10 g	Dextran sulphate

Make up to 50 ml with ddH_2O, and filter through Whatman paper.
Before use, take 10 ml aliquots and add 10 ml deionised formamide and 2 ml 10% SDS and mix well.

[$]**50X Denhardts**

1 ml	Ficoll 400
1 ml	Polyvinylpyrrolidone
1 ml	BSA Fraction V
93 ml	ddH_2O

[$$]**Sodium Chloride/ Sodium Citrate (SCC, 20x)**

175.3 g	NaCl
88.2 g	Tri-sodium citrate
800 ml	ddH_2O

pH 7.0 and make up to 1 L ddH_2O

<u>PCR</u>

5X TBE

27.5 g	Boric acid

54 g Tris
20 ml 0.5M EDTA (pH 8.0)
Make up to 1 litre with ddH_2O

DNA lysis buffer
11.7 g NaCl
0.6 g Tris-HCl
10 ml EDTA 0.5 M^
450 ml ddH_2O
pH 7.4 and make up to 500 ml with ddH_2O. Autoclave before use.

^0.5M EDTA
186.12 g EDTA
Stir in to 800 ml ddH_2O. Adjust pH to 8.0 and make up to 1L ddH_2O.

Sodium acetate (3M)
24.6 g Sodium acetate
100 ml ddH_2O

<u>Immunofluorescence</u>
IF blocking solution
10ml HINGS
0.05g BSA
Make up to 50 ml with 1x PBS. Filter and store at 4°C.

Table 1. Primary Antibodies

Target	Primary Antibody	Host	Cat code	Company	ChIP	IF	WB	kDa
E2	E2	Mouse		Thierry			1 in 1000	45
E4	1D11 (N-term)	Mouse		Sally		1 in 5		11
E4	R424 polyclonal	Rabbit		Sally		1 in 3000	1 in 5000	11
E6	(G-7): sc-365089	Mouse	610262	Santa Cruz			1 in 50	17
E7	Anti-HPV18 E7 antibody [8E2]	Mouse	ab100953	Abcam			1 in 1000	17
L1	HPV-18 L1 (5A3)	Mouse	NCL-HPV18	Nova Costra		1 in 100		60
BrDu	Anti-BrdU (BU1/75) ICR1	Rat	Ab6326	Abcam		1 in 500		
GAPDH	GAPDH Antibody (6C5): sc-32233	Mouse	sc-32233	Santa Cruz			1 in 5000	37
Rb2 (p130)	Purified Mouse Anti-Rb2	Mouse	610262	BD		1 in 250	1 in 1000	130
p53	p53 DO-1	Mouse		Roger Grand		1 in 50		53
Involucrin	Mouse anti-involucrin (mAb) Clone SY5	Mouse	I9018	Sigma-Aldrich			1 in 200	120
CTCF	CTCF antibody (pAb)	Rabbit	61311	Active motif	8 µl	1 in 3000	1 in 1000	120
H3K4me3	Histone H3K4me3 antibody (pAb)	Rabbit	39915	Active motif	5 µl		1 in 1000	17
H3K27me3	Histone H3K27me3 antibody (pAb)	Rabbit	39155	Active motif	5 µl		1 in 1000	17
RNAPolII	Total RNA pol II antibody (mAb) Clone: 4F8	Rat	61081	Active motif	5 µl		1 in 1000	240
YY1	YY1 antibody (pAb)	Rabbit	39071	Active motif	5 µl		1 in 1000	63
EZH2	EZH2 antibody (pAb)	Rabbit	39934	Active motif			1 in 1000	96
EZH2	EZH2 antibody (pAb)	Rabbit	39901	Active motif	5 µl			
EED	EED antibody	Rabbit	GTX33168	Gene Tex			1 in 1000	50
EED	EED antibody (mAb) Clone 41D	Mouse	61203	Active motif	4 µl			57
H2AK119Ub	Ubiquityl-Histone H2A (Lys119) (D27C4) XP®	Rabbit	8240	Cell Signalling	10 µl		1 in 1000	23
RING1B	Ring1B antibody (mAb)	Mouse	39663	Active motif	5 µl		1 in 1000	40
TEF-1	TEAD1 antibody (pAb)	Rabbit	61644	Active motif	5 µl		1 in 1000	60

Table 2. Secondary Antibodies

Secondary Abs	Host	Cat code	Company	ChIP	IF
Alexa-Fluor 488 Goat Anti-Mouse Alexa-Fluor 488 (Green)	Goat	A11029	ThermoFisher	1 in 1000	
Alexa-Fluor 488 Goat Anti-Rabbit Alexa-Fluor 488 (Green)	Goat	R37116	ThermoFisher	1 in 1000	
Alexa-Fluor 594 Goat Anti-Rabbit Alexa-Fluor 594 (Red)	Goat	A11012	ThermoFisher		
Goat anti-Mouse IgG (H+L) Secondary Antibody, HRP	Goat	31430	ThermoFisher		1 in 10,000
Goat anti-Rabbit IgG (H+L) Secondary Antibody, HRP	Goat	31460	ThermoFisher		1 in 10,000
Anti-Rat IgG (whole molecule)–Peroxidase antibody produced in rabbit	Rabbit	061M4805	Sigma-Aldrich		1 in 10,000

Table 3. HPV16 ChIP primers

Primer pair	Fw (5′ – 3′)		Rev (5′ – 3′)		Annealing Temperature (°C)
HPV16 W12E					
L2	(4419F)	CAGGGTCGGGTACAGGCGGA	(4542R)	GGATCGGAAGGGCCCACAGGA	50
L2 CTCTbs	(5123F)	AGGCGTACTGGCATTAGGTACAGT	(5309R)	AGGTAAGGCTGCATGTGAAGTGGT	50
L2/L1	(5646F)	TGGCTGCCTAGAGGCCACTGT	(5667R)	TGCGTGCAACATATTCATCCGTGC	50
L1 CTCFbs	(6039F)	TGCAGCAAATGCAGGTGTGGAT	(6157R)	TGGGGATCCTTTGCCCCAGTGT	50
L1	(7045F)	ACAAGCAGGATTGAAGGCCAAACCA	(7121R)	AGAGGTAGATGAGGTGGTGGGTGT	50
5'URR-Enh	(7419F)	TTTGTAGCGCCAGCGGCCATTT	(7552R)	GCATGGCAAGCAGGAAACGTACAA	50
Enh	(7555F)	CCAAATCCCTGTTTTCCTGA	(7681R)	CGTTGGCGCATAGTGATTTA	50
Earlyprom	(7854F)	GCAAACCGTTTTGGGTTACA	(65R)	ACTAACCGGTTTCGGTTCAA	50
E6 111-223	(111F)	AGGACCCACAGGAGCGACCC	(223R)	ACGTCGCAGTAACTGTTGCTTGCA	50
E6 427-506	(427F)	GCCACTGTGTCCTGAAGAAAAGCA	(506R)	GACCGGTCCACCGACCCCTT	50
Lateprom	(649F)	GACAGCTCAGAGGAGGAGGA	(765R)	GCACAACCGAAGCGTAGAGT	50
E1 1250-1368	(1250F)	GCGAAGACAGCGGGTATGGCA	(1368R)	GCAACCACCCCCACTTCCACC	50
E1 2158-2316	(2158F)	AGGGTAGATGATGGAGGTGATTGG	(2316R)	GATTTACCTGTGTTAGCTGCACCA	50
E2 (CTCF)	(2853F)	GGAAACACATGCGCCTAGAATGTGC	(2950R)	TGATACAGCCAGTGTTGGCACC	50
E4	(3407F)	CACTCCGCCGCGACCCATAC	(3514R)	GGTGTGGCAGGGGTTTCCGG	50
E5	(3936F)	ACGTCCGCTGCTTTTGTCTGTGT	(3958R)	ACCTAAACGCAGAGGCTGCTGT	50

Table 4. HPV18 ChIP primers

Primer pair	**Fw (5′ – 3′)**		**Rev (5′ – 3′)**		Annealing Temperature (°C)
HPV18	Nt	Seq	Seq		
L2	4440	GGGGTCGTACAGGGTACATT	GATGTTATATCAAACCCAGACGTG		56
L2 CTCTbs	5381	TCTGCCTCTTCCTATAGTAATGTAACG	GGAATAAAATAATATAATGGCCACAAA		56
L1 CTCFbs	5655	CCTCCTTCTGTGGCAAGAGT	GGTCAGGTAACTGCACCCTAA		56
L1	6659	AGTCTCCTGTACCTGGGCAA	AACACCAAAGTTCCAATCCTCT		58
5'URR-Enh	7301	GTGTGTTATGTGGTTGCGCC	GGATGCTGTAAGGTGTGCAG		58
Ori	7746	ACTTTCATGTCCAACATTCTGTCT	ATGTGCTGCCCAACCTATTT		56
Earlyprom/E6	155	TGTGCACGGAACTGAACACT	CAGCATGCGGTATACTGTCTC		58
Lateprom	751	CGAACCACAACGTCACACAAT	ACGGACACACAAAGGACAGG		58
E1	1500	GCAATGTATGTAGTGGCGGC	TACACTGCTGTTGTTGCCCT		58
E2	2819	TGCAGACACCGAAGGAAACC	CATTTTCCCAACGTATTAGTTGCC		58
E2 (CTCF)	2926	GGCAACTAATACGTTGGGAAAA	TGTCTTGCAGTGTCCAATCC		56
E2	3165	AGGTGGCCAAACAGTACAAGT	GCCGTTTTGTCCCATGTTCC		58
E4	3381	TGGGAAGTACATTTTGGGAATAA	TCCACAGTGTCCAGGTCGT		56
E5	3971	TATGTGTGCTGCCATGTCCC	CTGTGGCAGGGGACGTTATT		56

Table 5. HPV18 transcript primers

Primer pair	**Fw (5′ – 3′)**		**Rev (5′ – 3′)**		Annealing Temperature (°C)
Early transcripts	Nt	Seq	Nt	Seq	
121/295	121	ATCCAACACGGCGACCCTAC	295	GCAGCATGCGGTATACTGTCTCTA	56
121/3517	121	ATCCAACACGGCGACCCTAC	3517	ACGGACACGGTGCTGGAA	56
Late transcripts	Nt	Seq	Nt	Seq	
E1F1/L1R	909	CAACAATGGCTGATCCAGAAG	5759	CACCTGCAGGAACCCTAAAA	58

Relative gene expression primer	**Fw (5′ – 3′)**		**Rev (5′ – 3′)**		Annealing Temperature
Human TLR2 locus	1893	GCCAGCAAATTACCTGTGTGA	1957	GGCGGACATCCTGAACCT	58

Table 6. 3C HPV18 primers

Primer pair	**Fw (5′ – 3′)**	**Rev (5′ – 3′)**	Annealing Temperature (°C)
3C			
CTCF RV (E2 region)		CTTGCAGTGCCAATCCTCG	58
YY1 RV (L1 region)		GGTGCAGCATCCTTTTGACA	58
L2 region (Control)		CCCCGTACCAGAAGAACAA	58
Digestion Efficiency (Primer pair 2)	GCAAACGGGCTTCGGTAACT	TTATCTGCTAACGTGGTGCCC	58

2.2 Cell Culture

All cell culture protocols were carried out in sterile conditions in a Holton Laminair S2000 1.2 microbiological safety cabinet. Cell culture medium was prepared by filter sterilizing individual components through 0.22 μm membrane filters (Merck Millipore).

2.2.1 Cell counting

Disposable cell counting slides (Immune systems, UK) were used to determine cell concentrations. Firstly, 10 μl of cell suspension was added to one chamber of the cell counter slide. Cells were counted in four of the sub-dived squares, and then an average number was calculated. This average value was then multiplied by 10^4 in order to obtain the estimated number of cells per millilitre of cell suspension.

2.2.2 Mycoplasma tests

Cells were routinely tested for mycoplasma using the MycoAlert mycoplasma detection kit (Lonza, UK).

2.2.3 3T3-J2 fibroblasts

2.2.3.1 Maintenance of 3T3-J2 fibroblasts

3T3-J2 mouse fibroblast cells were maintained at 37°C and 5% CO2 in Dulbecco's modified Eagle's media DMEM (Sigma, D6171), containing bovine serum (10% v/v, Gibco) and 200 mM L-glutamine (2% v/v, Gibco). At 90% confluency cells were washed once with PBS and 1 ml 0.25% Trypsin-EDTA

(1X) (Gibco) was added to trypsinise cells, typically taking no longer than 5 minutes. Cells were checked for trypsination under the microscope. Cells were resuspended in 3T3-J2 media and centrifuged at 200 x *g* for 5 minutes. The cell pellet was resuspended in the appropriate volume of 3T3-J2 media for the intended number of plates, and then 1 ml of suspension was added to each new cell culture plate containing 9 ml of pre-warmed 3T3-J2 media. Alternatively, J2-3T3 pellets were resuspended in HFK complete E media, to give a final concentration of $2x10^6$/ml, and then irradiated with 30Gy of a caesium 137 source. Irradiated 3T3-J2 cells were stored at 4°C and used for up to five days. The 3T3-J2 cells were not used beyond passage 20, at which point lower passage cells were thawed for use.

2.2.3.2 Cryopreservation of 3T3-J2 fibroblasts

3T3-J2 mouse fibroblast cells were frozen down in freezing media consisting of DMEM (Sigma-Aldrich, D6171) supplemented with 20% (v/v) bovine serum (Gibco) and 10% (v/v) DMSO (Sigma-Aldrich). Typically between 1 and $2x10^6$ cells were frozen per cryovial.

2.2.4 Human Foreskin Keratinocytes (HFKs)

2.2.4.1 Maintenance of human foreskin keratinocytes (HFKs)

Normal human foreskin keratinocytes denoted as 'donor 1/Clonetics' were purchased from Lonza. HFKs denoted as 'donor 2/Georgie' were isolated from neonate foreskin by Dr Sally Roberts (Ethics number- Roberts:

06/Q1702/45). HFKs were cultured on collagen-coated tissue culture plates in serum-free keratinocyte media (SFM, Invitrogen, UK), with media change every two days. At 80% confluency the media was removed and cells were washed twice with 5 ml PBS, then 3 ml of TRypLE Express™ (Gibco) was added to trypsinise cells for 5 minutes at 37°C. Trypsin was inactivated by adding 1 ml trypsin inhibitor (0.25 mg/ml soybean trypsin inhibitor in PBS without calcium or magnesium, pH 7.2, Invitrogen). Cells were resuspended in SFM and transferred to a 15 ml tube prior to cell counting. HFKs were centrifuged at 200 x *g* for 5 minutes, then the supernatant was removed, and cells were resuspended in SFM. Typically $2x10^5$ cells were seeded in a 10 cm plate containing 10 ml SFM.

2.2.4.2 Maintenance of HPV genome containing HFKs

HFKs (Georgie and Clonetics) were cultured on irradiated J2-3T3 feeder cells and maintained at 37°C and 5% CO_2 in complete E media (CEM). CEM consists of Dulbecco's modified eagle media (DMEM 6171, 60% v/v, Sigma-Aldrich), Ham's F12 (32% v/v, Gibco) penicillin and streptomycin (2% v/v PenStrep, Gibco), hydrocortisone (0.1% v/v, Sigma-Aldrich), Hyclone FCS (5% v/v, Sigma-Aldrich), mouse epidermal growth factor (EGF, 0.5% v/v, BD Biosciences), cholera toxin A (0.1% v/v, Corning), insulin (0.2% v/v, Sigma-Aldrich), transferrin (0.2% v/v Sigma-Aldrich), tri-iodo-1L-thyronine (T3, 4 nM, Simga-Aldrich) and adenine (36 μM, Sigma-Aldrich). To split HFKs,

feeder cells were washed off twice with 5 ml PBS, then 3ml of TRypLE Express™ (Gibco) was added to trypsinise cells at 37°C, typically taking around 10 minutes. Cells were resuspended in CEM and transferred to a 15 ml tube prior to cell counting. Plates were washed again with CEM to retrieve any remaining cells. HFKs were centrifuged at 200 x *g* for 5 minutes, then the supernatant was removed, and cells were resuspended in CEM. HFKs were seeded onto approximately $2x10^6$ irradiated J2-3T3 feeder cells, which had been plated 24 hours previously in pre-warmed CEM. Typically $5x10^5$ HFKs seeded on to a 10 cm plate would be ready to split on the fifth day. Plates were gently moved from side to side in order to evenly distribute the cells and media was changed every other day. Maintenance of viral episomes in these HFK lines was verified by Southern blot.

2.2.4.3 Cryopreservation of HFKs

HPV containing HFKs were frozen down in cryovials in freezing media consisting of CEM, supplemented with 10% (v/v) Hyclone FCS (Thermo Fisher) and 20% (v/v) glycerol (Sigma-Aldrich). Untransfected HFKs were frozen down in SFM supplemented with 10% (v/v) FCS and 10% (v/v) DMSO (Sigma-Aldrich). Typically, between 1 and $2x10^6$ cells were frozen per cryovial.

2.2.5 Development of organotypic raft cultures

2.2.5.1 Metal grid preparation

To prepare raft grids 4x4 cm squares were cut from metal sheeting, then two opposite edges of the grid were bent by 0.5 cm on each side. A level was placed on top of each grid to ensure they were level. Metal grids were treated with concentrated sulphuric acid for 1 hour in a fume hood, then rinsed in a beaker under a running tap of H_20 overnight (O/N). Grids were rinsed in water for an additional 5 hours then autoclaved before use.

2.2.5.2 Collagen plug preparation

Each collagen plug was prepared using a mixture consisting of 2.4 ml of 4 mg/ml high concentration type 1 rat-tail collagen (BD Biosciences), 0.3 ml 10x reconstitution buffer (RCB, 260 mM $NaHCO_3$ and 200 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES) in 100 ml 50 mM NaOH; Sigma-Aldrich), 0.3 ml 10x DMEM (without $NaHCO_3$) and 1-2x10^6 J2 fibroblasts. All reagents and stripettes were kept on ice prior to use. A collagen master mix was prepared for the required number of collagen plugs, with enough mix for an additional two plugs due to the viscosity of the collagen. Firstly, J2-3T3 cells were harvested, resuspended in E-media, and counted to ensure there were approximately 2x10^6 cells for each intended collagen plug. The J2-3T3 cells were centrifuged at 200 x *g*. The supernatant was discarded and the remaining cell pellet was resuspended in the appropriate amount of 10x DME and 10x RCB. The collagen was the last

component added to the mixture. If the collagen mixture was a yellow colour 1 M filtered NaOH was added drop wise to obtain a red/orange colour— indicative of the correct pH. Next, 3 ml of the collagen mixture was transferred to a 3 cm petri dish and incubated for 30 minutes at 37°C to allow the gel to solidify. Then using a Pasteur pipette 2 ml of CEM was added to the surface of each plug and returned to the incubator at 37°C. Each plug was kept at 37°C for at least 1 day and used within 4 days of preparation.

2.2.5.3 Raft preparation

Each plug was seeded with approximately $2x10^6$ low passage HFKs (typically below passage 7), with culture media replaced daily. As soon as the media turned yellow within 24 hours of a media change the plugs were ready to use, typically taking between 2-4 days. Using sterile forceps the metal grid was placed in a 10 cm petri dish. Using a sterile spatula each plug was lifted on to a mesh grid and raft E media (excluding EGF) was added to each dish to create an air-liquid interface. It is important to minimize the creation of bubbles during this process. Rafts were incubated at 37°C and 5% CO_2 with media change every other day, then harvested on day 14.

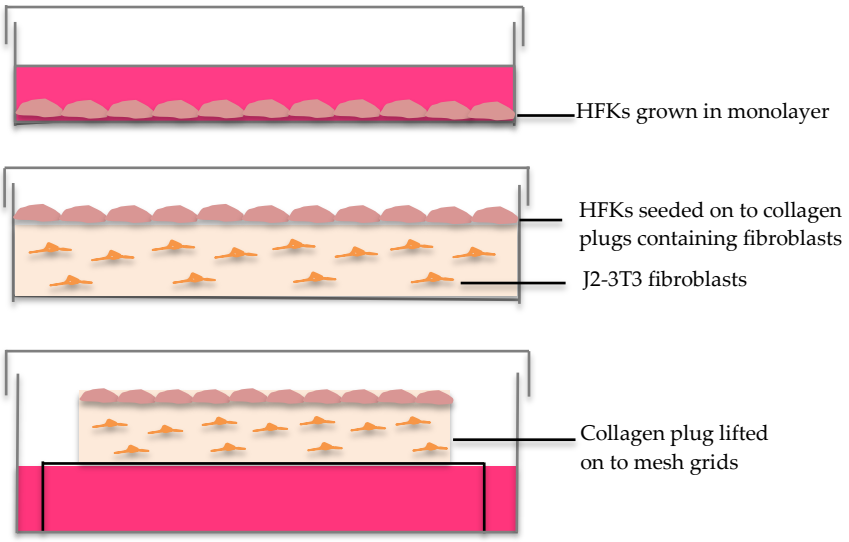


Figure 14. Preparation of an organotypic raft culture

A) Primary HFKs are grown in monolayer culture.

B) Schematic of a collagen plug embedded with irradiated J2s, seeded with HPV genome containing HFKs.

C) The collagen plug is lifted on to a metal grid, and cultured with E-media supplemented with L-glutamine. This creates an air-liquid interface and induces differentiation of the keratinocytes.

2.2.6 Processing of organotypic raft cultures

2.2.6.1 RNA extraction from rafts

Using sterile tweezers the top layers of the rafts were peeled away from the surface of the collagen plug and placed in a sterile 15 ml Falcon tube. RNA was extracted from 14-day-old HFK rafts using the RNA-STAT 60 protocol (AMS biotechnology, UK). Firstly, 500 µl RNA-STAT was added to each sample and homogenized using a 1 ml pipette, then incubated for 5 minutes at room temperature (RT). Next, 100 µl of chloroform was added and the tube was vigorously shaken for 15 seconds (s) then left at RT for 2-3 minutes. Samples were centrifuged at 12,000 x *g* for 15 minutes at 4°C. The aqueous

phase was transferred to a fresh Eppendorf and 250 µl of isopropanol was added, then samples were left at RT for 5 minutes. The Eppendorfs were centrifuged at 12,000 x *g* for 10 minutes. Each pellet was washed with 1 ml 75% ethanol then centrifuged at 7,500 x *g* for 5 minutes at 4°C. The supernatant was removed and discarded and the pellet was air dried for a few minutes. The RNA was dissolved in 40 µl nuclease free water, and incubated for 10 minutes at 60°C to solubilize the RNA. Next, 1.3 µl of RNA was used to measure the RNA concentration on the NanoDrop. RNA was stored at -80°C or used immediately for cDNA synthesis.

2.2.6.2 Preparation of protein lysate from organotypic rafts

Rafts were removed from collagen plugs with sterile tweezers and placed in a 15 ml tube. Firstly, 500 µl of urea lysis buffer (8M urea, 150 mM NaCl, 25 mM Tris-HCL,pH 7.2) supplemented with 150 mM β-mercaptoethanol (BME), phosphatase and protease inhibitors to a 1x final concentration were added, and rafts were homogenized using a plunger. A further 250 µl of lysis buffer was added to finish homogenizing the sample. Lysates were transferred to a fresh Eppendorf and left on ice for 20 minutes, then centrifuged at 16,000 x *g* for 20 minutes at 4°C. The supernatant (soluble fraction) was transferred to a fresh Eppendorf and diluted in 200 µl Llaemli buffer (+fresh BME), then sonicated for 10s at 50% amplitude. 500µl Llaemli buffer (+BME) was added

to the remaining pellet (insoluble fraction) and sonicated for 10s at 30% amplitude. Lysates were stored at -20°C.

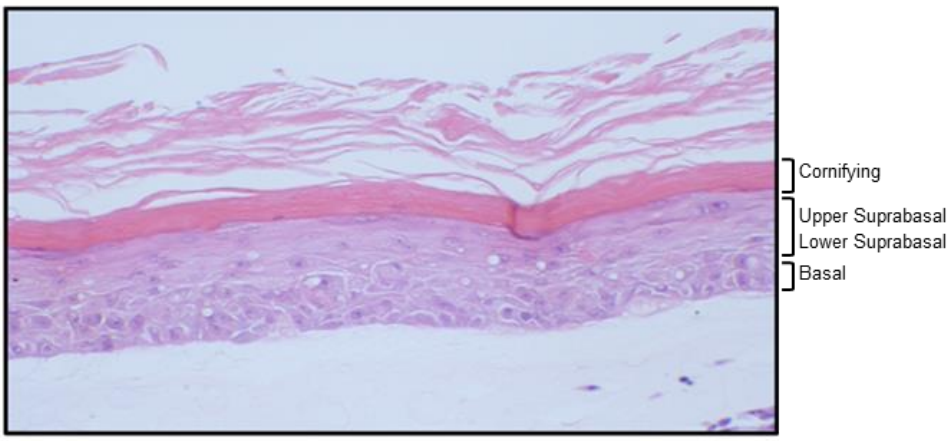


Figure 15. Cross-section of an organotypic raft. Organotypic raft culture derived from primary HFKs transfected with WT HPV18 genomes. The raft was fixed in 3.7% formaldehyde, then paraffin embedded and sectioned. Each layer of the raft is annotated as: basal, lower suprabasal, upper suprabasal and cornified layers. Image taken from Paris et al., 2015.

2.2.6.3 Immunofluorescence staining of organotypic raft sections

Rafts were fixed in 3.7% formaldehyde (Sigma-Aldrich), then paraffin embedded and sectioned (Propath Ltd., Hereford, United Kingdom). Immunofluorescence (IF) was carried out on the paraffin embedded organotypic rafts sections. Firstly, agitated low temperature epitope retrieval (ALTER) was performed in a series of sequential washes in histoclear and 100% IMS. Slides were incubated at 65°C overnight in 1 litre of buffer (1mM EDTA (pH 8.0), 0.1% Tween, Binding site (IC100)). Slides were blocked for 1 hour in a blocking solution of 20% HINGS and 0.1% BSA (Sigma-Aldrich, UK) in 1x PBS. Primary antibodies were added to the blocking solution and slides were incubated at 4°C overnight then washed three times in 1xPBS. Alexa

Fluor 488 goat anti-mouse secondary antibody (LifeTechnologies) or Alexa fluor 594 goat anti-rabbit (LifeTechnologies) were added to slides (1:1000 dilution in blocking solution). Slides were incubated at 37°C for 1 hour then washed four times in 1x PBS. Lastly, 10 µl of Hoescht (10 mg/ml stock) was added to the PBS filled chamber for 10 minutes in an additional step. A few drops of Fluoroshield mounting solution (Sigma-Aldrich, UK) were added to each slide and cover slips were placed on top. Slides were visualized on a Nikon inverted Epifluorescent microscope fitted with a 20x air or 40x oil objective. Images were captured using a Leica DC200 camera and software.

2.2.6.4 DNA extraction from rafts

Rafts were removed from each collagen plug and transferred to a 15 cm Falcon tube. Firstly, 3 ml of sterile DNA lysis buffer (400 mM NaCl, 10 mM Tris-HCl, 10 mM EDTA, pH 7.4) was added to each raft and homogenized using a douncer. Alternatively 3 ml lysis buffer was added to a previously stored HFK cell pellet and pipetted up and down with a 1 ml pipette. Then RNase A (400 ng/ml final), proteinase K (50 µg/ml final) and 30 µl of 20% SDS were added to each sample and incubated at 37°C O/N. The DNA was sheared using a 19G blunt needle by moving up and down ten times. Then 6 ml phenol:chloroform:isoamyl alcohol 25:24:1 (Saturated with 10 mM Tris, pH 8.0, 1 mM EDTA (Sigma-Aldrich, UK)) was added to each tube and inverted

ten times. Tubes were centrifuged at 1,800 x *g* for 5 minutes at 4°C. The supernatant was transferred to a fresh tube and the phenol extraction was repeated once more. Then 6 ml 24:1 chloroform/isoamyl alcohol (Sigma-Aldrich) was added to each tube and inverted 10 times. Tubes were centrifuged at 1,800 x *g* for 5 minutes at 4°C. The supernatant was transferred to a fresh tube. Finally 2 volumes of cold 100% ethanol and 1/10 volume of 3 M sodium acetate pH 5.2 were added to each sample. Samples were stored at -20°C O/N or alternatively -80°C for at least 1 hour. After incubation at -80°C or O/N -20°C samples were centrifuged 1,800 x *g* for 30 minutes at 4°C. The supernatant was removed and discarded and the pellet was washed with 5 ml 70% ethanol, then centrifuged for a further 15 minutes at 1,800 x *g* at 4°C. The supernatant was removed and discarded and the pellet was washed with 1 ml 70% ethanol then transferred to a 1.5 ml Eppendorf. The sample was centrifuged at 18,000 x *g* for 10 minutes at 4°C. The supernatant was removed and pellet air-dried for a few minutes. The DNA pellet was re-suspended in 50 µl TE buffer and incubated for 10 minutes at 60°C to completely solubilize the DNA pellet. A 1.3 µl sample of DNA was used to measure the DNA concentration and determine the 260/280 ratio. DNA was used immediately or stored at -20°C.

2.2.6.4.1 Spectrometry of nucleic acids

RNA and DNA concentrations were measured using a NanoDrop ND-1000

spectrophotometer. Before use, the NanoDrop was wiped clean with lens cleaning tissue. The NanoDrop was blanked with 1.3 µl of elution buffer, and wiped clean again, then 1.3 µl of sample was added to the NanDrop to measure RNA or DNA concentrations and the 260/280 ratios to assess sample purity. A ratio of ~1.8 is the ideal ratio for DNA and ~2.0 is ideal for RNA. A value higher or lower in either case indicates the presence of phenol, protein or contaminants.

2.2.7 Methylcellulose differentiation of HFKs

2.2.7.1 Preparation of methylcellulose

To prepare 1.5% methylcellulose, 3.75 g of methylcellulose was transferred to a 500 ml Duran bottle. Both the methylcellulose and a magnetic stirrer inside the bottle were autoclaved. After autoclaving, 125 ml of E media (without EGF) was transferred to the methylcellulose, which was then incubated for 20 minutes at 65°C, with swirling of the bottle at 5 minute intervals. A further 112.5 ml of E media was added to the mixture, which was left to stir at 600 rpm O/N at 4°C. The next day 12.5 ml of Hyclone FCS was added and the mixture was stirred for a further two hours at 600 rpm at 4°C. Methylcellulose bottles were covered in foil then stored at 4°C for up to six weeks.

2.2.7.2 Growth of keratinocytes in methylcellulose

Firstly, 30ml of pre-warmed methylcellulose was transferred to each bacterial plate required and placed in the incubator at 37°C for 2 hours. HFKs were

grown to approximately 70% confluency then harvested. The cell pellet was resuspended in enough E-media to distribute 1 ml per plate of methylcellulose using a 1 ml pipette; one plate of HFKs ($3x10^6$ cells) is sufficient to seed one plate of methylcellulose. The cells were dispersed throughout the methylcellulose by spreading with a 10 ml pipette. The plates were returned to the incubator at 37°C and 5% CO_2 for the intended incubation time (typically 24 and 48 hours).

2.2.7.3 Harvesting HFKs from methylcellulose

For each plate of methylcellulose, a cell scraper was used to evenly scrape the mixture between three 50 ml Falcon tubes. The tubes were filled to the 50 ml mark with ice cold PBS then spun at 244 x *g* for 10 minutes. The PBS was aspirated until the 10 ml mark. Again, the tubes were filled to the 50 ml mark with ice cold PBS and spun at 244 x *g* for 5 minutes. The supernatant was removed from each tube then using a 10 ml pipette one tube was combined to another tube, filled with PBS to 50 ml and spun for 5 minutes at 244 x *g*. This was repeated to end up with one final tube containing the cell pellet.

2.2.8 Growth of W12 keratinocytes in methylcellulose

2.2.8.1 Preparation of methylcellulose medium

To prepare 1.6% methylcellulose, 3.2g methylcellulose was added to a Duran bottle and autoclaved. Then 100 ml keratinocyte Glasgow minimal essential medium (GMEM) with ½ concentration of EGF was added to the

methylcellulose and incubated for 20 minutes at 60°C on a stir plate at 600 rpm. The remaining 100 ml media was added and the mixture was placed on a stir plate at 600 rpm O/N at 4°C.

2.2.8.2 Harvesting and fixation of monolayer W12 cells

Parental W12 cells were maintained in 10 cm plates in GMEM media with 3T3 irradiated feeders. At 80% confluency four monolayer plates at passage 13 representing 0 hr were fixed with 37% formaldehyde for 5 min at RT, then washed in ice cold PBS. Next, 5 ml glycine stop solution was added for 5 minutes at RT followed by another ice cold PBS wash. Cells were scraped and pooled in to a 15ml tube, followed by centrifugation at 720 x *g* for 10 minutes at 4°C. Next, 1 µl of PMSF and 1 µl of PIC were added to the cell pellet and stored at -80°C. One 10 cm plate at 80% confluency was washed with PBS to remove the feeders, then 1 ml of Tri reagent (Sigma-Aldrich) was added in the fume hood. Cells were washed off and transferred to an Eppendorf and stored at -80°C (ready for RNA extraction).

2.2.8.3 Growth of W12 cells in methylcellulose

Eight 10 cm plates (~$8x10^6$ cells/plate) of W12 cells at 80% confluency were trypsinised and counted. Cells were centrifuged at 200 x *g* and supernatant was discarded. Cells were resuspended in 10 ml of GMEM. Next, 1 ml of cell suspension (~$8x10^6$ cells) was added to 20 ml methylcellulose (x10 tubes), with five tubes allocated for 24 hour treatment and the remaining five tubes for 48

hr. Tubes were mixed by tapping the tube vigorously. The lids were unscrewed loosely and incubated at 37°C O/N.

2.2.8.4 Harvesting W12 cells from methylcellulose

After 24hr five tubes were removed from the incubator and diluted with 20 ml PBS then centrifuged at 244 x *g* for 10 minutes at 4°C. The media was aspirated leaving behind 5 ml in the tube. A further 20 ml of PBS was added to the tube and mixed thoroughly. Additional PBS was added up until the 50 ml mark then the tube was inverted ten times. The tubes were spun at 244 x *g* for 10 minutes at 4°C and then the supernatant was removed to leave behind a pellet. The same process was carried out for the tubes.

2.2.8.5 Fixing methylcellulose treated cell pellets

A double concentration of fixation solution was prepared (20 ml GMEM and 1.08 ml formaldehyde). Firstly, 5 ml media was added to break up each of the four cell pellets. One pellet was kept on ice for separate RNA extraction. The cell suspension was pooled in to one tube then 20 ml of the fixation solution was added. The tube was inverted for 3 minutes then spun at 244 x *g* for 2 minutes at 4°C followed by a PBS wash with the same centrifugation conditions, thus allowing a 5 minutes total fixation period. Next, 40 ml glycine stop solution was added and the tube was inverted for 3 minutes followed by a spin at 244 x *g* for 2 minutes at 4°C. The cells were washed in 40 ml PBS followed by the same centrifugation. Lastly, 5 ml PBS supplemented

with 30 µl PMSF was added to the tube, then the contents was transferred to a new 15 ml tube. The tube was centrifuged at 720 x *g* for 10 minutes at 4°C, then the cell pellet was supplemented with 1 µl PMSF and 1 µl PIC, and stored at -80°C. The same process was carried out for the 48 hr tubes.

2.2.8.6 Extraction of RNA

RNA was extracted with Tri-reagent (Sigma-Aldrich), following the manufacturer's protocol. Briefly, W12 cells were harvested from one 10 cm plate and centrifuged at 200 x *g*, and the supernatant was discarded. Next, 1ml of Tri-reagent was added to the cell pellet and mixed up and down several times to homogenize the sample. The sample was transferred to an eppendorf. The Eppendorf sample was vigorously shaken for 15 seconds, and allowed to stand for 10 minutes at RT. The sample was centrifuged at 12,000 × *g* for 15 minutes at 4°C. The supernatant was transferred to a new Eppendorf. Then, 500 µl of isopropanol was added and the sample was left to stand for 10 minutes at RT. The sample was centrifuged at 12,000 × *g* for 10 minutes at 4 °C. The supernatant was discarded the RNA pellet was washed in 1 ml of 75% ethanol step 1. The sample vortexed and centrifuged at 7,500 × *g* for 5 minutes at 4 °C. The RNA pellet was air dried for 5–10 minutes. The RNA pellet was resuspended in 50 µl RNase free water. To facilitate dissolution, mix by repeated pipetting with a micropipette at 55–60 °C for 10–15 minutes. RNA was stored at -80°C. RNA was converted to cDNA using the QuantiTect

Reverse Transcription Kit (Qiagen).

2.3 Protein analysis

2.3.1 Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE)

2.3.1.1 Preparation of protein lysate from cell pellets

Cell pellets containing 1-3x10^6 cells were stored at -20°C prior to protein extraction. Urea lysis buffer (ULB, supplemented with 1x protease and phosphatase inhibitors and 150 mM β-mercaptoethanol (BME)) was added to each cell pellet, and homogenized using a 1 ml pipette. Samples were left on ice for 15 minutes, then each sonicated for 10s at 30% amplitude. Samples were centrifuged for at 16,000 x *g* for 15 minutes, and the supernatant was transferred to new Eppendorfs, and labeled as the urea soluble fraction.

2.3.1.2 Protein quantification

Estimates of protein concentrations were determined using the Bradford assay (Biorad). Two-fold serial dilutions of BSA standards were prepared starting from 2 mg/ml. Protein lysates, ULB and Bradford reagent were all diluted 1:5 in water. Next, 1 ml of diluted Bradford reagent was added to 10 µl diluted ULB and transferred to a cuvette. The spectrophotometer was blanked with this ULB sample. Next, 1 ml of the diluted Bradford reagent was added to 10 µl of each of the standards or unknown samples, then transferred to individual cuvettes to measure absorbance readings at 495 nm.

Absorbance readings for each standard were plotted on to a graph and unknown sample concentrations were calculated accordingly.

x=(y-c)/m

Where: x= unknown concentration

y= absorbance reading

c= y-intercept

m= gradient

Figure 16. Calculation of protein concentration by Bradford assay..

2.3.1.3 Loading and running gels

Llameli loading buffer (+BME) was added to a final concentration of 1%. Samples were boiled for 5 minutes at 95°C on a heat block, then pulse spun in a microcentrifuge. Samples were loaded on to a 12% SDS-PAGE gel, along with 5µl of protein molecular weight marker (Pageruler plus, Thermo Scientific).

2.3.1.4 Gel transfer to the membrane

Gels were transferred to Polyvinylidene difluoride (PVDF) membrane using a semi-dry blotting machine at 25 V for 1 hour. Protein transfer to the membrane was checked using Ponceau Staining. The Ponceau Stain was washed off using 1x TBST before moving on to the blocking step.

2.3.1.5 Antibody staining

The membrane was blocked in 5% milk or 3% BSA in 1x TBST for 1hr at RT or O/N at 4°C. The membrane was then incubated with the appropriate dilutions of antibodies in 5% milk or 3% BSA at 4°C O/N on a plate shaker. The membrane was washed 4x 5 min in 1x TBST, and then incubated in secondary antibody (typically 1:10,000 in 5% milk) for 1 hr at RT. Lastly the membrane was washed for 6 x 5 minute washes in 1x TBST. For signal development excess buffer was drained off and then SuperSignal West Dura Extended Duration Substrate (Thermo Fisher) was transferred to the membrane and left for 5 minutes at RT. The membrane was placed in to a plastic wallet and imaged using a Fusion FX digital detection system.

2.3.1.6 Densitometry analysis

Western blots were saved as .tif files and bands were quantified using the Fusion FX digital detection system. Data were exported to Excel for normalization and analysis.

2.4 Southern Blot

2.4.1 Digestion of DNA

Restriction enzyme digests were typically performed on 5 μg genomic DNA, as outlined in table 1. Copy number controls were prepared, as outlined in table 2. *Eco*RI (Biolabs) was used to single cut and linearize the HPV18 genome and *Bgl*II (Biolabs, R01445) was used as a non-cutter. Tubes were

briefly vortexed and spun down to fully mix each reaction. Tubes were then incubated at 37°C O/N, then run on a gel or stored at -20°C for future use.

Southern blot digests:

Table 7. Southern blot restriction digests

DNA	Restriction Enzyme	Buffer (10x)	*Dpn*I	H_20
5 µg	1 µl	2 µl	1 µl	Up to 20 µl

Table 8. Copy number controls

Copies	pGEMII HPV18 (0.45 µg/µl)	Restriction Enzyme (*Eco*RI)	Buffer (10x)	Salmon sperm DNA (1 µg/µl)	H_20
50	1µl	1 µl	2 µl	2 µl	Up to 20 µl
100	2µl	1 µl	2 µl	2 µl	Up to 20 µl
200	4µl	1 µl	2 µl	2 µl	Up to 20 µl
500	10µl	1 µl	2 µl	2 µl	Up to 20 µl

2.4.2 Running gel and transfer to membrane

To prepare a 0.8% agarose gel, 4 g of agarose was dissolved in 500 ml 1x TBE (89 mM Tris (pH 7.6), 89 mM Boric acid, 2 mM EDTA) containing 5µl ethidium bromide. The gel tank was filled with 2.5 L 1x TBE containing 20 µl ethidium bromide. For each sample 5 µl of 5x DNA loading buffer was added

to each 20 µl reaction. The 25 µl sample was loaded to each well of the agarose gel and 5 µl of Hyperladder 1 kb plus (Invitrogen) was also loaded as a molecular size marker. The gel was run at 50 V O/N then UV images of each gel were obtained with an accompanying UV ladder (Syngene Bio Imaging, UK). The gel was washed for 2x 20 minute with 0.25 M HCL then 2x 30 minute washes with 0.4 M NaOH. The gel was flipped over and the transfer was assembled in the following order: glass plate placed over a dish, x4 filter paper (24x33 cm), agarose gel, gene screen membrane (20x22.5 cm), x4 filter paper (21x23.5 cm), two paper towel bundles, glass plate and weight bottle on top. A Strippette was used to roll the filter paper at each layer in order to exclude bubbles from the transfer. The transfer dish was filled with 3 litres of 0.4 M NaOH and all edges were sealed with cling film then left at RT O/N. The next day the gel was wrapped with edges of filter paper and inverted. Upon disassembly each well on the membrane was marked with a pencil. The membrane was auto UV cross-linked using a UV crosslinker 90 (Stratalinker, Stratagene). The membrane was then rinsed in 2x saline-sodium citrate (300 mM NaCl, 30 mM sodium citrate, pH 7.0), wrapped in cling film and stored at -20°C.

2.4.3 Preparation of a radioactive probe

2.4.3.1 Preparation of a radioactive probe using DNA labeling beads

To prepare the HPV18 probe the pGEMII HPV18 vector was linearized by *Eco*RI restriction digestion. The HPV18 DNA was separated from the plasmid

by agarose gel electrophoresis and the 8 kb HPV18 band was gel purified (Sigma gel extraction kit) and DNA was resuspended in 20 μl ddH_20. Next, 50 ng of DNA was diluted in to a final volume of 45 μl TE buffer (10mM Tris, 1mM EDTA, pH 8.0). The DNA was denatured at 95°C for 5 minutes followed by incubation on ice slurry for 2 minutes. In the radioactive room the denatured DNA was then added to DNA labeling beads from the Ready-to-go DNA labeling beads-dCTP kit (Amersham Biosciences, UK). Then 50 μCi of (alpha-^{32}P) dCTP (Perkin Elmer) was added to the probe behind a Perspex screen, and left to incubate at 37°C in the heat block for 30 minutes. Appropriate paperwork was completed detailing radioactive waste and usage. The probe was then purified using an Illustra Probe Quant G-50 microcolumn (Amersham Biosciences). The column was placed in an eppendorf and centrifuged at 700 x *g* for 1 minute at 4°C to wash buffer through the column. The probe sample was then transferred to the center of the resin column. The column was centrifuged at 700 x *g* for 2 minutes at 4°C and flow-through was retained. The Geiger counter was used to measure total CPS of the column and eluted probe sample, then CPS of the sample alone, in order to estimate total radioactive incorporation.

2.4.3.2 Preparation of a radioactive probe by nick translation

Firstly, 50 ng of HPV18 probe was added to water to make a final volume of 14 μl. The probe was boiled at 100°C for 5 minutes then placed on an ice

slurry for 2 minutes. Next, 5 µl of 5x oligonucleotide labeling buffer (1 mg/ml random deoxynucleotide primers, 1 M HEPES, pH 6.6) and 1.5 µl nuclear grade BSA was added. In the radioactivity room 3 µl of ^{32}P was added to the probe mixture followed by 1.5 µl DNA polymerase I Klenow, then left in a lead pot behind Perspex for 2-3 hours at RT. After the incubation period the probe was spin purified using an Illustra Probe Quant G-50 microcolumn (Amersham Biosciences), as previously mentioned.

2.4.4 Hybridization of radiolabelled probe to DNA

To prepare the hybridisation buffer, 2x hybridisation buffer (5x SSC; 750 nM NaCl, 1.5 M sodium citrate, pH 7.0), 10x Denharts (0.2% w/v ficoll 400, 0.2% w/v polyvinylpyrolilone), 0.2% w/v BSA fraction V and 20% dextran sulphate (Affymetrix) were mixed on a stir plate and separated in to 10 ml aliquots. The 2x hybridisation buffer was mixed 1:1 with deionized formamide (Sigma-Aldrich) and SDS was added to a final concentration of 0.1% (w/v). Next, 200 µl of salmon sperm DNA (10 mg/ml, Invitrogen) was boiled at 95°C for 5 minute followed by incubation on ice slurry for 2 minutes. The boiled salmon sperm was transferred to the hybridisation solution to constitute pre-hyb buffer. The membrane was removed from -20°C and placed on a damp square of gauze with the DNA facing upwards. The membrane and gauze were rolled tightly and evenly and placed in a hybridization canister (Hybaid, UK). The canister was rotated anti-clockwise to unravel the roll inside. The pre-hyb

solution was added to the canister which was rotated for 1 hr at 42°C in a hybridization oven. A separate 200 μl aliquot of salmon sperm DNA was added to the prepared radiolabelled probe (previously mentioned), then denatured at 95°C for 5 minutes, followed by cooling on an ice slurry for 2 minutes. Pre-hyb blocking buffer was removed and hybridization buffer containing the probe was added to the canister, which was returned to the 42°C oven and incubated O/N.

2.4.5 Stringency washes

All of the buffers used for washes are listed in table 3. The membrane was removed from the canister and placed in a glass dish. Approximately 100 ml buffer 1 was added on top of the membrane. The membrane was then cleaned using a sponge. A further 400 ml of buffer 1 was added to the membrane and the dish was shaken for 15 minutes at 100 revs/min on a plate shaker. Buffer four was pre incubated at 55°C in the oven ready for later use. Buffers were replaced every 15 minutes expect for buffer 4 (two 500 ml washes each for buffers 1- 3). For the wash with buffer 4 the membrane was incubated and shaken at 55°C for 30 minutes. Buffer 4 was removed and the membrane was drained and wrapped in cling film. X-ray film was placed over the membrane in a cassette in the dark room, and stored at -20°C to slow the β particles. Films were typically developed after 24 hours using a Compact x4 automatic

processor (Xograph Healthcare, UK), however some films required longer exposure.

Table 9. Constituents of Southern blot wash buffers

Buffer 1	2x SSC	0.1% SDS
Buffer 2	0.5x SSC	0.1% SDS
Buffer 3	0.1x SSC	0.1% SDS
Buffer 4	0.1x SSC	1% SDS

2.5 DNA Analysis

2.5.1 qPCR

SYBR green quantitative PCR (qPCR) was carried out to determine absolute and relative amounts of known DNA sequences, by detecting the fluorescence signal emitted by SYBR green dye bound to double stranded DNA. All qPCR reactions were set up using filter tips (Starlab, UK). Input samples were typically diluted 1:10 or 1:5 in nuclease free water, depending on initial concentrations. Ten-fold or five- fold serial dilutions of standards were prepared using one of the input samples. Forward and reverse primers were diluted to make a final working concentration of 10 mM. Typically for each well 2 µl DNA, 10 µl 2X SensiMix SYBR (Bioline, UK), 0.125 µl forward and 0.125 µl reverse primer and 7.75 µl nuclease free water were added to make a 20 µl total reaction volume. Samples were added in triplicate to a 96-well plate (Starlab) and either used immediately or stored in foil at 4°C for same day use or -20°C for next day use. Typical cycling parameters on a MxPro

3700 (Agilent Technologies) included conventional 3-step 40 cycle PCR with an initial denaturation step at 95°C for 10 minutes, followed by a further 95°C for 30 seconds, an annealing step for 30 seconds at an appropriate temperature for specific primers, 20 seconds extension step at 72°C and a final extension at 72°C for 10 minutes. Data were analysed using MxPro qPCR software (Agilent Technologies), to generate standards curves, amplification plots and dissociation curves. For the amplification plot the number of PCR cycles is plotted against the fluorescence values. The threshold cycle (CtS) values generated provide information on the amount of template DNA in each sample. The threshold is the first point at which the detected SYBR fluorescence is greater than the background signal, and is set in the log-linear region of the amplification plot.

2.5.1.1 Oligonucleotide primer design

Primers were designed using the NCBI Primer-Blast software and synthesized by Life Technologies.

2.5.1.2 Relative quantitation of gene expression

The delta-delta(ΔΔ) Ct method allows the quantitation of gene expression relative to an endogenous control (TLR). The Ct values of a specific gene are referred to as the target Ct and the Ct values of a housekeeping gene such as TLR are referred to as the endogenous Ct. The reference cells in this case are the WT HFKs and the target cells are the ΔCTCF HFKs. The target Ct minus

the endogenous Ct yields the dCt value. The reference Ct minus the target Ct yields the ddCt value. These values can be used to obtain a relative fold increase in gene expression.

Fold change in expression= $2^{-\Delta\Delta Ct}$

Where: ΔCt= target Ct – endogenous Ct

$\Delta\Delta Ct$= reference ΔCt – target ΔCt

Figure 17. Calculation of relative gene expression.

2.5.2 Agarose gel electrophoresis

Gels were prepared by dissolving molecular grade agarose powder (Bioline, UK) in 1x TBE buffer (89 mM Tris (pH 7.6), 89 mM Boric acid, 2 mM EDTA), containing 5 μl ethidium bromide. Typically a 1.8% gel was prepared for DNA fragments of approximately 100-200 bp in size and 1.2% gels were prepared for running sheared chromatin. Gels were run at constant 80 Amp for approximately 1-2 hours then visualized in a Gene Flash UV light box (Syngene Bio Imaging, UK).

2.5.2.1 Purification of DNA from agarose

DNA was purified from agarose using the GenElute™ gel extraction kit (Sigma-Aldrich). DNA bands were excised using a sterile scalpel and transferred to Eppendorf tubes. Gel excisions were weighed (typically 100 mg each) and 3 gel volumes of the Gel Solubilization Solution was added to each

tube. The gel mixture was incubated at 55°C for 10 minutes to dissolve the gel. Once the gel was dissolved it was important to check that the colour of the solution was yellow. Next, one gel volume of 100% isopropanol was added and the samples mixed. The binding column was placed in a collection tube and 500 µl of Column Preparation Solution was added. The column was centrifuged at 16,000 x *g* for 1 minute and the flow through was discarded. The gel solution was added to the binding column and centrifuged at 16,000 x *g* for 1 minute. The flow through liquid was discarded. Next, 700 µl of Wash Solution (containing ethanol) was added to the column and centrifuged at 16,000 x *g* for 1 minute. The flow through was discarded and the column was replaced in the tube for a repeat spin. The column was added to a new collection tube and 25 µl of elution buffer (preheated to 65°C prior to use) was added to the centre of the membrane, followed by centrifugation at 16,000 x *g* for 1 minute. The NanoDrop was used to measure DNA concentration as previously described.

2.5.2.2 DNA Sequencing

PCR products were gel purified and sent for DNA sequencing at the University of Birmingham Genomics Services Facility. In each sequencing reaction 10 ng DNA and 3.2 pmol forward or reverse primer were added to an Eppendorf and made up to a 10 µl total volume with nuclease free H_20. At the sequencing facility the 10 µl reaction was transferred to a well on the

sequencing plate. Prior to sequencing 10 µl of terminator reaction mix was added to make a total volume of 20 µl. DNA sequencing was performed on a capillary sequencer ABI 3730. Sequencing data were analyzed using FinchTV software and NCBI blast software was used to match DNA sequences to known library sequences.

2.6 RNA analysis

2.6.1 DNase treatment of RNA

RNA (up to 5 µg) was resuspended in nuclease free H_20 up to a total volume of 7 µl. The RNA was treated with 2 µl RQ1 DNase (Promega) and 1 µl 10x DNase buffer for 30 minutes at 37°C, then 1 µl of DNase stop solution was added for 10 minutes at 65°C for enzymatic inactivation. DNase treated RNA was used immediately for cDNA synthesis.

2.6.2 Reverse transcription/cDNA synthesis

RNA (5 µg) in water was reverse transcribed to cDNA in a 20 µl final volume using the Tetro cDNA kit (Bioline). Firstly a priming premix was prepared containing up to 5 µg RNA, 1 µl $Oligo(dT)_{18}$, 1 µl 10mM dNTP mix, 4 µl RT buffer, 1 µl RNase A inhibitor, 1 µl 200 u/µl Tetro Reverse Transcriptase and DEPC-treated water to make a final volume of 20 µl. Samples were mixed by pipetting, then incubated at 45°C for 30 minutes. The reaction was terminated by incubating at 85°C for 5 minutes, then samples were placed on ice. cDNA was used immediately for PCR or stored at -20°C. Negative control samples

without reverse transcriptase were also prepared. Typically, 2 µl of cDNA was PCR amplified in a 50 µl reaction volume, then 10 µl of 6x loading dye was added and 30 µl of PCR product was loaded on to a 1.2% agarose gel (containing ethidium bromide) for gel electrophoresis.

2.7 Formaldehyde isolation of regulatory elements (FAIRE)

2.7.1 Formaldehyde fixation of cells

The cell media from each cell culture plate was removed and 3T3-J2 feeder cells were washed off with 5ml PBS. Then 5 ml fixation solution (1% formaldehyde in 3T3-J2 D6171 media) was added to each plate. The plates were gently rocked for 10 minutes at RT on a plate shaker in the fume hood. The fixation solution was discarded and each plate was washed in 5 ml 1x PBS for 5s. PBS was removed and 5 ml glycine stop solution was added to each plate to quench the formaldehyde, then gently rocked for 5 minutes at RT. Glycine stop solution (Active Motif) was removed and each plate was washed again in 5 ml 1x PBS for 5s. Prior to use 10 µl of 100 mM PMSF was added to 2 ml PBS to make the cell scraping solution, with 2 ml sufficient to add to each plate. Cells were scraped off with a rubber policeman and pipetted in to a 15 ml tube. Tubes were centrifuged at 720 x *g* for 10 minutes at RT. Supernatant was discarded and 1 µl 100 mM PMSF and 1 µl PIC was added to each pellet. Pellets were used immediately or stored at -80°C.

2.7.2 Sonication & reverse cross-linking

Cell pellets were removed from -80°C and thawed on ice. For cell lysis 500 µl of DNA lysis buffer (Active Motif) (supplemented with 5 µl PIC and 5 µl 100 mM PMSF) was added to each pellet and incubated on ice for 30 minutes. For shearing, 350 µl of shearing buffer (Active Motif) was added to each sample and transferred to Eppendorfs. The samples were sonicated for 30s on and 30s off of for 15 minutes at 25% amplitude (30 minute total time). Two aliquots of chromatin were taken, one for '100%' input and one for 'FAIRE DNA'. To the 'FAIRE' samples 150 µl of water was added and samples were left on ice. To the '100%' input sample 150 µl of water was added followed by 10 µl of 5 M NaCl. Input samples were heated at 95°C for 15 minutes to reverse the DNA cross-links. Next, 1 µl of RNaseA (10 µg/µl) was added to each sample then incubated at 37°C in an incubator for 15 minutes. Finally, 2 µl of proteinase K (0.5 µg/µl) was added to each sample and incubated at 67°C in a heat block for 15 minutes.

2.7.3 DNA extraction

To the '100% Input' and 'FAIRE DNA' samples phenol chloroform extraction was performed. Firstly 200 µl of Phenol:Chloroform:Isoamyl Alcohol 25:24:1 (Saturated with 10 mM Tris, pH 8.0, 1 mM EDTA (Sigma-Aldrich, UK)) was added to the samples and tubes were inverted for 1 minute. Tubes were centrifuged at maximum speed (5917R) for 10 minutes at 4°C. The aqueous

supernatant that contains the non-protein associated DNA was transferred to a new Eppendorf. DNA precipitation was carried out on both 'input' and 'FAIRE' by adding 20 µl 3 M sodium acetate (pH 5.2) and 500 µl 100% ethanol, followed by vortexing and incubation at -20°C O/N or -80°C for 1 hour. Tubes were centrifuged at maximum speed for 10 minutes at 4°C. The supernatant was removed and discarded. Pellets were washed in 500 µl of 70% ethanol and centrifuged at full speed at for 5 minutes 4°C. Again, the supernatant was removed and discarded and the pellet was air dried for 5 minutes at RT. The pellets were resuspended in 50-150 µl TE buffer and stored at -20°C ready for use.

2.8 Chromosome Conformation Capture Assay (3C)

2.8.1 Single-cell preparation

The cell media was removed from the 10 cm cell culture plates and cells were washed twice with 5 ml PBS to remove feeders. Cells were trypsinised for 10 minutes at 37°C then resuspended in approximately 5 ml of E media per plate. Cells counts were obtained and were typically between 1-1.5x10^7 cells from three 90% confluent plates. Cell suspensions were centrifuged at 400g for 1 minute. The supernatant was discarded and pellets were resuspended in 1ml 10% (v/v) FCS/PBS, then filtered through a 70 µm cell strainer to make a single cell suspension.

2.8.2 Crosslinking and cell lysis

Next, 9.5 ml of 1% formaldehyde (in 10% FCS/PBS) was added to the cell suspension and incubated on an end-to-end rotator for 10 minutes at RT. The tubes were transferred to ice then glycine was added to a final concentration of 125 μM. Tubes were centrifuged at 225 x *g* for 8 minutes at 4°C. The supernatant was discarded and the pellet was resuspended in 5 ml of cold lysis buffer (10mM TRIS-HCl pH 7.7; 10mM NaCl; 5mM $MgCl_2$; 0.1mM EGTA; 1X complete protease inhibitor) and incubated on ice for 10 minutes, ensuring the solution was very homogenous. The tubes were centrifuged at 400 x *g* for 5 minutes at 4°C, supernatant removed and pellets stored at -80°C.

2.8.3 DNA digestion

Firstly, 0.5 ml 1.2X restriction enzyme buffer (NEB) was added to the nuclei pellet and transferred to an Eppendorf tube. The tubes were placed at 37°C and SDS was added to make a 0.3% final concentration. Tubes were incubated for 1 hr at 37°C while shaking at 900 rpm. 50 μl of 20% Triton X-100 was added then incubated for 1hr at 37°C while shacking at 900 rpm. Prior to digestion, 5 μl of DNA was removed and labeled as undigested genomic DNA. This control is necessary for checking digestion efficiency. Then 400U of the *Nla*III restriction enzyme (NEB) was added to the remaining sample and incubated at 37°C O/N whilst shaking at 900 rpm. After digestion, 5 μl of

DNA was removed and labeled as digested genomic DNA. Again this control is necessary for checking digestion efficiency.

2.8.4 Ligation

Firstly, 40 μl of 20% SDS was added to the sample and incubated for 20-25 minutes at 65°C, whilst shaking at 900 rpm. The digested nuclei was transferred to a 50 ml Falcon tube and 6.125 ml of 1.15x ligation buffer (704 μl of 10x Ligation buffer (NEB T4 Ligase kit) + 5420 μl H_2O) was added, followed by Triton X-100 to 1% final concentration. Tubes were incubated for 1 hr at 37°C whilst shaking gently. Next, 100U of T4 DNA ligase was added to each tube then incubated for 4 hr at 16°C in a water bath in the cold room, followed by 30 minutes at RT. Lastly, 300 μg proteinase K was added and incubated at 65°C O/N to de-crosslink the samples.

2.8.5 DNA purification

Firstly, 300 μg RNase A was added to the samples and incubated for 30-45 minutes at 37°C. Then 7 ml of Phenol:Chloroform:Isoamyl Alcohol 25:24:1 (Saturated with 10 mM Tris, pH 8.0, 1 mM EDTA, Sigma-Aldrich, UK) was added and samples were mixed vigorously by inverting ten times. Samples were centrifuged at 2,200 x *g* for 15 minutes at RT. Phenol-chloroform extraction was repeated if the supernatant remained turbid. The supernatant was transferred into a new 50 ml tube, followed by 7 ml of water, 1.5 ml of 2 M sodium acetate (pH 5.6), and 35 ml of ethanol. (It is necessary to increase

the volume before precipitation to dilute dithiothreitol (DTT) present in the ligation buffer and prevent it from precipitation). Samples were mixed by pipetting several times then stored at -80°C for ~ 1 hour or alternatively at -20°C O/N. Samples were centrifuged at 2,200 x *g* for 15 minutes at 4°C and then the supernatant was removed and the pellet was briefly air dried at RT. The DNA pellet was dissolved in 150 µl of 10 mM Tris-HCl pH 7.5 and stored at -20°C ready for PCR analysis.

2.8.6 Assessment of digestion efficiency

Previously aliquoted undigested (UND) and digested (D) DNA were thawed for processing. Firstly, 500 µl of 1x PK buffer (5 mM EDTA, pH 8.0; 10 mM Tris-HCl, pH 8.0; 0.5% SDS) and 20 µg proteinase K were added to the UND and D aliquots, then incubated for 30 min at 65 °C. Samples were equilibrated for a few minutes at 37 °C, then 1 µg RNase A was added and incubated for 2 hr at 37°C. Next, 500 µl of phenol:chloroform:isoamyl alcohol 25:24:1 (Saturated with 10 mM Tris, pH 8.0, 1 mM EDTA, Sigma-Aldrich, UK)was added to each sample and mixed vigorously. Samples were centrifuged at 16,100*g* for 5 minutes at RT. The supernatant was transferred into a new tube and 50 µl of 2 M sodium acetate pH 5.6 was added, followed by 1.5 ml of ethanol. Samples were mixed and placed at - 80°C until frozen (about 45 minutes). Samples were centrifuged at 16,100 x *g* for 20 minutes at 4°C. The supernatant was removed and pellets were washed with 50 µl of 70% ethanol,

then centrifuged at 16,100 x *g* for 4 minutes at RT. The supernatant was removed and the pellet was air dried at RT. The pellet was resuspended in 60 µl H_20.

2.9 Chromatin Immunoprecipitation (ChIP)

2.9.1 Formaldehyde fixation of cells

ChIP experiments were performed according to the Active Motif ChIP-IT express protocol (Active Motif 53008). The cell medium from each cell culture plate was removed and J2 feeder cells were washed off twice with 5ml PBS. Then, 10 ml fixation solution (1% formaldehyde in 10 ml D6171 media) was added to each plate. Plates were gently rocked for 5 minutes at RT on a plate shaker in the fume hood. The fixation solution was discarded and each plate was washed in 5 ml ice-cold 1x PBS for 5s. PBS was removed and 5 ml 1x glycine stop solution was added to each plate to quench the formaldehyde, then gently rocked for 5 minutes at RT. Glycine stop solution was removed and each plate was washed again in 5 ml ice cold 1x PBS for 5s. Just prior to use, 10 µl of 100 mM PMSF was added to 2 ml PBS to make the complete cell scraping solution, with 2 ml then transferred to each plate. Cells were scrapped off with a cell scraper and pipetted in to a 15 ml tube. Tubes were centrifuged at 720 x *g* for 10 minutes at 4°C. The supernatant were discarded and 1 µl 100mM PMSF and 1 µl PIC were added to each pellet. Pellets were stored at -80°C.

2.9.2 Shearing by sonication

The formaldehyde fixed cell pellets were thawed on ice. Each pellet was resuspended in 1 ml lysis buffer supplemented with 5 µl PMSF and 5 µl PIC (pellet from 5x 10 cm plates), and left on ice for 30 minutes. After the cell lysis step, the cells were dounced 30 times with a dounce homogenizer using a tight pestle. The lysed cells were transferred to new Eppendorf tubes and centrifuged at 2,500 x *g* for 10 minutes at 4°C. The supernatant was then removed and discarded. Each pellet of nuclei was resuspended in 350 µl of shearing buffer supplemented with 1.75 µl PMSF and 1.75 µl PIC. The samples were placed in to a pre-cooled heat block embedded in ice. The sonication probe was submerged into the sample to 5 mm above the base of the tube. The samples were sonicated at 25% amplitude for 30s on/30s off for 15 minute (30 minute total time). Samples were then centrifuged at full speed in a microcentrifuge (5917R) for 2 minutes at 4°C to pellet cell debris. The supernatant (chromatin) was removed and transferred to a fresh tube and 50 µl was taken to check the sonication efficiency. The remaining chromatin was stored at -80°C.

2.9.3 Immunoprecipitation

For each sample, 10 µl aliquots were taken and transferred to a fresh tubes to process later as input DNA. Each ChIP reaction was set up in siliconized tubes and each containing 7-25 µg chromatin, 25 µg protein G magentic

beads, 1 µl PIC, 10 µl ChIP buffer 1, 5 µg of target antibody or non-specific IgG antibody, made up to a total voulme of 100 µl with ddH_20. Tubes were placed on to an end-to-end rotator at 4°C O/N. Tubes were spun briefly to collect any liquid from the lids.

2.9.4 Magnetic bead wash

Tubes were placed on to a magentic tube rack to alllow the beads to pellet on to the side of the tubes. Beads were washed once with 800 µl ChIP buffer 1, pipetting up and down ten times. ChIP buffer 1 was aspirated then 800 µl of ChIP buffer 2 was added and pipetted up and down 10 times. The wash step was repeated once more with ChIP buffer 2. After the final wash step the remaining supernatant was aspirated using a 200 µl pipette. It was important to ensure that the beads did not dry out in between wash steps.

2.9.5 Chromatin elution, cross-link reversal and proteinase K treatment

The beads were resuspended in 50 µl of Elution Buffer AM2 and incubated on an ene-end rotator for 15 minutes at RT, then tubes were briefly spun. 50 µl of Reverse Cross-linking Buffer was added to the eluted chromatin and tubes were placed back on the maganetic rack to separate the beads from the chromatin (supernatant). The supernatant was transferred to a new Eppendorf tube. The 10 µl of input DNA previously set aside was processed by adding 88 µl of ChIP buffer two and 2 µl 5M NaCl to make a total volume

of 100 µl. Both the ChIP and input DNA were incubated for 65°C. Tubes were spun briefly then 2 µl of proteinase K was added and incubated at 37°C for 1 hour. Proteinase K stop solution was placed at RT 1 hour prior to use. Finally 2 µl of proteinase K stop solution was added to each sample. DNA was then stored at -20°C until required.

2.9.6 Quantitation

Fold binding over IgG= $(2^{\Delta Ct}$ target)/ $(2^{\Delta Ct}$ IgG)

Where: ΔCt target= input Ct – target Ct

ΔCt IgG= input Ct – IgG Ct

Figure 18. Calculation of fold binding over IgG.

Chapter 3: CTCF binds to the E2 open reading frame of high-risk HPV genomes

3 Chapter 3: CTCF binds to the E2 ORF of high-risk HPV genomes

3.1 Introduction

Previous research groups have identified and confirmed CTCF binding sites in the genomes of large DNA viruses, including EBV (Chau et al., 2006, Chen et al., 2013a, Tempera et al., 2010), HSV-1 (Lang et al., 2017, Amelio et al., 2006), CMV (Martinez et al., 2014) and KSHV (Kang et al., 2011, Stedman et al., 2008). Another group has also documented CTCF binding sites in the smaller adenovirus genome (Komatsu et al., 2013). The initial aim of this study was to determine if there were CTCF binding sites present within HPV genomes.

Previously in the Parish lab *in silico* screening methods were utlised to identify potenital CTCF binding sites in the genomes of both high and low-risk HPV genomes (Paris et al., 2015). The most striking observation was the identification of a predicted CTCF binding site around nucleotide 3000 in the E2 ORF of the three most prevalant high-risk types HPV16, 18 and 31. Interestingly, this predicted motif was not present in the two most common low-risk types HPV6 and 11. Therefore it was hypothesized that CTCF binding to this site in the E2 ORF is specifically important for the life cycle of the oncogenic high-risk HPV types. With this in mind it was important to determine if this CTCF binding site is conserved amongst the World Health Organisation's (WHO) 13 most high-risk HPV types; 16, 18, 31 33, 35, 39, 45,

51, 52, 56, 58, 59 and 68. Furthermore, experiments were carried out to determine if overall CTCF protein expression is upregulated in the presence of HPV episomes. Most importantly, by using chromatin immunoprecipitiation (ChIP) it was important to determine if CTCF protein binds to the HPV16 and HPV18 genomes *in vivo* at the predicted binding site in the E2 ORF. Furthermore, given that the viral life cycle is tightly linked to the differentiation of the host epithelium it was of interest to determine if CTCF binding is altered upon keratinocyte differentiation.

3.2 Generation of primary HFKs containing episomal HPV genomes

In order to study the interaction between CTCF and the HPV genome *in vivo* the primary human foreskin keratinocyte (HFK) cell model was utilised (Fig. 19), as this is one the most physiologically relevant cell systems for studying the HPV life cycle. Primary HFKs were first isolated from newborn foreskin tissue by Dr Sally Roberts, with previously obtained ethical consent (donor 2) (Ethics number: Roberts-06/Q1702/45). Additionally, a commerically available keratinocyte cell line was also purchased from Lonza (donor 1), to be used alongside donor 2. The HFKs were co-transfected with re-circularised HPV genomes and a plasmid carrying the neomycin resistant gene, followed by a G418 antibiotic selection process. Typically one month after selection keratinocyte colonies begin to appear on the cell culture plates which are then pooled and further expanded. The transfection of the HPV genomes in to

keratinocytes extends the lifespan of these primary cells than would normally occur in untransfected lines, thus allowing extended cell culture and generation of organotypic raft cultures. In this study a HFK cell line containing WT HPV18 episomal genomes was generated by Dr Jo Parish. Additionally, a HFK cell line containing WT HPV18 genomes with a three base pair mutation in the predicted CTCF binding site in the E2 region was also generated (ΔCTCF) (Paris et al., 2015). To account for possible donor variability, the transfection of either WT or ΔCTCF genomes was repeated in a second independent HFK cell donor (Dr Jo Parish). Once the cell lines were established the correct genome sequences were verified by DNA sequencing, and the presence of viral episomes was confirmed by Southern blot. Typically, transfected cells were not used beyond passage 12 as there is an increased occurence of viral genome integration beyond this passage. Therefore the primary cells used in subsequent experiments were used below passage 12—unless otherwise stated.

3.3 Confirmation of viral episomes in primary HFK cell lines

After generating the WT HPV18 and ΔCTCF HPV18 genome containing HFKs it was important to verify that the viral genomes were maintained episomally in both HFK cell lines. Initially, DNA was extracted from WT and ΔCTCF HFK monolayer cultures and a Southern blot was performed to detect viral episomes and estimate viral copy numbers. DNA extracted from WT

HPV18 and ΔCTCF HPV18 DNA from two individual donors was either digested with *Eco*RI, which linearizes the viral genome or with *Bgl*II restriction enzyme, which is a non-cutter of the viral genomes but cuts the host genome to allow identification of integrated genomes. The 8kb linearized viral genome (L) was detected in both WT HPV18 and ΔCTCF HPV18 HFKs from two independent donors (Fig. 20). Viral episomes were estimated to be between 25-50 copies per cell in donor 1 and approximately 50 copies per cell in donor 2. Uncut viral episomes are separated on an agarose gel as open circle (OC) and supercoiled (SC) bands, which were detected in both cells donors (Fig. 20). No differences were detected between WT and ΔCTCF HPV18 copy number in either donor. Of note, higher molecular weight bands that would be indicative of integrated HPV in to the host DNA were not detected in any of the cell lines (Fig. 20).

3.4 Identification of CTCF binding sites in high-risk HPV18 genomes

Firstly, the CTCFBSDS prediction tool was used to identify potential CTCF binding sites within the genomes of the 13 most high-risk HPV types (16, 18, 31 33, 35, 39, 45, 51, 52, 56, 59 and 68), as classified by the WHO (Table. 10). Previous studies have identified and mapped core sequences that bind to the CTCF protein and these core motifs have been represented by position weight matrices (PWM) (Bao et al., 2008, Ziebarth et al., 2013). The PWM scores were generated to determine the log odds of the observed sequence being

generated by the CTCF motif, with a score above three indicative of a match. Only the highest scoring sequences are displayed by the prediction tool. Ten of the high risk genomes—with the exception of HPV35 and 51—generated scores above three for potential CTCF binding sites within the E2 ORF between nucleotides 2700 and 2900. Moreover, HPV35 and HPV51 also displayed predicted binding sites within the E2 ORF but these were at postions 3365 and 3457, respectively. Interestingly, the predicted CTCF binding site in the E2 ORF between nucletodes 2800-3000 was not detected in the most prevelant low-risk types HPV6 and HPV11, or in the additional low-risk types 42 and 44. Furthermore, given that the beta HPV38 virus has also displayed transforming activity, it was interesting to determine if a CTCF binding site was also present in the HPV38 E2 ORF. A CTCF binding site in the E2 ORF of HPV38 was detected around nucleotide 3500, which is dissimlar to the conserved site observed in the high-risk alpha HPV types. Overall these data indicate that the CTCF binding site located at approximately 3000bp in the HPV genome is conserved amongst the high-risk HPV types.

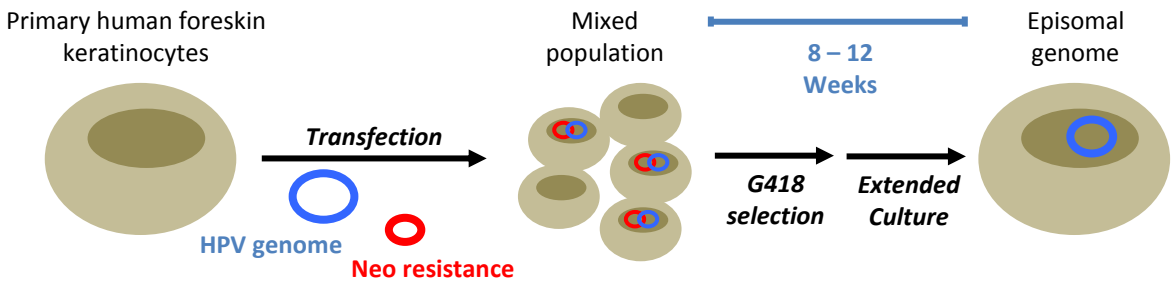


Figure 19. Generation of HFK cell lines containing episomal HPV genomes. Primary HFKs were transfected with re-circularized HPV genomes along with neomycin resistant plasmids. One day after transfection, the cells were grown on irradiated J2-3T3 feeder cells, followed by neomycin selection over a period of 8 days. Typically, one month after selection keratinocyte colonies would begin to appear on the cell culture plates, and would be pooled and further expanded over 8 weeks. Each cell should contain 50-100 of copies of the viral episomes. DNA sequencing confirmed the correct HPV genome sequence in each cell line and Southern blot analysis checked for viral integration and confirmed the presence of viral episomes.

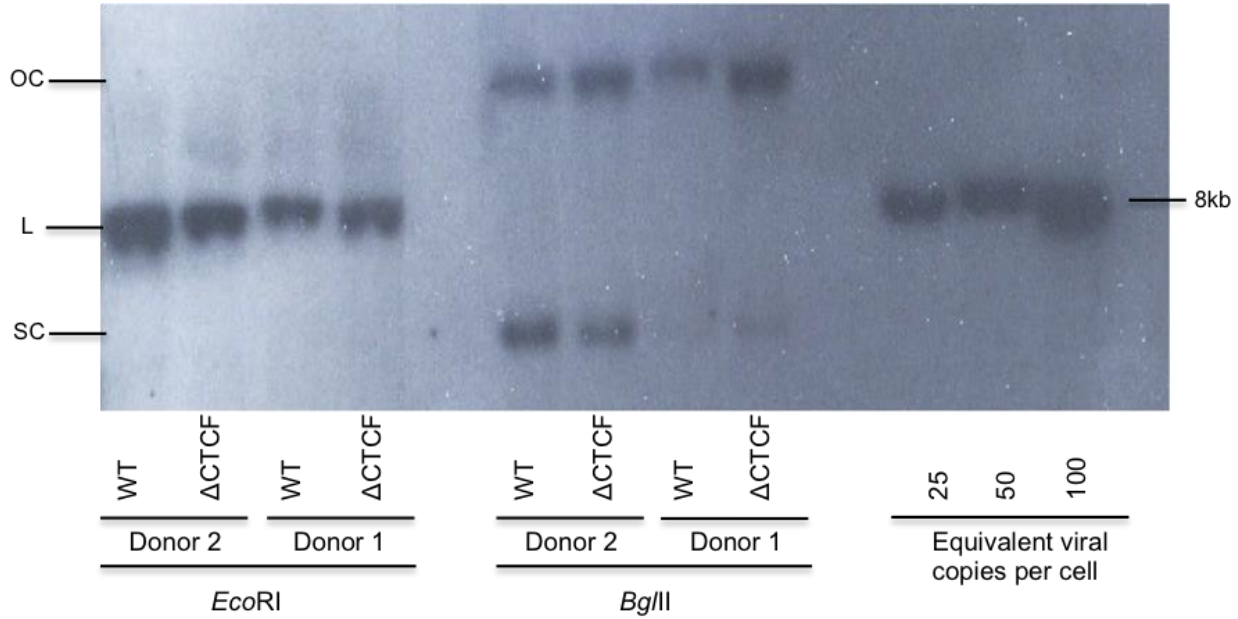


Figure 20. HPV18 genome status detected by Southern blotting. WT HPV18 and ΔCTCF HPV18 HFKs from two individual donors were grown in monolayer culture and DNA was extracted by phenol chloroform and ethanol precipitated. A Southern blot was performed using 5 μg DNA per condition. DNA was digested with *Eco*RI or *Bgl*II restriction enzymes and separated on a 0.8% agarose gel containing ethidium bromide. Copy number controls equivalent to 25, 50 and 100 viral episomes per cell were also separated on the agarose gel. DNA was transferred to a nitrocellulose membrane overnight through capillary action. The membrane was UV cross-linked and hybridized with a HPV18 probe labelled with P^{32}. The membrane was visualized on film by autoradiography.

HPV Type	Motif start site	Motif length	Motif Sequence	Motif Orientation	Motif PWM	Score
16	2845	20	TAACCACCAGGTGGTGCCAA	+	MIT_LM23	20.0598
18	2919	20	AAACCACCAGGTGGTGCCAG	+	MIT_LM23	23.4527
31	2783	20	TAACCACCAGGTGGTGCCAG	+	MIT_LM23	27.4756
33	2839	20	ATGCCACCAGGTGGTGCCTT	+	MIT_LM23	26.104
35	3365	20	ACGACTTCGAGGGGGTACCG	+	REN_20	9.56434
39	2903	20	TGACCACCAGGTGGTGCCAA	+	MIT_LM23	29.648
45	2877	20	AAACCACCAGGTGGTGCCTC	+	MIT_LM23	22.2343
51	3457	19	CTAACACTGGAGGGCACCA	+	MIT_LM2	17.3419
52	2833	20	AGGCCACCAGGTGGTGCCAC	+	MIT_LM23	25.7446
56	2837	20	AAACCACCAGATGGTGCCTT	+	MIT_LM23	20.9417
59	2838	20	AAACCACCAGGTGGTGCCAA	+	MIT_LM23	24.0063

Table 10. Predicted CTCF binding sites in high-risk HPV genomes

The CTCFBSDS prediction tool was used to identify potential CTCF binding sites i*n silico*. Complete genomes of the 12 WHO classified high-risk HPV types were imported from the Papillomavirus Episteme database (https://pave.niaid.nih.gov) and entered in to the prediction tool. Scores were generated to determine the log odds of the observed sequence being generated by the CTCF motif. A score above three is indicative of a CTCF binding. The prediction tool displays the highest scores generated per input sequence, accompanied with the predicted motif sequence and nucleotide start site. Scores shown are the highest obtained for each HPV genome. HPV35 and 51 are shaded grey they display a binding site at a different E2 region compared to the other high-risk HPV types.

3.5 The expression of CTCF protein in HFK cell lines

Western blot analysis was carried out to determine whether overall CTCF protein expression was altered upon the presence of HPV viral episomes (Fig. 21a). Western blots from two independent cell donors displayed a two-fold increase in CTCF protein expression in HFKs containing WT HPV18 episomes when compared to untransfected HFKs (Fig. 21b). This observed increase in CTCF protein expression indicates that CTCF may be upregulated by the virus to facilitate processes during the viral life cycle. Given that the life cycle of the virus is linked to epithelial differentiation CTCF western blots were also performed on both monolayer and methylcellulose differentiated WT and ΔCTCF HFKs (Fig. 22a). Growing HFKs in semi-solid methylcellulose medium for 48 hours is a routinely used method to induce the uniform differentiation of keratinocytes, coupled with viral genome amplification and induction of late gene expression. The expression of CTCF was similar between WT HPV18 and ΔCTCF HPV18 genome containing HFKs grown in monolayer culture, and remained similar between the two cell lines after 48 hours methylcellulose differentiation (Fig. 22b). Interestingly the overall level of CTCF expression in WT and ΔCTCF genome containing cells was not altered upon methylcellulose differentiation, which is similar to that observed when HPV31 containing CIN612 cells were differentiated (Mehta et al., 2015). On the other hand untransfected HFKs displayed a reduction in CTCF protein

expression upon methylcellulose differentiation (Mehta et al., 2015). However methylcellulose does not induce terminal differentiation of keratinocytes so a change in CTCF protein expression is more likely to be seen in the organotypic raft culture model.

In order to visualise the expression of CTCF in the differentiated epithelium, organotypic raft sections derived from either untransfected or HPV18 genome containing HFKs were stained with CTCF antibody and visualsed by immunofluroesence. Similar to previous findings, CTCF staining was reduced upon cellular differentiation (Delgado et al., 1999). Both untransfected and WT HPV18 rafts exibit CTCF nuclear staining in the basal and suprabasal layers, however there is a reduction in CTCF in some of the cells in the upper most granular layers of the rafts where the cells are fully differentiated. This was obsreved in two independent donors (Figs. 23 & 24).

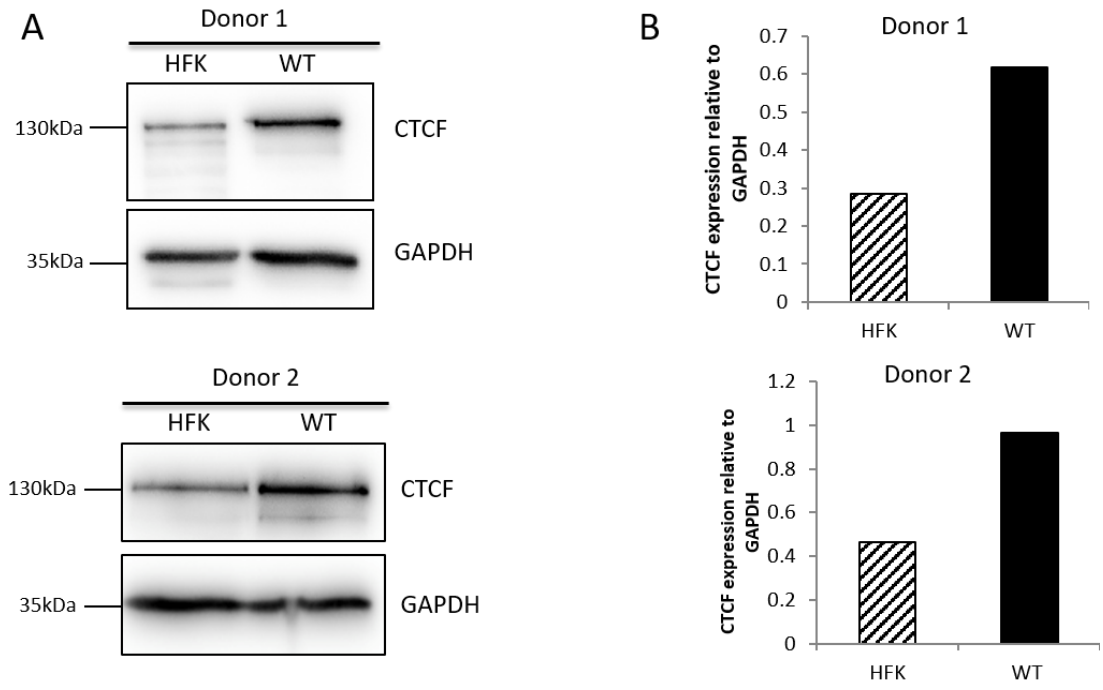


Figure 21. CTCF protein expression in untransfected and HPV18 genome containing HFKs. Protein lysates were extracted from monolayer untransfected HFKs and WT HPV18 genome containing HFKs from two cell donors. **A)** Western blots were carried out using 40μg of protein lysate to detect CTCF expression and GAPDH was used as a loading control. **B)** Western blot bands were quantified using the Fusion FX digital detection system. The expression of CTCF in each sample was calculated relative to GAPDH. The westerns shown are representative blots from two independent HFK donors.

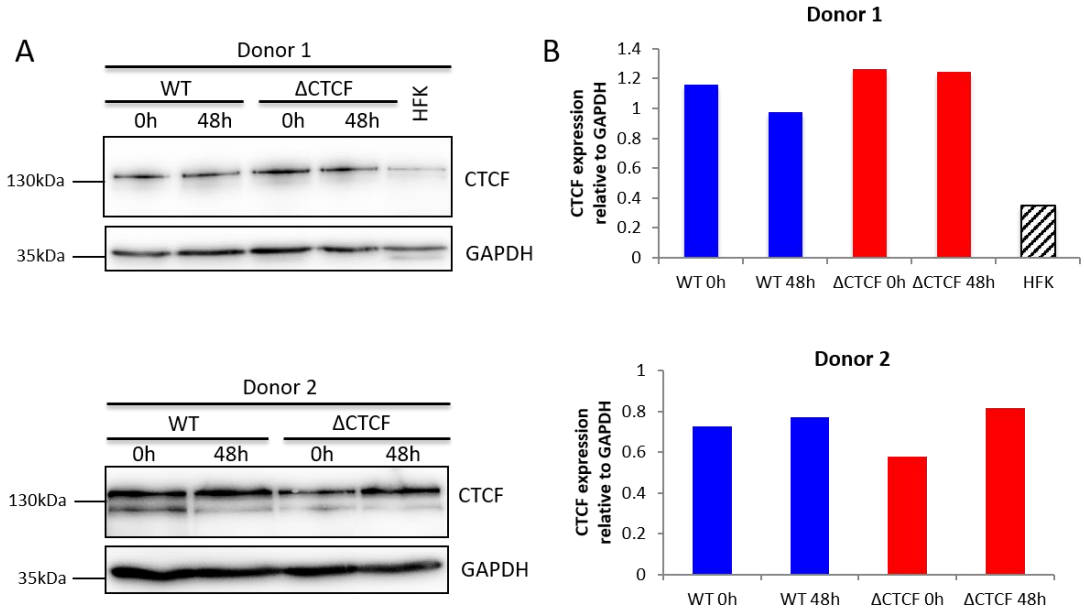


Figure 22. CTCF protein expression in monolayer and differentiated HFKs. Protein lysates were extracted from monolayer untransfected HFKs, WT HPV18 and ΔCTCF HPV18 genome containing HFKs grown in monolayer culture or differentiated in methylcellulose. **A)** Western blots were carried out using 30μg of protein to detect CTCF. GAPDH was used as a loading control. **B)** Western blot bands were quantified using the Fusion FX digital detection system. CTCF expression in each sample was calculated relative to GAPDH loading control. The westerns shown are representative blots from two independent HFK donors.

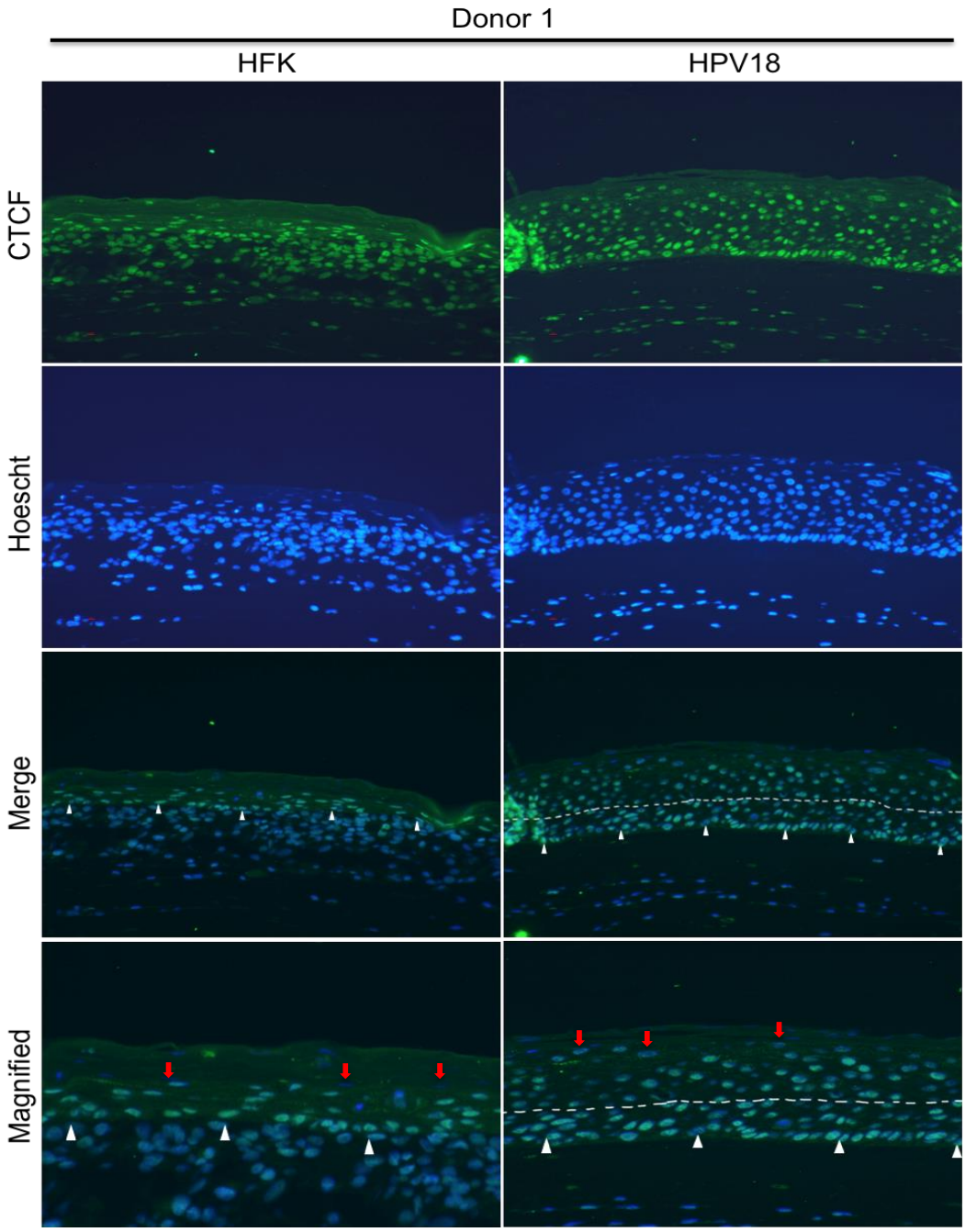


Figure 23. Immunofluorescence staining of CTCF in organotypic raft sections derived from untransfected HFKs and WT HPV18 genome containing HFKs. CTCF antibody was used at 1:5000 in BSA blocking solution with staining shown in the green panels. Hoescht 33342 was used to stain the nuclear DNA as shown in the blue panels. Slides were visualized on a Nikon inverted Epifluorescent microscope fitted with a 20x objective and images were captured using a Leica DC200 camera and software. The white arrows indicate the basal layer and the lower/upper suprabasal boundary is demarcated by the white dotted line. Images are representative of raft sections from cell donor 1.

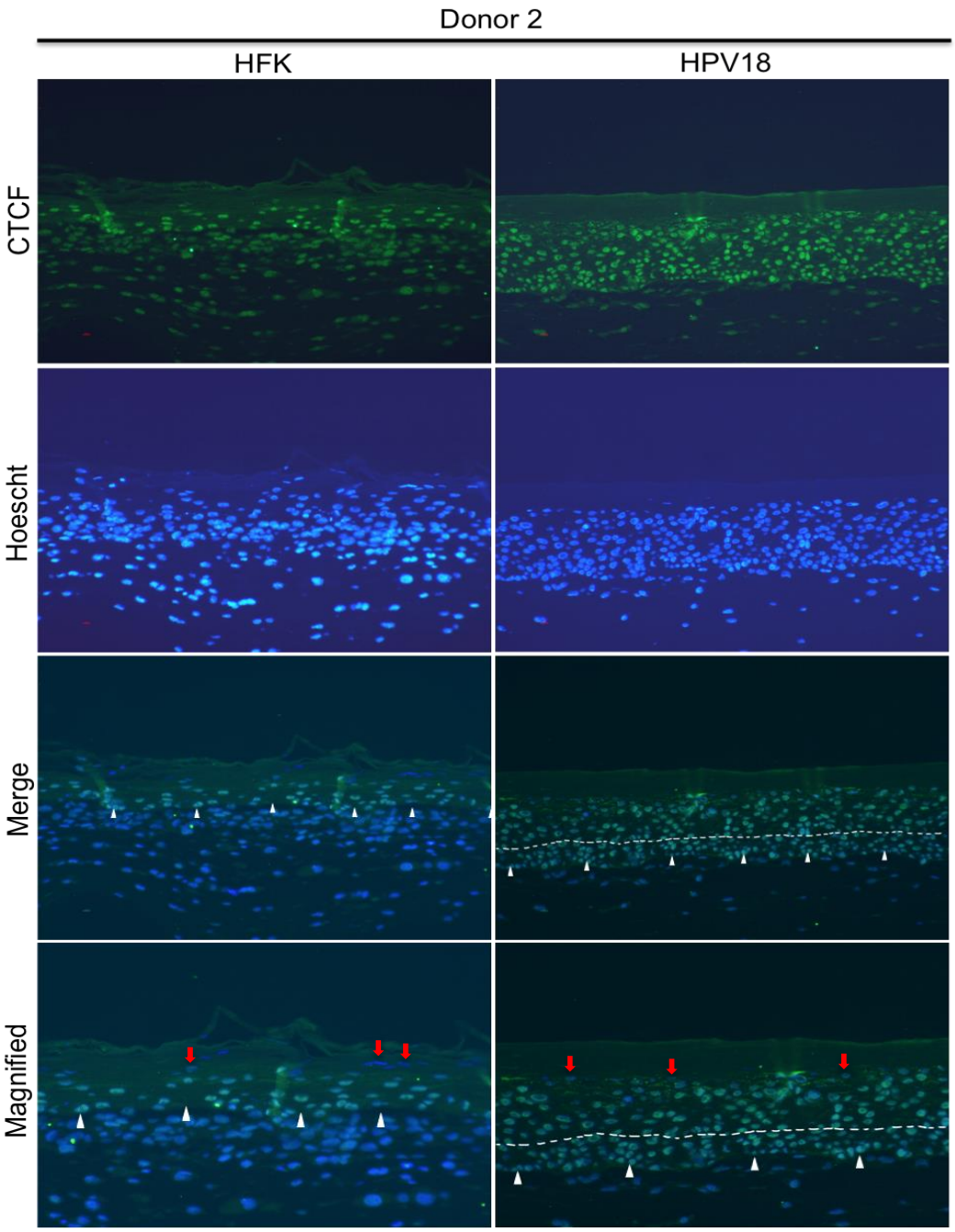


Figure 24. Immunofluorescence staining of CTCF in organotypic raft sections derived from untransfected HFKs and WT HPV18 genome containing HFKs. CTCF antibody was used at 1:5000 in BSA blocking solution with staining shown in the green panels. Hoescht 33342 was used to stain the nuclear DNA as shown in the blue panels. Slides were visualized on a Nikon inverted Epifluorescent microscope fitted with a 20x objective and images were captured using a Leica DC200 camera and software. The white arrows indicate the basal layer and the lower/upper suprabasal boundary is demarcated by the white dotted line. Images are representative of raft sections from cell donor 2.

3.6 Association of CTCF protein to high-risk HPV genomes

ChIP experiments were performed to determine if CTCF associates with the HPV18 genome in primary HFKs containing the transfected HPV18 viral episomes. ChIPs were performed on cells either grown in monolayer culture or differentiated in methylcellulose for 48 hours. Nuclei from the HFKs were pelleted and chromatin was sheared by sonication. DNA was extracted from chromatin samples and separated on an agarose gel to check the shearing efficiency, typically yielding fragments between 200 and 600bp (Figs. 25 & 28). ChIPs were performed using a CTCF or FLAG IgG antibody, followed by qPCR on eluted DNA using primers spanning the HPV18 genome. In three independent experiments a distinct binding peak was observed around nucleotide 3000 in the WT HPV18 genome, confirming the previously predicted binding site in this E2 region. In the representative ChIP graph there was a 1.4 fold enrichment of CTCF binding at the E2 site WT HPV18 genome containing cells from donor 1 (Fig. 26). As expected, the binding peak in the E2 region was abrogated in ΔCTCF HPV18 monolayer cells (Fig. 27). When the ChIP was repeated in a second HFK cell donor there was a 3.5 fold enrichment of CTCF binding to the E2 ORF in WT HPV18 monolayer cells (Fig. 29). Although the ΔCTCF cells displayed 2 fold binding at the E2 site, this level of binding was the same across the whole genome, indicating that this is in fact background noise and the binding peak is clearly lost in these cell lines (Fig. 30). ChIPs were also performed on WT and ΔCTCF genome

containing HFKs after 48 hours methylcellulose differentiation. Upon keratinocyte differentiation the level of involucrin expression is increased and the expression of HPV18 late E4 protein is induced, therefore western blots for involucrin and HPV18 E4 were performed to verify cells had undergone differentiation (Fig. 31). Upon 48 hours methylcellulose differentiation the enrichment of CTCF binding to this site was higher than in the monolayer HFKs, with 14-fold enrichment observed (Fig. 32). As expected, the binding peak in the E2 region was lost in the ΔCTCF HPV18 genome containing cells after 48 hours methylcellulose differentiation (Fig. 33). Overall, these results demonstrate that CTCF binds to the E2 ORF around nucleotide 3000 in the HPV18 genome, and upon methylcellulose differentiation this binding is further enriched.

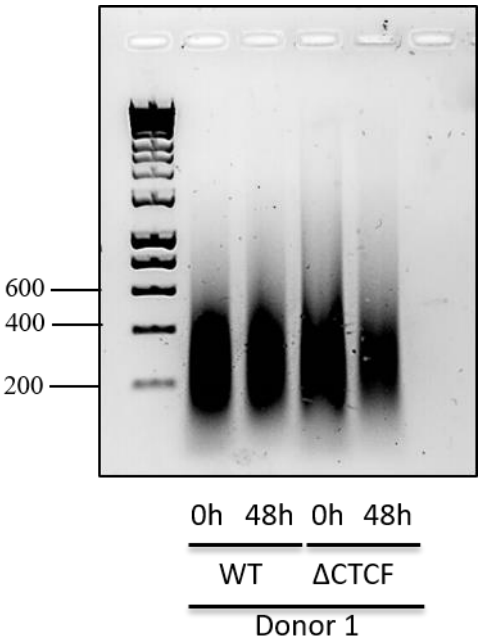


Figure 25. Sheared chromatin from HPV18 genome containing HFKs. Cells were fixed in 1% formaldehyde, lysed, and dounced with 40 strokes to aid nuclei release. Pelleted nuclei were resuspended in shearing buffer, then sonicated for 15 minutes on ice at 25% amplitude (30s on 30s off cycle). DNA was extracted from 50µl aliquots then separated on a 1.2% agarose gel containing ethidium bromide, and visualized on a Gene Flash UV light box to check shearing efficiency. Bioline 1kb DNA ladder was used as a molecular size marker. Optimal shearing typically yielded fragments between 200-600bp.

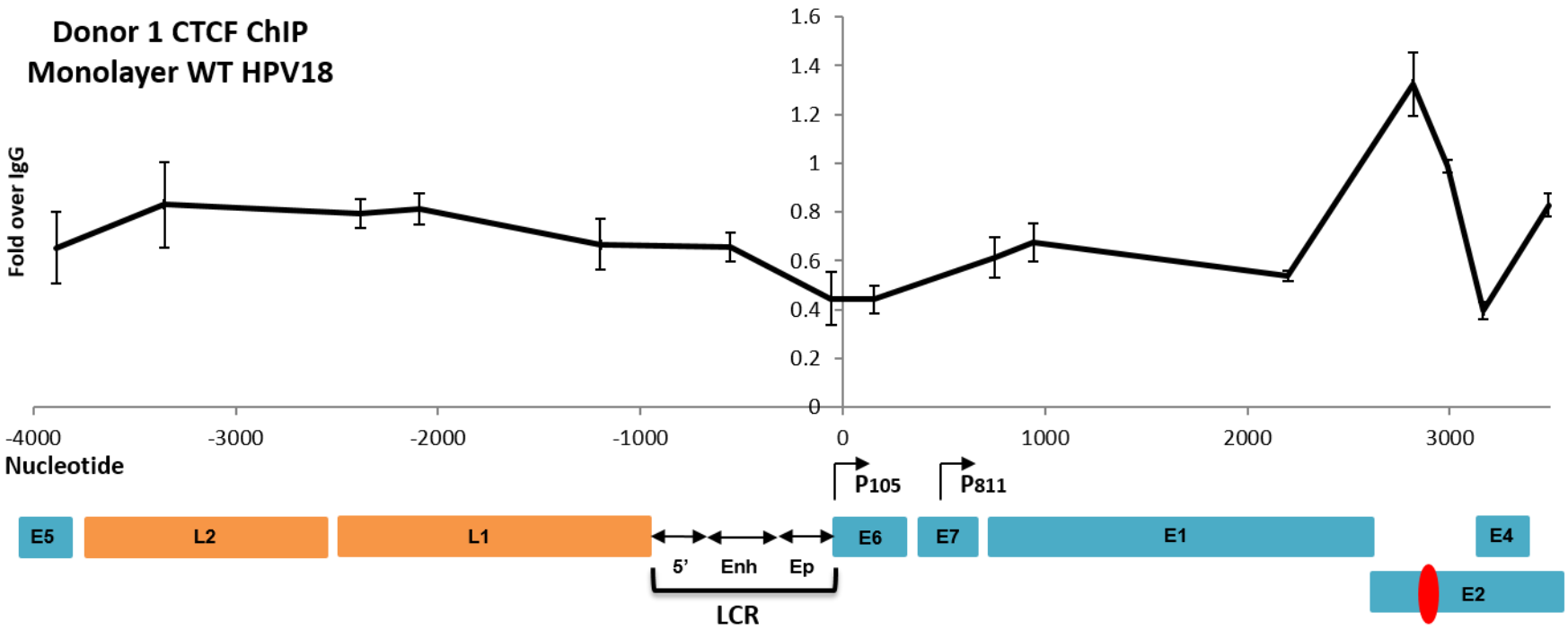


Figure 26. Association of CTCF with the WT HPV18 genome. CTCF ChIPs were performed using WT HPV18 genome contaning HFKs grown in monolayer culture. Cells were fixed in 1% formaldehyde, lysed, and dounced with 40 strokes to aid nuclei release. Chromatin was sheared by sonication on a cycle of 30s on and 30s for a total sonication time of 15 minutes. ChIP reactions were performed at 4°C overnight using 25µg of chromatin and 8µg of CTCF polyclonal antibody. ChIPs were also set up using FLAG as a negative control. DNA was de-crosslinked and eluted and used for qPCR analysis using primers spanning the HPV18 genome, with the mid-point of each primer displayed on the x-axis. The HPV18 genome is displayed below the x-axis in the linearized format for ease of presentation. The $2^{\Delta\Delta}$Ct method was used to calculate fold binding over FLAG ChIP (Y-axis) and error bars represent the s.d. for qPCR triplicate values from one individual experiment. The predicted CTCF binding site is displayed as a red marker. The viral LCR is comprised of the 5' section, enhancer (Enh) and the early promoter (Ep). The early transcripts are transcribed from P_{105} and the late transcripts from P_{811}. The ChIP graph shown is representative of three independent experiments for HFK cell donor 1.

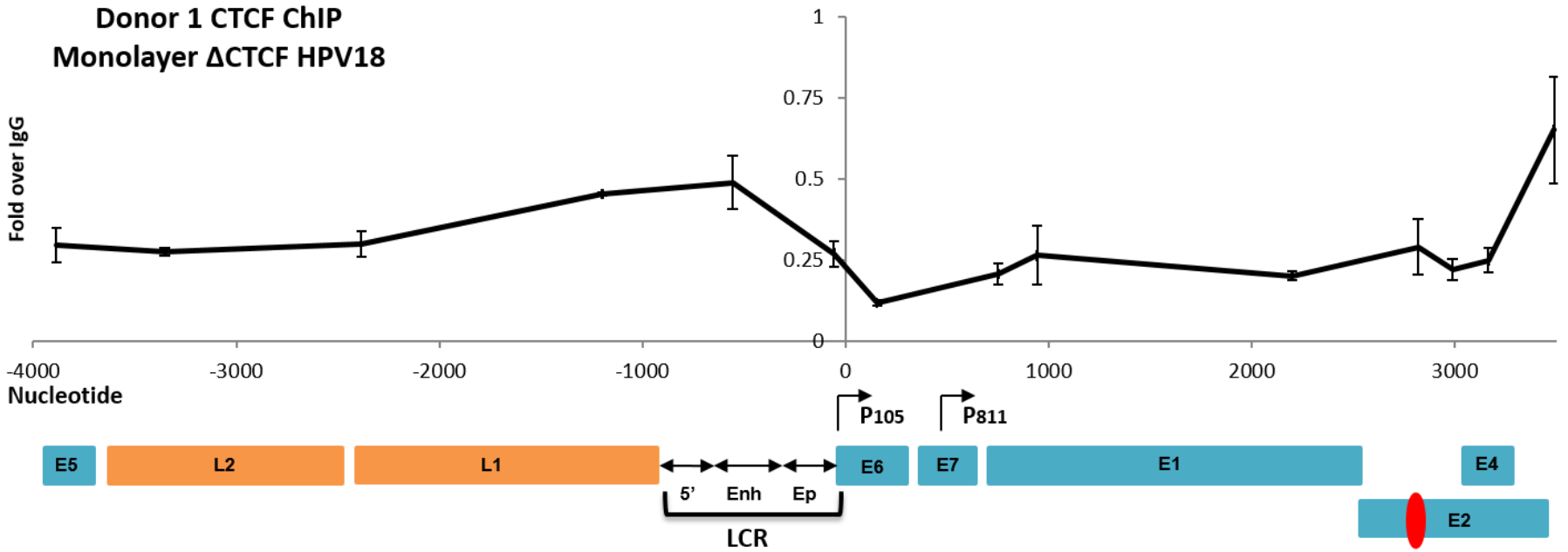


Figure 27. Association of CTCF with the ΔCTCF HPV18 genome. CTCF ChIPs were performed using ΔCTCF HPV18 genome containing HFKs grown in monolayer culture. Cells were fixed in 1% formaldehyde, lysed, and dounced with 40 strokes to aid nuclei release. Chromatin was sheared by sonication on a cycle of 30s on and 30s for a total sonication time of 15 minutes. ChIP reactions were performed at 4°C overnight using 25μg of chromatin and 8μg of CTCF polyclonal antibody. ChIPs were also set up using FLAG as a negative control. DNA was de-crosslinked and eluted and used for qPCR analysis using primers spanning the HPV18 genome, with the mid-point of each primer displayed on the x-axis. The HPV18 genome is dispalyed below the x-axis in the linearized format for ease of presentation. The $2^{\Delta\Delta}$Ct method was used to calculate fold binding over FLAG ChIP (Y-axis) and error bars represent the s.d. for qPCR triplicate values from one individual experiment. The predicted CTCF binding site is displayed as a red marker. The viral LCR is comprised of the 5' end, enhancer (Enh) and the early promoter (Ep). The early transcripts are transcribed from P_{105} and the late transcripts from P_{811}. The ChIP graph shown is representative of three independent experiments for HFK cell donor 1.

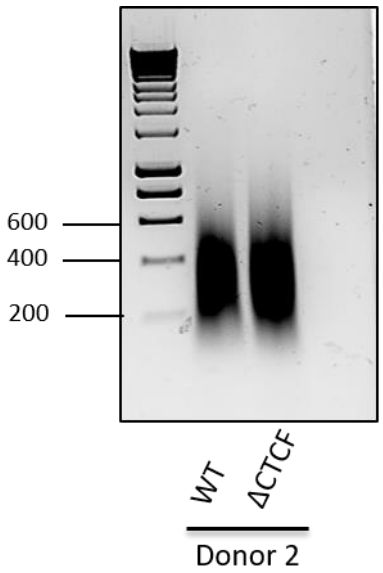


Figure 28. Sheared chromatin from WT HPV18 and ΔCTCF HPV18 genome containing HFKs. Cells were fixed in 1% formaldehyde, lysed, and dounced with 40 strokes to aid nuclei release. Pelleted nuclei were resuspended in shearing buffer, then sonicated for 15 minutes on ice at 25% amplitude (30s on 30s off cycle). DNA was extracted from 50µl aliquots then separated on a 1.2% agarose gel containing ethidium bromide, and visualized on a Gene Flash UV light box to check shearing efficiency. Bioline 1kb DNA ladder was used as a molecular size marker. Optimal shearing typically yielded fragments between 200-600bp.

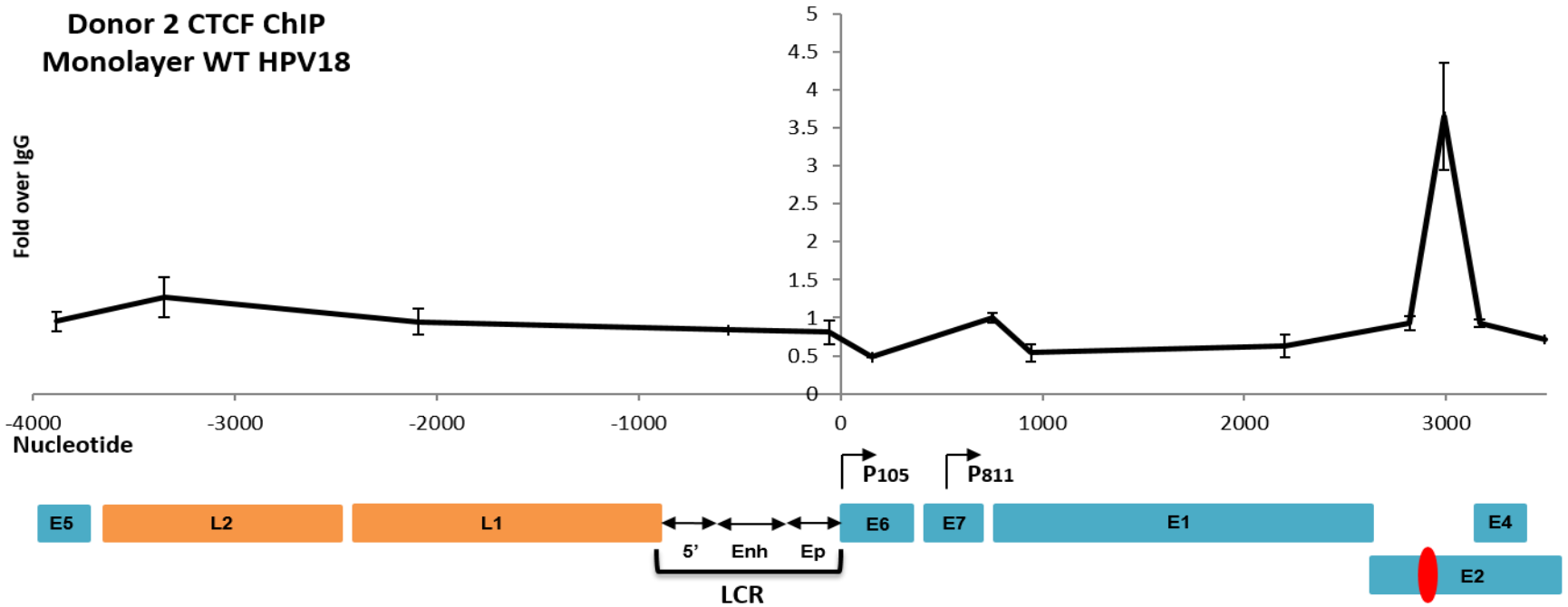


Figure 29. Association of CTCF with the WT HPV18 genome. CTCF ChIPs were performed using WT HPV18 genome containing HFKs grown in monolayer culture. Cells were fixed in 1% formaldehyde, lysed, and dounced with 40 strokes to aid nuclei release. Chromatin was sheared by sonication on a cycle of 30s on and 30s for a total sonication time of 15 minutes. ChIP reactions were performed at 4°C overnight using 25µg of chromatin and 8µg of CTCF polyclonal antibody. ChIPs were also set up using FLAG as a negative control. DNA was de-crosslinked and eluted and used for qPCR analysis using primers spanning the HPV18 genome, with the mid-point of each primer displayed on the x-axis. The HPV18 genome is dispalyed below the x-axis in the linearized format for ease of presentation. The $2^{\Delta\Delta}$Ct method was used to calculate fold binding over FLAG ChIP (Y-axis) and error bars represent the s.d. for qPCR triplicate values from one individual experiment. The predicted CTCF binding site is displayed as a red marker. The viral LCR is comprised of the 5' end, enhancer (Enh) and the early promoter (Ep). The early transcripts are transcribed from P105 and the late transcripts from P811. The ChIP graph shown is representative of one independent experiment for HFK cell donor 2.

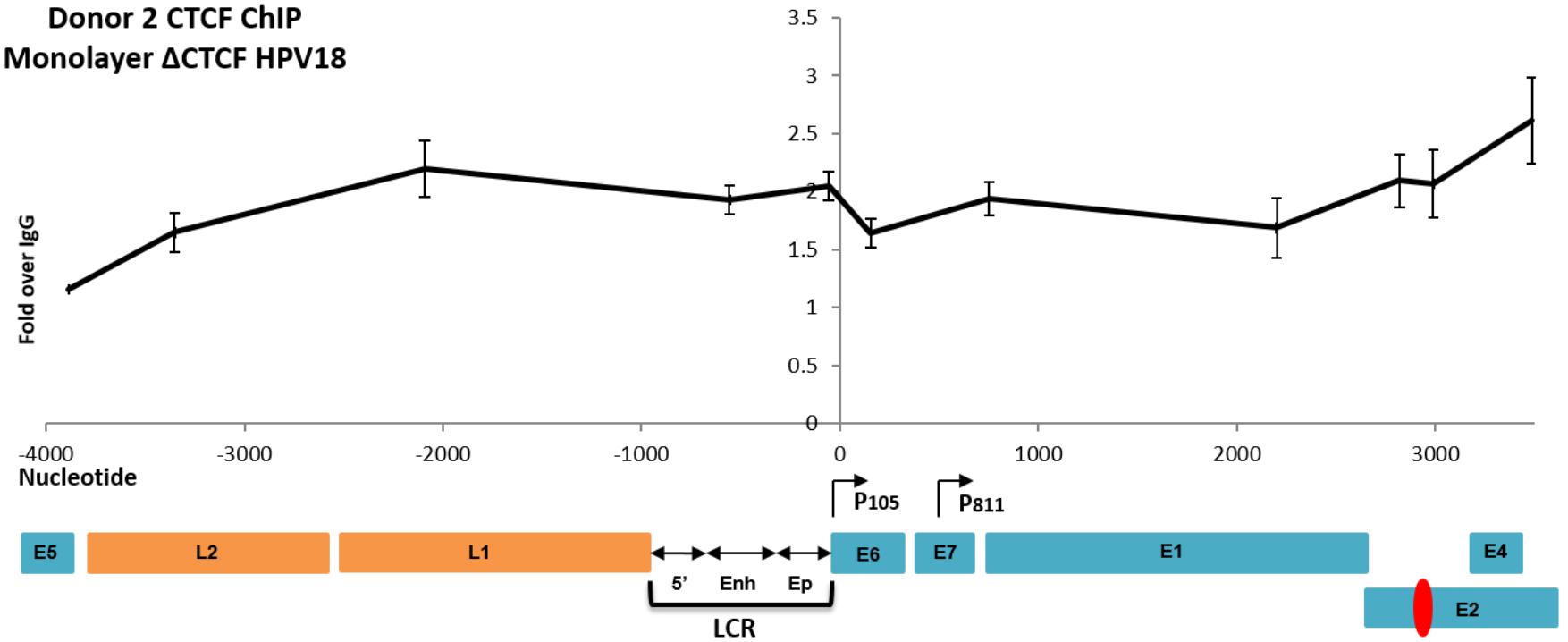


Figure 30. Association of CTCF with the ΔCTCF HPV18 genome. CTCF ChIPs were performed using ΔCTCF HPV18 genome containing HFKs grown in monolayer culture. Cells were fixed in 1% formaldehyde, lysed, and dounced with 40 strokes to aid nuclei release. Chromatin was sheared by sonication on a cycle of 30s on and 30s for a total sonication time of 15 minutes. ChIP reactions were performed at 4°C overnight using 25µg of chromatin and 8µg of CTCF polyclonal antibody. ChIPs were also set up using FLAG as a negative control. DNA was de-crosslinked and eluted and used for qPCR analysis using primers spanning the HPV18 genome, with the mid-point of each primer displayed on the x-axis. The HPV18 genome is dispalyed below the x-axis in the linearized format for ease of presentation. The $2^{\Delta\Delta}$Ct method was used to calculate fold binding over FLAG ChIP (Y-axis) and error bars represent the s.d. for qPCR triplicate values from one individual experiment. The predicted CTCF binding site is displayed as a red marker. The viral LCR is comprised of the 5' end, enhancer (Enh) and the early promoter (Ep). The early transcripts are transcribed from P105 and the late transcripts from P811. The ChIP graph shown is representative of one independent experiment for HFK cell donor 2.

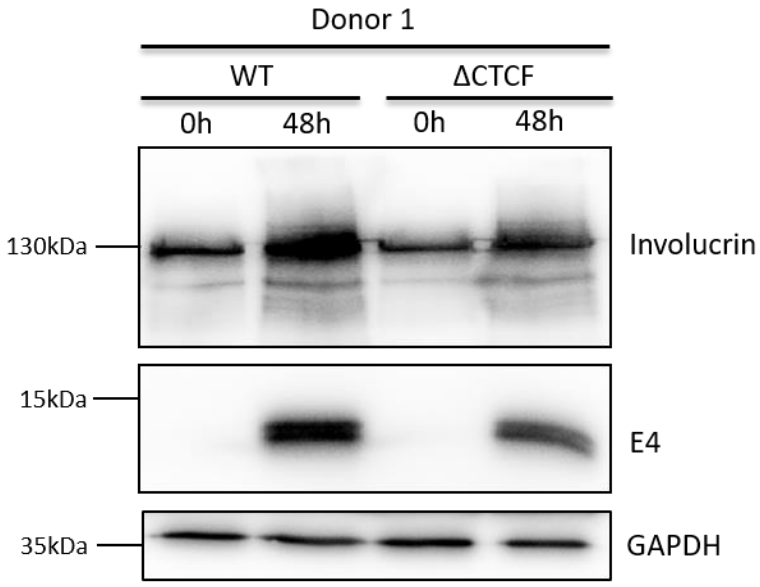


Figure 31. Involucrin and HPV18 E4 protein expression in monolayer and differentiated HPV18 HFKs. Protein lysates were extracted from WT HPV18 and ΔCTCF HPV18 genome containing HFKs grown in monolayer culture and after 48 hours methylcellulose differentiation. A western blot was carried out using 30 μg of protein lysate to detect HPV18 E4 and the differentiation marker involucrin. GAPDH was used as a loading control. The western shown is representative of three experimental repeats for donor 1.

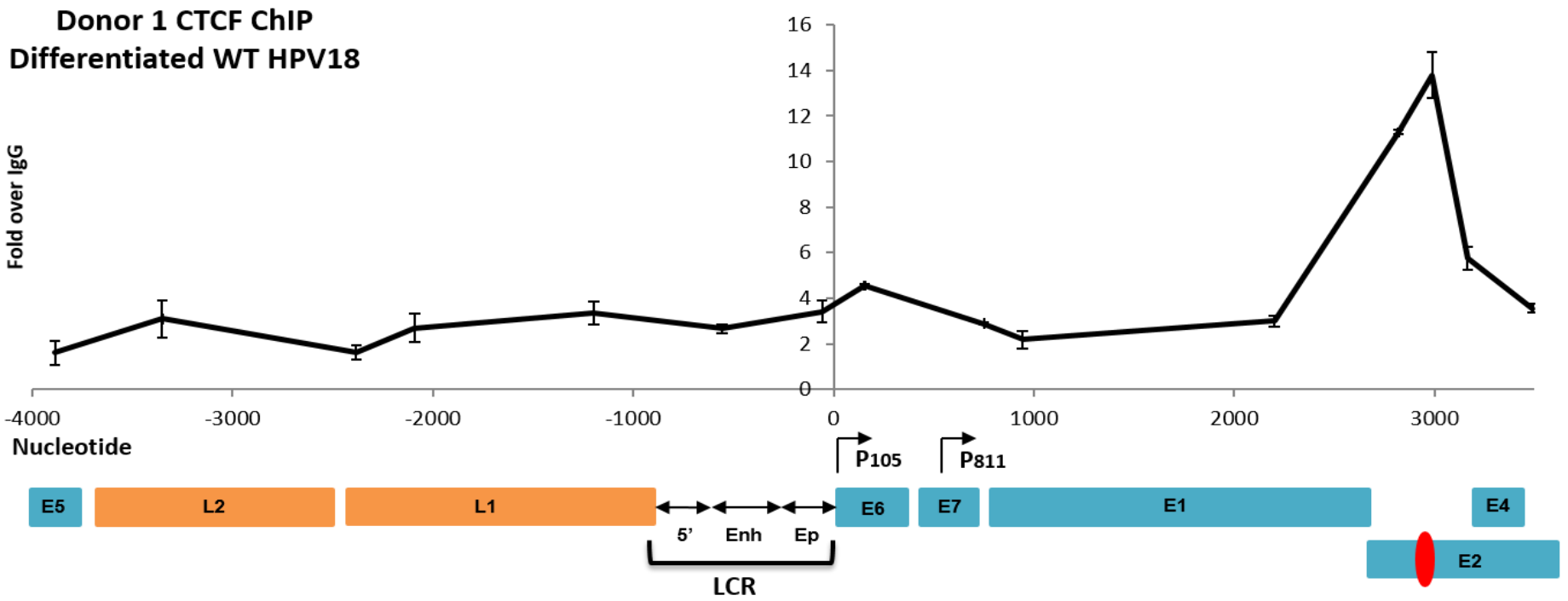


Figure 32. Association of CTCF with the WT HPV18 genome. CTCF ChIPs were performed using WT HPV18 genome containing HFKs differentiated in methylcellulose. Cells were fixed in 1% formaldehyde, lysed, and dounced with 40 strokes to aid nuclei release. Chromatin was sheared by sonication on a cycle of 30s on and 30s for a total sonication time of 15 minutes. ChIP reactions were performed at 4°C overnight using 25µg of chromatin and 8µg of CTCF polyclonal antibody. ChIPs were also set up using FLAG as a negative control. DNA was de-crosslinked and eluted and used for qPCR analysis using primers spanning the HPV18 genome, with the mid-point of each primer displayed on the x-axis. The HPV18 genome is dispalyed below the x-axis in the linearized format for ease of presentation. The $2^{\Delta\Delta}$Ct method was used to calculate fold binding over FLAG ChIP (Y-axis) and error bars represent the s.d. for qPCR triplicate values from one individual experiment. The predicted CTCF binding site is displayed as a red marker. The viral LCR is comprised of the 5' end, enhancer (Enh) and the early promoter (Ep). The early transcripts are transcribed from P105 and the late transcripts from P811. The ChIP graph shown is representative of three independent experiments for HFK cell donor 1.

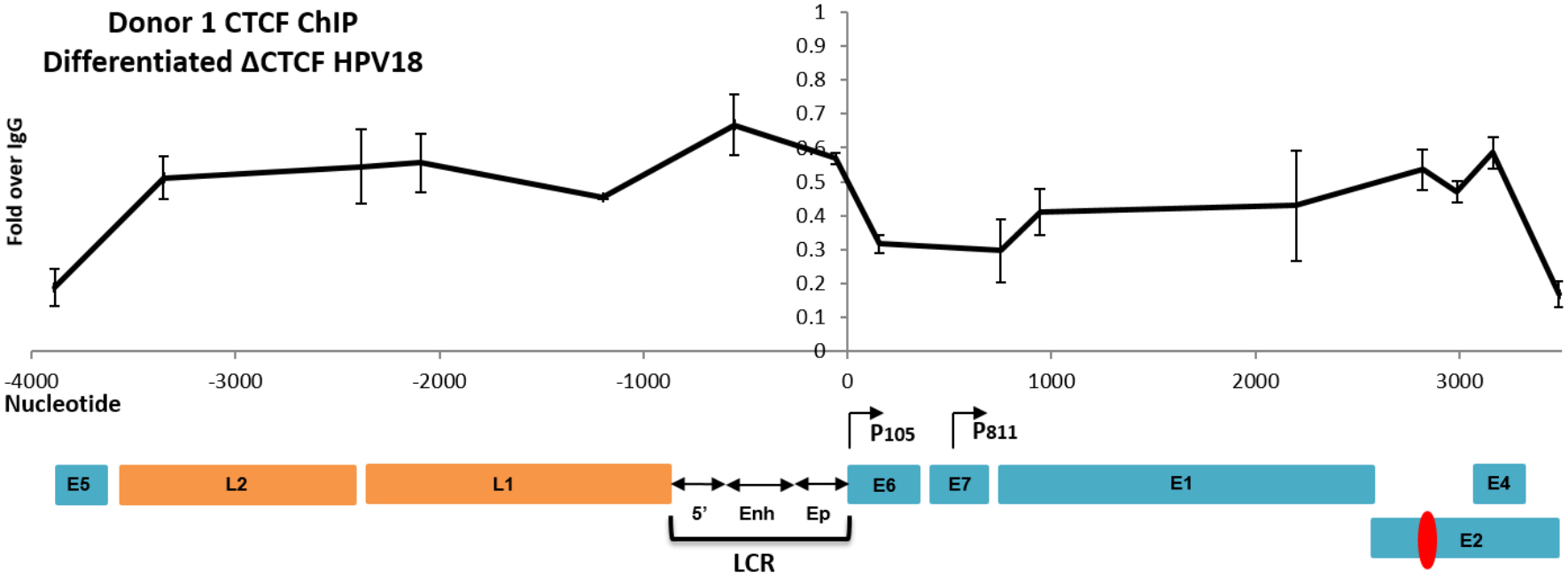


Figure 33. Association of CTCF with the ΔCTCF HPV18 genome. CTCF ChIPs were performed using ΔCTCF HPV18 genome containing HFKs differentiated in methylcellulose. Cells were fixed in 1% formaldehyde, lysed, and dounced with 40 strokes to aid nuclei release. Chromatin was sheared by sonication on a cycle of 30s on and 30s for a total sonication time of 15 minutes. ChIP reactions were performed at 4°C overnight using 25µg of chromatin and 8µg of CTCF polyclonal antibody. ChIPs were also set up using FLAG as a negative control. DNA was de-crosslinked and eluted and used for qPCR analysis using primers spanning the HPV18 genome, with the mid-point of each primer displayed on the x-axis. The HPV18 genome is dispalyed below the x-axis in the linearized format for ease of presentation. The $2^{\Delta\Delta}$Ct method was used to calculate fold binding over FLAG ChIP (Y-axis) and error bars represent the s.d. for qPCR triplicate values from one individual experiment. The predicted CTCF binding site is displayed as a red marker. The viral LCR is comprised of the 5′ end, enhancer (Enh) and the early promoter (Ep). The early transcripts are transcribed from P105 and the late transcripts from P811. The ChIP graph shown is representative of three independent experiments for HFK cell donor 1.

Similarly, HFKs containing episomal HPV16 were grown in monolayer or differentiated in methylcellulose for 48 hours and used for ChIP experiments with a CTCF or FLAG IgG antibody, followed by qPCR using primers spanning the HPV16 genome. DNA was extracted from chromatin samples and separated on an agarose gel to check the shearing efficiency, typically yielding fragments between 200 and 600bp (Fig. 34). Again, a distinct binding peak was observed around nucleotide 3000 in the WT HPV16 monolayer cells with approximately 14-fold binding at this site (Fig. 35). Interestingly, in the HPV16 genome containing monolayer HFKs CTCF binding peaks were also observed in the L1 (10 fold) and L2 (8 fold) gene regions, which was not observed in the HPV18 genome. There was also enrichment of CTCF binding across the viral LCR (8 fold) of the HPV16 genome. HFKs containing HPV16 were also grown in methylcelullose for 48 hours and a western blot for involucrin was performed to confirm cells had undergone differentiation (Fig. 36). CTCF enrichment was further increased to 50-fold when the cells had undergone differentiation (Fig. 37), reflecting the observation seen in WT HPV18 genome containing HFKS.

W12 cells derived from a low-grade cervical lesion and containing episomal HPV16 were also grown in monolayer or differentiated in methylcellulose for 24 and 48 hours, and used for CTCF ChIP experiments. As these cells are naturally infected with HPV16 they provide an additional model to confirm

CTCF binding to the E2 ORF. Chromatin was sheared as previously described (Fig. 39). A distinct binding peak was observed around nucleotide 3000 in the W12 monolayer cells, confirming the data obtained in the HPV16 and HPV18 HFK model system (Fig. 40). The cells were also differentiated in methylcelluose for either 24 or 48 hours and qPCR analysis was performed on the involucrin gene to confirm cells had undergone differentiation (Fig. 38). The binding of CTCF was enriched when cells had undergone differentiation at 24 hours and furthermore at the 48 hour time point (Fig. 40). Interestingly, distinct CTCF binding peaks were also observed within the LCR, L1 and L2 gene regions of the HPV16 genome upon cellular differentiation, which were previously observed in the HPV16 genome containing monolayer cells (Fig. 40).

Whilst the ChIP experiments have demonstrated CTCF binding at the E2 ORF in both HPV18 and HPV16 genome containing HFKs it was important to strengthen and validate these findings with an additional technique. ChIP-Seq couples ChIP with next generation DNA sequencing to determine genome-wide DNA binding sites for proteins. The main advantage over standard ChIP is the increased resolution of enrichment peaks across the genome obtained with ChIP-Seq. ChIP-Seq was performed using a CTCF antibody in HPV18 genome containing HFKs and untransfected HFKs to account for any background noise (in collaboration with Dr Adam Grundhoff, Heinrich Pette

Institute, Hamburg). ChIP-seq analysis confirmed the CTCF binding peak present in the WT HPV18 genome around nucleotide 3000, as previously identified by *in silico* screening and *in vivo* ChIPs, and there were no other distinct CTCF binding peaks across the HPV18 genome (Fig. 41).

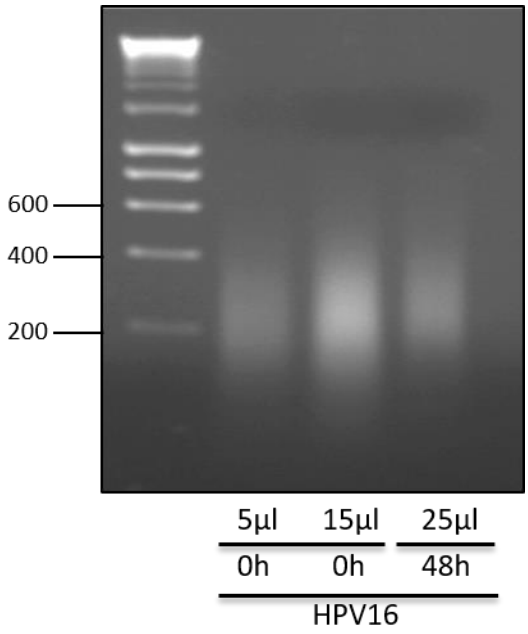


Figure 34. Sheared chromatin from HPV16 genome containing HFKs. Cells were fixed in 1% formaldehyde, lysed, and dounced with 40 strokes to aid nuclei release. Pelleted nuclei were resuspended in shearing buffer, then sonicated for 15 minutes on ice at 25% amplitude (30s on 30s off cycle). DNA was extracted from 50µl aliquots then separated on a 1.2% agarose gel containing ethidium bromide, and visualized on a Gene Flash UV light box to check shearing efficiency. Bioline 1kb DNA ladder was used as a molecular size marker. Optimal shearing typically yielded fragments between 200-600bp.

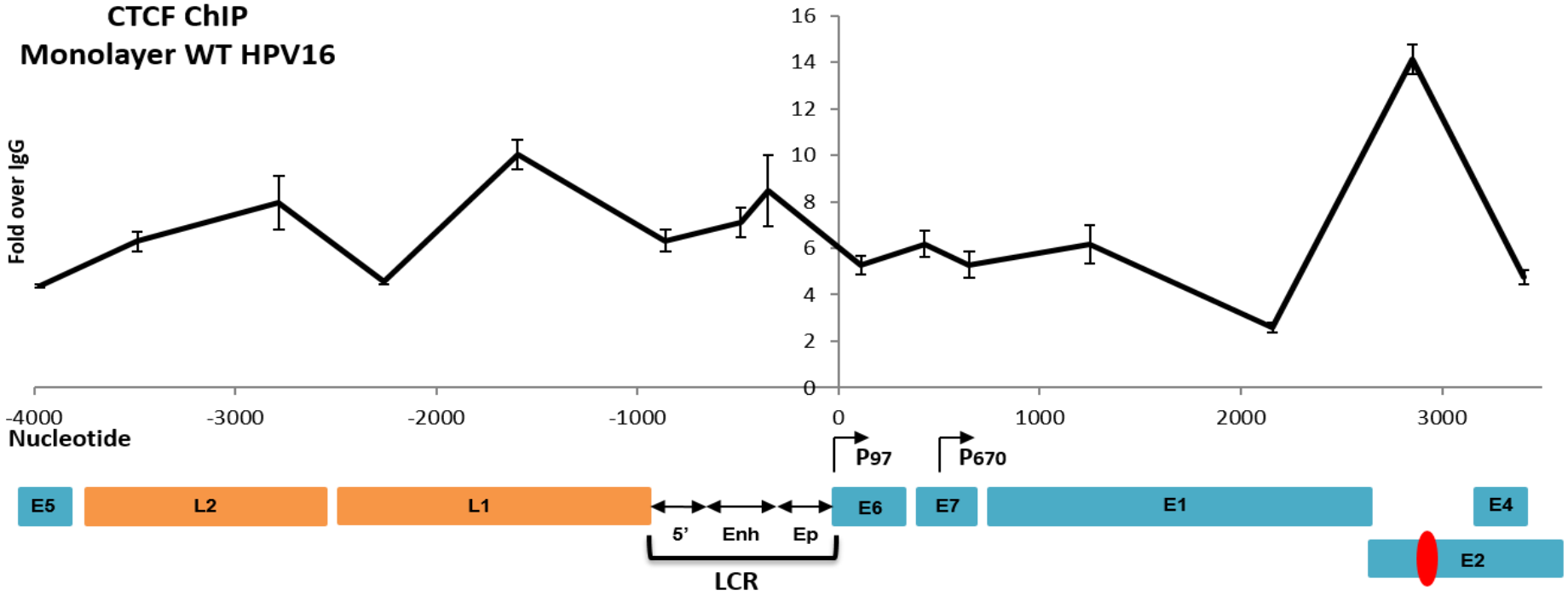


Figure 35. Association of CTCF with the HPV16 genome. CTCF ChIPs were performed using WT HPV16 genome containing HFKs grown in monolayer culture. Cells were fixed in 1% formaldehyde, lysed, and dounced with 40 strokes to aid nuclei release. Chromatin was sheared by sonication on a cycle of 30s on and 30s for a total sonication time of 15 minutes at 25% amplitude. ChIP reactions were performed at 4°C overnight using 25µg of chromatin and 8µg of CTCF polyclonal antibody. ChIPs were also set up using FLAG IgG as a negative control. . DNA was de-crosslinked and eluted and used for qPCR analysis using primers spanning the HPV18 genome, with the mid-point of each primer displayed on the x-axis. The HPV16 genome is dispalyed below the x-axis in the linearized format for ease of presentation. The $2^{\Delta\Delta}$Ct method was used to calculate fold binding over FLAG ChIP (Y-axis) and error bars represent the s.d. for qPCR triplicate values from one individual experiment. The predicted CTCF binding site is displayed as a red marker. The viral LCR is comprised of the 5' end, enhancer (Enh) and the early promoter (Ep). The early transcripts are transcribed from P97 and the late transcripts from P670. The ChIP graph shown is representative one independent experiment for HPV16 containing HFKs.

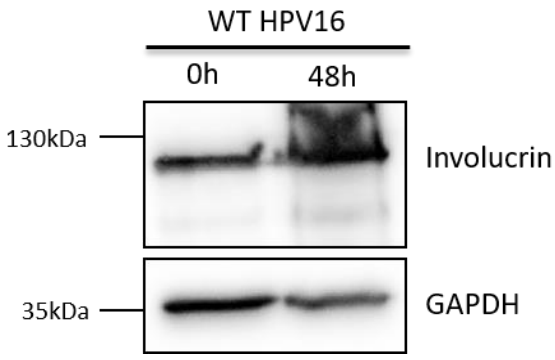


Figure 36. Involucrin protein expression in monolayer and differentiated HPV16 genome containing HFKs. Protein lysates were extracted from HPV16 containing HFKs in monolayer culture and after 48 hours methylcellulose differentiation. A western blot was carried out using 30μg of protein lysate to detect involucrin expression and GAPDH was used as a loading control.

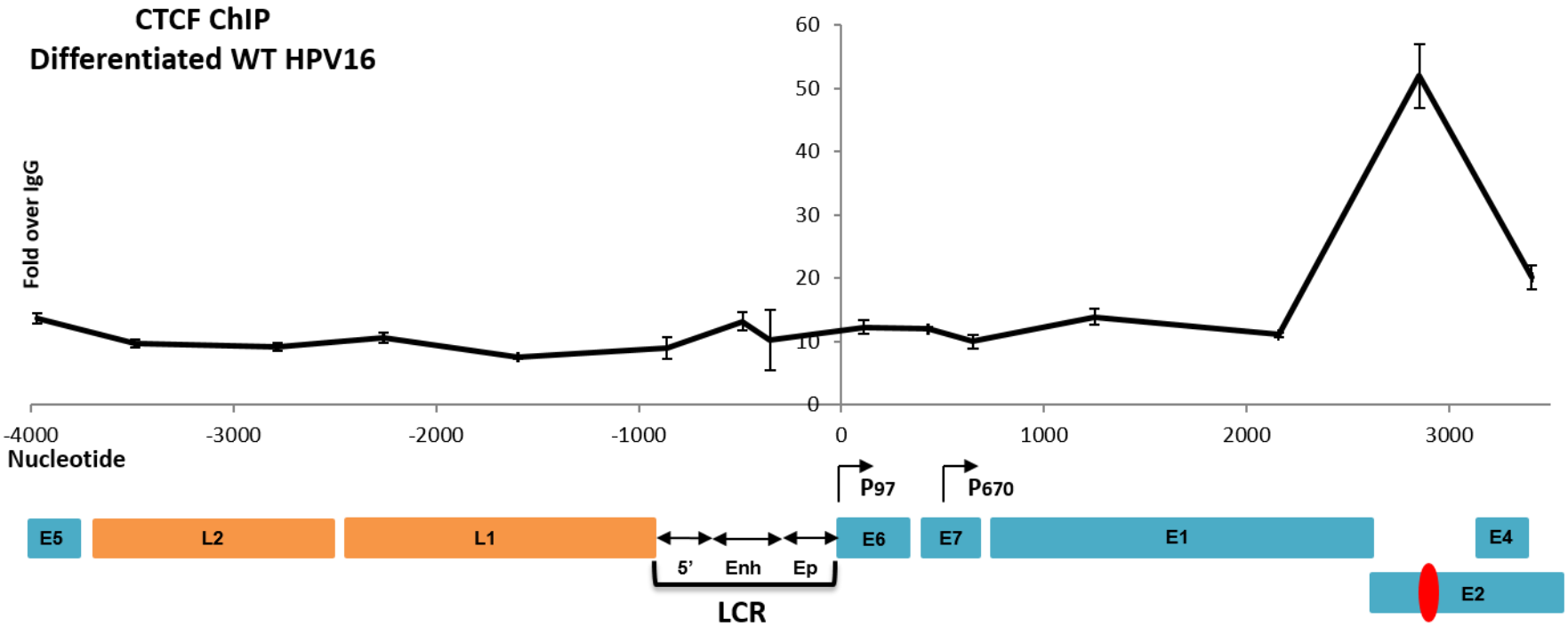


Figure 37. Association of CTCF with the WT HPV16 genome. CTCF ChIPs were performed using WT HPV16 genome containing HFKs differentiated in methylcellulose. Cells were fixed in 1% formaldehyde, lysed, and dounced with 40 strokes to aid nuclei release. Chromatin was sheared by sonication on a cycle of 30s on and 30s for a total sonication time of 15 minutes at 25% amplitude. ChIP reactions were performed at 4°C overnight using 25µg of chromatin and 8µg of CTCF polyclonal antibody. ChIPs were also set up using FLAG IgG as a negative control. DNA was de-crosslinked and eluted and used for qPCR analysis using primers spanning the HPV18 genome, with the mid-point of each primer displayed on the x-axis. The HPV16 genome is displayed below the x-axis in the linearized format for ease of presentation. The $2^{\Delta\Delta}$Ct method was used to calculate fold binding over FLAG ChIP (Y-axis) and error bars represent the s.d. for qPCR triplicate values from one individual experiment. The predicted CTCF binding site is displayed as a red marker. The viral LCR is comprised of the 5' end, enhancer (Enh) and the early promoter (Ep). The early transcripts are transcribed from P97 and the late transcripts from P670. The ChIP graph shown is representative one independent experiment for HPV16 containing HFKs.

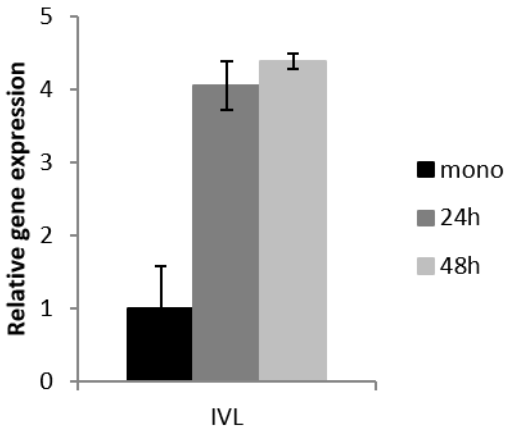


Figure 38. Involucrin gene expression in monolayer and differentiated W12 cells. RNA was extracted from monolayer W12 cells and W12 cells differentiated in methylcellulose for 24 or 48 hours. RNA was used to generate cDNA, followed by qPCR using primers within the involucrin gene. The $2^{\Delta\Delta}$Ct method was used to calculate gene expression relative to a housekeeping gene TLR2. Data are mean and s.d. of qPCR triplicates from one technical repeat.

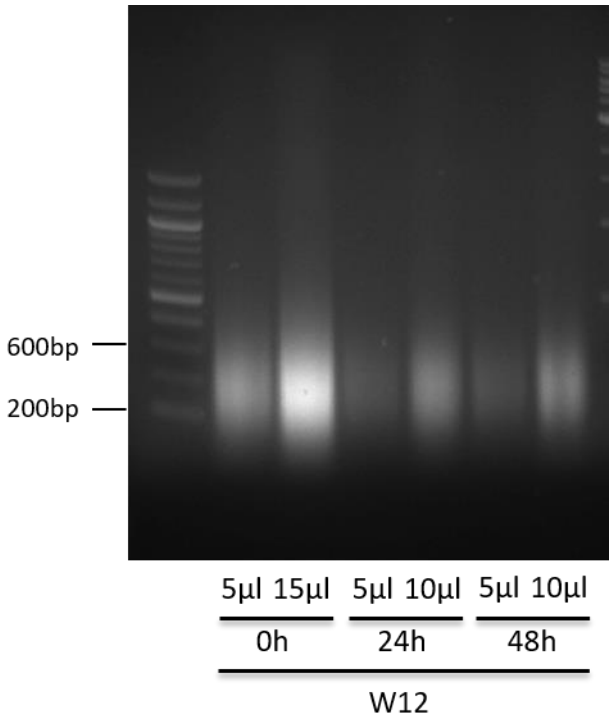


Figure 39. Sheared chromatin from HPV16 genome containing W12 cells. W12 cells were fixed in 1% formaldehyde, lysed, and dounced with 40 strokes to aid nuclei release. Pelleted nuclei were resuspended in shearing buffer, then sonicated for 15 minutes on ice at 25% amplitude (30s on 30s off cycle). DNA was extracted from 50µl aliquots then separated on a 1.2% agarose gel containing ethidium bromide, and visualized on a Gene Flash UV light box to check shearing efficiency. Bioline 1kb plus DNA ladder was used as a molecular size marker. Optimal shearing typically yielded fragments between 200-600bp.

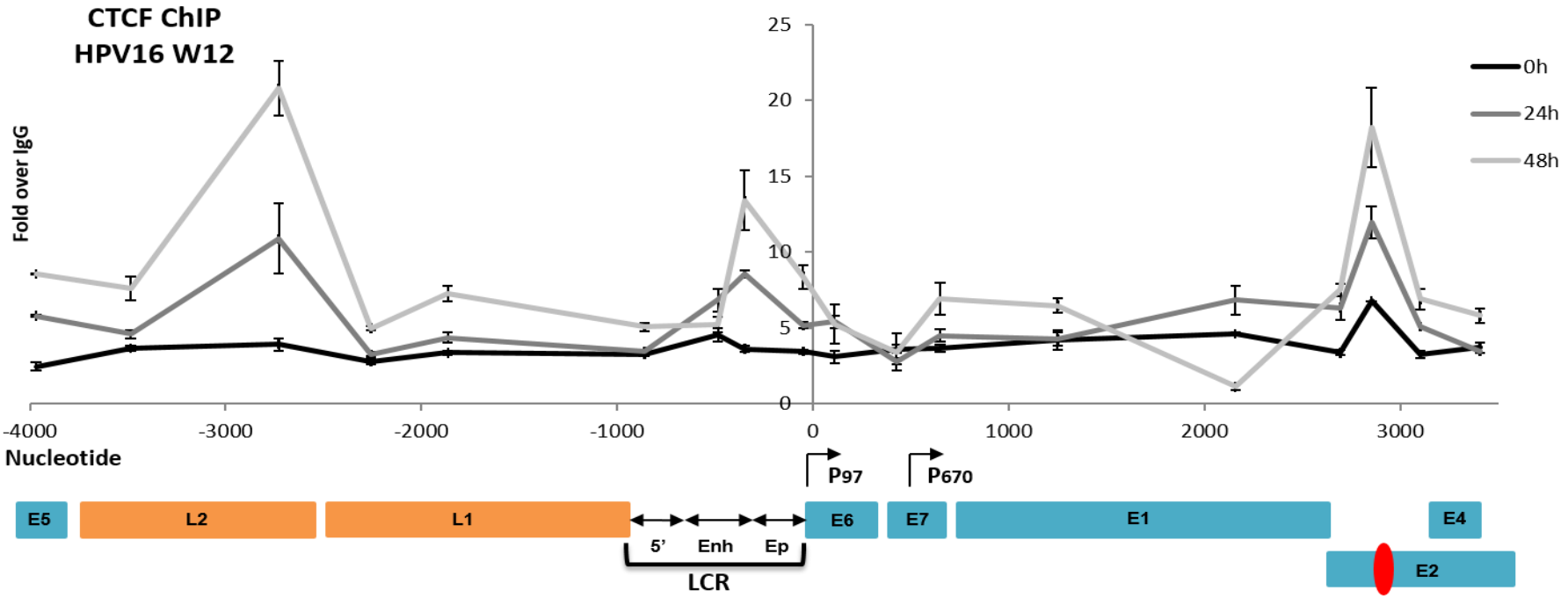


Figure 40. Association of CTCF with the HPV16 genome in W12 cells. CTCF ChIPs were performed using HPV16 containing W12 cells either in monolayer culture or differentiated in methylcellulose for 24hr or 48hr. Cells were fixed in 1% formaldehyde, lysed, and dounced with 40 strokes to aid nuclei release. Chromatin was sheared by sonication on a cycle of 30s on and 30s for a total sonication time of 15 minutes. ChIP reactions were performed at 4°C overnight using 25µg of chromatin and 8µg of CTCF polyclonal antibody. ChIPs were also set up using FLAG IgG as a negative control. DNA was de-crosslinked and eluted and used for qPCR analysis using primers spanning the HPV18 genome, with the mid-point of each primer displayed on the x-axis. The HPV16 genome is displayed below the x-axis in the linearized format for ease of presentation. The $2^{\Delta\Delta}$Ct method was used to calculate fold binding over FLAG ChIP (Y-axis) and error bars represent the s.d. for qPCR triplicate values from one individual experiment. The predicted CTCF binding site is displayed as a red marker. The viral LCR is comprised of the 5' end, enhancer (Enh) and the early promoter (Ep). The early transcripts are transcribed from P97 and the late transcripts from P670. The ChIP graph shown is representative one independent experiment for HPV16 containing W12 cells.

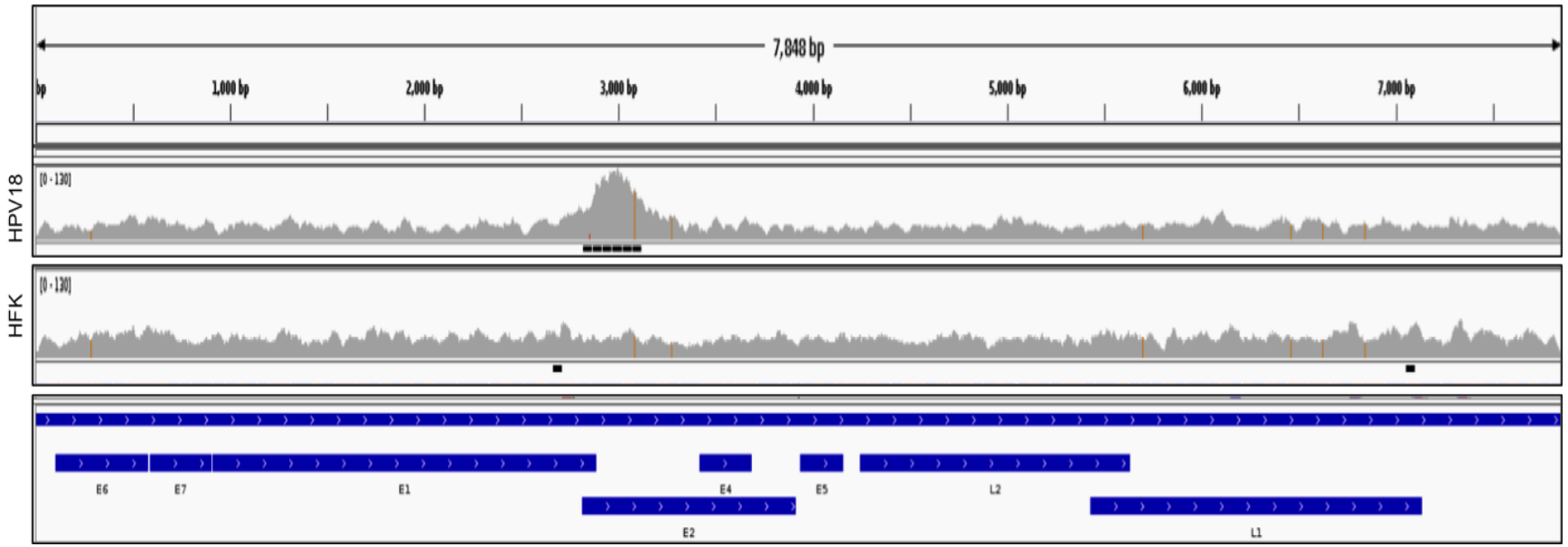


Figure 41. CTCF ChIP-Seq of untransfected HFKs and WT HPV18 genome containing HFKs. Untransfected HFKs and WT HPV18 genome containing HFKs were fixed in 1% formaldehyde and chromatin was sheared by sonication. ChIP was carried out using chromatin from $1x10^7$ cells and 4µg of CTCF antibody (Cell Signaling 3418), followed by next generation sequencing. The ChIP library was prepared using NEXTflex ChIP-seq library and 50 bp single read sequencing performed on an Illumina HiSeq 2500 with a read depth of 25-30 million reads. For data analysis CTCF ChIP-seq data was aligned to the HPV18 genome. Quality control and final data analysis were performed as described in (Gunther et al., 2016). Data panels represents the enrichment of CTCF binding peaks across the HPV genome in HPV18 containing HFKs and untrasnfected HFKs were used as a control to account for background noise. The alignment of the HPV ORFs are shown in blue.

3.7 Summary

In summary, by using the CTCFBSDS prediction tool a conserved CTCF binding site was observed around nucleotide 3000 within the E2 ORF in 10 out the 13 WHO high-risk HPV types, with a further E2 site identified in the remaining three types. In the primary HFK cell culture system and in agreement with studies in HPV31, CTCF protein expression was increased upon the presence of HPV18 viral episomes when compared to untransfected HFKs, indicating that the virus may up regulate CTCF protein levels to facilitate aspects of its life cycle. This increased CTCF expression was also maintained upon cellular differentiation in methylcellulose when compared to the HPV18 genome containing HFKs grown in monolayer culture. However when organotypic raft sections were stained for CTCF protein both untransfected and HPV18 WT rafts displayed CTCF in the basal and suprabsal layers but protein expression was reduced upon differentiation in some of the cells in the upper-most granular layers, in agreement with previous work. We have confirmed using ChIP that CTCF binds to the E2 ORF within the genomes of both high-risk HPV16 and HPV18 genome containing HFKs grown in monolayer, and upon cellular differentiation this binding is further enriched at this E2 site. Furthermore, additional CTCF binding sites were identified in the LCR, L1 and L2 regions in the HPV16 genome containing cells grown in monolayer culture. We have also shown that CTCF binds to the E2 ORF of episomal HPV16 in naturally infected

primary W12 cells. Upon differentiation CTCF binding was also detected in the viral LCR and L1 and L2 gene regions, indicating that there may be additional CTCF binding sites involved in the HPV life cycle and these may differ between HPV types. Indeed a CTCF binding site has been confirmed in the L2 region of the high-risk HPV31 genome (Mehta et al., 2015). To further strengthen the research findings ChIP-seq analysis was also used to confirm CTCF binding to the E2 ORF in the HPV18 genome, and detect any further sites that may have been overlooked. Overall, the CTCF ChIPs along with western blots demonstrating an increase in CTCF protein expression upon the presence of HPV18 episomes, lead to the hypothesis that CTCF is upregulated by HPV and recruited to the E2 ORF of high-risk HPV genomes, in order to facilitate processes during the viral life cycle.

Chapter 4- Phenotypic analysis of the HPV18 life cycle following abrogation of CTCF binding to the E2 open reading frame

4 Chapter 4: The role of CTCF binding to the HPV18 genome during the viral life cycle

4.1 Introduction

As previously discussed, the abrogation of CTCF binding to EBV and KSHV genomes has been shown to deregulate viral transcription. With this in mind it was important to determine the biological significance of CTCF recruitment to the HPV genome, specifically the role this interaction has on HPV18 viral gene and protein expression.

The complete life cycle of HPV is dependent on the differentiation of the epithelium that it infects. In order to model this experimentally the organotypic raft culture system was used, which recapitulates epithelial differentiation *in vivo* and supports viral genome amplification and differentiation dependent viral gene expression. For the following results, organotypic raft cultures were generated from HFKs containing wild type HPV18 genomes and rafts were also generated using HFKs containing ΔCTCF HPV18 genomes in which the CTCF binding site was mutated to abrogate CTCF binding (Paris et al., 2015). The use of this mutant cell line allows us to make comparisons to the WT HPV18 genome containing cells, and ultimately determine the biological significance of CTCF binding to this specific site during the complete viral life cycle.

Given that this binding site is conserved amongst the high-risk HPV types it was necessary to determine if abrogation of CTCF binding affects the expression of the viral oncoproteins, E6 and E7. We also wanted to determine if CTCF binding is required for controlling the expression of differentiation dependent late viral transcripts and proteins. Lastly we wanted to determine the role of CTCF in viral episome maintenance and viral amplification upon epithelial differentiation.

4.2 Abrogation of CTCF binding deregulates viral oncoprotein expression in primary HFK monolayer culture

As previously mentioned, the CTCF binding site in the E2 ORF was identified in 10 high-risk HPV types and further confirmed by ChIP in HPV16 and HPV18 genome containing HFKs, as well as the naturally infected W12 cell model. Initial western blot analysis revealed that monolayer HFKs containing ΔCTCF HPV18 genomes displayed an increase in both E6 and E7 onocprotein expression when compared to the WT HPV18 genome containing HFKs (Fig. 42). This was the first indication that the high-risk HPV types may need to recruit a host factor such as CTCF in order to control the balance of viral onocoprotein expression. To determine whether increased E6/E7 was recapitulated in the virus life cycle organotypic raft cultures were generated from the WT HPV18 and ΔCTCF HPV18 genome containing HFK cell lines.

4.3 Confirmation of viral episomes in primary HFK cell lines

Before generating the organotypic raft cultures it was important to verify that the viral genomes were maintained episomally in the HFK cell lines. Initially, DNA was extracted from WT and ΔCTCF HFK monolayer cultures and a Southern blot was performed to detect viral episomes, as explained in results chapter 1. DNA was extracted from either WT HPV18 and ΔCTCF HPV18 genome containing cells from two individual donors. The DNA was either digested with *Eco*RI, which linearizes the viral genome or with *Bgl*II restriction enzyme, which is a non-cutter of the viral genomes but cuts the host genome to allow identification of integrated genomes. The 8kb linearized viral genome (L) was detected in both WT HPV18 and ΔCTCF HPV18 HFKs from two independent donors (Fig. 43). Uncut viral episomes are separated on an agarose gel as open circle (OC) and supercoiled (SC) bands, which were detected in both cells donors (Fig. 43). Of note, higher molecular weight bands that would be indicative of integrated HPV in to the host DNA were not detected in any of the cell lines (Fig. 43).

4.4 Organotypic raft cultures generated from ΔCTCF HFKs display a more hyperproliferative phenotype

Organotypic raft cultures generated from untransfected HFKs and HFKs containing either WT or ΔCTCF HPV18 genomes were formaldehyde fixed, paraffin embedded and sectioned. Sections were stained with haematoxylin and eosin (H&E), in order to assess any morphological differences between

the cell lines (Fig. 44) (Propath Ltd, Hereford). These commonly used stains are able to differentiate between the cytoplasmic and nuclear regions of the epithelial cross-sections. Rafts derived from untransfected HFKs displayed a relatively thin morphology and a typical appearance for differentiated epithelium. On the other hand, rafts derived from WT HPV18 genome containing cells displayed an increased thickness of the suprabasal compartment of the epithelium. This was unsurprising considering the known co-operative actions of E6 and E7 to promote cell cycle entry in the lower suprabasal levels, which causes increased proliferation and thickening of the epithelium. Most interestingly, rafts derived from HFKs containing the ΔCTCF HPV18 genomes displayed an even more hyperproliferative phenotype compared to the WT HPV18 rafts, as evidenced by the increased thickness and expansion of the suprabasal compartment. These results indicate there may be an increase in the expression of the E6 and E7 oncoproteins in the rafts derived from the ΔCTCF HPV18 genome containing cells.

Rafts derived from WT HPV18 and ΔCTCF HPV18 genome containing HFKs were incubated with 20 μM BrdU sixteen hours prior to harvesting. The rafts were then formaldehyde fixed, paraffin embedded, and sectioned. Rafts were stained with BrdU specific antibody to assess cell cycle entry and cellular DNA replication, and Hoescht 33342 for nuclear visualization (Fig. 45). In

rafts derived from untransfected HFKs, BrdU incorporation was confined to the mitotically active basal layer of the epithelium (Fig. 45a). In rafts derived from WT HPV18 genome containing HFKs, there was an increase in BrdU incorporation in the basal and lower suprabasal layers of the raft (Fig. 45a). Most interestingly, rafts derived from ΔCTCF HPV18 genome containing HFKs displayed a significant reduction in BrdU incorporation in the basal layer compared to the WT HPV18 rafts and also a reduction in BrdU in the lower suprabasal layer, but this difference was not significant (Fig. 45a). Furthermore, ΔCTCF HPV18 rafts displayed a significant increase in BrdU positive cells in the upper suprabasal layers of the rafts when compared to the WT HPV18 rafts (Fig. 45b), indicating there is increased cell cycle entry and DNA replication.

4.5 Abrogation of CTCF binding deregulates viral oncoprotein expression in the organotypic raft culture model

Immunofluorescence was used to observe E6 expression in raft sections generated from untransfected, WT HPV18 and ΔCTCF HPV18 genome containing HFKs (Fig. 46). Unfortunately direct staining for E6 is unavailable with current antibodies; however, since E6 targets p53 for degradation staining for p53 serves as a suitable surrogate marker (Scheffner et al., 1990). Organotypic raft cultures were formalin fixed and paraffin embedded on day 14 of culture. Sections were cut from raft blocks and mounted on slides for immunofluorescence staining. Sections were stained with Hoescht 33342 for

nuclear visualization and p53 antibody to visualize cellular p53 protein. In the untransfected HFK rafts there was abundant p53 staining throughout the entire raft layers (Fig. 46a). However in the HPV18 WT HFK rafts there was a reduction in p53 staining, particularly in the upper layers of the raft (Fig. 46a). Interestingly, in the ΔCTCF HPV18 rafts there was a significant loss of p53 staining in the basal, lower and upper suprabasal layers (Fig. 46a & b). These differences in p53 protein levels were consistent between HFK donors. Increased loss of p53 in the ΔCTCF rafts suggests elevated E6 oncoprotein expression compared to the rafts derived from the WT HPV18 genome containing cells.

Direct staining of E7 protein is also not possible therefore rafts were stained for the Rb family member p130 as a surrogate marker for E7, as E7 is known to target p130 for degradation (Fig. 47) (Genovese et al., 2008). It is known that p130 is normally expressed in cells committed to differentiation, whereas the other Rb family members p107 and Rb are primarily expressed in the basal and suprabasal layers of epithelia (Paramio et al., 1998). In the rafts derived from untransfected HFK cells there was abundant p130 staining throughout the lower and upper suprabasal layers (Fig. 47a). In the rafts derived from WT HPV18 containing HFKs there was still some p130 staining visible, and this was confined to upper differentiated layers (Fig. 47a). Interestingly, in the ΔCTCF rafts there was a significant loss of p130 staining

in both the lower and upper suprabasal layers when compared to the WT HPV18 rafts (Fig. 46a & b). The reduction in p130 staining in the ΔCTCF HPV18 rafts suggests there may be elevated E7 oncoprotein expression compared to the WT HPV18 rafts.

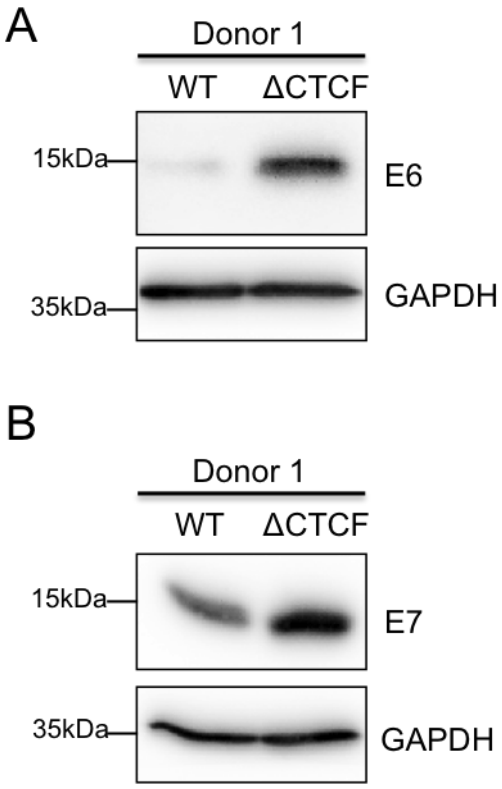


Figure 42. Expression of HPV18 E6 and E7 protein in HFKs. Protein lysates were extracted from WT HPV18 and ΔCTCF HPV18 genome containing HFKs grown in monolayer culture. Western blots were carried out using 30 μg of protein lysate to detect A) E6 and B) E7 expression and GAPDH was used as a loading control. Western blots are representative of one individual experiment for cell donor 1.

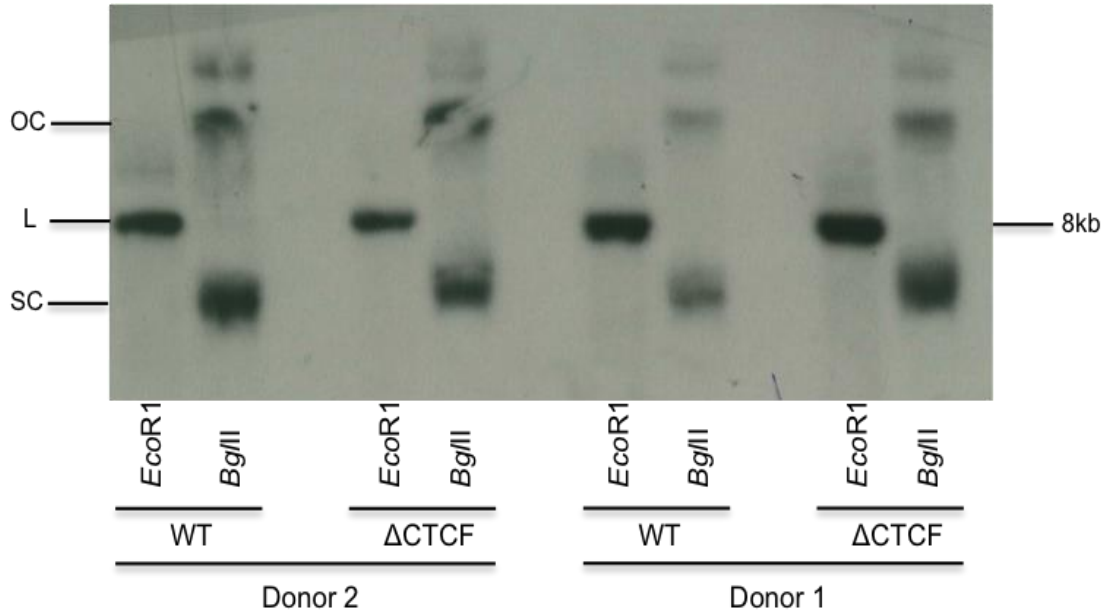


Figure 43. HPV18 genome status detected by Southern blotting. WT HPV18 and ΔCTCF HPV18 genome containing HFKs from two individual donors were grown in monolayer and DNA was extracted by phenol chloroform and ethanol precipitated. A Southern blot was performed using 5 μg DNA per condition. DNA was digested with *Eco*RI or *Bgl*II restriction enzymes and separated on a 0.8% agarose gel containing ethidium bromide. DNA was transferred to a nitrocellulose membrane overnight through capillary action. The membrane was UV crosslinked and hybridized with a HPV18 probe labelled with P^{32}. The membrane was visualized on film by autoradiography.

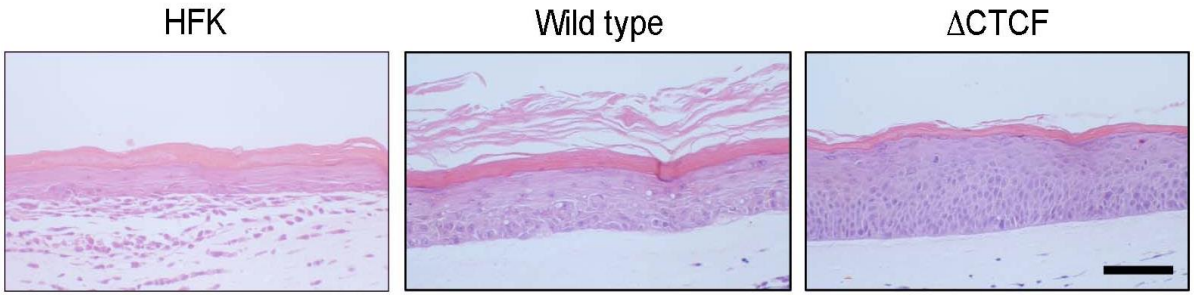


Figure 44. Hematoxylin and eosin staining of organotypic raft sections. Organotypic raft cultures were generated from untransfected HFKs and HFKs containing either WT HPV18 or ΔCTCF HPV18 genomes. On day 14 of culture the rafts were fixed in 3.7% formaldehyde, paraffin embedded then sectioned and mounted on to slides. The eosin dye stains the cyctoplasm pink and the haematoxylin stains the nucleus purple. Slides were visualized on a Nikon inverted Epifluorescent microscope fitted with a 20x objective and images were captured using a Leica DC200 camera and software. Scale bar, 10 μm. Image taken from Paris et al., 2015.

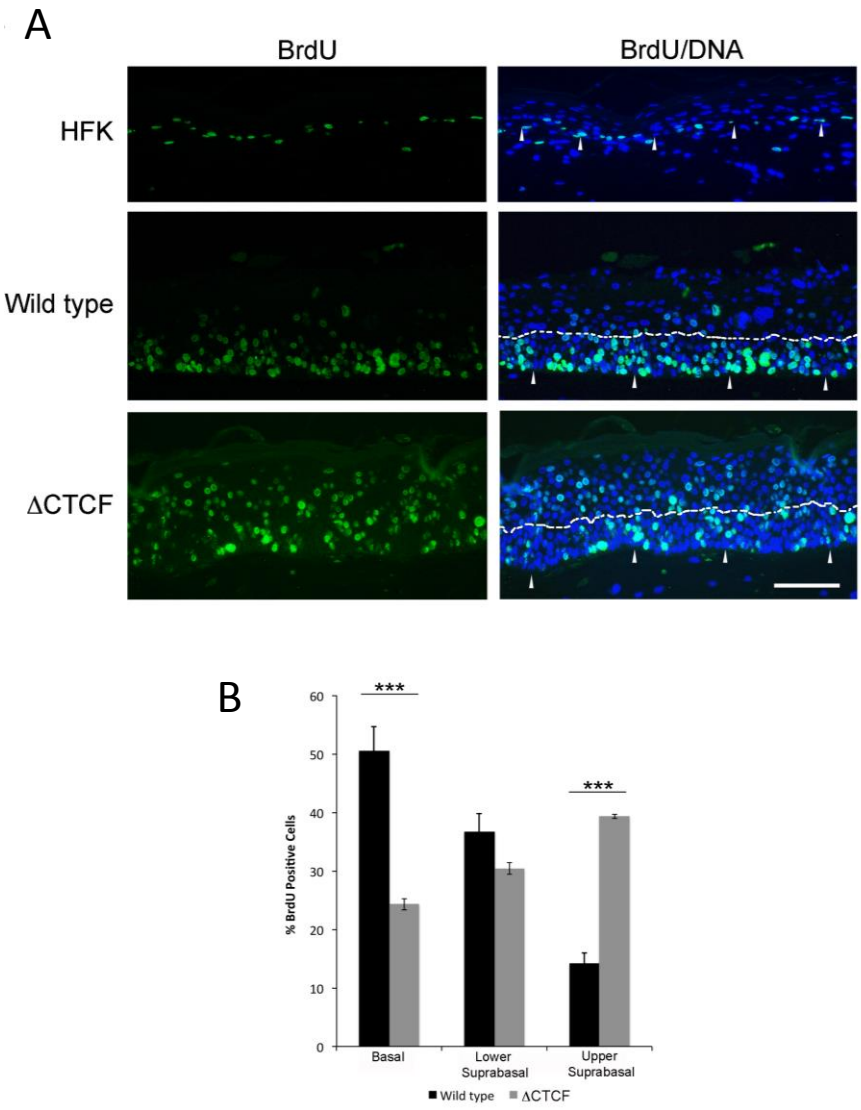


Figure 45. Immunofluorescence staining of BrdU incorporation in organotypic raft sections. Organotypic raft cultures were generated from untransfected HFKs and HFKs containing either WT HPV18 or ΔCTCF HPV18 genomes. Sixteen hours prior to harvesting, 20 μM BrdU was added to the growth medium. On day 14 of culture the rafts were fixed in 3.7% formaldehyde then sectioned and mounted on to slides. BrdU antibody was used in BSA blocking solution (green) and Hoescht 33342 was used to stain the nuclear DNA (blue). Slides were visualized on a Nikon inverted Epifluorescent microscope fitted with a 40x oil objective and images were captured using a Leica DC200 camera and software. Scale bar, 10 μm. The white arrows indicate the basal layer and the lower/upper suprabasal boundary is demarcated by the white dotted line. BrdU positive cells were quantified in the basal, lower suprabasal and upper suprabasal layers of the rafts and calculated as a percentage of the total cell count. Significance was determined by a t-test shown as ***= $p<0.001$. Image taken from Paris et al., 2015.

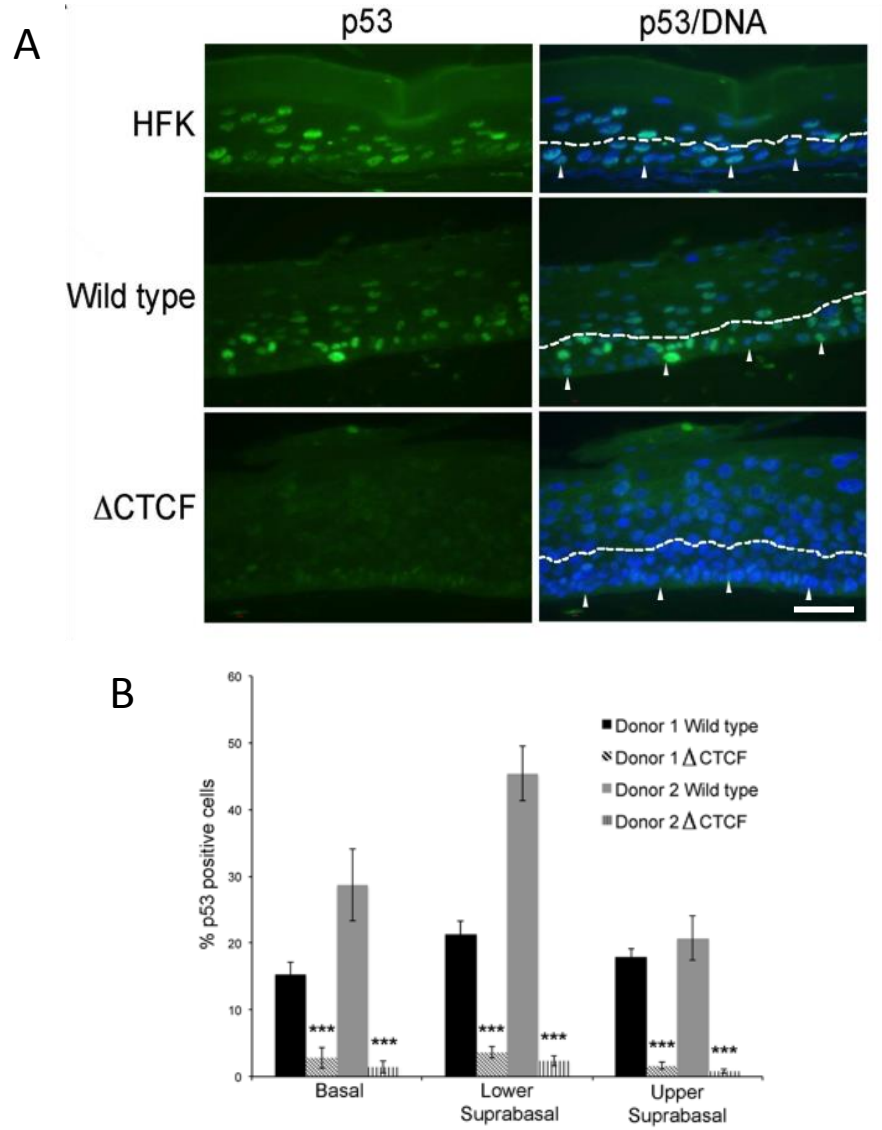


Figure 46. Immunofluorescence staining of p53 in organotypic raft sections. Organotypic raft cultures from two independent cell donors were generated from untransfected HFKs and HFKs containing either WT HPV18 or ΔCTCF HPV18 genomes. On day 14 of culture the rafts were fixed in 3.7% formaldehyde then sectioned and mounted on to slides. p53 antibody was used at 1:50 in BSA blocking solution and Hoescht 33342 was used to stain the nuclear DNA. Slides were visualized on a Nikon inverted Epifluorescent microscope fitted with a 40x oil objective and images were captured using a Leica DC200 camera and software. Scale bar, 10 μm. The white arrows indicate the basal layer and the lower/upper suprabasal boundary is demarcated by the white dotted line. p53 positive cells were quantified in the basal, lower suprabasal and upper suprabasal layers of the rafts and calculated as a percentage of the total cell count. Significance was determined by a t-test shown as ***= $p<0.001$. Image taken from Paris et al., 2015

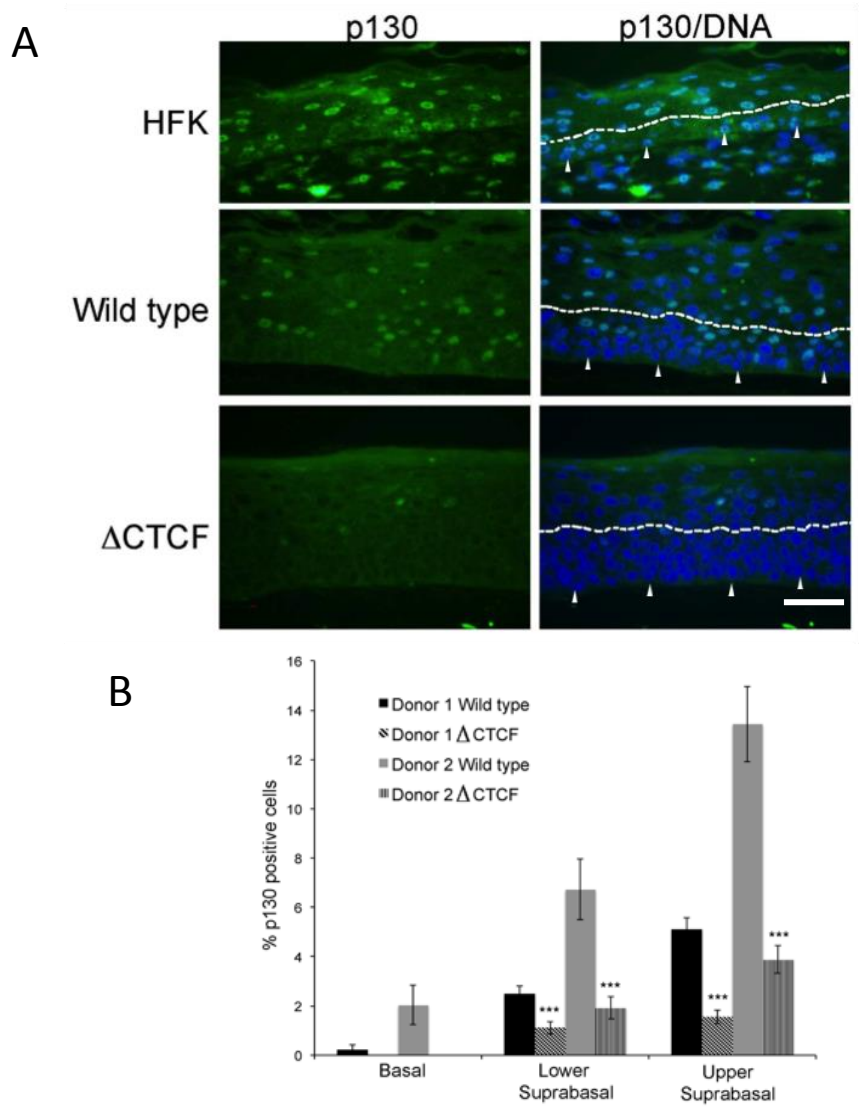


Figure 47. Immunofluorescence staining of p130 in organotypic raft sections. Organotypic raft cultures from two independent cell donors were generated from untransfected HFKs and HFKs containing either WT HPV18 or ΔCTCF HPV18 genomes. On day 14 of culture the rafts were fixed in 3.7% formaldehyde then sectioned and mounted on to slides. p130 antibody was used at 1:250 in BSA blocking solution and Hoescht 33342 was used to stain the nuclear DNA. Slides visualized on a Nikon inverted Epifluorescent microscope fitted with a 40x oil objective and images were captured using a Leica DC200 camera and software. Scale bar, 10 μm. The white arrows indicate the basal layer and the lower/upper suprabasal boundary is demarcated by the white dotted line. p130 positive cells were quantified in the basal, lower suprabasal and upper suprabasal layers of the rafts and calculated as a percentage of the total cell count. Significance was determined by a t-test shown as ***= $p<0.001$ Image taken from Image taken from Image taken from Paris et al., 2015.

4.6 Abrogation of CTCF binding causes an increase in HPV18 E6 and E7 transcript and protein expression

The immunofluorescence data has strongly indicated that there is an increase in both E6 and E7 expression in the ΔCTCF HPV18 rafts compared to the WT HPV18 rafts. To investigate this further, western blotting was performed using protein lysates extracted from three independent rafts derived from both WT HPV18 and ΔCTCF HPV18 genome containing cell lines from donor 1. Western blot analysis revealed an overall increase in E7 and E6 protein levels in the ΔCTCF HPV18 rafts relative to the WT HPV18 rafts (Fig. 48a & 49a). An increase in E6 and E7 was also observed in ΔCTCF HPV18 rafts derived from a second cell donor (Fig. 48b & 49b). Overall there was a 1.6 fold increase in E6 and a 1.8 fold increase in ΔCTCF HPV18 rafts compared to WT HPV18 rafts (Figs. 48c & 49c).

RT-PCR was performed to investigate E6 and E7 expression at the transcriptional level (Fig. 50). Firstly, RNA was extracted from both WT HPV18 and ΔCTCF HPV18 rafts on day 14 of culture then used to generate cDNA. The cDNA was amplified with primers for the unspliced early transcript region with the potential to encode both E6 and E7. The locations of the primers are indicated in the schematic shown (Fig. 50a). RT-PCR was also performed on raft cDNA using endogenous control primers for the large ribosomal protein gene (RPLPO) in order to quantify relative gene expression.

In both cell donors, the ΔCTCF HPV18 rafts showed a significant increase in the expression of the unspliced early product in comparison to the WT HPV18 rafts (Fig. 50b). Additionally, qRT-PCR was also used to quantify E6/E7 transcript levels in the HPV18 ΔCTCF rafts compared to WT HPV18 levels, with a 44.08-fold increase observed in donor 1 (±26.95-fold s.e.m.) and 21.19-fold increase in donor 2 (±10.48-fold [s.e.m.]) (Fig. 51). These results indicate that loss of CTCF binding in the E2 ORF in the HPV18 genome leads to an increase in the production of unspliced transcripts with the potential to encode the E6 and E7 oncoproteins.

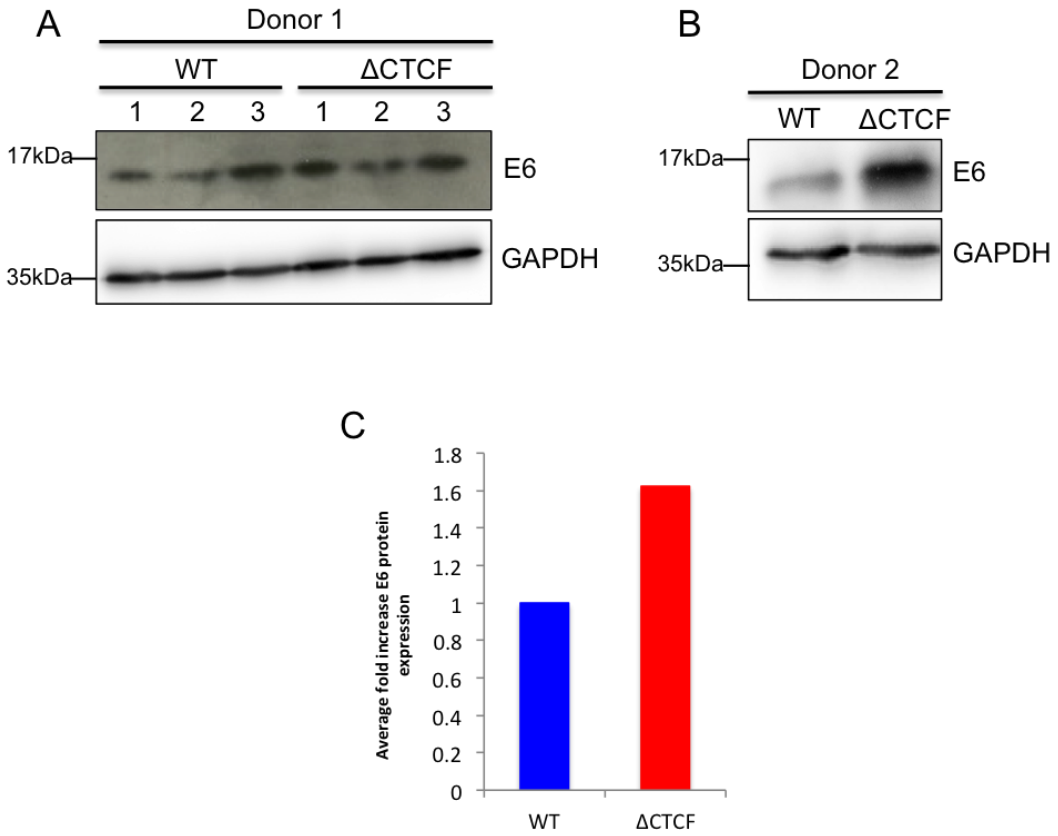


Figure 48. Expression of HPV18 E6 protein in WT HPV18 and ΔCTCF HPV18 HFK organotypic raft cultures. A) Three independent raft cultures were generated from either WT HPV18 or ΔCTCF HPV18 genome containing HFKs from donor 1 or **B)** single WT HPV18 or ΔCTCF HPV18 rafts from donor 2. Protein lysates were extracted on day 14 of raft culture. Western blotting was carried out using 30 μg of protein to detect HPV18 E6. GAPDH was used as a loading control. **C)** Western blot bands were quantified using the Fusion FX digital detection system. The average E6 expression of all four rafts relative to GAPDH was calculated for WT and ΔCTCF. The average fold increase in E6 expression relative to WT was calculated. Image B taken from Paris et al., 2015.

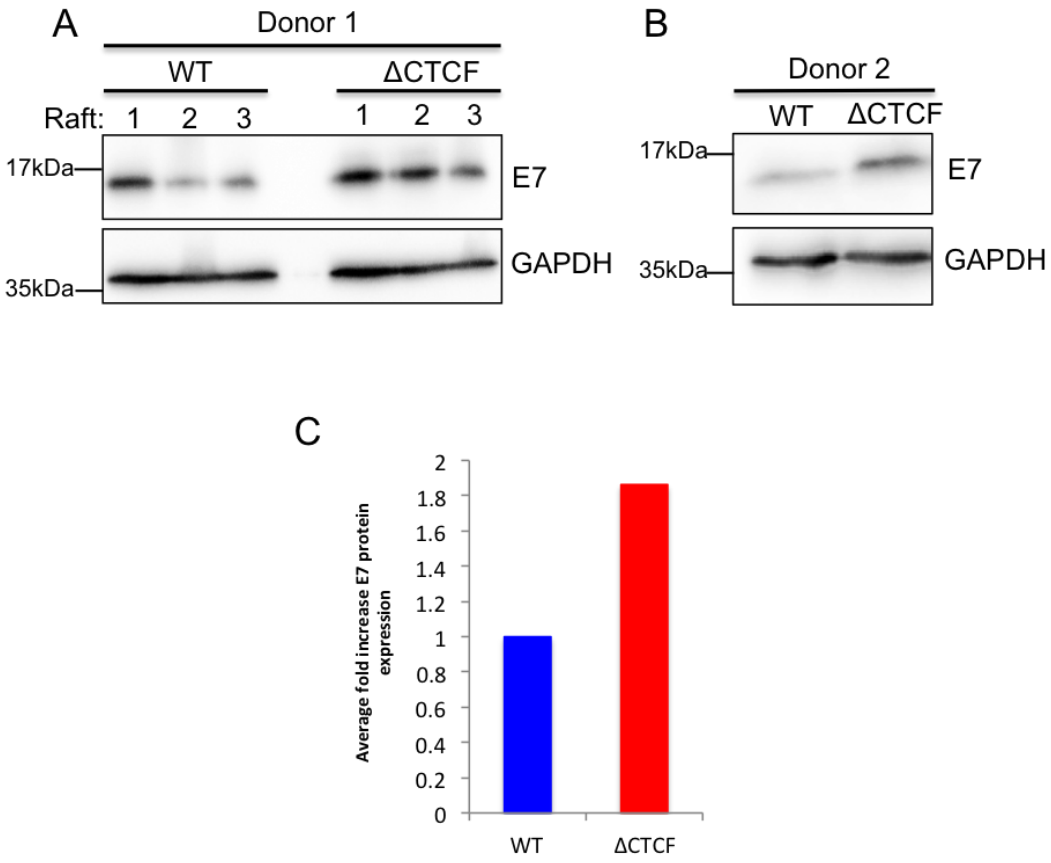


Figure 49. Expression of HPV18 E7 protein in WT HPV18 and ΔCTCF HPV18 HFK organotypic raft cultures. A) Three independent raft cultures were generated from either WT HPV18 or ΔCTCF HPV18 genome containing HFKs from donor 1 or **B)** single WT HPV18 or ΔCTCF HPV18 rafts from donor 2. Protein lysates were extracted on day 14 of raft culture. Western blotting was carried out using 30 μg of protein to detect HPV18 E6. GAPDH was used as a loading control. **C)** Western blot bands were quantified using the Fusion FX digital detection system. The average E7 expression of all four rafts relative to GAPDH was calculated for WT and ΔCTCF. The average fold increase in E6 expression relative to WT was calculated. Image B taken from Paris et al., 2015.

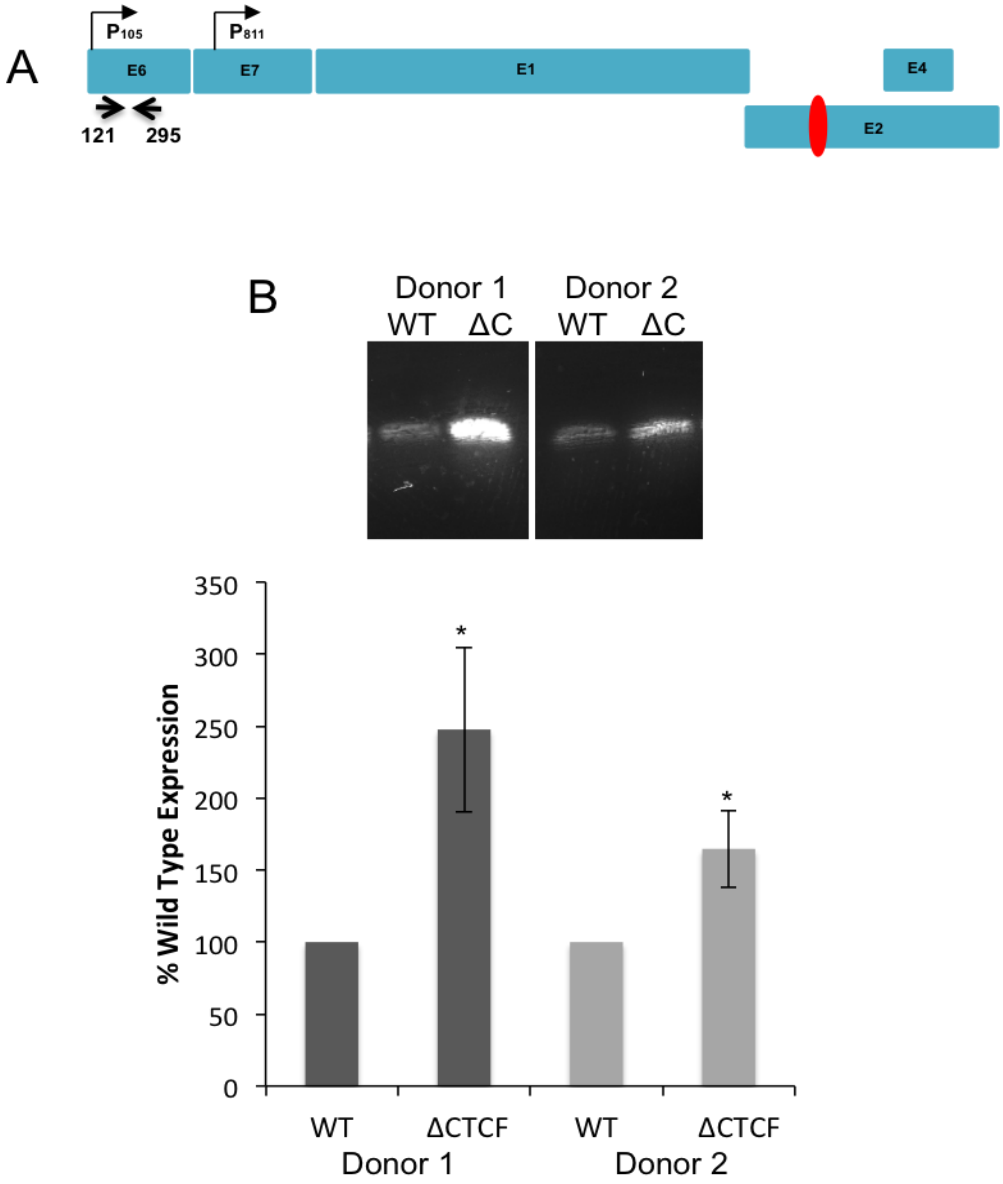


Figure 50. RT-PCR analysis of E6E7 unspliced transcripts. A) Schematic depicting the location of primer annealing in the unspliced early region at nucleotides 121 (Fw) and 295 (Rv). **B)** Three independent raft cultures were generated from either WT HPV18 or ΔCTCF HPV18 genome containing HFKs from two independent donors. RNA was extracted on day 14 of raft culture and used to generate cDNA, which was amplified by PCR using indicated primers. PCR products were separated on a 1.8% agarose gel containing ethidium bromide. DNA products were purified using a gel extraction kit (Sigma) and sent for DNA sequencing. Sequencing chromatograms were viewed using FinchTV software and aligned to known sequences using the NCBI blast database. Bands were quantified using imageJ and data represented as a percentage of E6/E7 expression in WT HPV18 cells. Significance was determined by a t-test shown as *= $p<0.05$. Image taken from Paris et al., 2015.

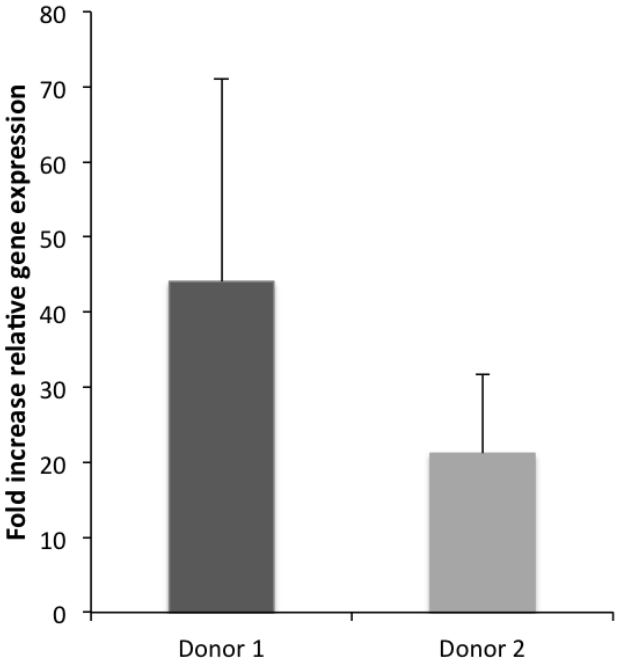


Figure 51. qRT-PCR analysis of E6E7 unspliced transcripts. Three independent raft cultures were generated from either WT or ΔCTCF HFKs from two cell donors. RNA was extracted on day 14 of raft culture and used to generate cDNA, which was amplified by SYBR qPCR using primers at nucleotides 121 (Fw) and 295 (Rv). qPCR was also carried out on the same cDNA using primers targeting the housekeeping gene RPLPO. The $2^{\Delta\Delta}$Ct method was used to calculate the relative gene expression. Data shown represent the fold increase in E6E7 gene expression in ΔCTCF HFKs compared to WT HFKs. Image taken from Paris et al., 2015.

4.7 Increased E6 and E7 expression is not a result of reduced HPV18 E2 expression

The CTCF binding site mutation was generated in the E2 ORF of the HPV18 genome. Although this mutation did not alter the amino acid sequence encoding E2 it was important to confirm that E2 protein expression remained similar in both WT HPV18 and ΔCTCF HPV18 genome containing cell lines, especially considering the known role of E2 in mediating repression of E6 and E7 (Bernard et al., 1989). Therefore, western blotting was performed using protein lysates extracted from three independent rafts derived from both WT HPV18 and ΔCTCF HPV18 genome containing HFKs from donor 1 (Fig. 52a) and individual rafts for donor 2 (Fig. 53). Western blots and accompanying quantifications showed that the E2 protein levels remained similar in both the WT HPV18 and ΔCTCF HPV18 rafts from two independent donors (Fig. 52a & 52b), and any small differences were not statistically significant (Fig. 52c). These findings demonstrate that the observed increase in oncoprotein expression is not due to altered E2 protein expression. Furthermore, rafts derived from untransfected, WT HPV18 or ΔCTCF HPV18 genome containing HFKs were also stained with E2 specific antibody (Fig. 54). As expected, E2 protein was undetectable in the untransfected rafts. In the intermediate and upper layers of both the WT HPV18 and ΔCTCF HPV18 rafts there was a similar intensity of E2 protein expression (performed by JP). However a noticeable delay in E2 expression was visible in the ΔCTCF HPV18 rafts,

consistent with the observed increase in E6 and E7 expression in these cells, which modulate expansion of the lower suprabasal cells and a delay in differentiation.

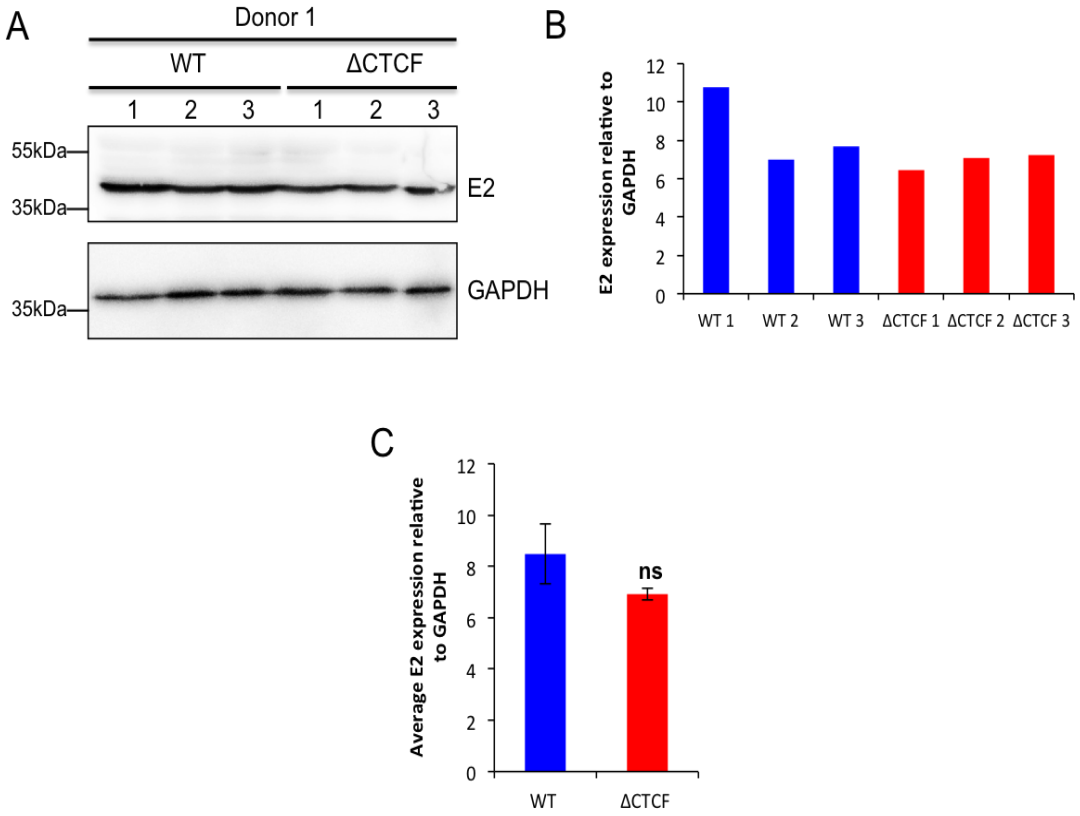


Figure 52. Expression of HPV18 E2 protein in WT HPV18 and ΔCTCF HPV18 organotpyic raft cultures. Three independent raft cultures were generated from either WT HPV18 or ΔCTCF HPV18 genome containing HFKs from donor 1. Protein lysates were extracted on day 14 of raft culture. **A)** Western blotting was carried out using 30ug of protein to detect HPV18 E2 and GAPDH was used as a loading control. Western blot bands were quantified using the Fusion FX digital detection system. **B)** E2 expression was calculated relative to GAPDH for each of the three independent rafts for WT and ΔCTCF. **C)** The average E2 expression of all three rafts relative to GAPDH was calculated for WT and ΔCTCF. Error bars are representative of SEM and significance was determined by a t-test. No statistical difference in E2 protein expression was observed.

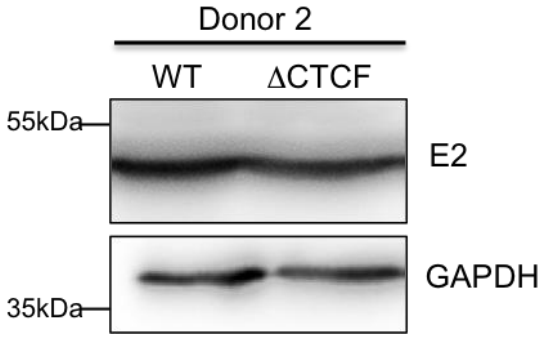


Figure 53. Expression of HPV18 E2 protein in WT HPV18 and ΔCTCF HPV18 HFK organotypic raft cultures. Raft cultures were generated from either WT HPV18 or ΔCTCF HPV18 genome containing HFKs from donor 2. Protein lysates were extracted on day 14 of raft culture. Western blotting was carried out using 30μg of protein to detect HPV18 E2 and GAPDH was used as a loading control. Image taken from Paris et al., 2015.

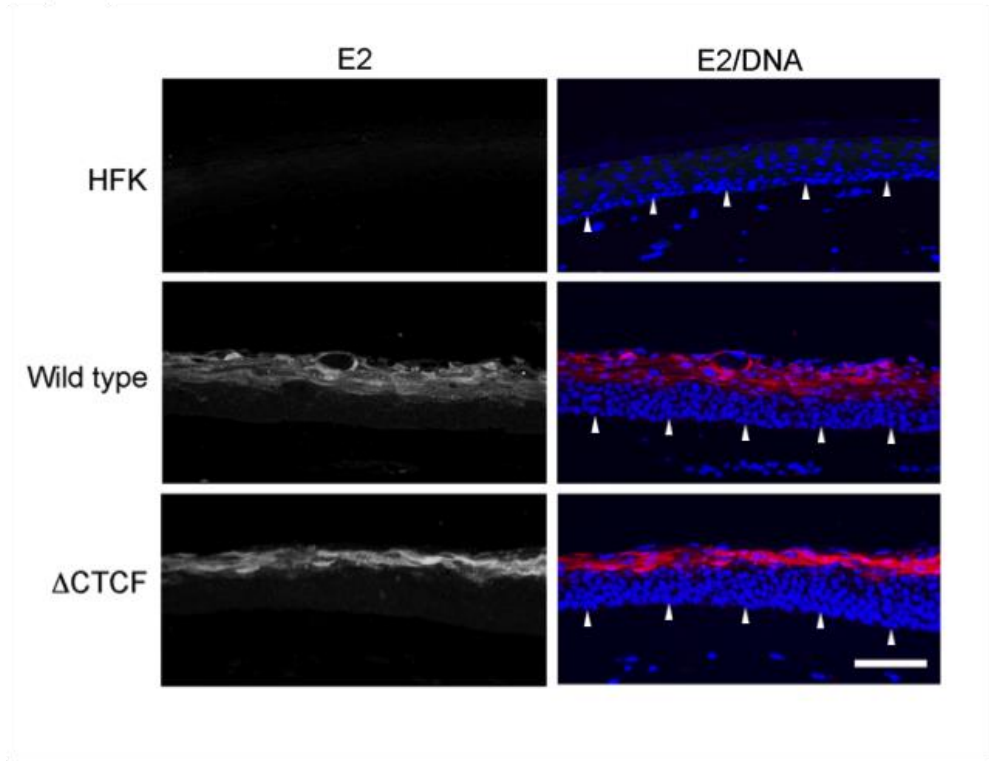


Figure 54. Immunofluorescence staining of HPV18 E2 in WT HPV18 and ΔCTCF HPV18 organotypic raft sections. Rafts were generated from either untransfected HFKs or WT HPV18 and ΔCTCF HPV18 genome containing HFKs. On day 14 of culture the rafts were fixed in 3.7% formaldehyde then sectioned and mounted on to slides. HPV18 E2 antibody was used at 1:100 in BSA blocking solution and Hoescht 33342 was used to stain the nuclear DNA. Sides were visualized on a Nikon inverted Epifluorescent microscope fitted with a 20x objective and images were captured using a Leica DC200 camera and software. Scale bar, 10 μm. The white arrows arrows indicate the basal layer. E2 protein staining was performed by Dr Jo Parish. Image taken from Paris et al., 2015.

4.8 Abrogation of CTCF binding deregulates HPV18 late transcript and protein expression

Given that abrogation of CTCF binding in monolayer and differentiated HPV18 genome containing HFKs deregulated early viral gene expression it was also important to address the impact on late gene expression. In order to investigate late HPV protein expression organotypic raft sections derived from WT HPV18 and ΔCTCF HPV18 genome containing HFK were stained with HPV18 L1 specific antibody (Fig. 55a). L1 is one of the viral capsid proteins so it is expressed exclusively during the late stages of the viral life cycle in the upper granular layer of the epithelium. In the WT HPV18 rafts L1 staining was clearly visible in the upper cornfield layer. However in the ΔCTCF HPV18 rafts there was a significant loss of L1 staining compared to the WT HPV18, with an overall 60% reduction in L1 positive cells quantified from 3 independent rafts from two independent donors $p<0.001$ (Fig. 55b).

It was not possible to detect HPV18 L1 via a western blot as the keratin present in the raft cultures prevents detection of the L1 protein. However, HPV18 E1^E4 is also a late viral protein as it is expressed in the upper differentiated layers of the epithelium. Western blot analysis of HPV18 E1^E4 revealed a decrease in both the full length and cleaved form of E1^E4 in the ΔCTCF HPV18 rafts in comparison to the WT HPV18 rafts (Fig. 56). HFKs were also differentiated in methylcellulose for 48 hours, and E1^E4 protein

detected by western blot (Fig. 57a). Again there was a significant reduction in E1^E4 protein expression in the ΔCTCF HPV18 HFKs from four independent methylcellulose differentiation experiments (Fig. 57b). Furthermore, there was a 2-fold reduction in E1^E4 expression in ΔCTCF HPV18 genome containing cells in a second HFK donor (Fig. 57b).

Given the observed reduction in late E1^E4 and L1 protein expression in the ΔCTCF HPV18 rafts compared to WT HPV18 rafts, it was important to investigate expression of the transcripts known to encode these proteins. Firstly, RNA was extracted from methylcellulose differentiated WT HPV18 and ΔCTCF HPV18 HFKs and used to generate cDNA for RT-PCR analysis. Previously published primers were used to identify transcripts with the potential to encode E1^E4 and L1 (Fig. 58). Unsurprisingly, the late transcripts were undetectable in the WT HPV18 monolayer HFKs, and upon differentiation the expected L1 (**1.** 187bp) and E1^E4, L1 (**2.** 450bp) transcripts were expressed. The other WT HPV18 product (**3.** 419bp) is thought to be a previously identified L1 encoding transcript but the DNA concentration was too low to confirm this by DNA sequencing. Interestingly, a novel or previously undetected spliced transcript (4.) was detected upon differentiation of the ΔCTCF HPV18 HFKs, which was not detected in the WT HPV18 HFKs. This transcript contained additional splicing between nt 3547(E1^E4) and nt 4680(L2) and also between nt 4775(L2) and nt 5609(L2),

which caused an L1 frame shift. Furthermore, the ΔCTCF HPV18 genome containing HFKs expressed this novel transcript (**3.** 187bp) whilst in monolayer culture, indicating that there may be loss of early polyadenylation site usage in these ΔCTCF genomes The alterations in late transcript expression and splicing provide an explanation for the observed loss of late E1^E4 and L1 protein expression when CTCF binding is abrogated. Overall these results offer a glimpse in to the potential role of CTCF in regulating late transcript splicing, polyadenylation site usage and gene expression through binding to the E2 ORF in HPV18 genomes.

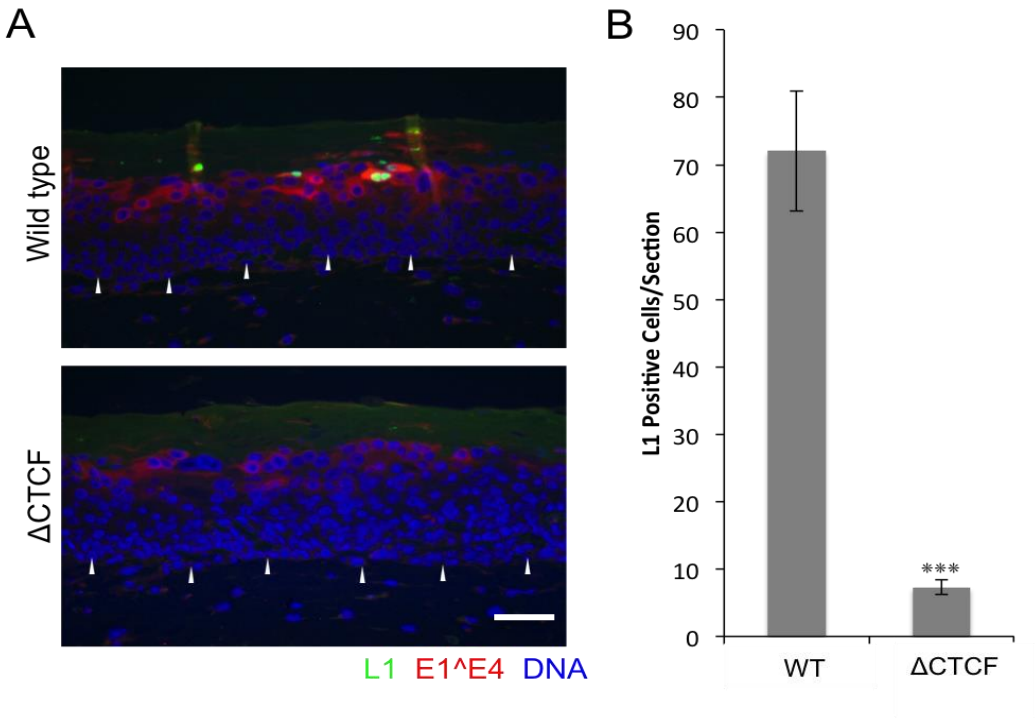


Figure 55. Immunofluorescence staining of HPV18 L1 protein in WT HPV18 and ΔCTCF HPV18 organotypic raft sections. Rafts were generated from either WT HPV18 or ΔCTCF HPV18 genome containing HFKs. On day 14 of culture the rafts were fixed in 3.7% formaldehyde then sectioned and mounted on to cover slips**. A)** L1 antibody was used at 1:100 in BSA blocking solution (green). Hoescht 33342 was used to stain the nuclear DNA (blue). Sides were visualized on a Nikon inverted Epifluorescent microscope fitted with a 40x oil objective and images were captured using a Leica DC200 camera and software. Scale bar, 10 μm. The white arrows indicate the basal layer. **B)** L1 positive cells were quantified in the basal, lower suprabasal and upper suprabasal layers of the rafts and calculated as a percentage of the total cell count. Significance was determined by a t-test shown as ***= $p<0.001$.

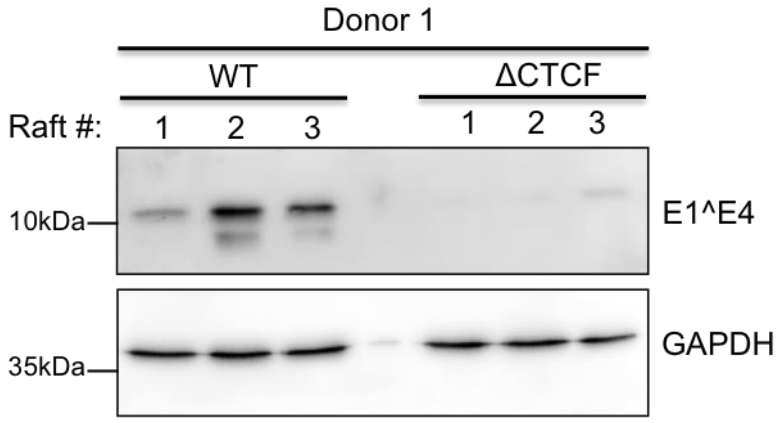


Figure 56. HPV18 E1^E4 and involucrin protein expression in WT HPV18 and ΔCTCF HPV18 HFK organotypic raft cultures. Three independent raft cultures were generated from either WT HPV18 or ΔCTCF HPV18 genome containing HFKs from donor 1. Protein lysates were extracted on day 14 of raft culture. Western blotting was carried out using 30 μg of protein to detect HPV18 E1^E4 and GAPDH was used as a loading control.

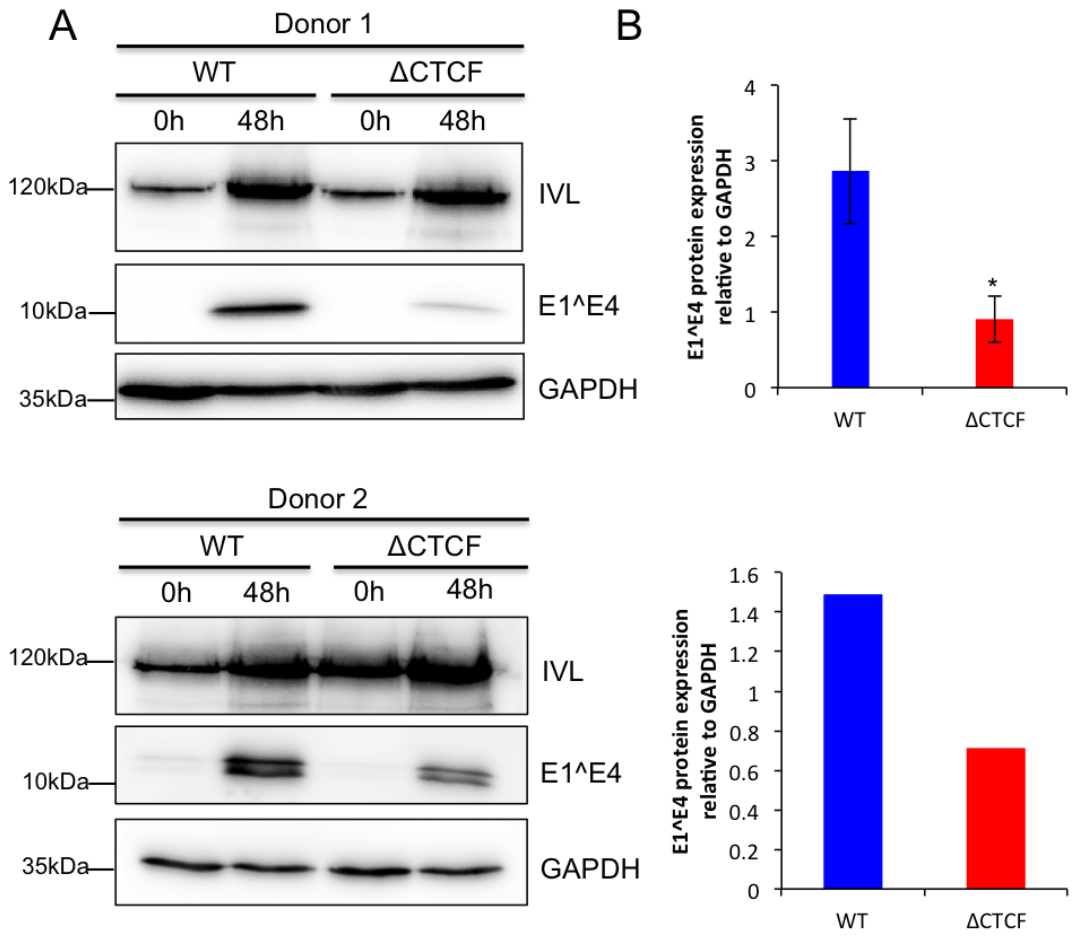


Figure 57. HPV18 E1^E4 and involucrin protein expression in WT HPV18 and ΔCTCF HPV18 genome containing HFKs. A) HFKs were grown in monolayer or differentiated in methylcellulose for 48 hours. Protein lysates were extracted and 30 μg was used to detect involucrin and HPV18 E1^E4. GAPDH was used as a loading control. **B)** Western blot bands were quantified using the Fusion FX digital detection system. E1^E4 expression was calculated relative to GAPDH for 4 independent experiments for HFK donor 1. Error bars are representative of SEM and significance was determined by a t-test shown as *= $p<0.05$. The western blots shown are representative of four experimental repeats for donor 1 (n=4) and two for donor 2 (n=2).

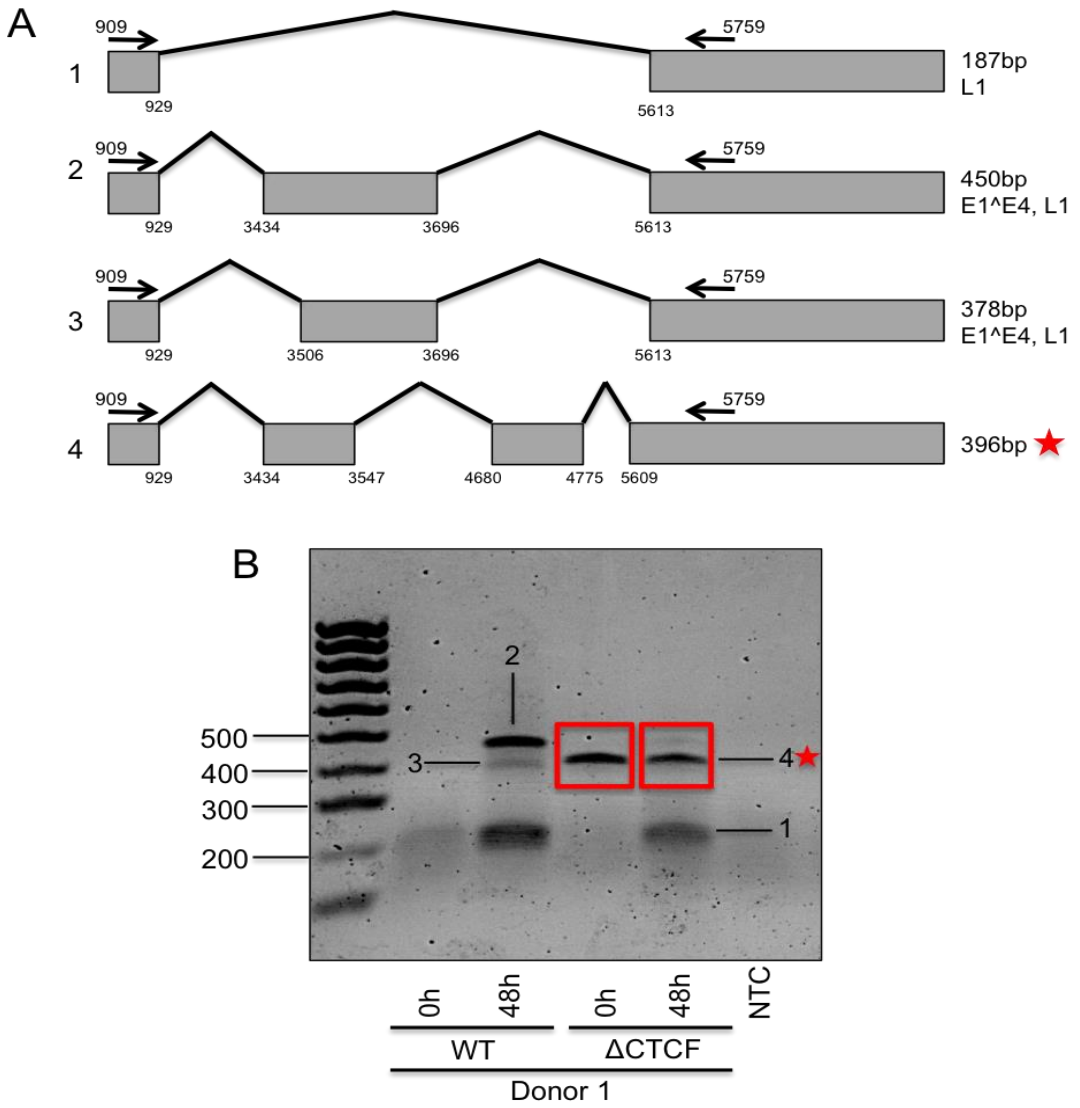


Figure 58. RT-PCR analysis of late viral transcripts. A) Schematic depicting the location of primer annealing and known splice sites in the HPV18 genome. The annotated PCR products encode: **1.** L1 (187bp), **2.** E1^E4, L1 (450bp), **3.** E1^E4,L1 (378bp) and **4.** novel spliced product (396bp). **B)** WT HPV18 and ΔCTCF HPV18 HFKs were grown in monolayer or differentiated in methylcellulose for 48 hours. RNA was extracted from HFKs and used to generate cDNA. Primers within the E1 (Fw) and L1 (Rv) regions were used to amplify PCR products. PCR products were separated on a 1.8% agarose gel containing ethidium bromide with an accompanying Bioline 100bp ladder, and then detected using a Gene Flash UV light box (Syngene Bio Imaging, UK). A control sample containing Taq Polymerase and primers was also included (NTC). Bands were cut from the gel and purified using a gel extraction kit (Sigma). Eluted DNA was sequenced and the sequencing chromatograms were viewed using FinchTV software and aligned to known sequences using the NCBI blast software. The PCR image shown is representative of three experimental repeats for donor 1 (n=3).

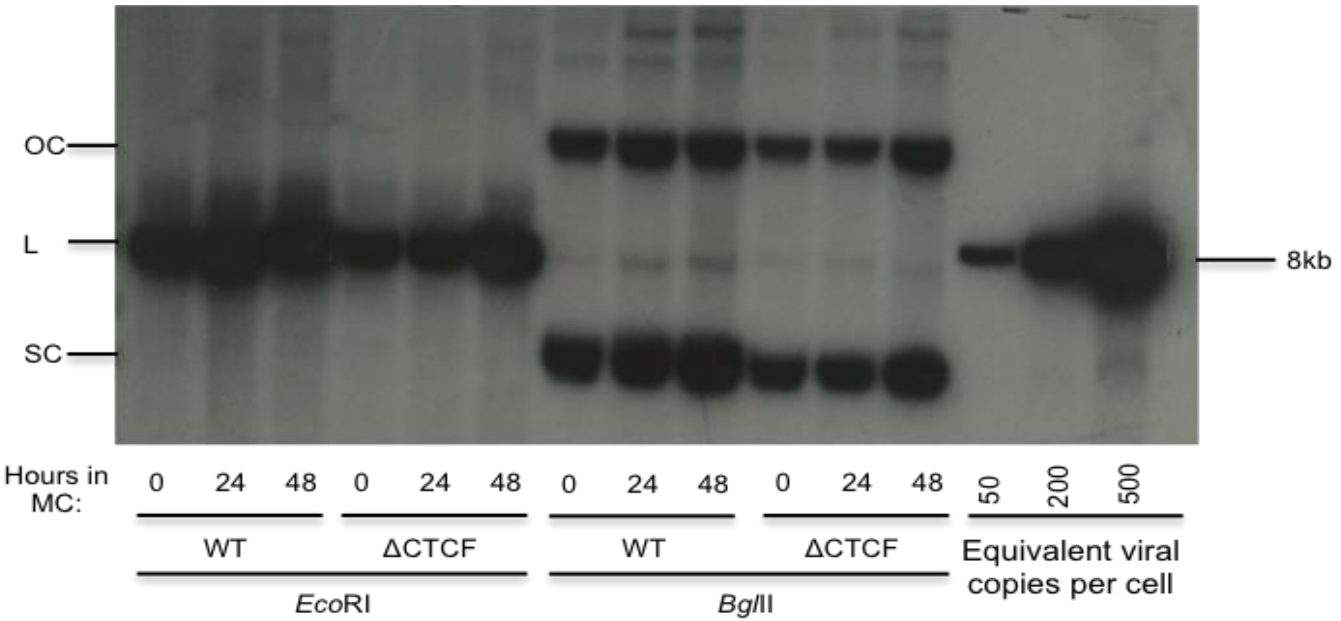


Figure 59. Detection of HPV18 genome amplification by Southern blotting. WT HPV18 and ΔCTCF HPV18 HFKs were differentiated in methylcellulose and DNA was extracted at 0h (monolayer), 24h and 48h time points. Southern blotting was performed using 5 μg DNA per condition. DNA was digested with *Eco*RI or *Bgl*II restriction enzymes and separated on a 0.8% agarose gel containing ethidium bromide. Copy number controls equivalent to 50, 200 and 500 viral episomes per cell were also separated on the agarose gel. DNA was transferred to a nitrocellulose membrane overnight through capillary action. The membrane was UV cross-linked and hybridized with a HPV18 probe labelled with p^{32}. The membrane was visualized on x-ray film by autoradiography. The Southern blot shown is representative of an individual experiment for cell donor 1.

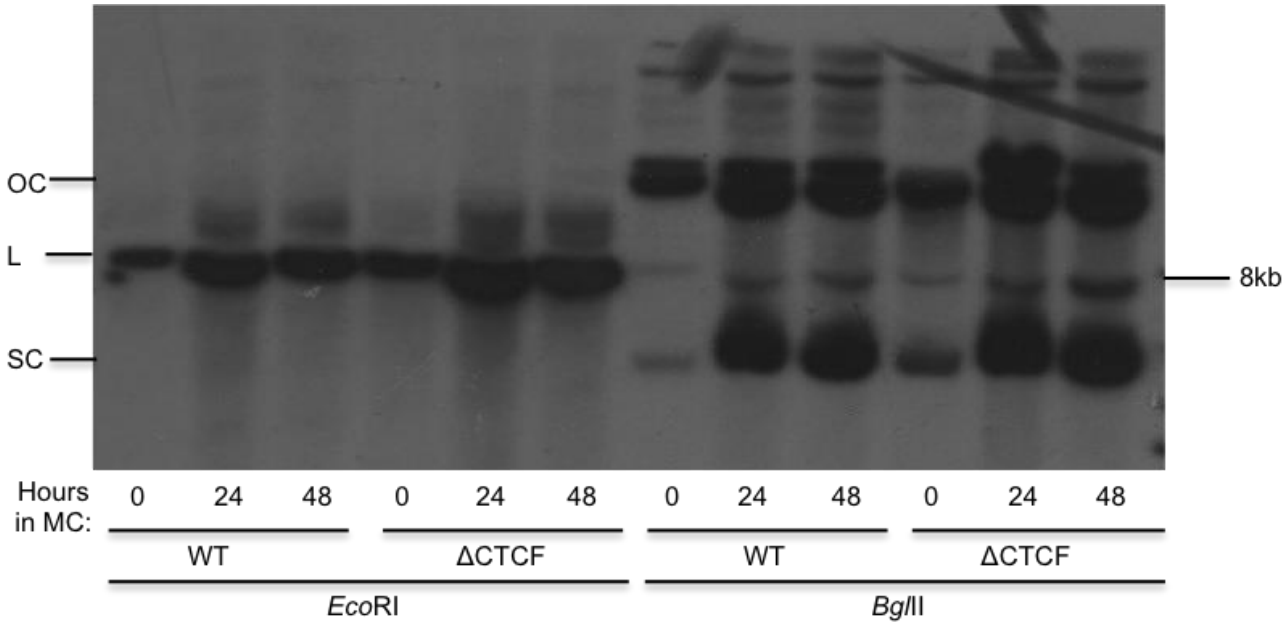


Figure 60. Detection of HPV18 genome amplification by Southern blotting. WT and ΔCTCF HFKs were differentiated in methylcellulose and DNA was extracted at 0h (monolayer), 24h and 48h time points. A southern blot was performed using 5 μg DNA per condition. DNA was digested with *Eco*RI or *Bgl*II restriction enzymes and run on a 0.8% agarose gel containing ethidium bromide. Copy number controls equivalent to 50, 200 and 500 viral episomes per cell were also separated on the agarose gel. DNA was transferred to a nitrocellulose membrane overnight through capillary action. The membrane was UV cross-linked and hybridized with a HPV18 probe labeled with p^{32}. The membrane was visualized on film by autoradiography. The southern blot shown is representative of an individual experiment for cell donor 2.

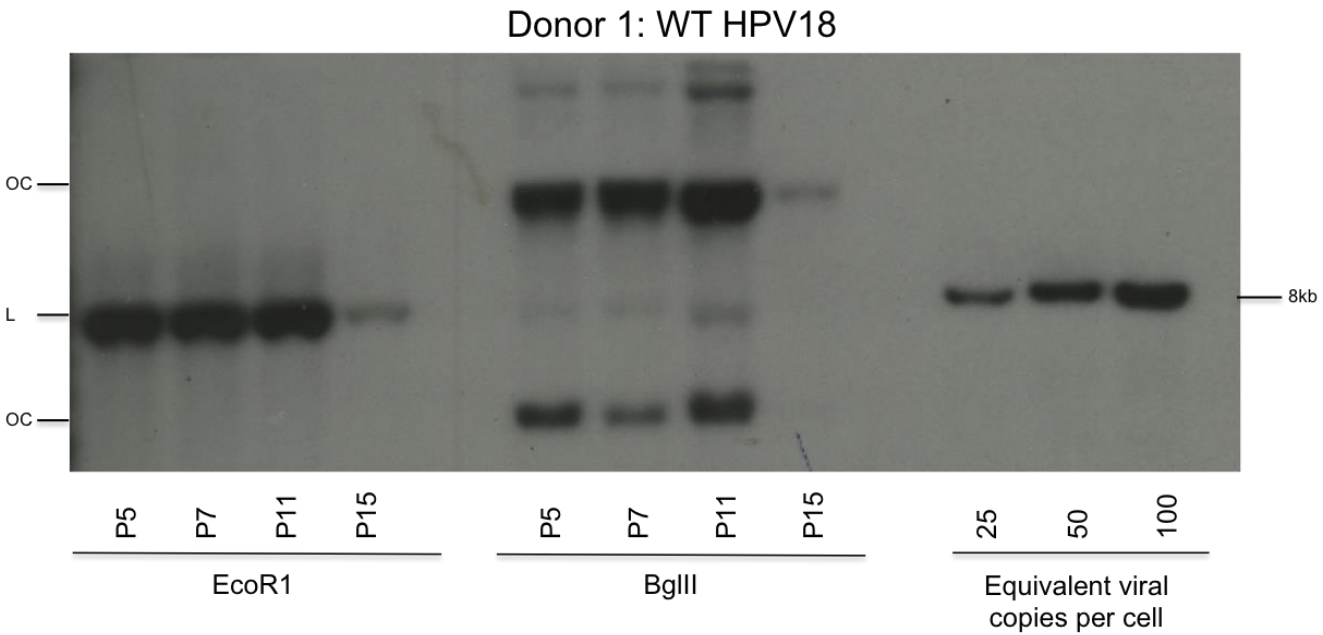


Figure 61. Detection of HPV18 episome maintenance by Southern blotting. WT HPV18 HFKs were grown in monolayer culture and DNA was extracted at passages 5, 7,11 and 15. Southern blotting was performed using 5 μg DNA per condition. DNA was digested with *Eco*RI or *Bgl*II restriction enzymes and separated on a 0.8% agarose gel containing ethidium bromide. Copy number controls equivalent to 25, 50 and 100 viral episomes per cell were also separated on the agarose gel. DNA was transferred to a nitrocellulose membrane overnight through capillary action. The membrane was UV cross-linked and hybridized with a HPV18 probe labelled with p^{32}. The membrane was visualized on x-ray film by autoradiography. The Southern blot shown is representative of an individual experiment for cell donor 1.

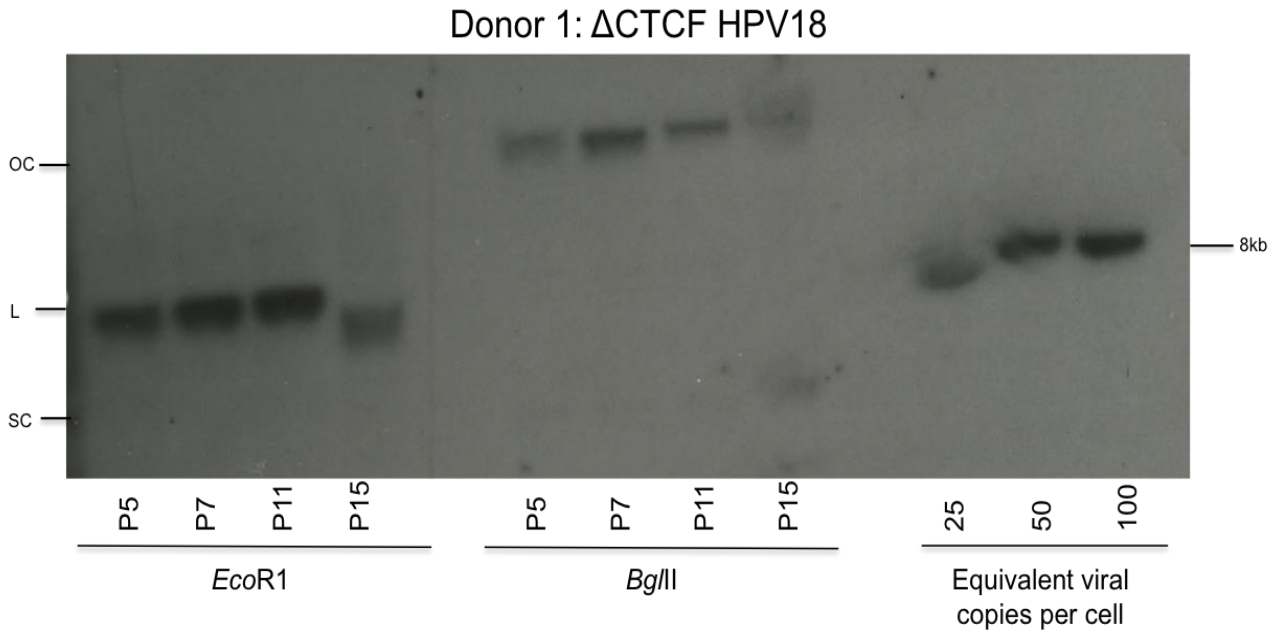


Figure 62. Detection of HPV18 episome maintenance by Southern blotting. ΔCTCF HPV18 genome containing HFKs were grown in monolayer culture and DNA was extracted at passages 5, 7,11 and 15. Southern blotting was performed using 5 μg DNA per condition. DNA was digested with *Eco*RI or *Bgl*II restriction enzymes and separated on a 0.8% agarose gel containing ethidium bromide. Copy number controls equivalent to 25, 50 and 100 viral episomes per cell were also separated on the agarose gel. DNA was transferred to a nitrocellulose membrane overnight through capillary action. The membrane was UV cross-linked and hybridized with a HPV18 probe labelled with p^{32}. The membrane was visualized on x-ray film by autoradiography. The Southern blot shown is representative of an individual experiment for cell donor 1.

4.9 Abrogation of CTCF binding does not have a major impact on viral genome amplification

Upon epithelial differentiation the expression of late viral transcripts and proteins coincides with viral genome amplification in the upper layers of the epithelium. Therefore, Southern blot analysis was performed to determine whether the abrogation of CTCF binding to the E2 ORF also reduces HPV18 genome amplification (Figs. 59 & 60). When HFK cells are grown in methylcellulose media they are induced to differentiate, and cells containing viral episomes undergo amplification of these viral genomes. DNA was extracted from WT HPV18 and ΔCTCF HPV18 genome containing monolayer HFKs from two independent donors to represent the '0hr' undifferentiated state. DNA was also extracted from HFKs that were differentiated in methylcellulose for 24hr and 48hr time points. As previously described, DNA was digested with either *Eco*RI to linearize the viral genome or *Bgl*II which cuts the host DNA. There was a slight reduction in amplification between 0hr and 24hr in the donor 1 ΔCTCF cells when compared to the same time points in the WT HPV18 cells (Fig. 59), but this was actually due to less DNA loaded on to the agarose gel in the 0hr and 24hr wells. However, similar amplification of viral genomes was observed at the 24hr and 48hr time points between WT HPV18 and ΔCTCF HPV18 genome containing cells from donor

2 (Fig. 60). These results indicate that CTCF binding does not have a major effect on HPV18 genome amplification.

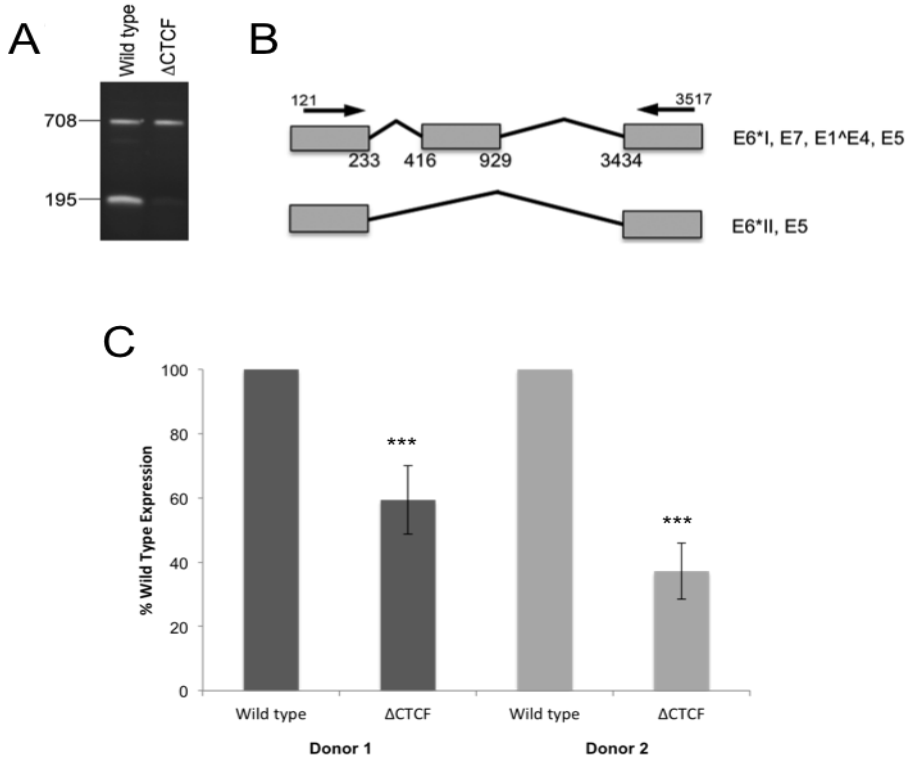


Figure 63. Analysis of early transcript splicing. Three independent raft cultures were generated from either WT or ΔCTCF HFKs from two cell donors. A) RNA was extracted on day 14 of raft culture and used to generate cDNA, which was amplified by PCR using primers that annealed at nucleotides 121 (Fw) and 3517 (Rv). PCR products were separated on a 1.8% agarose gel containing ethidium bromide. DNA products were purified using a gel extraction kit (Sigma) and DNA sequenced. Sequencing chromatograms were viewed using FinchTV software and aligned to known sequences using the NCBI blast software. B) Schematic depicting the location of primer annealing and known splice sites in the HPV18 genome. The 708bp product encodes E6*I, E7, E1^E4 and E5. The 195bp encodes E6*II and E5. C) The 195bp bands were quantified using imageJ and data represented as percentage of wild type expression. Significance was determined by a t-test shown as *= $p<0.05$. The PCR image shown is representative of three experimental repeats for each donor (n=3). Figure taken from (Paris et al., 2015).

4.10 Abrogation of CTCF binding does not have a major impact on viral genome maintenance

Southern blot analysis was used to determine whether the abrogation of CTCF binding to the E2 ORF affects viral episome maintenance (Figs. 61 & 62). The DNA used for the Southern blot was extracted from WT HPV18 and ΔCTCF HPV18 genome containing HFK cell pellets of increasing passage numbers. HPV episomes were detectable in p5, 7, 11, and 15 HFKs harboring either the WT HPV18 and ΔCTCF HPV18 genomes in two individual donors, however episomes at 25 copies or above were not observed beyond p15 in both cell lines and donors. Overall a similar level of viral episome maintenance was observed between the WT HPV18 and ΔCTCF HPV18 genome containing cell lines in both donors.

4.11 Abrogation of CTCF binding deregulates HPV18 splicing events

Given that HPV mRNA is extensively alternatively spliced (Wang et al., 2011) the next step was to determine if CTCF is involved in this splicing process, particularly the splicing of additional early transcripts. RNA was extracted from WT HPV18 and ΔCTCF HPV18 raft cultures and used to generate cDNA. Raft cDNA was amplified using a forward primer at nucleotide 121 (E6) upstream of the first splice donor site (233) and reverse primer at 3517 (E2), which is downstream of five known splice acceptor sites in the early region of HPV18 at nucleotides 416, 2779, 3434, 3465, and 3506 (Fig. 63b).

From the WT HPV18 rafts two PCR products of 708bp and 195bp were identified. Sequencing revealed that these two products were spliced between nucleotides 233^416 and 929^3434 to generate the 708bp product, which is known to encode E6*I, E7, E1^E4 and E5, and spliced at 233^343 to obtain the smaller 195bp product, which is known to encode E6*II and E5 (Fig. 63a). Interestingly, when PCR was carried out with ΔCTCF HPV18 raft cDNA only the 708bp was clearly present, as there was a significant reduction and in some cases complete loss of the 195bp transcript (Fig. 63a and b). Overall these data indicate that abrogation of CTCF binding in the E2 ORF alters splice site usage and ultimately deregulates major HPV splicing events, which may favour the expression of the E6E7 transcripts.

4.12 Summary

It was initially postulated that the CTCF protein is recruited to the E2 ORF of high-risk HPV genomes in order to regulate processes throughout the viral life cycle. Furthermore, due to the conservation of this CTCF binding site specifically amongst the oncogenic HPV types, it was hypothesized that CTCF recruitment is needed to regulate and co-ordinate oncoprotein expression. Indeed, both haematoxylin and eosin staining and immunofluorescence staining for BrdU incorporation revealed that rafts derived from ΔCTCF HPV18 genome containing cells displayed a more

hyperproliferative phenotype compared to WT HPV18 rafts. Moreover, western blotting revealed an increase in E6 and E7 protein expression in monolayer ΔCTCF HPV18 genome containing cells in comparison to WT HPV18 monolayer cells. Furthermore, using the organotypic raft culture system we have demonstrated that abrogation of CTCF binding to the E2 ORF in HPV18 causes an increase in both E6 and E7 expression at both the transcriptional and protein level, and importantly this not due to a reduction in overall HPV18 E2 protein expression. The use of this organotypic raft culture system as well as methylcellulose differentiation has also revealed alterations in late protein expression, with an observed decrease in both L1 and E1^E4 protein expression in ΔCTCF HPV18 genome containing cells. Interestingly, when the expression of the late transcripts was investigated further, we identified a novel spliced transcript in the ΔCTCF HPV18 genome containing cells containing additional splicing between nt 3547(E1^E4) and nt 4680(L2) and also between nt 4775(L2) and nt 5609(L2), which may explain the observed reduction in late protein levels. We also identified alterations in additional early HPV splicing events, specifically the loss of the transcript encoding E5 and E6*I in the ΔCTCF HPV18 genome containing cells, indicating that CTCF is required to bind in the E2 ORF in order to co-ordinate gene splicing events. Abrogation of CTCF binding did not have a major effect on HPV18 viral genome amplification or HPV18 episome maintenance.

Overall, it is clear that upon CTCF abrogation there are perturbations in HPV splicing events, subsequently altering HPV gene and protein expression. However it still remained unclear how exactly CTCF binding in the E2 ORF regulates early gene expression, which is controlled by factors binding to the viral LCR region 3kb upstream.

Chapter 5- Understanding the mechanism for increased E6 and E7 oncoprotein expression in ΔCTCF HPV18 genome containing HFKs

5 Chapter 5: Understanding the mechanism for increased E6 and E7 oncoprotein expression in ΔCTCF HPV18 genome containing HFKs

5.1 Introduction

The results so far have demonstrated that abrogation of CTCF binding in the E2 ORF of the high-risk HPV18 genome results in an increase in both E6 and E7 viral oncoprotein expression. The aim of this next chapter was to determine the mechanism underpinning this increase in oncoprotein expression. Given that CTCF plays a major role in chromatin organisation in the human genome, it was hypothesised that CTCF binding is required to regulate the chromatin structure in the HPV genome and therefore modulate transcriptional control of the early viral genes.

CTCF is involved in nucleosome positioning and chromatin modeling, therefore it was important to look at the chromatin landscape of the viral genome. A technique termed formaldehyde isolation of regulatory elements (FAIRE) was used to compare areas of open and transcriptionally active regions, to more transcriptionally repressed regions between the WT HPV18 and ΔCTCF HPV18 genome containing cells. Another aim was to determine the association of specific histone modifications that are correlated with either transcriptional activation or repression across the viral genome, and ultimately determine if CTCF binding can modulate these epigenetic signatures. Furthermore, it is well established that CTCF plays an important

role in mediating long range chromosomal interactions via DNA looping, therefore the final aim was to determine if CTCF binding in the E2 ORF of HPV18 could mediate DNA loop formation in order to regulate viral gene expression.

The HFKs used for the subsequent experiments were cultured between passages 9 and 11. It was important to verify that the viral genomes were maintained episomally in the HFK cell lines used for these experiments. As previously discussed, DNA was extracted from WT HPV18 and ΔCTCF HPV18 genome containing HFK monolayer cultures and Southern blotting was performed to detect viral episomes and estimate viral copy numbers in each cell line (Fig. 64). DNA was either digested with the *Eco*RI restriction enzyme, which linearizes the viral genome, or incubated with *Bgl*II, which is a non-cutter of the genome but digests the host DNA. The 8kb linearized viral genome was detected in the WT HPV18 and ΔCTCF HPV18 genome containing cell lines from donor 2, and there was no visible difference in copy number between the WT HPV18 and ΔCTCF HPV18 genome containing cells (Fig. 64). The copy number was estimated to be around 50-100 copies in both WT HPV18 and ΔCTCF HPV18 genome containing cells. A similar result was obtained using WT HPV18 and ΔCTCF HPV18 genome containing cells from donor 1 (Fig. 64). Open circle and supercoiled bands corresponding to the

uncut genome were detected in WT HPV18 and ΔCTCF HPV18 genome containing cells from both donors, verifying the presence of viral episomes (Fig. 64). Of note, higher molecular weight bands that are indicative of integrated HPV in to the host DNA were not detected in any of the cell lines (Fig. 64).

It is well established that the double stranded DNA of the HPV genome is chromatinized within the host cell (Stunkel and Bernard, 1999; (Swindle and Engler, 1998). Chromatin is comprised of nucleosome units, which themselves are made up of a protein core of four duplicated histones: H2A, H2B, H3 and H4. The eight histones that compromise the nucleosome core are wrapped around by 147bp of DNA and each core is separated by linker DNA region. The positioning of these nucleosomes along the genome is tightly linked to transcriptional regulation; areas of nucleosome depletion or destabilization are associated with transcriptional activation, as this conformation allows binding access for specific transcription factors and regulatory proteins (Lee et al., 2004). Furthermore, a previous study has demonstrated that CTCF binding can provide an anchor for positioning surrounding nucleosomes (Fu et al., 2008). With this in mind it was important to determine if abrogation of CTCF binding to the E2 ORF in the HPV18 genome could alter nucleosome positioning and chromatin remodeling. The FAIRE technique was used to

identify and compare nucleosome-depleted areas of the HPV18 genome. Here, HFKs were formaldehyde fixed, lysed and chromatin was sheared by sonication. Sonication typically yields fragments between 200-600bp, similar to those obtained for ChIP experiments (Fig. 65). Cells were either subjected directly to phenol chloroform extraction (FAIRE) or reverse-crosslinked prior to extraction (INPUT). Phenol chloroform extraction separates nucleosome depleted DNA in to the aqueous phase, whereas nucleosome-bound regions are trapped at the aqueous/solvent interface. Therefore, if an area of the chromatin is nucleosome-depleted and more accessible then more DNA will be extracted from the FAIRE samples compared to the more nucleosome bound regions of the chromatin. Extracted DNA from both 'FAIRE' and 'INPUT' samples was amplified by qPCR using primers spanning the HPV18 genome, and a ratio was calculated to determine the proportion of open DNA relative to input DNA at specific regions across the viral genome (Giresi et al., 2007, Simon et al., 2014). When FAIRE experiments were carried out in WT HPV18 genome containing HFKs grown in monolayer culture, there was an enrichment of DNA extracted from within the LCR, corresponding to an area of more open chromatin (Fig. 66). This is unsurprising considering that the viral LCR region contains an array of transcription factor and regulatory protein binding sites, which are differentially bound in order to control viral transcription. However, the most striking observation was that ΔCTCF

HPV18 genome containing monolayer HFKs displayed a more open chromatin landscape across the LCR compared to the WT HPV18 genome containing cells (Fig. 66). The more open LCR region in the ΔCTCF HPV18 HFKs was maintained after cells were differentiated in methylcellulose for 48 hours (Fig. 67). Overall these results demonstrate that abrogation of CTCF binding in the E2 ORF of HPV18 can alter nucleosome occupancy within the viral LCR region (3kb upstream) by creating a more open and nucleosome depleted chromatin landscape.

Furthermore, each histone protein contains an N-terminal tail that can undergo multiple post-translational modifications at different amino acid residues, including methylation, acetylation, ubiquitinylation and phosphorylation. The modifications of specific histone tails can recruit chromatin-remodeling factors and other co-factors, which confer different transcriptional outcomes. For example, H3K27me3 is a well-known hallmark of repressed chromatin, whereas H3K4me3 (Bernstein et al., 2002) and H3K9ac are signatures commonly detected at the promoter regions of actively transcribed genes (Barski et al., 2007a). Indeed, the association of specific histone marks within the viral genome has previously been demonstrated in HPV31 containing CIN612 cells, which compared histone modifications at the

viral promoters in monolayer and differentiated cells (Wooldridge and Laimins, 2008).

5.2 Increased enrichment of H3K4me3 at the early promoter in ΔCTCF HPV18 genome containing HFKs

Given the observed increase in early gene expression in the ΔCTCF HPV18 genome containing HFKs, the first step was to investigate the association of histone modifications throughout the viral genome that are associated with active gene transcription. Firstly, the expression of the H3K4me3 histone modification was mapped across the viral genome, which is a post-translational marker of active promoter regions. ChIP experiments using a H3K4me3 specific antibody revealed an increase in enrichment across the early promoter and early gene body in the ΔCTCF HPV18 genome containing HFKs when compared to the WT HPV18 genome containing cells (Fig. 68). Western blotting revealed that ΔCTCF HPV18 genome containing cells in monolayer culture displayed higher levels of H3K4me3 protein than WT HPV18 cells (Fig. 69a). However the biggest difference in H3K4me3 expression was most apparent after cells had undergone methylcellulose differentiation, where H3K4me3 expression was 4-fold higher in ΔCTCF HPV18 genome containing cells compared to WT HPV18 genome containing cells from two independent donors (Fig. 69a and b).

5.3 Increased enrichment of total RNA Polymerase II at the early promoter in ΔCTCF HPV18 genome containing HFKs

It has been reported that the H3K4 methyltransferase protein can interact with RNA polymerase II, indicating that H3K4 deposition occurs co-transcriptionally (Zhang et al., 2015). Indeed, ChIP experiments revealed an increase in the enrichment of total RNA polymerase II binding across the early promoter region in the ΔCTCF HPV18 genome containing HFKs compared to WT HPV18 genome containing HFKs (Fig. 70). Given that RNA polymerase II is an essential protein involved in transcriptional initiation and elongation these findings support the hypothesis that the early genes are more transcriptionally active in the ΔCTCF HPV18 genome containing cells compared to the WT HPV18 genome containing cells. Western blotting showed that the total level of RNA polymerase II protein was similar between WT HPV18 and ΔCTCF HPV18 genome containing HFKs in monolayer culture (Fig. 71). Taken together, the increase in both H3K4me3 and RNA polymerase II enrichment further supports our previous findings of an increase in early gene expression when CTCF binding to the E2 ORF is abrogated.

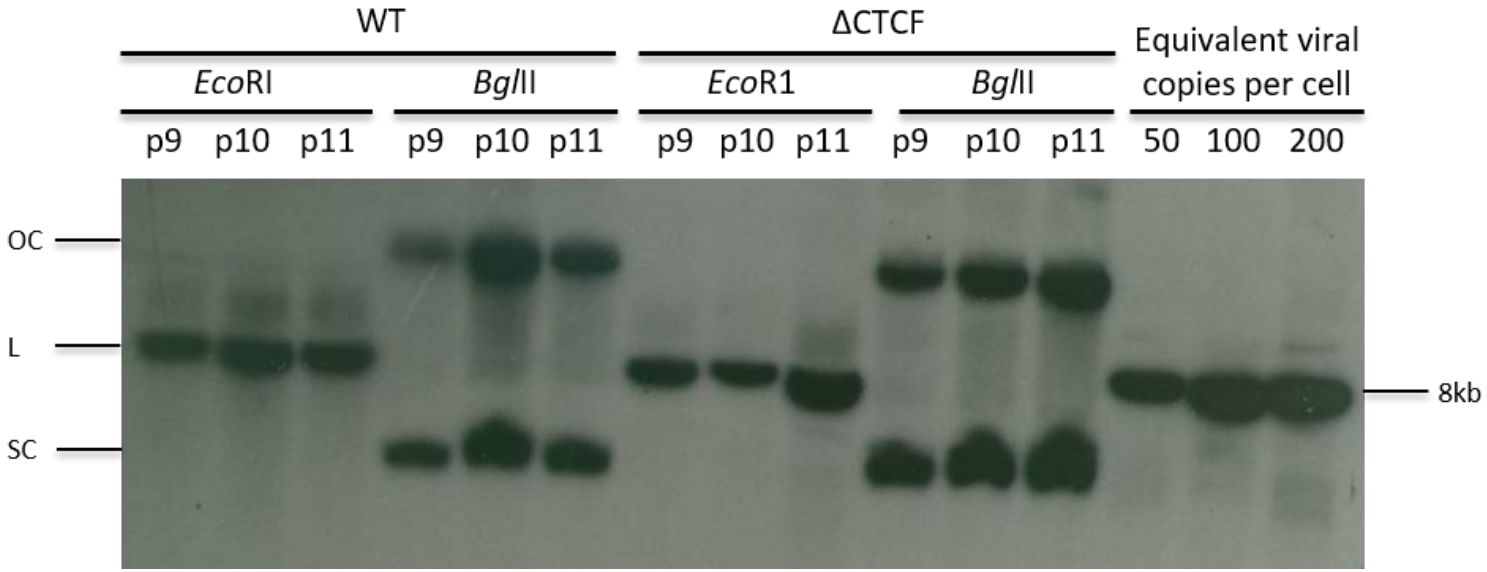


Figure 64. Detection of viral episomes by Southern blotting. WT HPV18 and ΔCTCF HPV18 genome containing HFKs were grown in monolayer culture and DNA was extracted at passages 9, 10 and 11. A Southern blot was performed using 5μg DNA per condition. DNA was digested with *Eco*RI or *Bgl*II restriction enzymes and separated on a 0.8% agarose gel containing ethidium bromide. Copy number controls equivalent to 50, 100 and 200 viral episomes per cell were also separated on the agarose gel. DNA was transferred to a nitrocellulose membrane overnight through capillary action. The membrane was UV cross-linked and hybridized with a HPV18 probe labeled with P^{32}. The membrane was visualized on film by autoradiography. The Southern blot shown is from donor 2. Results were consistent in both donors.

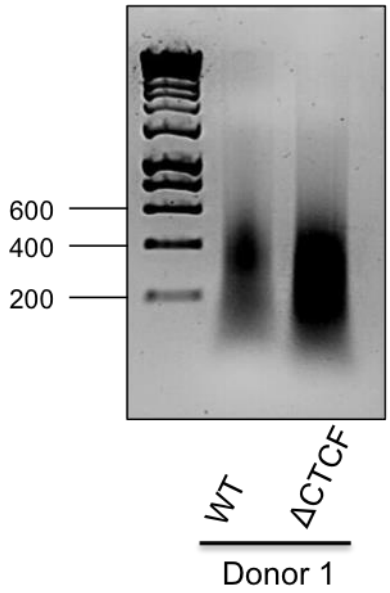


Figure 65. Sheared chromatin from WT HPV18 and ΔCTCF genome containing HFKs. HFKs containing either WT HPV18 or ΔCTCF HPV18 genomes were fixed in 1% formaldehyde, lysed, and dounced with 40 strokes to aid nuclei release. Pelleted nuclei were resuspended in shearing buffer, then sonicated for 15 minutes on ice at 25% amplitude (30s on 30s off cycle). DNA was extracted from 50µl aliquots then separated on a 1.2% agarose gel containing ethidium bromide, and visualized on a Gene Flash UV light box to check shearing efficiency. Bioline 1kb DNA ladder was used as a molecular size marker. Optimal shearing typically yielded fragments between 200-600bp.

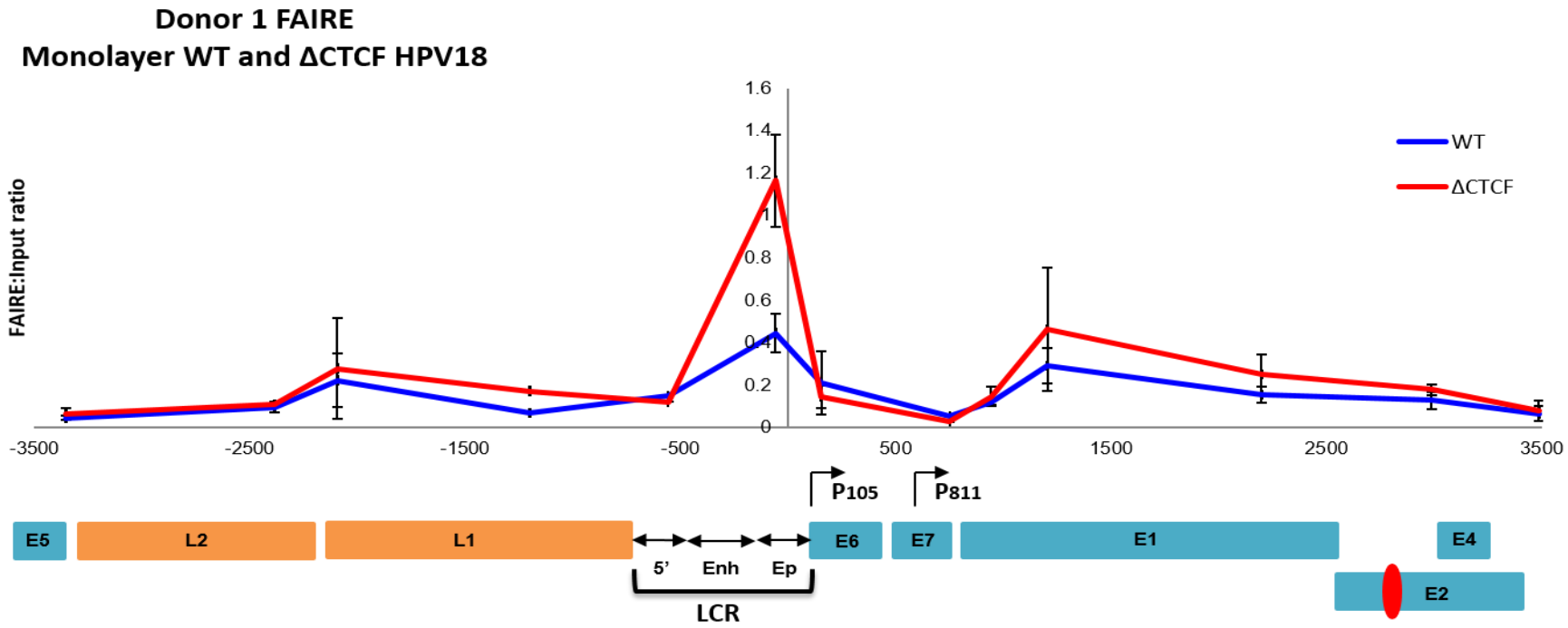


Figure 66. Formaldehyde isolation of regulatory elements. FAIRE was performed using WT HPV18 and ΔCTCF HPV18 genome containing HFKs grown in monolayer culture. Cells were fixed in 1% formaldehyde, lysed and dounced with 40 strokes to aid nuclei release. Chromatin was sheared by sonication on a cycle of 30s on and 30s off for a total sonication time of 15 minutes. Phenol chloroform DNA extraction was immediately performed on the 'FAIRE' chromatin, whereas reverse DNA cross-linking was performed on 'Input' chromatin prior to DNA extraction. qPCR was carried out on both 'Input' and 'FAIRE' DNA using primers spanning the HPV18 genome. The $2^{\Delta\Delta ct}$ values were calculated using the 'Input' and 'FAIRE' Ct values and a ratio was obtained (y-axis value). Error bars are SEM from three indepdent experiments from cell donor 1. The predicted CTCF binding site is displayed as a red marker. The viral LCR is comprised of the 5' end, enhancer (Enh) and the early promoter (Ep). The early transcripts are transcribed from P_{105} and the late transcripts from P_{811}.

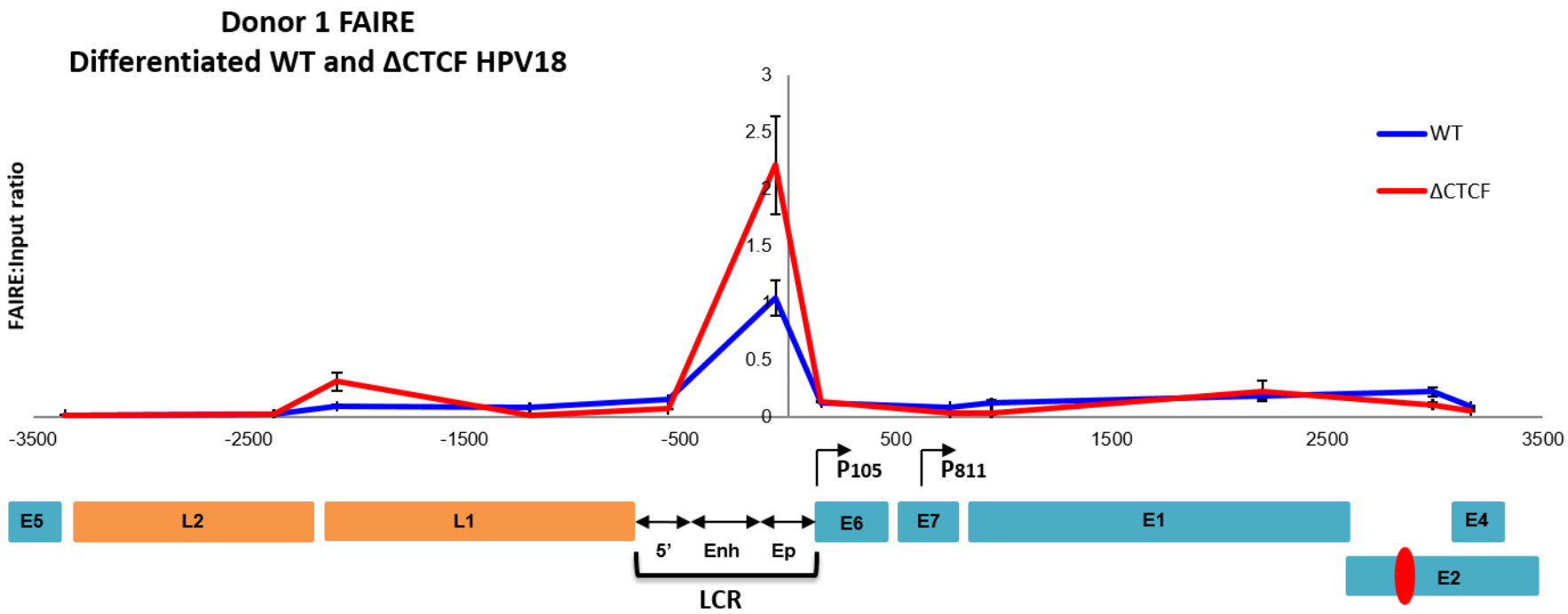


Figure 67. Formaldehyde isolation of regulatory elements. FAIRE was performed using WT HPV18 and ΔCTCF HPV18 genome containing HFKs after 48 hours differentiation in methylcelullose. Cells were fixed in 1% formaldehyde, lysed and dounced with 40 strokes to aid nuclei release. Chromatin was sheared by sonication on a cycle of 30s on and 30s off for a total sonication time of 15 minutes. Phenol chloroform DNA extraction was immediately performed on the 'FAIRE' chromatin, whereas reverse DNA cross-linking was performed on 'Input' chromatin prior to DNA extraction. qPCR was carried out on both 'Input' and 'FAIRE' DNA using primers spanning the HPV18 genome. The $2^{\Delta\Delta ct}$ values were calculated using the 'Input' and 'FAIRE' Ct values and a ratio was obtained (y-axis value). Error bars are SEM from three indepdent experiments from cell donor 1. The predicted CTCF binding site is displayed as a red marker. The viral LCR is comprised of the 5' end, enhancer (Enh) and the early promoter (Ep). The early transcripts are transcribed from P_{105} and the late transcripts from P_{811}.

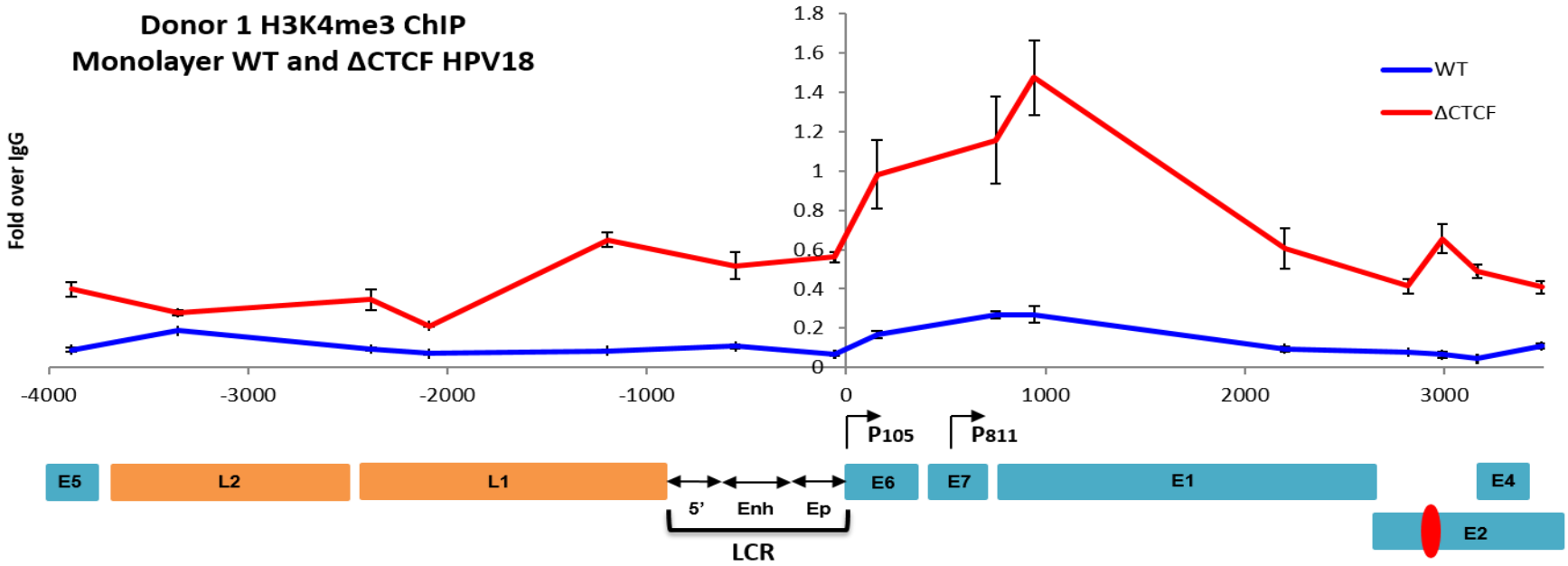


Figure 68. Characterisation of H3K4me3 enrichment throughout the HPV18 genome. HFKs containing either WT or ΔCTCF HPV18 genomes were grown in monolayer culture. Cells were fixed in 1% formaldehyde, lysed, and dounced with 40 strokes to aid nuclei release. Chromatin was sheared by sonication on a cycle of 30s on and 30s off for a total sonication time of 15 minutes. ChIP reactions were performed at 4°C overnight using 25µg of chromatin and 5µg of H3K4me3 antibody. ChIPs were also set up using FLAG antibody as a negative control. DNA was de-crosslinked and eluted and used for qPCR analysis using primers spanning the HPV18 genome. Protein binding was calculated as fold binding over FLAG ChIP. The ChIP shown is representative of three experimental repeats for donor 1 (n=3). Error bars represent the s.d of triplicate qPCR reactions. The predicted CTCF binding site is displayed as a red marker. The viral LCR is comprised of the 5' end, enhancer (Enh) and the early promoter (Ep). The early transcripts are transcribed from P_{105} and the late transcripts from P_{811}.

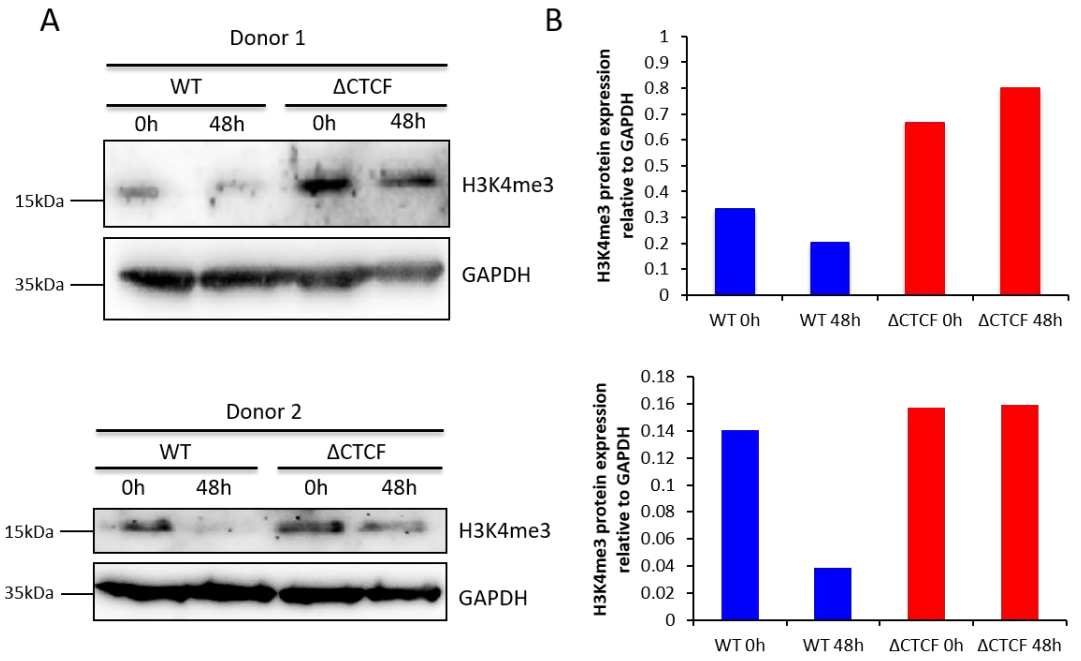


Figure 69. H3K4me3 expression in HFKs. A) HFKs containing either WT HPV18 or ΔCTCF HPV18 genomes were grown in monolayer or differentiated in methylcellulose for 48 hours. Protein lysates were extracted and 40 μg was used to detect for H3K4me3 and GAPDH was used as a loading control. **B)** Western blot bands were quantified using the Fusion FX digital detection system. H3K4me3 expression was calculated relative to GAPDH. The western shown is representative of two experimental repeats for donor 1 (n=2) and one experiment for donor 2 (n=1).

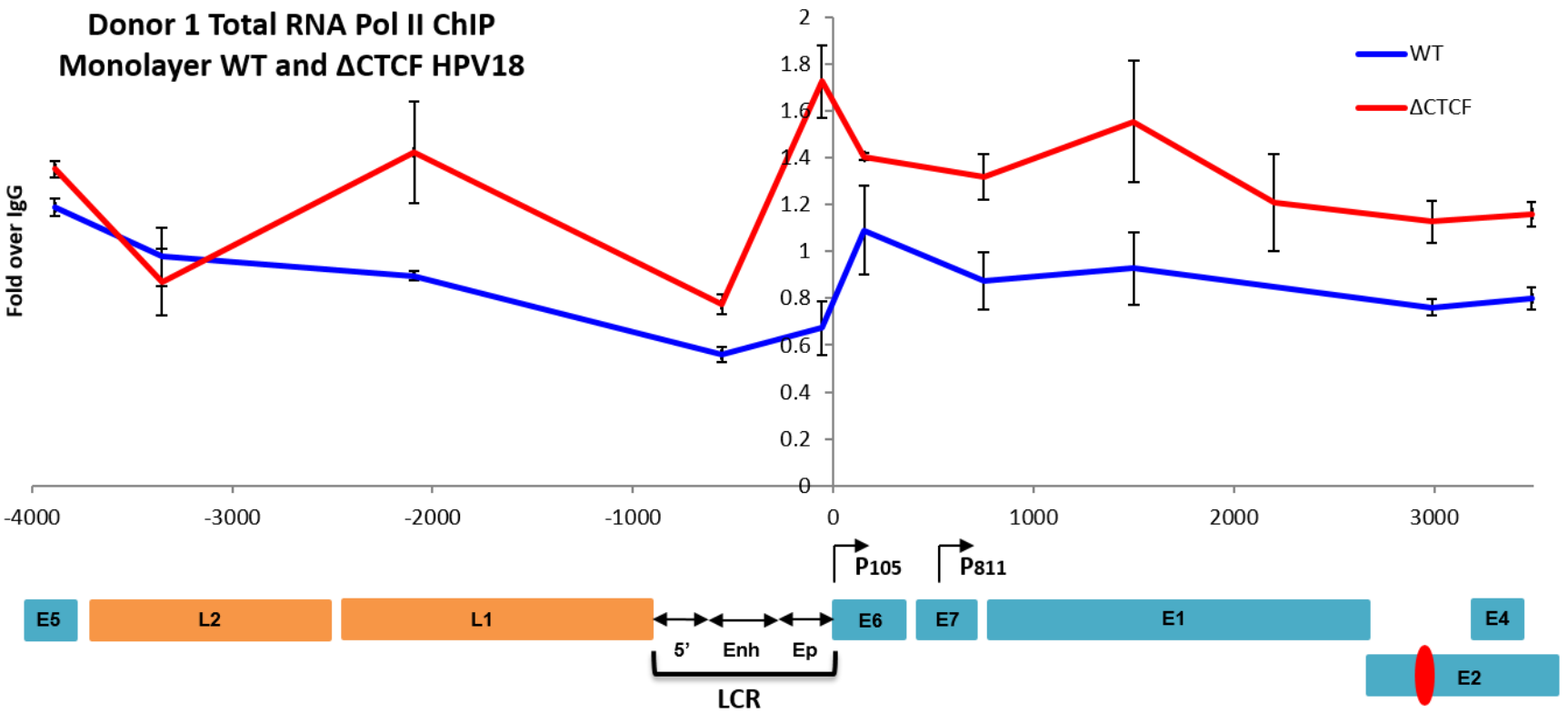


Figure 70. Binding of Total RNA polymerase II throughout the HPV18 genome. HFKs containing either WT HPV18 or ΔCTCF HPV18 genomes were grown in monolayer culture. Cells were fixed in 1% formaldehyde, lysed, and dounced with 40 strokes to aid nuclei release. Chromatin was sheared by sonication on a cycle of 30s on and 30s off for a total sonication time of 15 minutes. ChIP reactions were performed at 4°C overnight using 25µg of chromatin and 5µg of Total RNA Pol II antibody. ChIPs were also set up using FLAG antibody as a negative control. DNA was de-crosslinked and eluted and used for qPCR analysis using primers spanning the HPV18 genome. Protein binding was calculated as fold binding over FLAG. Error bars represent the s.d of triplicate qPCR reactions.The ChIP shown is representative of one experimental repeat for donor 1 (n=3).

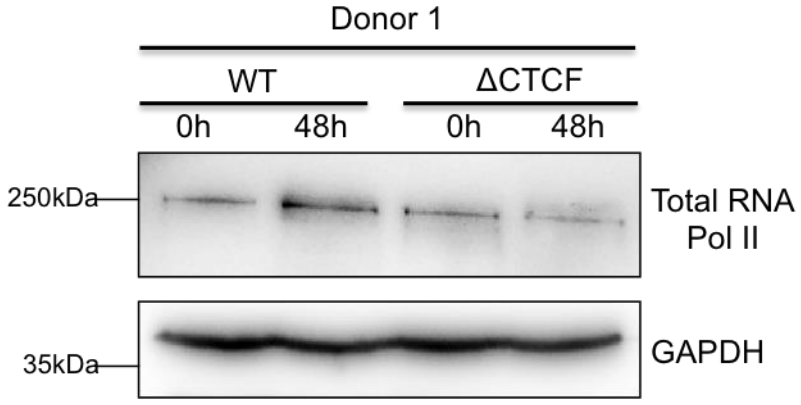


Figure 71. Total RNA polymerase II protein expression in HFKs. HFKs containing either WT HPV18 or ΔCTCF HPV18 genomes were grown in monolayer or differentiated in methylcellulose for 48 hours. Protein lysates were extracted and 30 μg was used to detect for total RNA Polymerase II and GAPDH was used as a loading control. The western shown is representative of two experimental repeats for donor 1 (n=2)

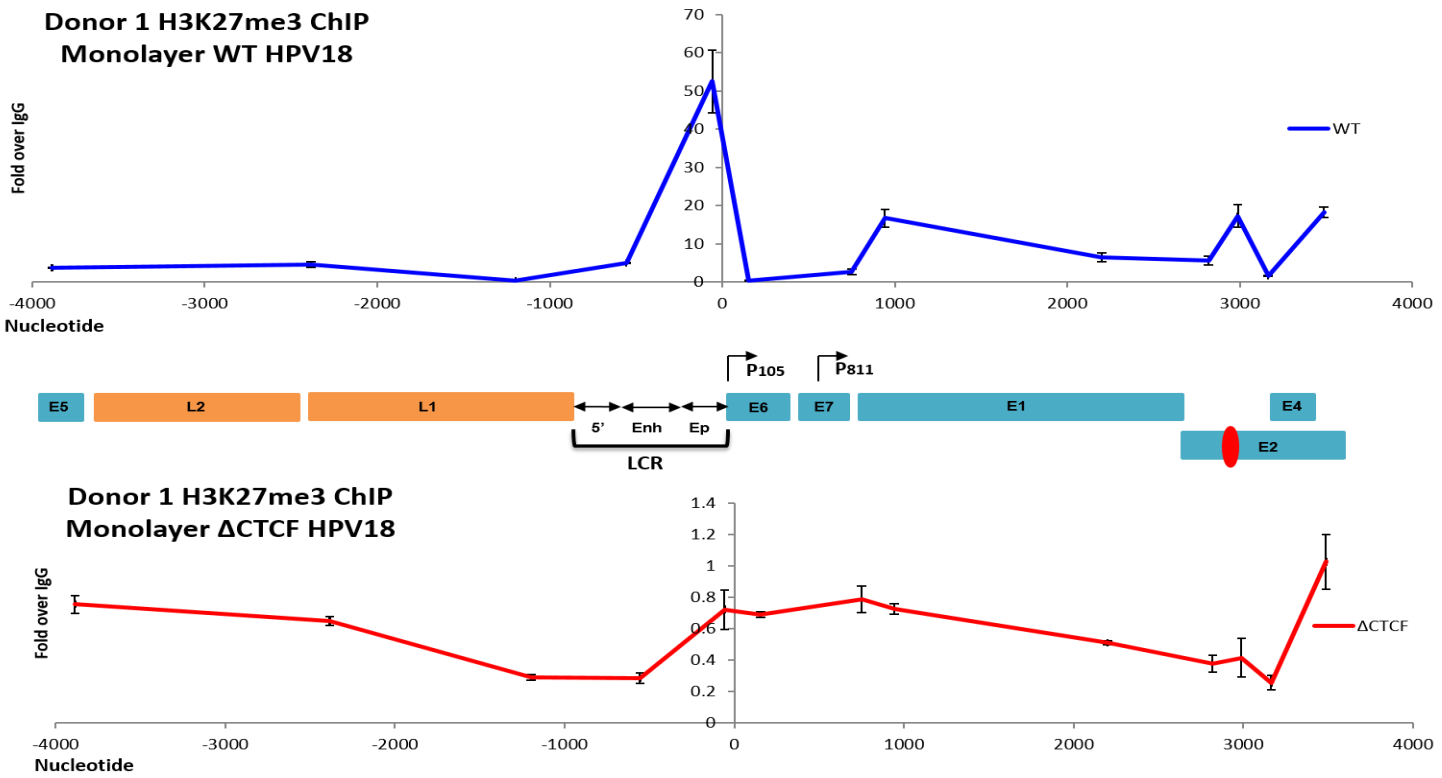


Figure 72. Characterisation of H3K27me3 enrichment throughout the HPV18 genome. HFKs containing either WT HPV18 or ΔCTCF HPV18 genomes were grown in monolayer culture. Cells were fixed in 1% formaldehyde, lysed, and dounced with 40 strokes to aid nuclei release. Chromatin was sheared by sonication on a cycle of 30s on and 30s off for a total sonication time of 15 minutes. ChIP reactions were performed at 4°C overnight using 25µg of chromatin and 5µg of H3K27me3 antibody. ChIPs were also set up using FLAG antibody as a negative control. DNA was de-crosslinked and eluted and used for qPCR analysis using primers spanning the HPV18 genome. Protein binding was calculated as fold binding over FLAG ChIP. Error bars represent the s.d of triplicate qPCR reactions.The ChIP shown is representative of three experimental repeats for donor 1 (n=3).

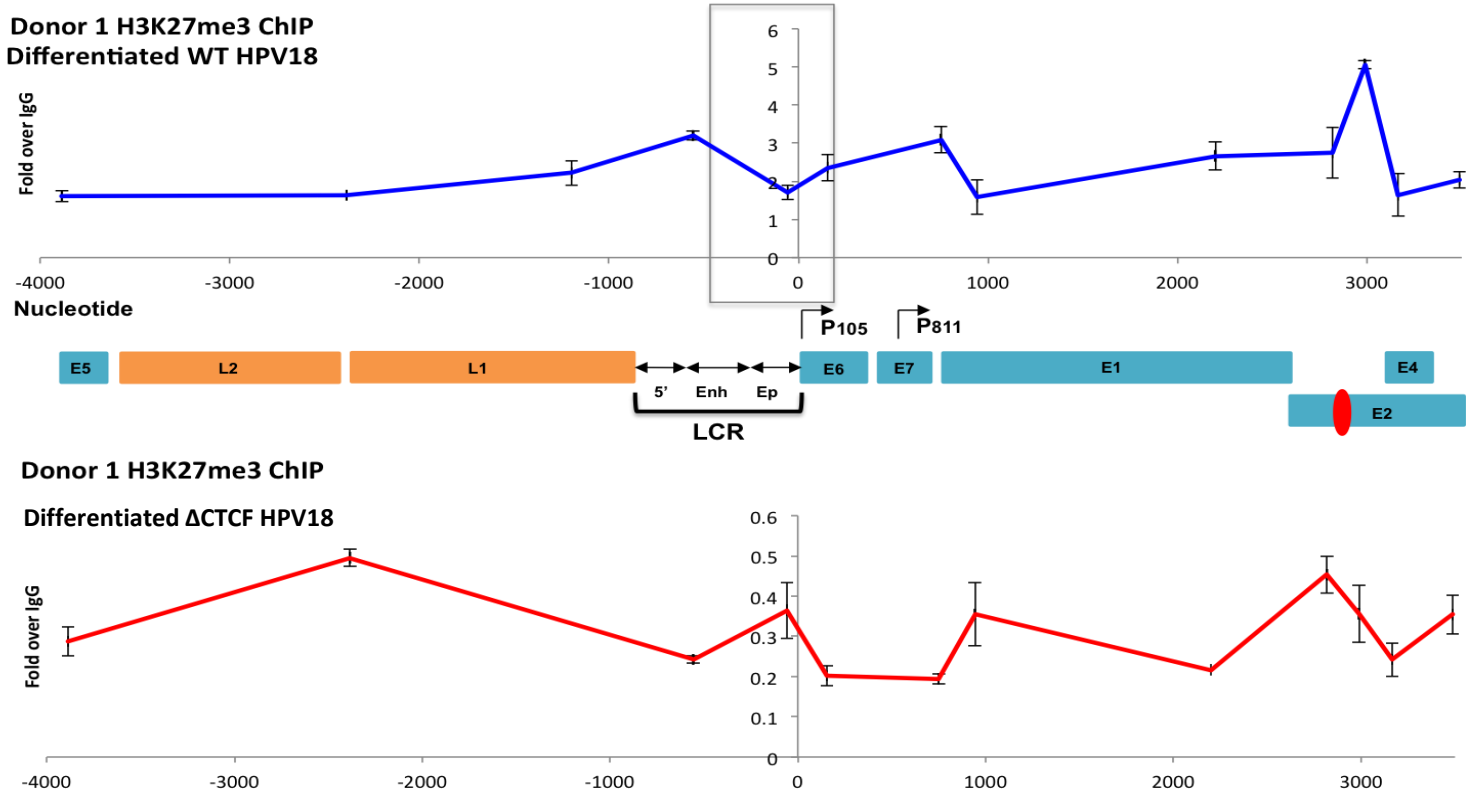


Figure 73. Characterisation of H3K27me3 enrichment throughout the HPV18 genome. HFKs containing either WT HPV18 or ΔCTCF HPV18 genomes were differentiated in methylcellulsoe for 48 hours. Cells were fixed in 1% formaldehyde, lysed, and dounced with 40 strokes to aid nuclei release. Chromatin was sheared by sonication on a cycle of 30s on and 30s off for a total sonication time of 15 minutes. ChIP reactions were performed at 4°C overnight using 25µg of chromatin and 5µg of H3K27me3 antibody. ChIPs were also set up using FLAG antibody as a negative control. DNA was de-crosslinked and eluted and used for qPCR analysis using primers spanning the HPV18 genome. Protein binding was calculated as fold binding over FLAG ChIP. Error bars represent the s.d of triplicate qPCR reactions.The ChIP shown is representative of three experimental repeats for donor 1 (n=3).

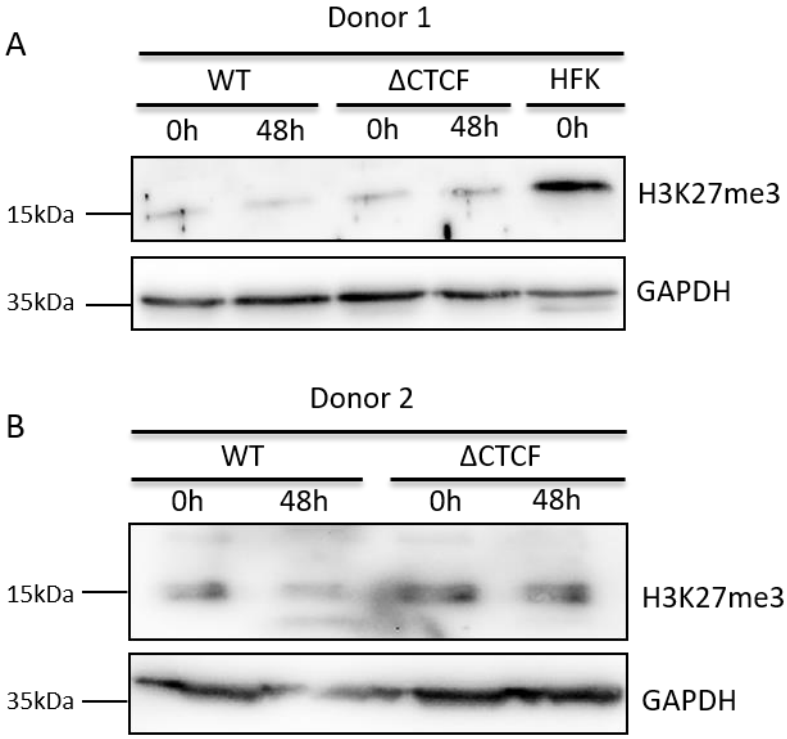


Figure 74. H3K27me3 expression in HFKs. HFKs containing either WT HPV18 or ΔCTCF HPV18 genomes for **A)** donor 1 and **B)** donor 2 were grown in monolayer or differentiated in methylcellulose for 48 hours. Protein lysates were extracted and 30 μg was used to detect for H3K27me3 and GAPDH was used as a loading control. The western shown is representative of two experimental repeats for donor 1 (n=2) and one experiment for donor 2 (n=1).

5.4 Loss of the repressive H3K27me3 mark in the LCR in ΔCTCF HPV18 genome containing HFKs

Conversely, it was also important to investigate the association of a repressive histone mark across the viral genome and determine if its distribution was altered between WT HPV18 and ΔCTCF HPV18 genome containing cells. ChIP analysis revealed peak H3K27me3 enrichment within the viral LCR in the WT HPV18 genome containing cells (Fig. 72); however, there was a loss of this epigenetic signature across the LCR region in the ΔCTCF HPV18 genome containing HFKs (Fig. 72). The observed enrichment of this repressive mark in the viral LCR correlates with previous studies that have demonstrated that H3K27me3 is enriched at transcriptional start sites of promoters (Pan et al., 2007). Another noticeable H3K27me3 enrichment peak was observed within the late promoter region in both WT HPV18 and ΔCTCF HPV18 HFKs in monolayer culture (Fig. 72). Interestingly, after methylcellulose differentiation, WT HPV18 genome containing cells displayed a loss of this H3K27me3 repressive mark across the LCR region (Fig. 73). Furthermore, the ΔCTCF HPV18 genome containing HFKs also maintained a loss of this H3K27me3 mark with the LCR region (Fig. 73). Overall, total H3K27me3 levels were reduced in HPV genome containing cells when compared to untransfected keratinocytes (Fig. 74a & 74b), which has been observed previously (McLaughlin-Drubin et al., 2011). Nevertheless, H3K27me3 levels remained similar between WT HPV18 and ΔCTCF HPV18 cells in both the

monolayer and differentiated state in two independent cell donors (Fig. 74a & 74b).

5.5 Increase in TEF-1 binding across the LCR in ΔCTCF HPV18 genome containing HFKs

The FAIRE experiments have revealed that the ΔCTCF HPV18 genome containing cells display a more open chromatin landscape in the LCR region. Therefore, it was initially hypothesized that the LCR region would be more accessible for binding of transcription factors associated with activation of early viral genes. ChIP experiments were performed to determine the binding of TEF-1, which had previously been shown to activate HPV16 early gene expression (Ishiji et al., 1992a). ChIP experiments revealed no difference in TEF-1 enrichment at the previously identified TEF-1 binding site in the viral LCR at position nt 7807 to 7815. However there was increased enrichment of TEF-1 binding within the viral enhancer and early promoter region in the ΔCTCF HPV18 cells compared to the WT HPV18 cells (Fig. 75a). Furthermore total TEF-1 protein expression was elevated in ΔCTCF HPV18 genome containing monolayer cells in comparison to the WT HPV18 genome containing cells, but expression was similar between both lines after methylcellulose differentiation (Figure. 75b). These experiments demonstrate that increased TEF-1 binding to the HPV18 LCR may contribute to the

observed increase in early gene expression in the ΔCTCF HPV18 genome containing HFKs.

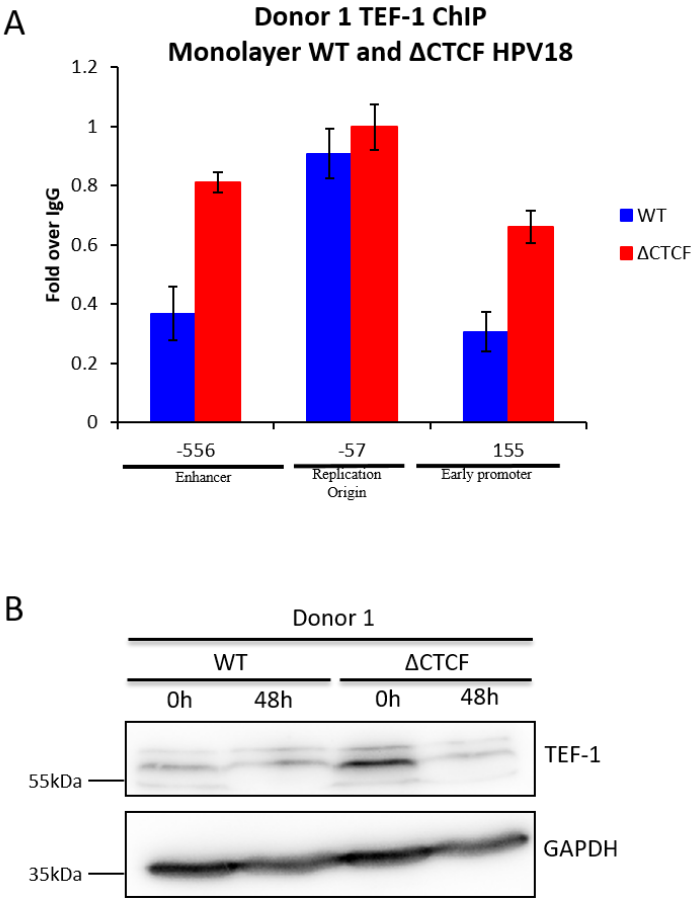


Figure 75. Association of TEF-1 protein within the HPV18 LCR and early promoter. HFKs containing either WT HPV18 or ΔCTCF HPV18 genomes were grown in monolayer or differentiated in methylcellulose for 48 hours. **A)** Monolayer cells were fixed in 1% formaldehyde, lysed, and dounced with 40 strokes to aid nuclei release. Chromatin was sheared by sonication on a cycle of 30s on and 30s off for a total sonication time of 15 minutes. ChIP reactions were performed at 4°C overnight using 25µg of chromatin and 5µg of TEF-1 antibody. ChIPs were also set up using FLAG antibody as a negative control. DNA was de-crosslinked and eluted and used for qPCR analysis using primers at sites in the HPV18 LCR and early promoter. Protein binding was calculated as fold binding over FLAG. The ChIP shown is representative of one experiment for donor 1 (n=1). Error bars represent the standard deviation of triplicate qPCR reactions. **B)** Protein lysates were extracted from monolayer and differentiated WT HPV18 and ΔCTCF HPV18 cells and 30 µg was used to detect for TEF-1. GAPDH was used as a loading control. The western shown is representative of two experimental repeats for donor 1 (n=2).

5.6 Loss of YY1 binding to the viral LCR in ΔCTCF HPV18 genome containing HFKs

There are also transcription factors that are known to bind to the LCR and repress HPV early gene expression, so it was also important to map the association of these factors across the viral LCR. Initial experiments investigated the association of the transcription factor Ying Yang 1 (YY1), which is already known to bind to the viral LCR and repress early gene expression (OConnor et al., 1996, Lace et al., 2009). ChIP experiments were performed to determine the association of YY1 across the viral LCR of WT HPV18 and ΔCTCF HPV18 genome containing HFKs in monolayer culture and after 48 hours methylcellulose differentiation. Interestingly, whilst there was a 2.5-fold enrichment of YY1 in the LCR of WT HPV18 monolayer cells, there was loss of binding in the ΔCTCF HPV18 cells (Fig. 76a). A similar pattern was observed in a second cell donor (Fig. 76c). After 48 hours methylcellulose differentiation, YY1 binding was reduced to 2-fold in the WT HPV18 cells and binding was still abrogated in differentiated ΔCTCF HPV18 cells (Fig. 76b). The loss of YY1 binding was not due to the loss of overall YY1 protein expression, as this clearly remained similar between WT HPV18 and ΔCTCF HPV18 monolayer cells (Fig. 77a & 77b). Moreover, the level of YY1 protein expression was reduced upon differentiation in both WT HPV18 and ΔCTCF HPV18 genome containing HFKs, however the relative expression was similar between the two cell lines (Fig. 77a & 77b). Overall, it is apparent

that upon abrogation of CTCF binding in the E2 ORF there is a loss of YY1 binding in viral LCR, which is approximately 3kb away in distance from the CTCF binding site. Moreover, many of the YY1 binding sites in the HPV18 LCR overlap with the TEF-1 binding sites. Therefore, it is possible that in the ΔCTCF HPV18 monolayer HFKs that express higher levels of TEF-1 protein, the TEF-1 protein may be competitively binding with YY1 on the overlapping sites in the LCR, and thus reducing access for YY1 binding.

Sections of organotypic raft cultures were also stained with YY1 specific antibody to determine the expression of YY1 protein throughout the differentiated epithelium (Fig. 78). In each of the three rafts derived from either untransfected, WT HPV18 or ΔCTCF HPV18 genome containing HFKs there was abundant YY1 expression in the basal and lower suprabasal layers; however, there was a clear reduction in the upper most layers of each of the rafts (Fig. 78). These staining results support the western blot results that showed a reduction in YY1 protein expression upon methylcellulose differentiation (Fig. 77).

YY1 is a particularly interesting protein as it has been identified as the functional homolog of the Drosophila Pho protein, which is known to recruit the polycomb group proteins (PcG) to DNA and mediate transcriptional

silencing (Brown et al., 1998). A previous study has demonstrated that knockdown of YY1 protein in HeLa cells caused a reduction in polycomb recruitment and a concomitant increase in Hox gene expression (Basu et al., 2014a). Therefore it was hypothesized that the observed reduction in YY1 binding in the LCR of ΔCTCF HPV18 genome containing HFKs may cause a reduction in polycomb recruitment to the viral LCR and early promoter regions, and cause a concomitant increase in early gene expression. The aim of subsequent experiments was to map the association of polycomb proteins across the viral LCR and compare the binding profile between WT HPV18 and ΔCTCF HPV18 genome containing cells.

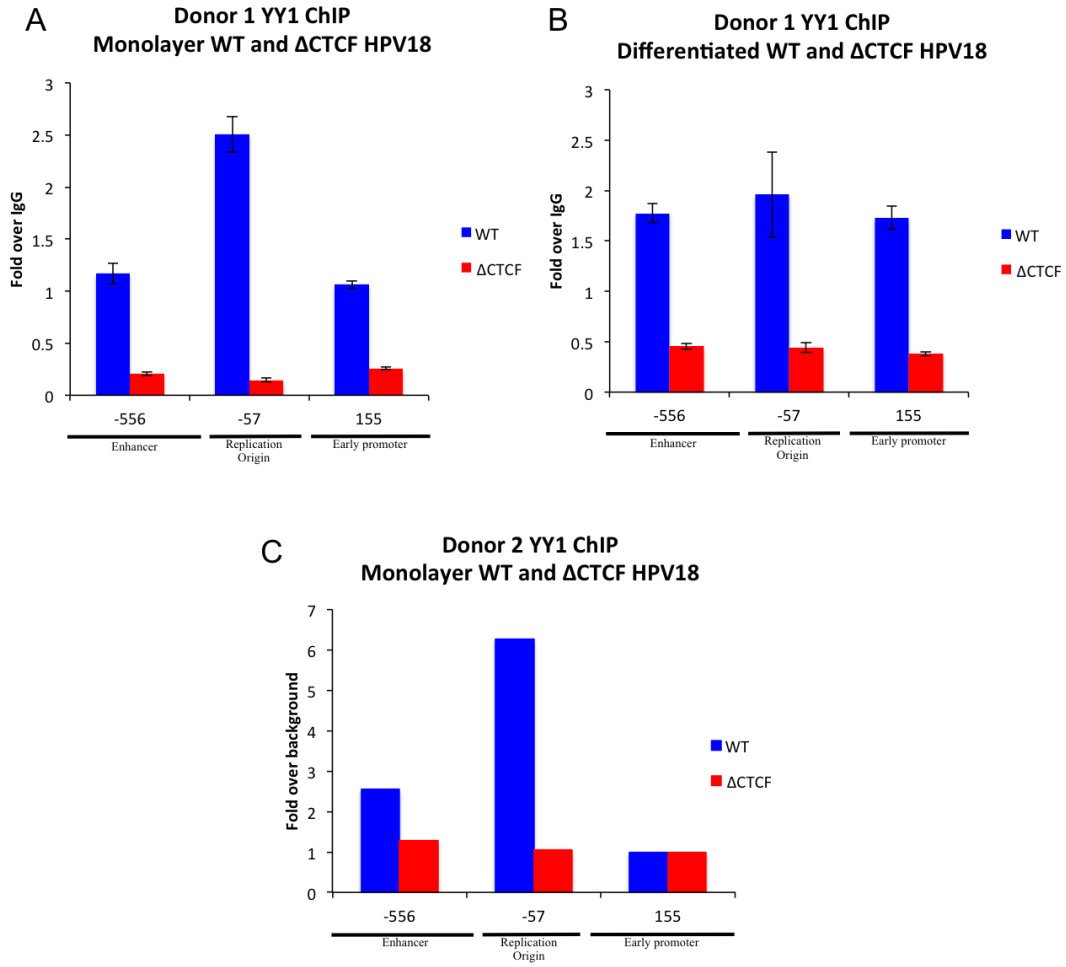


Figure 76. Association of YY1 protein within the HPV18 LCR and early promoter. Donor 1 HFKs containing either WT HPV18 or ΔCTCF HPV18 genomes were grown in **A)** monolayer culture or **B)** differentiated in methylcellulose for 48 hours. **C)** Donor 2 HFKs containing either WT HPV18 or ΔCTCF HPV18 genomes were grown in monolayer. Cells were fixed in 1% formaldehyde, lysed, and dounced with 40 strokes to aid nuclei release. Chromatin was sheared by sonication on a cycle of 30s on and 30s off for a total sonication time of 15 minutes. ChIP reactions were performed at 4°C overnight using 25µg of chromatin and 5µg of YY1 antibody. ChIPs were also set up using FLAG antibody as a negative control. DNA was de-crosslinked and eluted and used for qPCR analysis using primers at sites in the HPV18 LCR and early promoter. Protein binding was calculated as fold binding over FLAG. **A)** and **B)** ChIPs shown are representative of three experimental repeats for donor 1 (n=3). Error bars represent the standard deviation of triplicate qPCR reactions. **C)** One experimental ChIP from donor 2 (n=1).

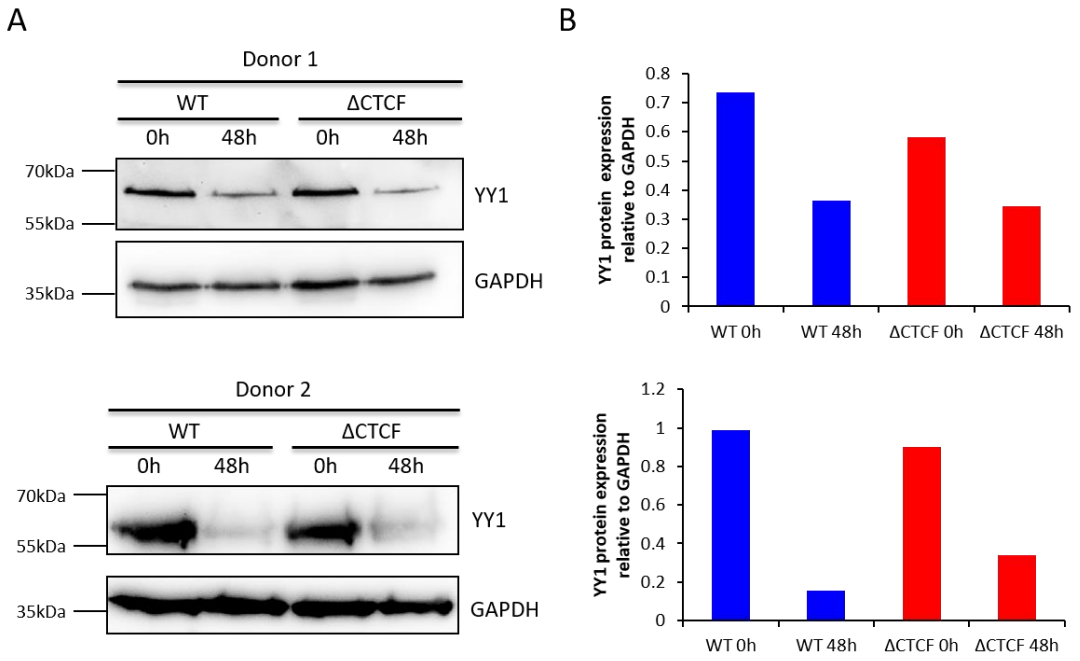


Figure 77. YY1 protein expression in HFKs. A) HFKs containing either WT HPV18 or ΔCTCF HPV18 genomes were grown in monolayer or differentiated in methylcellulose for 48 hours. Protein lysates were extracted and 30 μg (donor 1) or 40 μg (Donor 2) was used to detect YY1 and GAPDH was used as a loading control. **B)** Western blot bands were quantified using the Fusion FX digital detection system. YY1 expression was calculated relative to GAPDH. The western shown is representative of two experimental repeats for donor 1 (n=2) and one experiment for donor 2 (n=1).

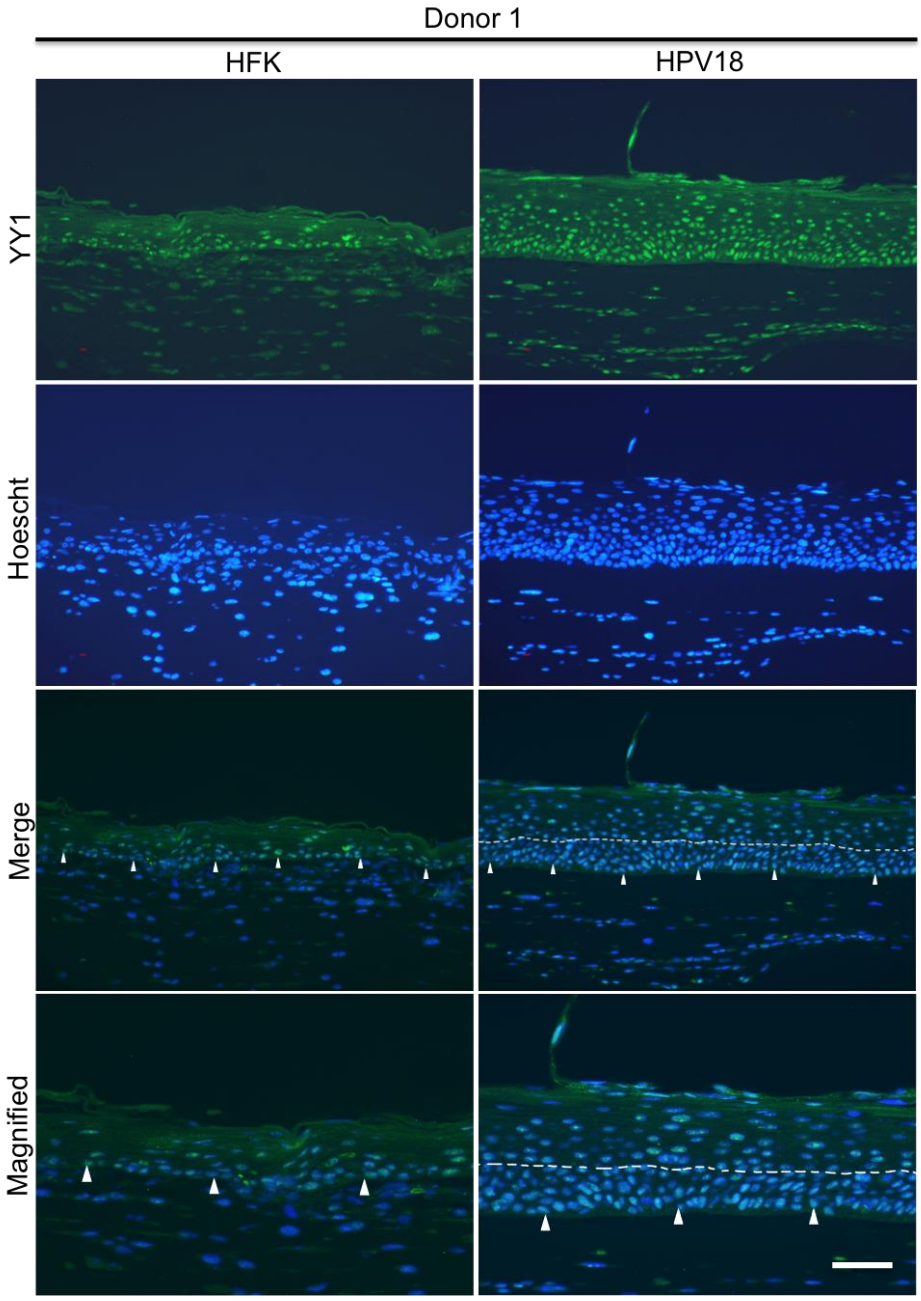


Figure 78. Immunofluorescence staining of YY1 in untransfected and WT HPV18 organotypic raft sections. YY1 antibody was used at 1:2000 in BSA blocking solution with staining shown in the green panels. Hoescht 3342 was used to stain the nuclear DNA as shown in the blue panels. Sides were visualized on a Nikon inverted Epifluorescent microscope fitted with a 20x objective and images were captured using a Leica DC200 camera and software. Scale bar, 10 μm The white arrows indicate the basal layer and the lower/upper suprabasal boundary is demarcated by the white dotted line. Images are representative of raft sections from cell donor 1.

5.7 Association of the PRC2 complex protein EZH2 within the HPV18 LCR region

Enhancer of zeste homolog 2 (EZH2) is a methyltransferase protein that forms part of the PRC2 polycomb complex along with SUZ12 and EED. EZH2 is the only known methyltransferase responsible for catalyzing the methylation of Histone 3 at lysine 27 and enrichment of this H2K27me3 mark correlates with repressive chromatin (Kuzmichev et al., 2002, Barski et al., 2007a, Cao et al., 2002). ChIP experiments were performed to determine the binding of EZH2 protein within the viral LCR of WT HPV18 and ΔCTCF HPV18 HFKs. Surprisingly, there was a slight increase in the amount of EZH2 binding to the LCR region in ΔCTCF HPV18 HFKs compared to WT HPV18 in monolayer culture (Fig. 79a) and binding was similar between WT HPV18 and ΔCTCF HPV18 genome containing cells after methylcellulose differentiation (Fig. 79b). However, EZH2 is able to bind DNA independently of the PRC2 complex, so the observed binding of EZH2 may not be acting as part of the PRC2 complex. Additionally, whilst the EZH2 ChIP antibody was able to detect total EZH2 protein binding within the viral LCR it was not able to distinguish between the phosphorylated forms of EZH2. It is known that the phosphorylated form of EZH2 at Y224 can also act as a transcriptional activator (Yan et al., 2016), so the observed binding within the LCR may be a combination of the PRC2 EZH2 and/or the activating phosphorylated form. Therefore further ChIPs need to be carried out to decipher which form of

EZH2 is binding within the viral LCR. Overall, EZH2 protein expression was similar between WT HPV18 and ΔCTCF HPV18 monolayer cells, and upon differentiation EZH2 western blotting detected an additional band in both WT HPV18 and ΔCTCF HPV18 cells, indicating that additional post-translational modifications occur on the EZH2 protein upon cellular differentiation (Fig. 79c). Furthermore, the overall level of EZH2 protein in HPV18 or ΔCTCF HPV18 genome containing cells was higher than untransfected HFKs, which has previously been observed in HPV16 E6/E7 transfected cell lines (Hyland et al., 2011, McLaughlin-Drubin et al., 2011) (Fig. 79c).

5.8 Loss of EED binding to the LCR in ΔCTCF HPV18 genome containing HFKs

Given that EZH2 can bind to DNA independently of the PRC2 complex and positively regulate transcription, it was important to investigate the binding of another protein in the PRC2 complex. EED is not thought to bind to DNA and function independently from the PRC2 complex, so it is an appropriate indicator of PCR2 complex binding (Margueron et al., 2009). ChIP experiments were performed to determine the association of EED across the viral LCR of WT HPV18 and ΔCTCF HPV18 genome containing HFKs (Fig. 80). Interestingly, whilst there was binding of EED in the LCR of WT HPV18 cells there was loss of binding in the ΔCTCF HPV18 cells, indicating that there is loss of the PRC2 complex within the ΔCTCF HPV18 LCR region (Fig. 80a).

Overall, total EED protein expression was similar between WT HPV18 and ΔCTCF HPV18 monolayer cells, and upon cellular differentiation there was a similar increase in EED levels in both WT HPV18 and ΔCTCF HPV18 genome containing cells (Fig. 80b and 80c).

5.9 Loss of RING1B binding to the LCR in ΔCTCF HPV18 genome containing HFKs

Events involving the PCR2 complex are thought to be sequentially followed by the activity of the PRC1 complex, which always contains the E3 ligase protein RING1B. However there is also emerging evidence that initial PRC1 activity actually recruits PRC2 to target sites (Blackledge et al., 2014). The most understood role of the PRC1 complex is to recognize the H3K27me3 mark laid down by PRC2 and RING1B is responsible for catalyzing H2A monoubiqtinylation, which leads to chromatin compaction and silencing (Wang et al., 2004a). ChIP experiments were performed to determine the binding of RING1B protein across the viral LCR of WT HPV18 and ΔCTCF HPV18 genome containing HFKs (Fig. 81). Interestingly, whilst there was binding of RING1B in the LCR of WT HPV18 cells, there was loss of binding in the ΔCTCF HPV18 cells, indicating a loss of the PRC1 complex to the ΔCTCF HPV18 LCR region (Fig. 81a). Overall, RING1B protein levels were similar between WT HPV18 and ΔCTCF HPV18 genome containing monolayer HFKs, and whilst expression of this protein was reduced upon

differentiation it remained similar between the WT HPV18 and ΔCTCF HPV18 genome containing cell lines (Fig. 81b).

5.10 Loss of H2AK119Ub within the early gene region in ΔCTCF HPV18 genome containing HFKs

Ubiquitinylation of histone 2A (H2A) at lysine 119 by RING1B is associated with the formation of compact and silenced chromatin within the genome (Wang et al., 2004a); therefore it was important to determine the association of total H2AK119Ub within the viral LCR of WT HPV18 and ΔCTCF HPV18 genome containing HFKs (Fig. 82). ChIPs revealed abundant H2AK119Ub enrichment within the early and late promoters of the WT HPV18 HFKs; however, in the ΔCTCF HPV18 HFKs, there was a reduction of this ubiquitinylated H2A mark within the early gene region (Fig. 82a). These results indicate that the early promoter region is less compact in the ΔCTCF HPV18 cells compared to the WT HPV18 cells and subsequently more transcriptionally active. Overall, H2AK119Ub expression was similar between WT HPV18 and ΔCTCF HPV18 genome containing cells in monolayer and differentiated culture (Fig. 82b and c).

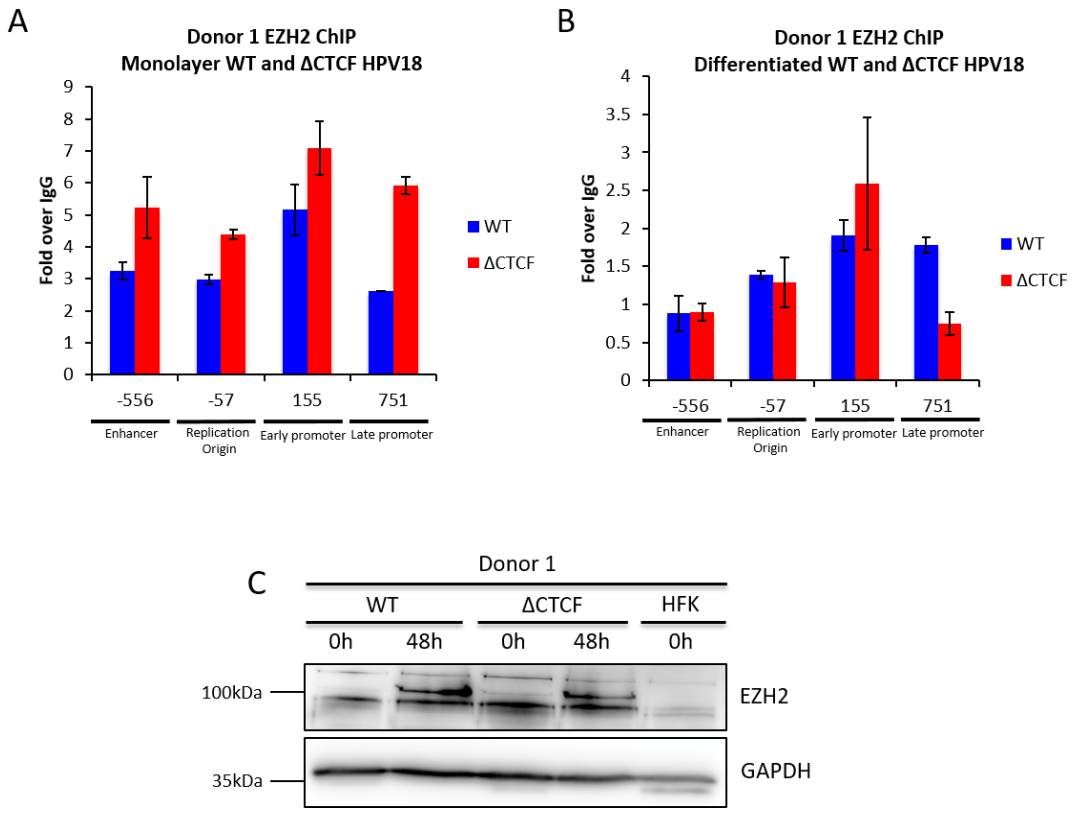


Figure 79. Association of EZH2 protein within the HPV18 LCR and early promoter. HFKs containing either WT HPV18 or ΔCTCF HPV18 genomes were grown in **A)** monolayer culture or **B)** differentiated in methylcellulsoe for 48 hours. Cells were fixed in 1% formaldehyde, lysed, and dounced with 40 strokes to aid nuclei release. Chromatin was sheared by sonication on a cycle of 30s on and 30s off for a total sonication time of 15 minutes. ChIP reactions were performed at 4°C overnight using 25µg of chromatin and 5µg of EZH2 antibody. ChIPs were also set up using FLAG antibody as a negative control. DNA was de-crosslinked and eluted and used for qPCR analysis using primers at sites in the HPV18 LCR and early promoter. Protein binding was calculated as fold binding over FLAG. The ChIP shown is representative of 2 experimental repeats for donor 1 (n=2). Error bars represent the standard deviation of triplicate qPCR reactions. **C)** Protein lysates were extracted from WT, ΔCTCF and untransfected cells (HFK) and 30 µg was used to detect EZH2 and GAPDH was used as a loading control.

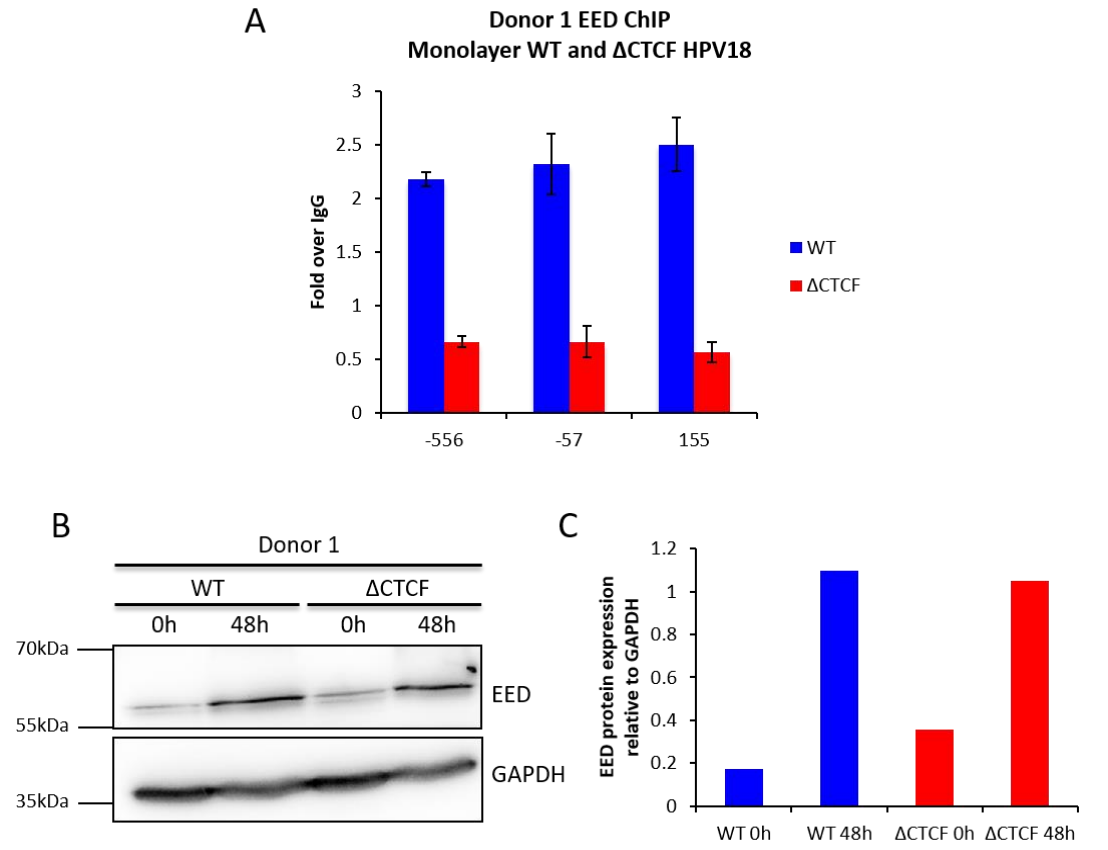


Figure 80. Association of EED protein with the HPV18 LCR and early promoter. HFKs containing either WT HPV18 or ΔCTCF HPV18 genomes were grown in monolayer culture. Cells were fixed in 1% formaldehyde, lysed, and dounced with 40 strokes to aid nuclei release. Chromatin was sheared by sonication on a cycle of 30s on and 30s off for a total sonication time of 15 minutes. ChIP reactions were performed at 4°C overnight using 25µg of chromatin and 5µg of EED antibody. ChIPs were also set up using FLAG antibody as a negative control. DNA was de-crosslinked and eluted and used for qPCR analysis using primers at sites in the HPV18 LCR and early promoter. Protein binding was calculated as fold binding over FLAG. The ChIP shown is representative of 2 experimental repeats for donor 1 (n=2). Error bars represent the standard deviation of triplicate qPCR reactions. **B)** Protein lysates were extracted from WT and ΔCTCF HFKs and 30 µg of lysate was used to detect EED and GAPDH was used as a loading control. **C)** Western blot bands were quantified using the Fusion FX digital detection system. EED expression was calculated relative to GAPDH.

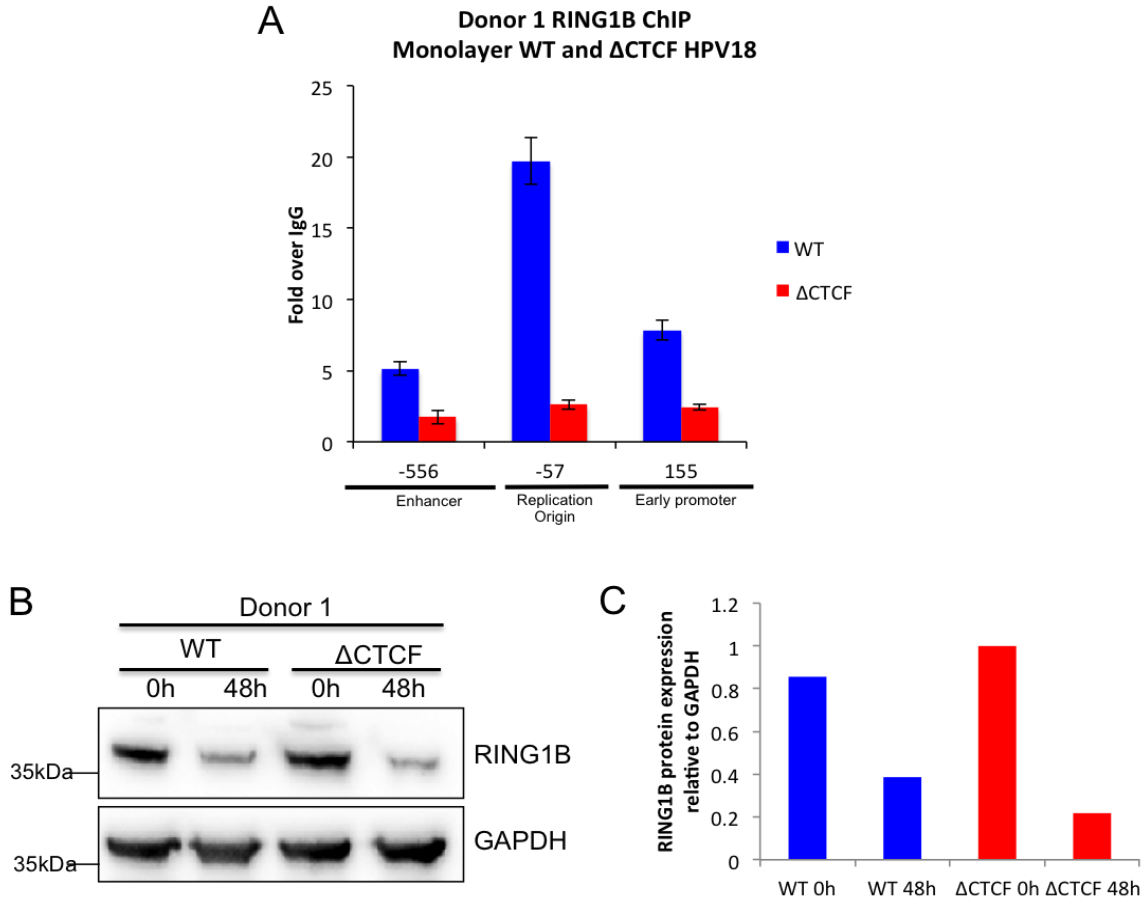


Figure 81. Association of RING1B protein with the HPV18 LCR and early promoter. HFKs containing either HFKs containing either WT HPV18 or ΔCTCF HPV18 genomes were grown in monolayer or differentiated in methylcellulose for 48 hours. **A)** Monolayer cells were fixed in 1% formaldehyde, lysed, and dounced with 40 strokes to aid nuclei release. Chromatin was sheared by sonication on a cycle of 30s on and 30s off for a total sonication time of 15 minutes. ChIP reactions were performed at 4°C overnight using 25µg of chromatin and 5µg of RING1B antibody. ChIPs were also set up using FLAG antibody as a negative control. DNA was de-crosslinked and eluted and used for qPCR analysis using primers at sites in the HPV18 LCR and early promoter. Protein binding was calculated as fold binding over FLAG. The ChIP shown is representative of two experimental repeats for donor 1 (n=2). Error bars represent the standard deviation of triplicate qPCR reactions. **B)** Protein lysates were extracted from monolayer and differentiated WT HPV18 and ΔCTCF HPV18 cells and 30 µg was used to detect for RING1B. GAPDH was used as a loading control. The western shown is representative of two experimental repeats for donor 1 (n=2). **C)** Western blot bands were quantified using the Fusion FX digital detection system. RING1B expression was calculated relative to GAPDH.

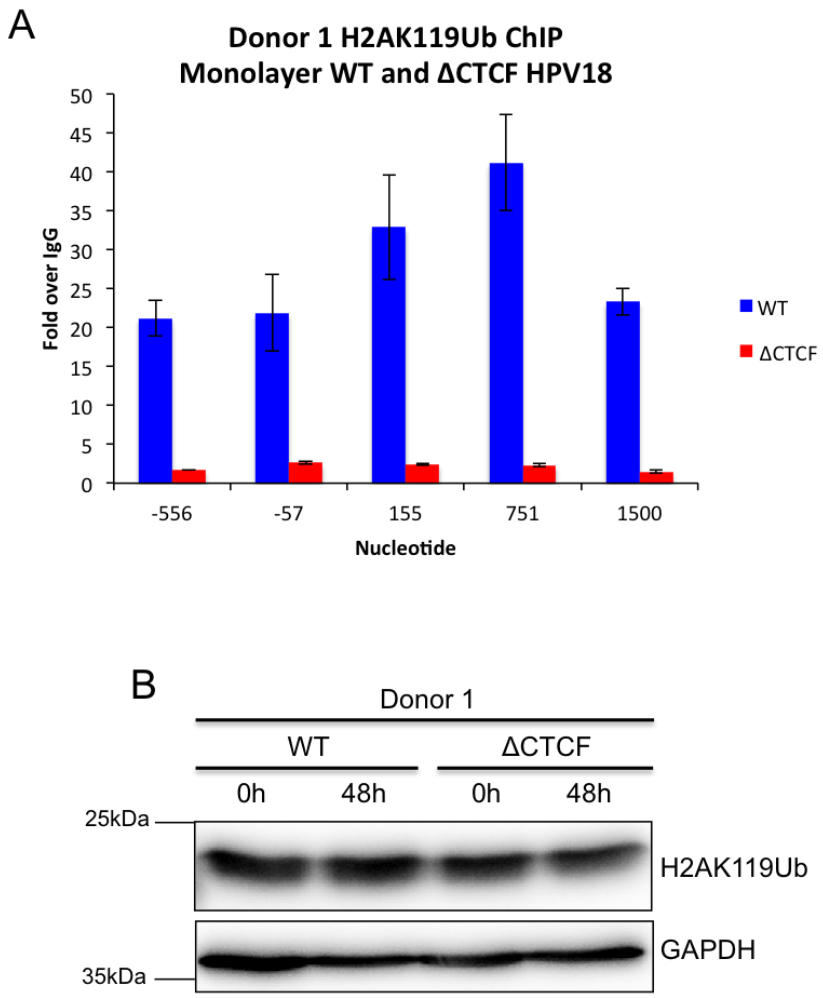


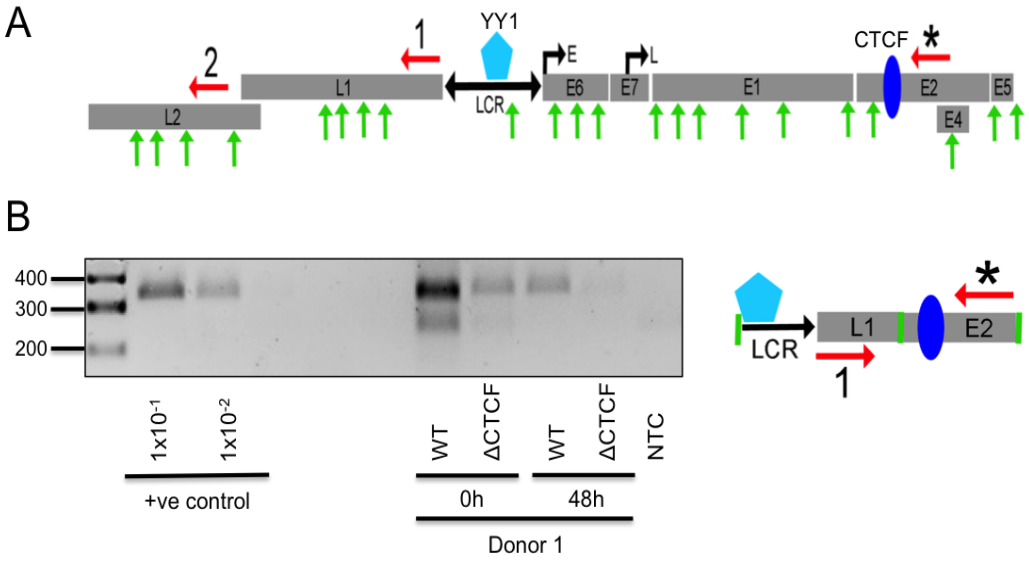
Figure 82. Association of H2AK119Ub within the HPV18 LCR and early promoter. HFKs containing either HFKs containing either WT HPV18 or ΔCTCF HPV18 genomes were grown in monolayer or differentiated in methylcellulose for 48 hours. **A)** Monolayer cells were fixed in 1% formaldehyde, lysed, and dounced with 40 strokes to aid nuclei release. Chromatin was sheared by sonication on a cycle of 30s on and 30s off for a total sonication time of 15 minutes. ChIP reactions were performed at 4°C overnight using 25µg of chromatin and 5µg of H2AK119Ub antibody. ChIPs were also set up using FLAG antibody as a negative control. DNA was de-crosslinked and eluted and used for qPCR analysis using primers at sites in the HPV18 LCR and early promoter. Protein binding was calculated as fold binding over FLAG. The ChIP shown is representative of two experimental repeats for donor 1 (n=2). Error bars represent the standard deviation of triplicate qPCR reactions. **B)** Protein lysates were extracted from monolayer and differentiated WT HPV18 and ΔCTCF HPV18 genome containing HFKs and 30 µg was used to detect H2AK119Ub. GAPDH was used as a loading control. The western shown is representative of two experimental repeats for donor 1 (n=2).

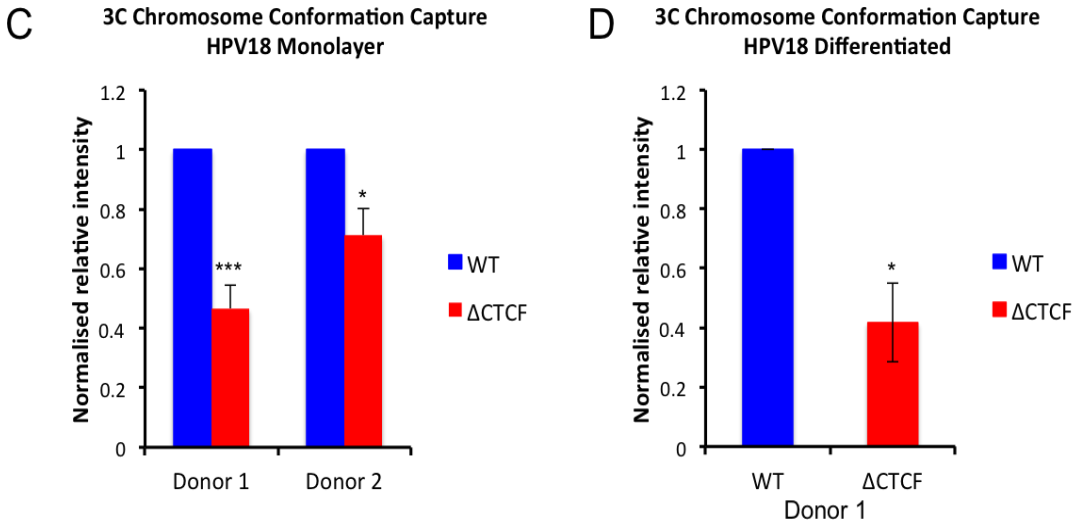
5.11 Analysis of chromatin loop formation between the CTCF binding site in the HPV18 E2 ORF and YY1 in the HPV18 LCR region

When CTCF binding was abrogated in the E2 ORF, ChIP experiments revealed a loss of YY1 binding in the viral LCR, concomitant with a loss of the repressive mark H3K27me3 in the LCR and an increase in enrichment of the H3K4me3 mark within the early promoter. However it remained unclear how the YY1 protein in the LCR was functionally associated with the CTCF binding site 3kb away in the E2 ORF. Moreover, CTCF and YY1 are known protein-binding partners (Donohoe et al., 2007), and have been shown to mediate chromatin loop formation during neural lineage commitment (Beagan et al., 2017). With this in mind it was hypothesised that a repressive DNA loop was formed between CTCF in the E2 ORF and YY1 in the LCR, which was required to recruit polycomb complex proteins and attenuate early gene expression in monolayer culture. To test this hypothesis, chromosome conformation capture (3C) was employed to determine the presence of a potential DNA loop between the CTCF binding site in the E2 ORF and YY1 in the viral LCR. The restriction enzyme *Nla*III was used to digest HPV18 DNA and multiple sites throughout the HPV18 genome. After restriction enzyme digestion, the DNA was re-ligated, thus allowing cut ends held in close proximity to form ligation products. In this experiment reverse primers in the E2 and L1 regions were designed to detect a novel PCR product that would only form if the CTCF and YY1 binding regions were in close enough proximity for re-ligation (Fig. 83a). This PCR product was successfully detected in WT HPV18 monolayer cells and to a much lesser extent in

the ΔCTCF HPV18 genome containing cells (Fig. 83b). The ligation product was reduced upon methylcellulose differentiation of WT HPV18 cells and barely detectable in ΔCTCF HPV18 differentiated cells (Fig. 83b). A DNA fragment that was the same sequence of the predicted ligation product was also amplified as a positive control. Each PCR product was sequenced to ensure the presence of the predicted product. Both the WT HPV18 and ΔCTCF HPV18 PCR product band intensities were calculated relative to the positive control band. These normalized values were used to calculate the amount of ligation product in ΔCTCF HPV18 cells relative to WT HPV18 cells. Overall, there was a ~55% reduction in ligation product detected in monolayer ΔCTCF HPV18 genome containing cells in donor 1 ($p<0.001$) and a ~30% reduction in donor 2 ($p<0.05$) (Fig. 83c). Furthermore, there was a ~60% reduction in ligation product in ΔCTCF HPV18 genome containing cells after 48 hours methylcellulose differentiation, when compared to the differentiated WT HPV18 genome containing cells ($p<0.05$) (Fig. 83d). However, the most interesting result was a ~68% reduction in ligation product in the WT HPV18 genome containing cells after differentiation when compared to the WT HPV18 genome containing cells in monolayer culture ($p<0.001$) (Fig. 84e). Overall, these 3C results provide strong evidence that there is a long-range DNA interaction (3kb) between CTCF in the E2 ORF and YY1 in the viral LCR in monolayer culture, and formation of this loop is reduced upon methylcellulose differentiation.

During the 3C experiment it is possible that random interactions can occur between digested sections of the viral genome. Therefore the primer within the same E2 region and new primer located in the L2 region were used for a negative control PCR reaction, where a long range DNA interaction was not predicted to occur between these regions. PCR reactions confirmed that there were no ligation products detected in the WT HPV18 or ΔCTCF HPV18 genome containing cells with this control primer pair (Fig. 84). This provides evidence that the ligation product formed between the E2 ORF and the LCR has not occurred from random genomic interactions. Additionally, restriction enzyme digestion efficiencies were calculated to ensure equal DNA digestion between WT HPV18 and ΔCTCF HPV18 samples. An internal control primer pair in the E1 gene region and a primer pair spanning a selected *Nla*III digest site were used in separate qPCR reactions. This quantitative qPCR approach allowed the calculation of the $2^{\Delta\Delta}$Ct values and digestion efficiencies for each donor line. Digestion efficiencies were calculated to be over 70% for both WT HPV18 and ΔCTCF HPV18 samples, so the observed reduction in the PCR ligation product in ΔCTCF HPV18 cell lines was not due to different digestion efficiencies.


A
1
2
YY1
E
L
CTCF
L1
LCR
E6
E7
E1
E2
E5
L2
E4
B
400
300
200
1x10-1
1x10-2
WT
ΔCTCF
WT
ΔCTCF
NTC
+ve control
0h
48h
Donor 1
LCR
L1
E2
1


C
3C Chromosome Conformation Capture
HPV18 Monolayer
Normalised relative intensity
1.2
1
0.8
0.6
0.4
0.2
0
***
*
WT
ΔCTCF
Donor 1
Donor 2
D
3C Chromosome Conformation Capture
HPV18 Differentiated
Normalised relative intensity
1.2
1
0.8
0.6
0.4
0.2
0
*
WT
ΔCTCF
WT
ΔCTCF
Donor 1

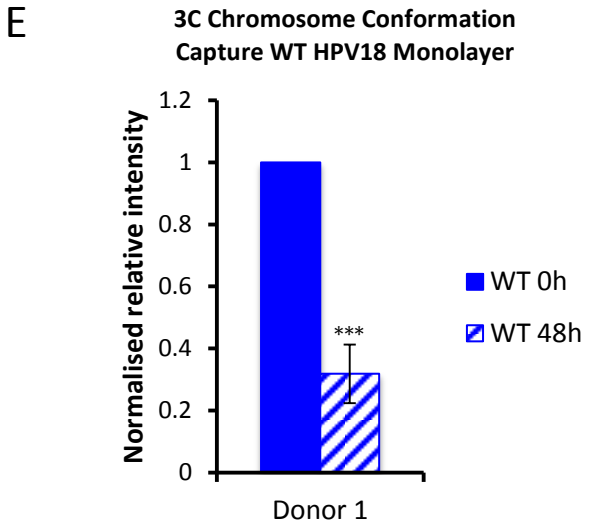

E
3C Chromosome Conformation
Capture WT HPV18 Monolayer
Normalised relative intensity
1.2
1
0.8
0.6
0.4
0.2
0
***
WT 0h
WT 48h
Donor 1

Figure 83. 3C Chromosome conformation capture analysis of WT HPV18 and ΔCTCF HPV18 genomes in HFKs. A) Schematic depicting the location of 3C primers on the HPV18 genome. **B)** HFKs containing either WT HPV18 or ΔCTCF HPV18 genomes were grown in monolayer culture or differentiated in methylcellulose for 48 hours. Cell pellets were harvested and resuspended in 10% (v/v) FBS/PBS. Cell suspensions were filtered through a 70 μm cell strainer to make a single cell suspension, then fixed in 1% formaldehyde and lysed. Cell nuclei were enzymatically digested with *Nla*III then re-ligated with T4 ligase. DNA was isolated by phenol chloroform extraction and ethanol precipitated. The 3C DNA was used in a PCR reaction using primers in the E2 **(*)** and L1 **(1)** gene region. PCR products were separated on a 1.8% agarose gel containing ethidium bromide. A mastermix containing primers and Taq polymerase only (NTC) was included in the PCR reaction to eliminate the possibility of contamination. The PCR shown is representative of 3 experimental repeats for donor 1. PCR bands from **C)** monolayer culture or **D) and E)** differentiated culture were quantified using the Fusion FX digital detection system. WT and ΔCTCF band intensities were calculated relative to the positive control band. These normalized values were used to calculate the amount of product in ΔCTCF cells relative to WT. Error bars represent the SEM, donor 1 monolayer and differentiated (n=4) and donor 2 monolayer (n=2). Significance was determined by a t-test shown as ***= $p<0.001$ and *= $p<0.05$.

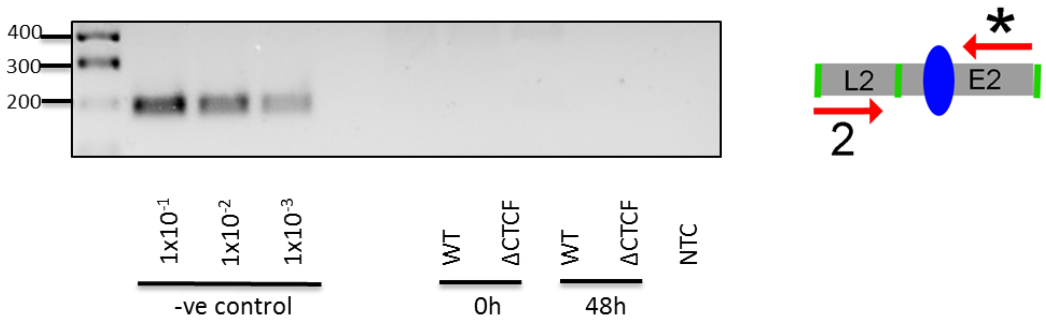


Figure 84. 3C Chromosome conformation capture analysis of WT HPV18 and ΔCTCF HPV18 genomes in HFKs. HFKs containing either WT or ΔCTCF HPV18 genomes were grown in monolayer culture or differentiated in methylcellulose for 48 hours. Cell pellets were harvested and resuspended in **10% (v/v)** FBS/PBS. Cell suspensions were filtered through a 70 μm cell strainer to make a single cell suspension, then fixed in 1% formaldehyde and lysed. Cell nuclei were enzymatically digested with *Nla*III then re-ligated with T4 ligase. DNA was isolated by phenol chloroform extraction and ethanol precipitated. The 3C DNA was used in a PCR reaction using primers in the E2 and L2 region. PCR products were separated on a 1.8% agarose gel containing ethidium bromide. A mastermix (NTC) containing primers and Taq polymerase only was included in the PCR reaction to eliminate the possibility of contamination. The PCR shown is representative of 4 experimental repeats for donor 1 and two experimental repeats for donor 2.

5.12 Summary

The main aim of this chapter was to elucidate the mechanisms underpinning the observed increase in E6 and E7 oncoprotein expression when CTCF binding was abrogated in the HPV18 E2 ORF. Given the prominent role of CTCF in genome organization and transcriptional regulation, it was hypothesized that CTCF binding was required to regulate chromatin remodeling in the HPV18 genome and subsequently modulate early gene expression.

Initial experiments using FAIRE revealed that the area of chromatin within the viral LCR region was more open in ΔCTCF HPV18 genome containing cells in comparison to the WT HPV18 cells. Furthermore, abrogation of CTCF binding resulted in epigenetic switching of histone marks across the HPV18 genome. In ΔCTCF HPV18 genome containing cells there was increased enrichment of the active H3K4me3 mark across the early promoter region as well as an increase in RNA polymerase II binding within the early gene region when compared to the WT HPV18 genome containing cells. Furthermore, whilst the WT HPV18 genome containing cells exhibited H3K27me3 binding within the viral LCR there was an apparent depletion of this repressive epigenetic mark within the LCR in ΔCTCF HPV18 genome containing cells.

The investigation of factors binding to the viral LCR revealed a loss of YY1 binding in the LCR of ΔCTCF HPV18 monolayer cells when compared to the WT HPV18

cells. This loss of YY1 binding was maintained upon cellular differentiation of the ΔCTCF HPV18 genome containing cells. However the loss of YY1 binding was not due to reduced YY1 protein expression as total YY1 protein levels remained similar between WT HPV18 and ΔCTCF HPV18 cells in monolayer culture, and a similar decrease in YY1 protein expression was observed upon differentiation. However, ChIPs were also performed to determine the association of the TEF-1 transcription factor across the LCR of WT HPV18 and ΔCTCF HPV18 cells, which is known to activate early viral gene expression. Interestingly, ΔCTCF HPV18 genome containing cells displayed increased enrichment of TEF-1 across the viral enhancer in comparison to the WT HPV18 genome containing cells. Furthermore, overall TEF-1 protein levels were higher in the ΔCTCF HPV18 cells in monolayer culture when compared to the WT HPV18 monolayer cells. Given that YY1 and TEF-1 have overlapping binding motifs in viral LCR it is tempting to speculate that increased TEF-1 protein expression in ΔCTCF HPV18 cells resulted in competitive binding with the YY1 protein, and an overall loss of YY1 binding in the ΔCTCF HPV18 LCR region.

The loss of the H3K27me3 mark within the LCR of ΔCTCF HPV18 cells highlighted the possibility that additional co-factors were involved in regulating this histone modification across the HPV18 genome. Indeed the PRC2 complex is a well-known factor involved in regulating H3K27me3 deposition and mediating repressive chromatin formation. Furthermore, the YY1 protein is thought to facilitate the

recruitment of the polycomb complexes to chromatin. ChIP experiments were performed to determine the association of PRC2 complex proteins to the viral LCR in WT HPV18 and ΔCTCF HPV18 genome containing cells. There was a slightly higher enrichment of EZH2 within the viral LCR in ΔCTCF HPV18 cells compared to WT HPV18 cells in monolayer culture and similar enrichment after differentiation. However, the EZH2 ChIPs could not distinguish between phosphorylated subsets of EZH2 that can have an activating role or whether the protein was binding independently of the repressive PRC2 complex. Therefore ChIPs were also performed to determine the association of EED, which is another protein in the PRC2 complex. There was a reduction in EED binding across the viral LCR in ΔCTCF HPV18 cells compared to the WT HPV18 cells, indicating a loss of the PRC2 complex in the ΔCTCF HPV18 HFKs. The PRC2 complex is closely linked to the activity of the PRC1 complex, which is known to be involved in the ubiquitinylation of H2A, which induces compact chromatin formation. Indeed there was abrogation of RING1B binding in the LCR of ΔCTCF HPV18 cells, concomitant with a loss of H2AK119Ub, which is a known hallmark of compact and silenced chromatin.

Whilst it was clear that ΔCTCF HPV18 genome containing cells displayed abrogation of both YY1 and polycomb complex binding in the viral LCR and a loss of the repressive H3K27me3 histone mark, it remained unclear how this LCR region of the viral genome was functionally associated with CTCF binding site in the E2 ORF located 3kb downstream. Nevertheless, it was hypothesized that YY1 and CTCF-

known protein binding partners- formed a repressive DNA loop, which was required to recruit the PRC1 and PRC2 complexes to repress E6 and E7 oncoprotein expression in monolayer culture. Using 3C conformation capture it was possible to demonstrate the formation of a DNA loop between CTCF in the E2 ORF and YY1 in the viral LCR. In ΔCTCF HPV18 genome containing cells this DNA loop formation was abrogated, as there was a clear reduction in the PCR ligation product. Upon methylcellulose differentiation the PCR product was reduced in WT HPV18 cells when compared to WT HPV18 cells in monolayer culture, and nearly absent in ΔCTCF HPV18 differentiated cells. These results indicate that a repressive loop formation is required in monolayer culture in order to control the balance early gene expression; however, upon methylcellulose differentiation the formation of the repressive loop is reduced to allow E6 and E7 expression and viral genome replication.

Overall, these results suggest that the high-risk HPV18 genome recruits CTCF to the E2 ORF, which is required to form a repressive DNA loop with YY1 bound to the viral LCR. In theory, the formation of this DNA loop may facilitate the YY1-dependent recruitment of the PRC1 and PRC2 complexes that modulate repression of the early viral genes. However, when CTCF binding is abrogated in the E2-ORF, this disrupts the formation of the DNA loop with YY1 in the viral LCR. Subsequently this may favour the displacement of YY1 in the LCR by TEF-1 binding,

resulting in a loss of PRC1 and PRC2 recruitment, and subsequent loss of early viral gene repression.

Chapter 6:
Discussion

6 Chapter 6: Discussion

CTCF is a well-studied architectural protein involved in organising chromatin and mediating long-range chromosomal interactions (Splinter et al., 2006, Li et al., 2008). Several researchers have highlighted a role for CTCF in mediating long-range interactions in viral DNA genomes, in order to co-ordinate and regulate viral transcription (Kang et al., 2011, Tempera et al., 2011). Of note, CTCF binding sites have been identified in the large DNA genome viruses, including EBV, HSV and KSHV. Due to the large sizes of these viral genomes it is unsurprising that they would recruit a host protein in order to facilitate long-range DNA interactions to support viral processes. Nevertheless, it was also postulated that smaller DNA viruses also need to recruit a chromatin organiser to regulate viral transcriptional processes; therefore, it was hypothesized that CTCF is recruited to the HPV genome to facilitate and regulate aspects of the HPV life cycle.

6.1 Identification of potential CTCF binding sites

Initial bioinformatics analysis was carried out to identify potential CTCF binding sites in HPV genomes. Genome sequences of the 13 classified high-risk and two most common low-risk HPV types were exported from the PAVE database and transferred to the CTCF binding site database (CTCFBSDB) (Bao et al., 2008, Ziebarth et al., 2013). Previous studies have identified and mapped core sequences that bind to the CTCF protein and these core motifs have been represented by position weight matrices (PWM). The CTCFBSDB determines if there are any core CTCF motifs in the

imported HPV sequence and displays a PWM score, which represents the log-odds ratio of the sequence generated by the CTCF motif. However, this tool only displays the highest scoring motifs, so any other potential CTCF binding sites may be overlooked. Nevertheless, a highly conserved CTCF binding site was observed around nucleotide 3000 in the E2 ORF of 10 high-risk HPV types, including the most prevalent: HPV16, HPV18 and HPV31. Interestingly, this predicted CTCF binding site was not observed in the E2 ORF of the beta HPV types or HPV6 and HPV11, which are two of the most prevalent low-risk types. Taken together, these findings lead to the hypothesis that CTCF recruitment to the E2 ORF of the viral genome is specific to the high-risk oncogenic HPV types.

Whilst the predicted E2 ORF fragments in HPV16, HPV18 and HPV31 were shown to bind to CTCF protein *in vitro* by EMSA (Paris et al., 2015), it was important to confirm and characterise this interaction *in vivo*. Keratinocytes are a natural host for HPV infection, therefore the primary HFK cell culture system serves as a physiologically relevant model for studying the HPV life cycle and HPV-host interactions. Moreover, transfection of high-risk HPV viral episomes in to primary HFKs extends the lifespan of these primary cells than would normally occur in untransfected lines, thus allowing extended cell culture and generation of organotypic raft cultures. The cell lines contain between 50-100 copies of viral episomes per cell, which are maintained over several passages, and integration does not occur until much later passages. Furthermore, it is possible to generate desired

mutations in the viral episomes prior to transfection, which eventually yields a population of HFKs containing the mutated viral episomes. In this case, a three base pair mutation was generated in the CTCF binding motif in the E2 ORF of HPV18 prior to transfection in to HFKs, leading to the generation of the ΔCTCF HPV18 genome containing cell line. ChIP experiments successfully demonstrated that CTCF binding to the E2 ORF was abrogated in the ΔCTCF HPV18 cell line. By using the ΔCTCF HPV18 cell line in parallel to the WT HPV18 cells it was possible to determine the biological significance of CTCF binding during the complete HPV life cycle.

6.2 CTCF protein expression in HFK cell lines

Interestingly, western blotting demonstrated a two-fold increase in CTCF protein expression in HPV18 genome containing cells in comparison to untransfected HFKs. Similarly, Mehta *et al* also observed an increase in CTCF protein expression in HPV31 genome containing CIN612 cells compared to untransfected HFKs (Mehta et al., 2015). These results were the first indication that HPV genome containing cells need to up regulate the expression of host CTCF protein in order to facilitate processes throughout the viral life cycle. However, given that the complete life cycle is dependent on epithelial differentiation it was also important to consider changes in CTCF protein expression in organotypic raft cultures or upon methylcellulose differentiation.

Previous studies have demonstrated that CTCF protein expression decreases upon differentiation of myeloid cell lines (Delgado et al., 1999) and corneal epithelial cells (Tsui et al., 2016, Bravo and Felez-Sanchez, 2015) so it was hypothesised that CTCF protein expression would decrease upon keratinocyte differentiation. Surprisingly, upon methylcellulose differentiation of WT HPV18 and the ΔCTCF HPV18 genome containing HFKs the level of CTCF protein expression was similar to the monolayer cultures, but still remained higher than untransfected HFKs. However, incubation of HFKs in methylcellulose does not induce the terminal differentiation observed from generating organotypic raft cultures. Therefore, any potential differences in CTCF protein expression in the methylcellulose differentiated HFKs may be undetectable because the cells have not undergone terminal differentiation. In order to generate organotypic raft cultures, primary HFKs were seeded on to collagen plugs and transferred to metal grids suspended at the air-liquid interface. Over a period of 2 weeks the epithelial cells are induced to undergo differentiation and stratification. After generating organotypic raft cultures they can be homogenised and used for DNA, RNA and protein extraction. One of the main advantages over methylcellulose differentiation is the ability to fix and stain raft cross-sections for expression of either host or viral proteins throughout the differentiated epithelium. Organotypic raft cultures derived from either WT HPV18 or ΔCTCF HPV18 genome containing HFKs were stained with CTCF specific antibody to visualise the localisation of CTCF protein throughout the differentiated epithelium. Immunofluorescence staining revealed a reduction in nuclear CTCF staining in some of the cells in the upper-most

layers of the WT HPV18 and ΔCTCF HPV18 rafts, which have undergone terminal differentiation. These findings are in corroboration with previous results that demonstrated a reduction in CTCF protein expression upon differentiation of myeloid cells and calcium differentiation of corneal epithelial cells (Delgado et al., 1999, Tsui et al., 2016).

6.3 Mapping of CTCF binding sites *in vivo*

Although the CTCF prediction tool gave insight in to potential CTCF binding site positions, it was important to verify occupancy of these sites *in vivo*, as different cellular factors or the cellular state can influence CTCF binding (Holwerda and de Laat, 2013). ChIP experiments were performed to determine the binding of CTCF throughout the viral genome in WT HPV18 and ΔCTCF HPV18 genome containing HFKs. Initially, WT HPV18 and ΔCTCF HPV18 genome containing HFKs were fixed in 1% formaldehyde to cross-link the DNA and associated proteins. The fixation time is a very important aspect of the ChIP protocol and can heavily influence the efficiency of the ChIP reaction. If the cells are incubated in formaldehyde for too long then it is more difficult to shear the chromatin evenly; however, if the incubation is too short then the efficiency of protein cross-linking is reduced. In these experiments the chromatin was sheared by probe sonication as opposed to enzymatic digestion. Probe sonication often yields less sample-to-sample variation compared to enzymatic digestion, and in this case sonication consistently sheared chromatin fragments between 200-600bp in each of the ChIP experiments. However, the probe

sonication procedure does generates heat that can degrade the chromatin samples or reverse the protein cross-links; therefore, samples were sonicated on ice to minimise chromatin degradation and maintain sample integrity.

Three independent ChIP experiments confirmed the binding of CTCF to the E2 ORF of WT HPV18 genome containing HFKs grown in monolayer culture, and this was further confirmed in a second cell donor. Interestingly, upon methylcellulose differentiation, the binding of CTCF to the E2 ORF was further enriched. Moreover, in monolayer HFKs containing the ΔCTCF HPV18 genomes, ChIP experiments confirmed that there was abrogation of CTCF binding to the E2 ORF, as expected. Furthermore, HFKs containing WT HPV16 genomes also displayed CTCF enrichment at the predicted binding site in the E2 ORF, and again this enrichment was further increased upon 48 hours methylcellulose differentiation. Whilst CTCF binding was confirmed in the HFK transfection model it was also shown that CTCF binds to the HPV16 E2 ORF in the HPV16 genome containing W12 cells, which are derived from a low-grade cervical lesion. Interestingly, the W12 cells also displayed CTCF binding in the L1 and L2 gene regions as well as the viral LCR upon methylcellulose differentiation. This binding profile was not as apparent in the WT HPV18 genome containing HFKs, but a similar enrichment in the L1, L2 and LCR regions was observed in the HPV16 genome containing HFKs in monolayer culture. Consensus CTCF binding sites have also been identified in the E2, L1 and L2 regions of the HPV31 genome in naturally derived CIN612 cells and ChIP experiments

confirmed that the HPV31 genome contains functionally important CTCF binding sites in the L2 region (Mehta et al., 2015). Essien *et a*l highlighted that high affinity CTCF binding sites are occupied regardless of tissue type or the differentiation state of the cells (Essien et al., 2009), which corroborates with the conserved binding of CTCF in the E2 ORF of HPV16, HPV18 and HPV31. However, whilst HPV16, 18, and 31 all contain CTCF binding sites in the E2 ORF, there seems to be a discrepancy with CTCF binding sites in the late gene regions. However, HPV16 and HPV31 are sister taxa and more closely related with each other than HPV18 (Bravo and Felez-Sanchez, 2015), so it is plausible that HPV16 and HPV31 have co-evolved to contain additional CTCF binding sites in the late gene region, which are not present in the HPV18 genome. Binding of CTCF to additional binding sites may also be influenced by differentiation and other cellular factors, but the HPV16 and HPV18 genome containing cells were grown in the same cell culture conditions and both differentiated using the same methylcellulose method. CTCF binding to the late regions identified in the naturally occurring W12 and CIN612 cell lines may however be due to differences in cellular factors compared to the HFK transfected cell lines. It is also possible that additional CTCF binding sites are present within the HPV genome, but were not detected using the consensus CTCF prediction tool software or the selected qPCR primers. However, the use of ChIP-Seq analysis provided deeper coverage of CTCF binding throughout the HPV18 genome and confirmed that the only significant CTCF binding peak was located in the E2 ORF. Further

ChIP-Seq analyses with other HPV genomes would need to be carried out to identify other potential CTCF binding sites.

The results from the ChIP experiments are represented as fold binding over FLAG IgG. These values represent how many times more enriched a specific protein is at a genomic location compared to the background noise of non-specific FLAG antibody binding. With each experiment there are both advantages and caveats to the methodology. The main advantage of using a negative control antibody in the ChIP reaction is the elimination of background noise caused by non-specific binding. However the fold binding values can vary considerably between independent ChIP experiments due to the differences in sample preparation and experimental variability. Another method for calculating ChIP binding involves calculating the percentage input, which calculates the percentage of DNA bound by the protein of interest. Whilst this method is supposed to normalise starting material across independent samples, the input chromatin is extracted and processed quite early in the process, so total levels across different samples may actually decrease throughout the course of the experiment.

6.4 The role of CTCF in the HPV life cycle

The co-operative actions of both E6 and E7 are known to promote the malignant conversion of infected cells. The E7 oncoprotein interacts with the Rb protein and related pocket proteins p130 and p107, and perturbs the normal differentiation of the epithelium, which ultimately promotes an environment permissive for cell division

and viral replication (Munger et al., 1989, Dyson et al., 1989). The tumour suppressor p53 is expressed in response to cellular stress and will induce apoptosis, cell-cycle arrest or senescence. The main role of high-risk E6 protein is to circumvent the p53-mediated cellular stress responses by promoting the degradation of p53 via the ubiquitin ligase E6AP, which leads to increased cellular proliferation (Scheffner et al., 1990). Current E6 and E7 antibodies are not suitable for detecting protein expression in organotypic raft cultures by immunofluorescence. Nevertheless, given that E6 targets p53 for degradation and E7 preferentially targets p130 for degradation, both p53 and p130 serve as appropriate surrogate markers. Organotypic raft cultures derived from untransfected HFKs, HPV18 WT or ΔCTCF HPV18 genome containing HFKs were stained with antibodies to detect p53 and p130 throughout the raft epithelium. The rafts derived from the ΔCTCF HPV18 genome containing HFKs displayed a significant decrease in p53 expression throughout the layers of the epithelium when compared to the untransfected and WT HPV18 rafts. Furthermore, ΔCTCF HPV18 rafts also displayed a significant reduction in p130 expression in the upper differentiated layers of the rafts in comparison to the untransfected and WT HPV18 rafts. These results indicated that the ΔCTCF HPV18 genome containing cells displayed an increase in E6 and E7 expression in comparison to the WT HPV18 genome containing cells.

Haematoxylin and eosin staining was also performed on organotypic raft cultures derived from untransfected HFKs, HPV18 WT or ΔCTCF HPV18 genome containing

HFKs in order to determine any morphological differences between the cell lines. The untransfected rafts displayed a relatively thin morphology, which is unsurprising considering the absence of HPV episomes. The WT HPV18 rafts displayed increased thickness of the suprabasal layer of the epithelium, which is due to the increased proliferation induced by the viral oncoproteins E6 and E7. Furthermore, the rafts derived from the ΔCTCF HPV18 genome containing HFKs displayed an even more hyperproliferative phenotype as indicated by the increased thickness of the suprabasal compartment, suggesting that these cells are indeed expressing more E6 and E7 than the WT HPV18 genome containing cells.

Indeed, initial western blotting experiments revealed an increase in both E6 and E7 oncoprotein expression in the ΔCTCF HPV18 genome containing HFKs grown in monolayer in comparison to the WT HPV18 genome containing cells. Furthermore, protein lysates were extracted from organotypic raft cultures and western blotting demonstrated an increase in E6 and E7 protein expression in rafts derived from ΔCTCF HPV18 genome containing HFKs when compared to the rafts derived from WT HPV18 genome containing cells. These results indicate that high-risk HPV types need to recruit CTCF to bind to the E2 ORF in order to control and attenuate the expression of the viral oncoproteins.

As mentioned, the predicted CTCF binding site lies within the E2 ORF of the HPV18 genome. To generate the ΔCTCF HPV18 genome containing HFKs a three base pair

mutation was made in this predicted CTCF binding site in the E2 ORF of HPV18. Disruption of E2 expression by viral integration is known to result in derepression of the oncoproteins (Berumen et al., 1994), and several studies have highlighted a key role for E2 in repressing early gene expression. Although the three nucleotide mutations did not alter the amino acid sequence that encodes the E2 protein it was still important to ensure similar E2 expression between ΔCTCF HPV18 and the WT HPV18 genome containing HFKs. Protein lysates were extracted from organotypic raft cultures derived from either WT HPV18 or ΔCTCF HPV18 genome containing HFKs, and western blotting confirmed similar E2 expression between three independent rafts for each cell line. This provided strong evidence that the observed increase in E6 and E7 expression was not due to an overall decrease in E2 protein expression and loss of E2 mediated repression of E6 and E7. Furthermore, raft sections were also stained with E2 specific antibody and revealed similar intensity of E2 expression between WT HPV18 and ΔCTCF HPV18 rafts, but also revealed delayed expression of E2 protein in the upper layers of the ΔCTCF HPV18 raft sections. This is interesting because we know that E2 expression is increased in the upper layer of the epithelium, and E2 protein binds to the early viral promoter and displaces transcription factors TBP and SP1, leading to the repression of E6 and E7 expression (Thierry and Howley, 1991, Tan et al., 1992, Demeret et al., 1994). Therefore, the delayed expression of E2 protein in the upper layer of the epithelium corresponds to the increase in E6 and E7 observed in the ΔCTCF HPV18 genome containing HFKs.

Whilst it was clear that abrogation of CTCF binding caused aberrant early protein expression it was also important to assess the effect on late protein expression. Organotypic raft staining revealed a significant decrease in L1 protein expression in the upper layers of rafts derived from ΔCTCF HPV18 genome containing HFKs, when compared to rafts derived from WT HPV18 genome containing HFKs. Western blotting is not possible with current L1 antibodies because the keratin in the upper raft layers interferes with L1 detection; however, the late E1^E4 protein is readily detectable by western blotting. Both WT HPV18 and the ΔCTCF HPV18 genome containing HFKs were differentiated in methylcellulose for 48 hours and protein lysates were extracted for western blotting. Results consistently showed a decrease in E1^E4 protein expression in the ΔCTCF HPV18 HFKs compared to the WT HPV18 HFKs. Conversely, organotypic rafts stained for E1^E4 did not show an obvious difference in E1^E4 protein expression between rafts derived from either the WT HPV18 or the ΔCTCF HPV18 HFKs, which is surprising given the western blot results. However due to the high expression of E1^E4 in the orgnaotypic raft sections, any subtle differences in E1^E4 may be difficult to detect. Overall it seems that abrogation of CTCF binding to the E2 ORF is also causes a reduction in late viral protein expression.

Abrogation of CTCF binding to the E2 ORF in the ΔCTCF HPV18 genome containing cell line has been shown to alter HPV gene splicing events. The abrogation of CTCF

binding resulted in an increase in the unspliced transcripts with the potential to encode E6 and E7, which corroborates with the observed increase in E6 and E7 protein expression. The investigation of additional viral transcripts revealed a loss of the mRNA species with the potential to encode E6*II and E5. The loss of the E6*II transcript is particularly interesting because the high-risk HPV E6* splice variants are believed to have anti-oncogenic roles (Filippova et al., 2014) and can inhibit the functions of the full length E6 protein (Pim et al., 1997). Overall it seems likely that the virus recruits CTCF to the viral genome in order to control the equilibrium of expression of both the full length E6 and E6* splice variants. Abrogation of CTCF binding to the E2 ORF can also alter the expression of late viral transcripts. It is apparent that ΔCTCF HPV18 genome containing cell lines display expression of late viral transcripts after methylcellulose differentiation but also in monolayer culture. The fact that the late transcripts are expressed in monolayer culture indicates that there is read-through of the early polyadenylation site in the ΔCTCF HPV18 genome containing HFKs. It is likely that when CTCF binds to the E2 ORF its main role is to facilitate RNA polymerase II pausing and regulate splicing events of all of the HPV transcripts. Therefore, upon abrogation of CTCF binding there may be a subsequent increase in RNA polymerase II read-through along the genome and aberrant expression of the late transcripts. Indeed there is evidence that CTCF can physically interact with the largest subunit of RNA polymerase II (Chernukhin et al., 2007), and Shulka *et al* previously demonstrated how CTCF binding can cause RNA polymerase stalling and allow splicing and inclusion of weak upstream exons (Shukla et al.,

2011). With this in mind it is likely that CTCF binding in the HPV genome plays a major role in regulating gene splicing, through the modulation of RNA polymerase elongation.

WT HPV18 and ΔCTCF HPV18 genome containing HFKs were differentiated in methylcellulose for 24hr and 48hr time-points and Southern blotting was performed to determine any differences in viral genome amplification between the cell lines. Given that the ΔCTCF HPV18 genome containing cells displayed a decrease in late E1^E4 protein expression compared to the WT HPV18 genome containing cells, it was predicted that the ΔCTCF HPV18 cells would have impaired genome amplification. There does appear to be a decrease in viral amplification in the ΔCTCF HPV18 cells between 0 hour and 24 hour time points, however this is due to a smaller quantity of DNA loaded on to the agarose gel in for the 24hr time-point. Taken together, there was similar amplification of WT HPV18 and ΔCTCF HPV18 genomes after 24 hours and 48 hours in two independent cell donors. Conversely, Mehta *et al* demonstrated that either mutation of the CTCF binding sites in the L2 region of HPV31 or shRNA knockdown of CTCF protein resulted in the inhibition of HPV31 genome amplification upon methylcellulose differentiation (Mehta et al., 2015). However, these findings may reflect that the CTCF binding sites in the late region are involved in different viral processes compared to the CTCF site in the E2 ORF. The E2 ORF binding site may be imperative for attenuating early gene expression and regulating gene splicing events, but is not required for modulating

viral genome amplification. Given that the HPV16 genome containing cells also displayed CTCF binding in the L1 and L2 ORFs it would be interesting to determine if the mutation of these CTCF binding sites in the HPV16 late region would affect viral genome amplification in the same way as the HPV31 genome containing cells.

6.5 The role of CTCF in regulating HPV18 chromatin structure

Stunkel and Bernard have previously mapped the positioning of nucleosomes associated with the HPV16 and HPV18 viral genomes, and identified dynamically associated nucleosomes across the LCR region (Stunkel and Bernard, 1999). During this current study the FAIRE technique was used to identify and compare regions of open chromatin throughout the WT HPV18 and ΔCTCF HPV18 genomes. This is a relatively simple and efficient method for isolating open regulatory regions of the chromatin, which are depleted of nucleosomes and other proteins. Both WT HPV18 and ΔCTCF HPV18 genome containing HFKs were grown in monolayer culture or differentiated in methylcellulose then fixed in formaldehyde. The fixation time is longer than for ChIP experiments, as it allows more stringent isolation of open regions and nucleosome depleted areas of the chromatin. The samples are sonicated using the same conditions as the ChIP samples, but are separated in to 'INPUT' and 'FAIRE' aliquots. The 'INPUT' samples are reverse-cross-linked prior to DNA isolation with phenol-chloroform. 'FAIRE' samples are directly subjected to phenol chloroform extraction after fixation, which allows isolation of areas of the chromatin that are depleted of protein, as this will be trapped in the aqueous/solvent interface.

The DNA isolated from 'INPUT' and 'FAIRE' samples was amplified by qPCR using primers spanning the HPV18 genome. The results revealed an area of open chromatin across the LCR of the WT HPV18 genome containing cells, which is consistent with previous observations that nucleosomes are depleted at actively transcribed regions genomes (Lee et al., 2004). Most strikingly, the FAIRE data also revealed an even more open chromatin region within the viral LCR of the ΔCTCF HPV18 genome containing HFKs in comparison to the WT HPV18 cells, indicating that the ΔCTCF HPV18 genomes are more transcriptionally active than the WT HPV18 genomes. Moreover, Schwartz *et al* proposed that nucleosomes themselves could serve as speed-bumps for RNA polymerase II progression, thus slowing down the rates of transcription (Schwartz et al., 2009, Schwartz and Ast, 2010). Kang *et al* also observed alterations in nucleosome phasing in the KSHV genome when the CTCF binding site was mutated, however this increased chromatin accessibility was observed directly at the region containing the CTCF binding site (Kang et al., 2013).

Taking this further, the use of FAIRE-Seq would provide even wider coverage of the viral genome and detect more subtle changes in nucleosome occupancy throughout the genome that may have gone undetected using FAIRE-qPCR (Tsompana and Buck, 2014). FAIRE-Seq still has several limitations as it relies heavily on fixation efficiency and still requires a large number of cells ($1x10^5$-10^7). Recent studies have highlighted that the assay for transposase accessible chromatin (ATAC-Seq) as an important technique to simultaneously map open chromatin, transcription factor

binding and nucleosome positioning. This technique requires much fewer cells (500-500,000) and involves a simple two-step protocol (Buenrostro et al., 2015, Buenrostro et al., 2013). The methodology involves using unfixed nuclei that are tagged *in vitro* with adapters using Tn5 transposase (Goryshin and Reznikoff, 1998, Adey et al., 2010). The adapters are inserted in to open regions of the chromatin and mapped by high-throughput genome wide sequencing (Buenrostro et al., 2013, Buenrostro et al., 2015). In the future, ATAC-Seq would provide more efficient analysis of nucleosome positioning and transcription factor binding in HPV studies.

Given the observed changes in nucleosome occupancy in the HPV18 genome it was important to determine any alterations in histone modifications associated with the viral genome, which may confer with the aberrant expression of the viral oncoproteins. The association of specific histone modifications within the HPV31 genome in naturally derived CIN612 cells has previously been studied (Wooldridge and Laimins, 2008). Here, Wooldridge *et al* demonstrated that both the early and late HPV31 promoters were in the active state in monolayer culture, as indicated by the presence of the histone modifications H3K4me3, H3Ac and H4Ac. Upon methylcellulose differentiation, the association of these histone modifications was further enriched at both the early and late promoters (Wooldridge and Laimins, 2008). Given that the ΔCTCF HPV18 genome containing HFKs displayed an increase in E6 and E7 expression it was hypothesised that histone modifications associated with active transcription were more enriched in ΔCTCF HPV18 genomes compared

to the WT HPV18 genome. H3K4me3 was a specific target as it is a well-known hallmark of active chromatin and is enriched at promoter regions and the TSS of active genes (Santos-Rosa 2002). Indeed, ChIP experiments demonstrated an increase in the enrichment of the histone modification H3K4me3 within the early promoter region of the ΔCTCF HPV18 genome containing cells. Similar to the Wooldridge study, the WT and ΔCTCF HPV18 late promoters still displayed active histone enrichment whilst cells were grown in monolayer culture. Furthermore, the H3K4 methyltransferase protein can directly interact with RNA polymerase II, indicating that deposition of H3K4me3 occurs co-transcriptionally (Zhang et al., 2015). Indeed, ChIP experiments also revealed an increase in total RNA polymerase II enrichment at the early promoter in ΔCTCF HPV18 genome containing HFKs when compared to the WT HPV18 cells. Taken together, the observed increase in H3K4me3 and RNA polymerase II enrichment at the viral LCR and early promoter region corroborates with the increase in early gene expression in the ΔCTCF HPV18 genome containing cells.

Whilst it was clear that ΔCTCF HPV18 genome containing HFKs displayed increased enrichment of the active histone mark H3K4me3 at the early promoter, it was interesting to investigate the association of the repressive H3K27me3 modification throughout the WT HPV18 and ΔCTCF HPV18 genomes. Interestingly, WT HPV18 genome containing cells displayed peak enrichment of H3K27me3 within the LCR, and this binding peak was lost in the LCR of ΔCTCF HPV18

genome containing HFKs. Overall, western blotting showed that total H3K27me3 expression was similar between WT HPV18 and ΔCTCF HPV18 genome containing HFKs in monolayer and differentiated state, but levels were consistently lower than untransfected HFKs grown in monolayer. McLaughlin-Drubin *et al* have also demonstrated that E7 expressing HFKs displayed a decrease in H3K27me3 expression compared to untransfected HFKs, which is due to the increase expression of the H3K27-specific demethylases KDM6A and KDM6B (McLaughlin-Drubin et al., 2011). Hyland *et al* also demonstrated via immunofluorescence and western blotting that E6/E7 expressing HFKs displayed reduced H3K27me3 expression compared to untransfected HFKs (Hyland et al., 2011).

Regulation of the early viral promoter involves the co-ordinated and differential binding of a host of transcription factors to the LCR region, which can either activate or repress transcription. The complexity of early promoter regulation may reflect the need to express the viral oncoproteins at precise times during the viral life cycle, particularly in the cells that would normally undergo differentiation in the suprabasal compartment of the epithelium. Therefore any perturbations in regulating the early promoter may lead to aberrant expression of the early genes. A previous study has demonstrated a role for YY1 in mediating repression of the early promoter in HPV16 by quenching AP-1 activity in the viral LCR (O'Connor et al., 1996). Several studies have shown that cervical cancers containing non-integrated HPV16 episomes display mutations in YY1 binding sites in the viral LCR (Dong et

al., 1994, May et al., 1994), providing evidence that YY1 is involved in repression of early gene expression. ChIPs from this study have revealed a loss of YY1 binding in the LCR of ΔCTCF HPV18 genome containing HFKs both in monolayer and differentiated culture, indicating that these cells have lost the YY1-mediated repression of early viral gene expression. Interestingly, the HPV LCR region in each HPV type contains a different number and distribution of potential YY1 binding sites. For example the high-risk HPV16 LCR contains 10 potential YY1 sites, HPV31 has 8 and HPV18 has 7. On the other hand, the low-risk HPV6 only has 2 potential YY1 binding sites and HPV11 contains 4 YY1 sites (O'Connor et al., 1995). It is possible that high-risk types such as HPV16 and 18 may have evolved to include more YY1 binding sites to mediate repression of early gene expression. On the other hand the low-risk HPV types contain fewer YY1 binding sites in the viral LCR, which may reflect a lesser need to repress early gene expression compared to their high-risk counterparts.

The drosophila PHO protein binds to polycomb response elements (PREs) and anchors polycomb complexes to mediate transcriptional repression. PHO was also shown to share sequence homology with the mammalian YY1 protein (Brown et al., 1998). A previous genome-wide analysis of the human genome has demonstrated a poor overlap of YY1 DNA binding sites at regions targeted by the SUZ12 polycomb protein, suggesting that in humans YY1 is not necessarily a key regulator of polycomb recruitment (Xi et al., 2007). However, due to the large abundance of YY1

binding sites in the HPV18 LCR, which is the key regulatory region for transcriptional control, it is likely that YY1 binding in this region is involved in polycomb recruitment.

The HPV18 LCR is known to bind a plethora of transcription factors and HPV proteins in order to regulate viral transcription. Whilst it has been demonstrated that HFKs containing the ΔCTCF HPV18 genomes show a loss of YY1 binding to the viral LCR, it is possible that there is also differential binding of other factors in this LCR region, which haven't been investigated in this study. Indeed transcription factors including AP-1, NF1 and Oct-1 have been implicated in regulating viral transcription (Baldwin et al., 2007, Chan et al., 1990), O'Connor et al., 1995) so the binding profiles of these proteins may also be altered in the ΔCTCF HPV18 genome containing HFKs.

Western blotting revealed that the ΔCTCF HPV18 genome containing cells grown in monolayer culture displayed an increase in TEF-1 protein expression in comparison to the WT HPV18 genome containing cells. Interestingly, a previous study has demonstrated that TEF-1 is upregulated by HPV E6 (Mori et al., 2017), which is in agreement with our observed increase in E6 expression in the ΔCTCF HPV18 cell lines. Furthermore, TEF-1 has been shown to activate HPV16 E6 and E7 gene expression (Ishiji et al., 1992a). Interestingly, ChIP experiments demonstrated an increase in TEF-1 binding to the viral LCR in the ΔCTCF HPV18 genome containing cells when compared the WT HPV18 genome containing HFKs. This increase in

TEF-1 binding to the viral enhancer in the ΔCTCF HPV18 genome containing cells may be a result of the increase in overall TEF-1 protein expression that was detected. Interestingly, several of the TEF-1 binding sites overlap with YY1 binding sites; therefore, the increase in TEF-1 protein mediated by the augmented E6 expression may cause the displacement or competition with YY1 binding in the viral LCR.

Given that the WT HPV18 and ΔCTCF HPV18 genome containing cells displayed reduced H3K27me3 expression in comparison to untransfected HFKs, one would predict a reduction in PRC2 complex protein expression, specifically EZH2, which is known to catalyse methylation at lysine 27 (Cao et al., 2002, Czermin et al., 2002, Kuzmichev et al., 2002, Muller et al., 2002). However in this study, the level of EZH2 protein was actually increased in both WT HPV18 and ΔCTCF HPV18 genome containing HFKs grown in monolayer culture in comparison to untransfected HFKs. Similarly in another study, HFKs expressing E7 displayed increased EZH2 mRNA expression compared to untransfected HFKs (McLaughlin-Drubin et al., 2011). Furthermore, Hyland *et al* also demonstrated an increase in both EZH2 mRNA and protein levels in E6/E7 expressing HFKs compared to the untransfected HFKs (Hyland et al., 2011). They also showed that organotypic rafts derived from E6/E7 expressing HFKs displayed higher levels of nuclear P-EZH2-Ser21 staining compared rafts derived from untransfected HFKs. Furthermore reduced H3K27me3 staining in high-grade squamous intraepithelial lesions (HSILs) correlated with increased levels of EZH2 expression (Hyland et al., 2011).

The ChIP experiments revealed a slight increase in EZH2 binding within the viral LCR and early promoter region in ΔCTCF HPV18 genome containing cells compared to the WT HPV8 cells. The antibody used for the ChIP experiment detects total EZH2 protein and is unable to discriminate between the phosphorylated forms of EZH2. Emerging research has demonstrated that phosphorylation of EZH2 can modulate different functions of the protein. One study demonstrated that the Akt signalling pathway was involved in the phosphorylation of EZH2 at serine-21, which subsequently caused a reduction in PRC2-mediated histone methylation (Cha et al., 2005). Given that Hyland *et al* detected an increase in EZH2 phosphorylated at serine 21 in E6/E7 expressing raft cultures (Hyland et al., 2011), it would be interesting to determine if the observed increase in EZH2 binding across the ΔCTCF HPV18 LCR is attributable to the P-EZH2-Ser21 form.

Given that EZH2 function is modulated by phosphorylation events and can act as a transcriptional activator, it was important to consider other PRC2 proteins that are not thought to function independently of the PRC complex. ChIP experiments were carried out to determine the binding of another PRC2 complex protein, EED, within the viral LCR of WT HPV18 and ΔCTCF HPV18 genome containing cells. These experiments revealed a loss of EED binding in the LCR of ΔCTCF HPV18 genome containing cells when compared to the WT HPV18 LCR, which displayed enrichment of EED binding. There was no difference in EED protein expression between WT HPV18 and ΔCTCF HPV18 genome containing cells grown in

monolayer culture. Furthermore, Hyland et al found no detectable difference in EED protein expression between E6/E7 expressing HFKS and untransfected HFKs (Hyland et al., 2011).

The H3K27me3 repressive mark is recognised by chromobox proteins within the PRC1 complex (Cao et al., 2002). The PRC1 complex contains the E3 ligase protein RING1B, which catalyses monoubiquitinylation of H2A on lysine 119 (H2AK119ub), resulting in the formation of compact and silenced chromatin. ChIP experiments showed that WT HPV18 genome containing HFKs displayed peak RING1B enrichment in the viral LCR and enrichment of H2AK119ub across the early gene region. However, ΔCTCF HPV18 genome containing cells displayed a loss of RING1B binding in the viral LCR, concomitant with a decrease in H2AK119ub enrichment within the early gene region.

Overall, these experiments have demonstrated that upon abrogation of CTCF binding to the E2 ORF there are alterations in polycomb protein recruitment to the viral genome. Interestingly, polycomb proteins have been implicated in chromatin loop formation at various cellular sites that are bound by CTCF protein (Li et al., 2008), therefore it was important to investigate the presence of chromatin loops in the HPV18 genome.

6.6 Identification of a CTCF-mediated chromatin loop in the HPV18 genome

Handoko *et al* have demonstrated that a diverse number of chromatin looping interactions in murine embryonic cells (mES) are anchored and connected by the CTCF protein. They have further demonstrated that CTCF-mediated chromatin loops form functional domains that can be either be transcriptionally active gene clusters or silenced domains (Handoko et al., 2011). Furthermore, Beagan *et al* demonstrated YY1-mediated loop interactions in neural progenitor cells that were anchored by constitutively occupied CTCF sites (Beagan et al., 2017), and Donohoe *et al* have already postulated that the zinc finger binding domain of YY1 can directly interact with the N-terminus of the CTCF protein (Donohoe et al., 2007). Taken together, it was hypothesised that there was a repressive chromatin loop formed between CTCF in the E2 ORF and YY1 bound in the viral LCR, and upon abrogation of CTCF binding this repressive loop would be lost, leading to the observed increase in early gene expression. Several techniques are available to study 3D chromatin interactions; however, in this case two regions of the genome were hypothesised to be interacting, so 3C was the most appropriate method to investigate this interaction.

The 3C method involves cross-linking interacting chromatin regions, followed by restriction enzyme digestion and re-ligation of the cross-linked fragments. The re-ligated fragments are subsequently identified by PCR. The principle is that if two

chromatin regions form an interaction, then upon re-ligation the ends of the digested fragments will be in close enough proximity to re-ligate and form a novel product (Dekker et al., 2002). In this study, reverse PCR primers were designed in the CTCF-bound E2 region and L1 region proximal to the YY1 bound LCR of the HPV18 genome. If CTCF and YY1 form an interaction then, upon re-ligation, the E2 primer would have the potential to become the forward primer and amplify a PCR product with the L1 reverse primer. However, if CTCF and YY1 do not form an interaction then the cut ends would not be in close enough proximity to form a ligation product. The predicted ligation product is 350bp; therefore, PCR was used to amplify the product, as opposed to qPCR which amplifies smaller products below 200bp. Additionally, a commercially synthesised DNA fragment with the same sequence as the predicted ligation product was also amplified in the PCR reaction as the positive control and separated on an agaorse gel with the unknown samples. All of the PCR products were sequenced to ensure specificity of the reaction. Indeed, the 3C experiments have demonstrated the presence of a chromatin loop between CTCF in the E2 ORF and YY1 in the viral LCR, as indicated by the presence of a ligation product in the WT HV18 genome containing HFKs, which was significantly reduced in the ΔCTCF HPV18 genome containing cells. Interestingly, the ligation product was further reduced in the WT HPV18 genome containing HFKs after methylcellulose differentiation, and was barely detectable in differentiated ΔCTCF HPV18 genome containing cells. This indicates that this chromatin loop is dynamically regulated throughout the viral life cycle and formation is regulated by

cellular differentiation. Interestingly, SUZ12 which is also member of the PRC2 complex, can directly interact with the CTCF protein to mediate the suppression of the maternal Igf2 promoters (Li et al., 2008), so it possible that there is a complex of CTCF, YY1 and SUZ12 that is required to mediate repression and attenuation of the early promoter. It is possible that random interactions occur in the HPV genome and may be detected in the 3C experiment. To address this issue, a second primer was designed in the L2 region to use in a PCR reaction with the E2 primer. A ligation product is not expected to arise from using these two primers. Indeed, PCR reactions failed to amplify an L2-E2 ligation product in either the WT HPV18 or ΔCTCF HPV18 genome containing cells using the L2 and E2 primer set. This provided strong evidence that the PCR product observed using the L1 and E2 primers was not due to a random ligation event in the HPV18 genome.

Further work is needed to identify the specific YY1 sites involved in this chromatin loop formation, which could involve the mutation of individual YY1 binding sites in a plasmid containing the LCR and E2 regions, followed by 3C chromatin loop analysis. However, individual mutations of the YY1 binding sites are time consuming, and formation of the chromatin loop may be dependent on the presence of the whole viral genome. Alternatively it would be interesting to knockdown YY1 protein expression, and determine the presence of the chromatin loop formation. This would provide strong evidence that the CTCF mediated loop formation is dependent on YY1 protein binding in the LCR region.

Additional work has been carried out to confirm that the E2-LCR chromatin loop is dependent on CTCF. A doxycycline inducible HFK cell line was created in order to induce the stable knockdown of CTCF protein. The completion of 3C experiments using this inducible cell line confirmed that upon knockdown of CTCF protein there was also a significant reduction in the formation of the chromatin loop product (Dr. Campos Leon, unpublished). This provides further evidence that formation of this chromatin loop is dependent on CTCF protein.

The aim of the 3C experiment was to determine if there was an intramolecular chromatin loop within viral episomes. These experiments focussed on the interaction between the E2 ORF and YY1 in the LCR, but it is possible that additional chromatin loops are present within the viral genome. Alternatively, unbiased 3C walking could be used to determine other sequences involved in long-range interactions with CTCF or other regions of interest. Furthermore, chromatin interaction analysis by paired-end tag sequencing (ChIA-PET) would be the most comprehensive experiment to carry out in order to identify additional chromatin interactions associated with the CTCF protein. This technique involves an initial ChIP reaction in order to enrich for the specific protein, followed by proximity ligation and next generation sequencing of interacting fragments (Goh et al., 2012). Furthermore, the HFK cell lines in these experiments contain multiple copies of the viral episomes, so the current 3C method doesn't differentiate between intra- or inter-molecular interactions. Therefore, it is also possible that chromatin interactions form between neighbouring viral episomes

and not just within a single episome. Interestingly, Kang *et al* have already postulated that CTCF bound to the KSHV genome may mediate inter-molecular interactions important for sister-chromatid cohesion (Kang et al., 2011). It is also possible that CTCF bound to the HPV genome may facilitate tethering of viral episomes to the host chromatin. Previous work has already elucidated the role of ChlR1 in tethering viral episomes to host chromatin and mediating viral genome segregation (Parish et al., 2006b). ChlR1 has also been shown to interact with SMC1 and cohesion proteins (Parish et al., 2006a). Furthermore, genome wide studies have found that occupancy of the cohesin subunits RAD21 and SMC3 are highly correlated with CTCF-occupied sites (Hsu et al., 2017) (Wendt et al., 2008, Parelho et al., 2008). Mehta *et al* have already demonstrated that the cohesion protein pSMC1 and CTCF co-localise to distinct nuclear foci in HPV31 containing CIN612 cells after calcium-induced differentiation. They have also shown that mutation of three CTCF binding motifs in the L2 region of the HPV31 genome resulted in the loss of both CTCF and SMC1 binding to the L2 gene region. Furthermore, mutation of the three CTCF binding motifs resulted in a loss of viral episome maintenance in monolayer culture (Mehta et al., 2015). Therefore, CTCF bound to the viral genome may recruit the cohesion complex and ChlR1 in order to facilitate the maintenance of viral episomes in monolayer culture; however, mutation of the CTCF binding site in HPV18 does not affect genome maintenance and therefore argues against this hypothesis. In addition, Tempera *et al* have demonstrated that mutation of CTCF binding sites can result in a reduction in stable episome copy number of EBV

genomes, so a CTCF-mediated tethering mechanism is also likely to occur in this viral genome as well (Tempera et al., 2010).

6.7 Comparisons to the role of CTCF binding in other DNA viruses

Findings from our study correlate with some of the data obtained from the research studying CTCF involvement in large DNA virus life cycles. Our results have demonstrated that CTCF binding in the E2 ORF of HPV18 is required to form a repressive chromatin loop to attenuate early gene expression. In CMV, depletion of CTCF caused an increase in the major immediate early (MIE) and early gene expression and a subsequent 50-fold increase in virion production, supporting the role of CTCF as a transcriptional repressor. The same group also proposed that CTCF binding in the first intron can block RNA polymerase II elongation during the initial stages of transcription (Martinez et al., 2014). Similarly, CTCF binding between the OriP and C promoter in the EBV genome in type I latency cells was required for enhancer blocking repression of EBNA2 expression (Chau et al., 2006). Furthermore, Kang *et al* identified a CTCF binding site cluster in the first intron of the KSHV latency control region, and subsequent mutation of this site caused elevated expression of the lytic genes K14 and ORF74, again indicating CTCF has a role to play in maintaining transcriptional repression (Stedman et al., 2008).

The role of CTCF in mediating long-range chromatin loops has also been investigated in the EBV and KSHV genomes. Tempera *et al* demonstrated that CTCF

mediated distinct chromatin loops in different EBV latency types. Here, they demonstrated that in type I latency cells, where the Qp promoter is active, there was a chromatin loop formed between Qp and OriP (40kb). Furthermore, in type III cells, where the Cp promoter is active, there was a chromatin loop formed between Cp and OriP (2.2kb), which is a similar distance to our chromatin loop formed in the HPV genome (3kb) (Tempera et al., 2011). Chen *et al* demonstrated that abrogation of CTCF binding in the overlapping 3′ region of LMP1 and the first intron of LMP2A, resulted in the loss in loop formation between OriP and the LMP1 and LMP2A regions. This subsequently resulted in repression of the LMP1 and LMP2A promoter regions, and an alternate increase in LMP2B expression (Chen et al., 2013a). Kang *et al* also used 3C to demonstrate a role for CTCF in mediating DNA loop formation in the KSHV genome to regulate viral transcription. Here, they identified a DNA loop formed between the CTCF binding sites in the latency control region and ORF50 (60kb apart), and loop between the 5′ end of latency control region and the 3′ end (18kb). When the three CTCF binding site clusters in the latency control region were mutated there was a 2-3 fold reduction in loop formation between the ORF50 region, as well as a 2-3 fold reduction in ORF50 lytic mRNA and even derepression of some latent genes (Kang et al., 2011). They also hypothesised that PRC2 recruitment and H3K27 methylation may be directly involved in regulating KSHV loop formation, and this is highly possible given our observed alterations in H3K27me3 enrichment in the HPV genome. Abrogation of CTCF binding did not completely abrogate loop formation in our monolayer cultures. Similarly, siRNA depletion of CTCF did not

completely abolish KSHV chromatin loop formation between the latency control region and ORF50, which Kang *et al* suggested is due to the involvement of other cellular factors in mediating loop formation (Kang et al., 2011). Moreover, abrogation of CTCF binding to the HPV18 E2 ORF is not 100% efficient, so any residual CTCF binding it still able to mediate some loop formation.

Overall these results demonstrate that CTCF binds to multiple sites in the genomes of DNA viruses, in order to control aspects of these viral life cycles. It is apparent that CTCF is required to mediate distinct chromatin loops that are specific to EBV latency types. Furthermore, CTCF is also required to mediate long-range chromatin loops, in order to maintain the balance of latent and lytic expression in the KSHV genome. Similarly, CTCF is needed to bind to the high-risk HPV genome, to mediate repressive chromatin loop formation and to maintain the balance of early viral oncoprotein expression.

6.8 Future Directions

6.8.1 The role of CTCF in transcriptional control

Whilst some of these experiments have successfully demonstrated that abrogation of CTCF binding in the E2 ORF of HPV18 causes aberrant gene splicing events in the viral genome, it would be advantageous to obtain a more comprehensive analysis of viral transcript expression in WT HPV18 and ΔCTCF HPV18 genome containing cells. The application of RNA-Seq would permit higher resolution sequencing of RNA transcripts produced in the WT HPV18 and ΔCTCF HPV18 genome containing

HFKs. This technique would allow the comparison of RNA species between the WT HPV18 and ΔCTCF HPV18 genome containing cells, in terms of both novel spliced transcripts in the mutant cell line but also the altered ratios of known HPV transcripts. Indeed, preliminary RNA-Seq has been carried out using RNA extracted from WT HPV18 and ΔCTCF HPV18 genome containing HFKs grown in monolayer culture (Dr Campos-Leon). RNA-Seq has confirmed the previously identified increase in E6 and E7 encoding transcripts in the ΔCTCF HPV18 genome containing cells. Furthermore, both the WT HPV18 and ΔCTCF HPV18 genome containing cells expressed a subset of previously uncharacterised transcripts and alterations in the ratio of known mRNA species. In terms of known viral transcripts, ΔCTCF HPV18 cells displayed a decrease in the transcript with splicing between 1357^3434, which is has the potential to encode E8^E2 and E5. Conversely, ΔCTCF HPV18 cells displayed an increase in the transcript with splicing between 929^2779, which has the potential to encode E6, E7, E2 and E5. Furthermore, the transcript with splicing between 929^3465 that encodes E6, E7 and E5 was undetectable in the WT HPV18 cells, whilst there were 391 detectable reads in ΔCTCF HPV18 cells.

Both the WT HPV18 and ΔCTCF HPV18 genome containing cells expressed a previously uncharacterised transcript with splicing between 3284^3434, but with more expressed in the WT HPV18 cells. The ΔCTCF HPV18 cells displayed a complete loss of a novel mRNA product with splicing between 3165^3434, which was abundantly expressed in the WT HPV18 cells. The loss of this particular

transcript is interesting because the splice donor site at nucleotide 3165 is close to the CTCF binding site at nucleotide 2989. In the ΔCTCF HPV18 cells where CTCF binding is abrogated, the RNA polymerase complex will not be paused on the transcript and therefore this splicing event may be skipped. Furthermore, two uncharacterised transcripts spliced between 929^6628 and 1509^6740 were expressed in the ΔCTCF HPV18 cells and completely absent in the WT HPV18 genome containing cells. It is clear that CTCF binding in the E2 ORF of the HPV18 genome is required to co-ordinate splicing events and the subsequent abrogation of CTCF binding results in aberrant gene splicing to occur. Previous genome wide ChIP-Seq studies have shown that CTCF binding sites co-localise with RNA polymerase II enriched regions (Wada et al., 2009), and Chernukhin *et al* have demonstrated that CTCF binds to the large subunit of RNA polymerase II (Chernukhin et al., 2007). Furthermore, CTCF has been shown to bind to the CD45 locus and cause RNA polymerase II pausing to allow inclusion of weak upstream exons (Shukla et al., 2011). It is likely that the interaction and interplay between RNA polymerase II protein and CTCF is involved in modulating HPV transcript splicing and expression. The ChIPs shown in Chapter 3, used an antibody that detects total RNA polymerase II and was unable to discriminate between the different phosphorylated forms of the protein. Further ChIP experiments need to be carried out to confirm the binding of the elongating or paused forms of RNA polymerase II throughout the viral genome, to give a more detailed understanding of CTCF binding and RNA polymerase II processing. Overall, the pausing of RNA polymerase II in the E2 ORF may facilitate

the splicing of all the potential HPV transcripts, so it is something that needs further investigation. RNA polymerase II is known to recruit splicing factors to pre-mRNA (de Almeida and Carmo-Fonseca, 2008, Howe, 2002, Goldstrohm et al., 2001a, Goldstrohm et al., 2001b) so it is possible that differential binding of RNA polymerase II in the WT or ΔCTCF HPV18 genome may be modulating splicing factor recruitment. It would also be interesting to investigate the recruitment of chromatin modifying proteins to the viral genome, as it has previously been shown that RNA polymerase II can alter chromatin organization to facilitate splicing factor recruitment (Schwartz and Ast, 2010, Luco et al., 2010).

6.8.2 The role of CTCF in viral genome tethering

Wu *et al* demonstrated that BRD4 is implicated in the E2-mediated repression of early HPV gene expression. Here they demonstrated that BRD4 mediates E2 recruitment to the HPV chromatin and blocks binding of TFIID and RNA polymerase II, which prevents assembly of the preinitiation complex (Wu et al., 2006). BRD4 has also been shown to bind to BPV1 E2 protein and tether BPV1 DNA to mitotic host DNA (You et al., 2004). Given that BET proteins are also widely involved in chromatin organization (Muller et al., 2011), it was hypothesized that these proteins may associate with CTCF to facilitate its role as a chromatin insulator. In a genome wide study the BET protein BRD2 was shown to co-localize with CTCF sites in the human genome. Furthermore, binding of BRD2 and cohesin displayed strong correlation at these CTCF-occupied sites, which suggests that many of the

CTCF binding sites are occupied by both BRD2 and cohesin. Overall the binding of BRD2 to CTCF sites is thought to maintain the integrity of insulating domain boundaries. Conversely, the BET protein BRD4 was not strongly correlated with genome-wide CTCF sites. With this in mind the CTCF bound to the E2 ORF may not bind to host BRD4 to mediate viral tethering to the host chromatin. On the other hand, E2 protein has also been shown to bind to ChlR1, which facilitates the tethering of viral genomes to the host chromatin (Parish et al., 2006a). Interestingly, Mehta *et al* have already demonstrated that the cohesin protein pSMC1 and CTCF co-localise to distinct nuclear foci in HPV31 containing CIN612 cells. Mutation of CTCF binding motifs in the L2 region of the HPV31 genome resulted in the loss of both CTCF and SMC1 binding to the L2 gene region, and a loss of viral episome maintenance in monolayer culture (Mehta et al., 2015). Therefore, it would be interesting to determine if CTCF bound to the viral genome forms an interaction with the cohesion complex and ChlR1 in order to facilitate viral genome tethering and the maintenance of viral episomes in monolayer culture.

6.8.3 Clinical implications

The results presented in this thesis have demonstrated that upon abrogation of CTCF binding to E2 ORF of high-risk HPV18, there is a concomitant increase in the expression of the E6 and E7 viral oncoproteins. It is clear that the high-risk virus needs to recruit host proteins to the viral genome in order to control the equilibrium of viral oncoprotein expression and abrogation of this interactions leads to aberrant

oncoprotein expression. Therefore it would be interesting to experimentally determine if CTCF binding to the E2 ORF of viral genomes correlates with clinical outcome. Interestingly, there is evidence from our collaborators that reduced CTCF binding in the HPV16 genome correlates with increased E6 mRNA expression in W12 derived clones (Groves and Coleman). It is hypothesised that loss of CTCF binding to the E2 ORF correlates with increased malignant progression. Fanelli *et al* have optimised the PAT-ChIP technique which combines ChIP with high-throughput sequencing from paraffin embedded tissue (Fanelli et al., 2011). Therefore, to test this hypothesis, clinical sections derived from CIN1, CIN2, CIN3 and carcinoma *in situ* cervical tissue samples would be used in CTCF ChIP experiments followed by qRT-PCR analysis of the E6 and E7 encoding transcripts. The amount of CTCF bound to the E2 ORF of HPV18 would be calculated for the tissue stage of each sample, and used to determine any significant correlation between CTCF binding and clinical pathology. One would predict that CIN1 samples would contain higher enrichment of CTCF binding to the E2 ORF of HPV18 compared to the CIN2 and CIN3 samples. Furthermore, CTCF binding to the E2 ORF could be compared between samples with the same histological grade, and determine if there is a correlation between CTCF binding and progression to SCC.

Frequent insertion/deletion mutations have been identified in CTCF coding sequences expressed in endometrial cancers with microsatellite instability (Zighelboim et al., 2014), with one study estimating that CTCF mutations occur in

around 25% of endometrial tumours (Walker et al., 2015). Taken together, these findings demonstrate that the loss of function of CTCF as a tumour suppressor is a frequent event in endometrial tumourigenesis. Data from a TCGA study identified CTCF gene deletion or mutation in 2.6% of 191 cervical cancer samples. Overall survival was not significantly altered between patients with normal CTCF expression and patients expressing alterations in CTCF (P=0.935). However, the same data revealed that relapse occurred in 28% of cases with altered CTCF expression, whilst relapse only occurred in 16% of cases with unaltered CTCF expression, although overall disease free survival was not significantly different between patients with normal CTCF expression and patients expressing alterations in CTCF (P=0.179). Another TCGA study identified alterations in the CTCF gene in 4.3% of 279 of head and neck squamous cell carcinoma samples; however, the proportion of HPV positive samples is unknown in this cohort (Cancer Genome Atlas Network 2015). When data were stratified according to levels of CTCF mRNA expression in cervical SCC, the overall survival data revealed a significantly higher survival rate in patients with high CTCF expression, when compared to those with low CTCF expression (P=0.047) (Fig. 85). When the same stratification was used for head and neck SCC, the data revealed that high CTCF protein expression did not correlate with higher survival rates in HPV negative head and neck SCC (Fig. 86). Most interestingly, in HPV positive head and neck SCC there appears to be higher survival rates in patients with high CTCF expression compared to low CTCF expression, although this was not significant (P=0.14) (Fig. 86). Nevertheless, this

strengthens our case that high-risk HPV types need to recruit CTCF to the viral genome in order to attenuate early gene expression. In the head and neck example, lower expression of CTCF and thus reduced binding to the E2 ORF will cause an increase in early gene expression, and may contribute to HPV induced carcinogenesis, and the lower survival rates.

Several studies have identified mutations in CTCF binding sites in certain cancers. In fact CTCF binding sites are emerging as mutation hot spots in the non-coding genome sequences in many cancers (Katainen et al., 2015). Although HPV genomes are very stable and rarely undergo mutations in the DNA sequence, viral integration is known to disrupt and inactivate the E2 gene (Cullen et al., 1991, Pett and Coleman, 2007). Equally, disruption of the E2 ORF can interrupt the binding of CTCF to the E2 ORF binding site. Indeed, CTCF does not bind to the HPV18 genome in HeLa cells, which were originally derived from a high-grade cervical carcinoma (Johannsen and Lambert, 2013). We hypothesise that there is a correlation between CTCF binding site integrity in the HPV18 genome and clinical outcome; cells displaying disrupted CTCF binding sites within the HPV genome may be linked to a worse prognosis.

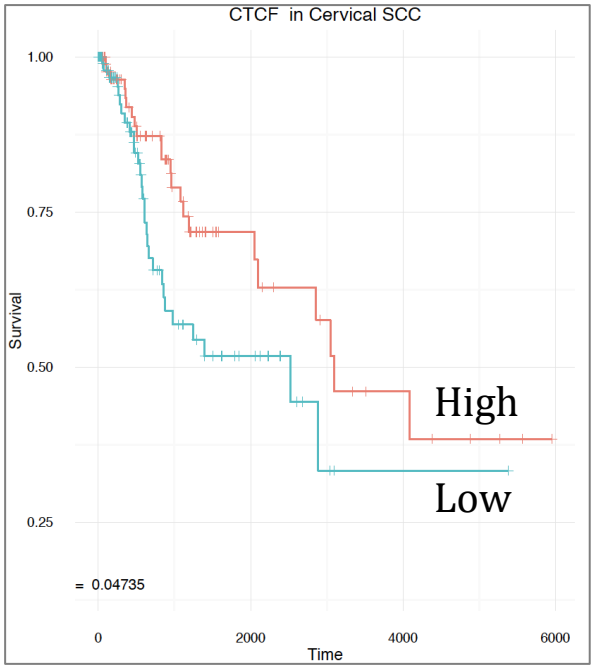


Figure 85. Kaplan-Meier survival plot for cases of cervical SCC with high or low CTCF expression. Cervical SCC survival data was taken form the TCGA data base and stratified according high or low CTCF mRNA expression. Survival is signiciantly better with high CTCF protein expression compared to low expression. P=0.04735.

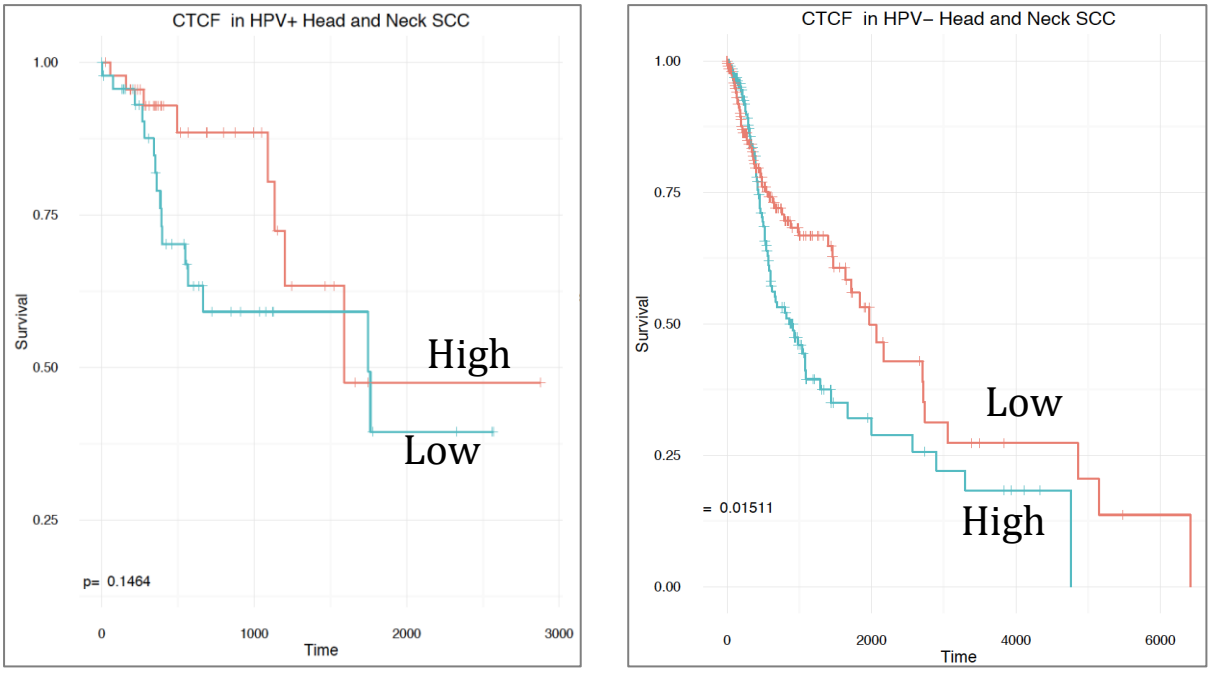


Figure 86. Kaplan-Meier survival plot for cases of HPV positive or HPV negative SCC with high or low CTCF expression. HPV positive or HPV negative head and neck SCC survival data was taken from the TCGA database and stratified according high or low CTCF mRNA expression. Significance was determined. Survival is better with high CTCF expression in HPV positive head and neck SSCs but this is not signiciant, P=0.1464. Survival is significantly better with low CTCF protein expresison in HPV negative head and neck SCCs, P=0.01511.

6.9 Summary and final suggestions

In conclusion, we have demonstrated that the host protein, CTCF, is recruited to the E2 ORF of high-risk HPV genomes in order to regulate aspects of the viral life cycle. Abrogation of CTCF binding to the E2 ORF resulted in an increase in E6 and E7 oncoprotein expression, as well as alterations in late viral gene and protein expression. Loss of CTCF binding was shown to alter both the epigenetic landscape and nucleosome positioning across the viral genome. Overall, we propose a model whereby CTCF binds to the E2 ORF of HPV18 and forms a repressive chromatin loop with YY1 bound in the LCR region. The formation of this chromatin loop is needed to recruit the polycomb repressive complexes for attenuation of early gene expression in undifferentiated cells. Whilst the PRC2 complex is known to be recruited to DNA by YY1, the SUZ12 component of PRC2 is also a known binding partner of CTCF, so this may form part of a larger complex (Fig. 87). Upon epithelial differentiation we see a reduction in both YY1 and CTCF protein expression, but surprisingly an increase in CTCF enrichment to the E2 ORF. Therefore, it is most likely that loss of YY1 binding to the LCR upon cellular differentiation results in the reduction of repressive loop formation. This in turn would allow the increased E6 and E7 expression required for viral genome replication in the suprabasal layers of the epithelium. Importantly, by employing a mechanism to attenuate early gene expression this will not only facilitate long-term HPV infection in basal epithelial cells but will also help the virus evade immune surveillance.

Further work is needed to understand the role of CTCF in mediating viral splicing events; however, it is clear from our data that abrogation of CTCF binding to the E2 ORF causes aberrant gene splicing of both early and late transcripts. Further work is needed to determine the binding profile of the elongating or paused forms of RNA polymerase across the WT HPV18 genome and compare the binding in the ΔCTCF HPV18 genome containing cells. It would also be interesting to investigate the recruitment of splicing factors and chromatin modifying proteins and determine their role in HPV splicing events.

We have successfully demonstrated that there is a chromatin loop between CTCF in the E2 ORF and YY1 in the LCR; however, CTCF bound to the E2 ORF may be forming additional chromatin loops in the viral genome. Furthermore, upon differentiation there was increased enrichment of CTCF binding to the E2 ORF, but a reduction in the E2-LCR loop formation. Therefore it is possible that a different loop is formed between CTCF in the E2 ORF and another region of the viral genome. It is also possible that the other identified CTCF binding sites in L1 and L2 late regions of the HPV16 genome may be forming distinct chromatin loops. A higher resolution technique such as ChIA-PET would allow the identification of any further long-range interactions mediated by CTCF.

Another area to investigate further is the possibility that CTCF binding in the viral genome is involved in viral genome tethering to the host chromatin. As mentioned

previously, it would be interesting to determine if CTCF forms a complex with cohesion and ChlR1.

Lastly, it would be interesting to determine if CTCF binding is correlated with clinical outcome. PAT-ChIP experiments could be performed on clinical tissue with different histological grades to determine the enrichment of CTCF bound to the E2 ORF of the viral genome in the infected tissue. Ultimately, the determination of CTCF binding patterns could be used as a potential diagnostic and prognostic tool in the clinical setting. .

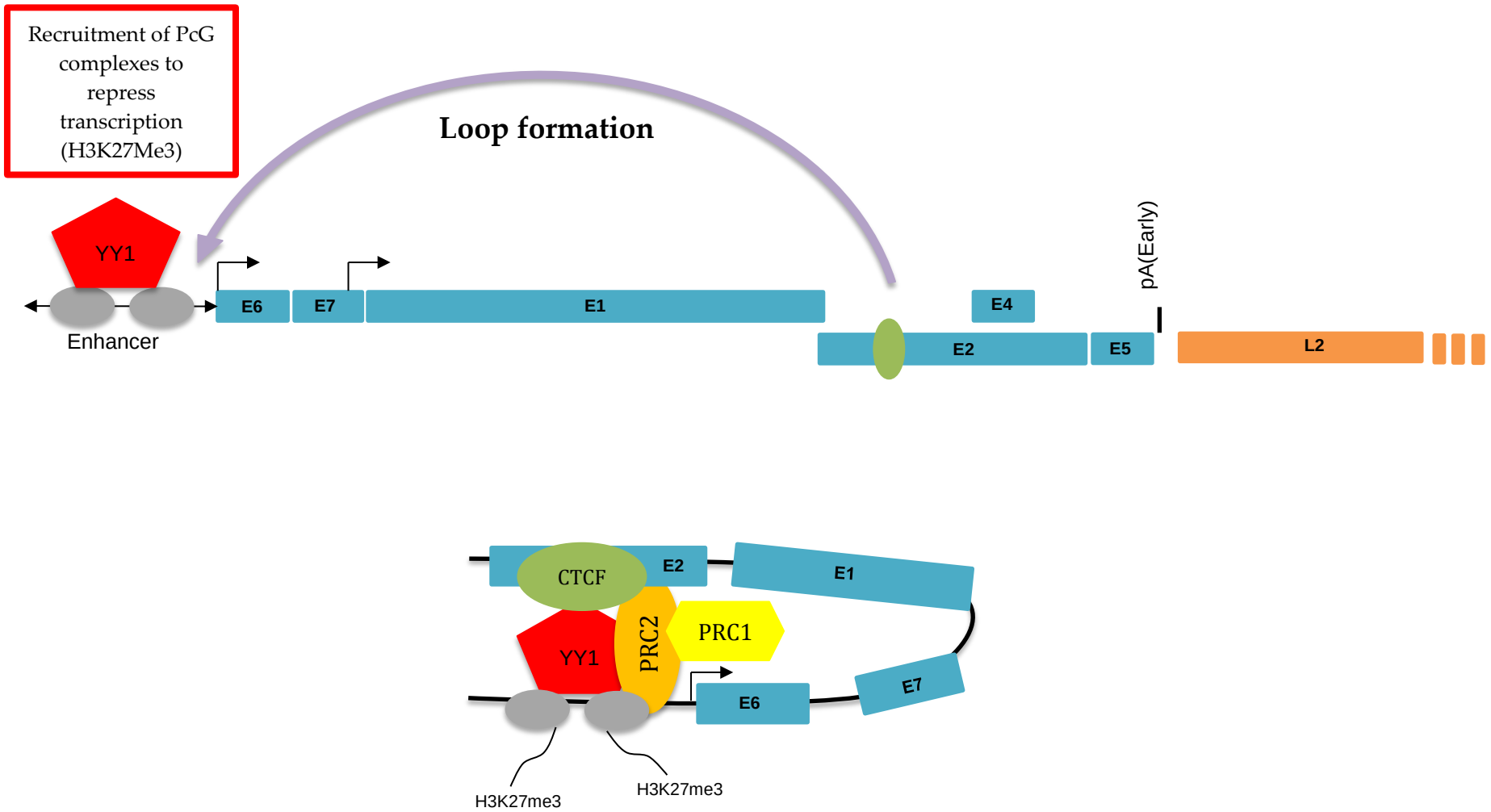


Figure 87. Proposed model for CTCF-mediated repressive loop formation in the HPV18 genome. CTCF binds in the E2 ORF and forms a chromatin loop with YY1 bound in the viral LCR. YY1 can recruit the PRC2 complex, which catalyses the methylation of lysine 27 on histone 3. PRC2 can also recruit the PRC1 complex, which catalyses the ubiuitiniation of H2A at lysine 119. The recruitment of the PRC complexes mediates repression of E6 and E7 expression from the early promoter.

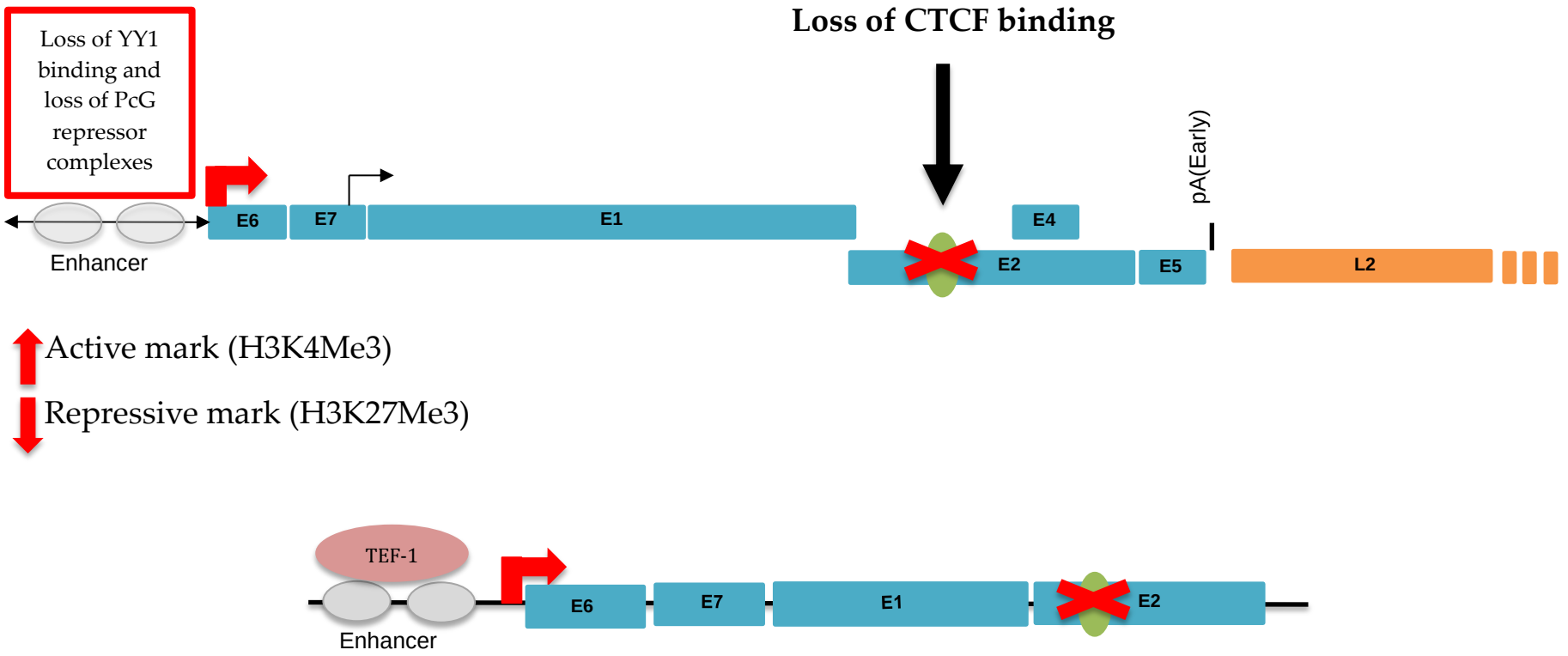


Figure 88. Loss of repressive loop formation upon mutation of the CTCF binding site in the HPV18 genome. Mutation of the CTCF binding site in the E2 ORF causes abrogation of CTCF binding. There is also a loss of YY1 binding in the LCR and a loss of chromatin loop formation between the E2 ORF and the LCR region. There is loss of both PRC1 and PRC2 repressive complex recruitment and depletion of H3K27me3 enrichment within the LCR region. The increase in early promoter activity is indicated by enrichment of H3K4me3and increased E6 and E7 expression. There is also increased binding of TEF-1 within the viral enhancer of the LCR region.

7 References

2014. HPV Vaccine Works against Nine Viral Types. *Cancer Discov,* 4**,** OF2.

ADEY, A., MORRISON, H. G., ASAN, XUN, X., KITZMAN, J. O., TURNER, E. H., STACKHOUSE, B., MACKENZIE, A. P., CARUCCIO, N. C., ZHANG, X. & SHENDURE, J. 2010. Rapid, low-input, low-bias construction of shotgun fragment libraries by high-density in vitro transposition. *Genome Biol,* 11**,** R119.

ALCOCER-GONZALEZ, J. M., BERUMEN, J., TAMEZ-GUERRA, R., BERMUDEZ-MORALES, V., PERALTA-ZARAGOZA, O., HERNANDEZ-PANDO, R., MORENO, J., GARIGLIO, P. & MADRID-MARINA, V. 2006. In vivo expression of immunosuppressive cytokines in human papillomavirus-transformed cervical cancer cells. *Viral Immunol,* 19**,** 481-91.

AMELIO, A. L., MCANANY, P. K. & BLOOM, D. C. 2006. A chromatin insulator-like element in the herpes simplex virus type 1 latency-associated transcript region binds CCCTC-binding factor and displays enhancer-blocking and silencing activities. *Journal of Virology,* 80**,** 2358-2368.

AMMERMANN, I., BRUCKNER, M., MATTHES, F., IFTNER, T. & STUBENRAUCH, F. 2008. Inhibition of transcription and DNA replication by the papillomavirus E8(Lambda)E2C protein is mediated by interaction with corepressor molecules. *Journal of Virology,* 82**,** 5127-5136.

ARBEIT, J. M., MUNGER, K., HOWLEY, P. M. & HANAHAN, D. 1994. Progressive Squamous Epithelial Neoplasia in K14-Human Papillomavirus Type-16 Transgenic Mice. *Journal of Virology,* 68**,** 4358-4368.

ARBYN, M., CASTELLSAGUE, X., DE SANJOSE, S., BRUNI, L., SARAIYA, M., BRAY, F. & FERLAY, J. 2011. Worldwide burden of cervical cancer in 2008. *Ann Oncol,* 22**,** 2675-86.

ARNEY, A. & BENNETT, K. M. 2010. Molecular Diagnostics of Human Papillomavirus. *Labmedicine,* 41**,** 523-530.

ASHRAFI, G. H., BROWN, D. R., FIFE, K. H. & CAMPO, M. S. 2006. Down-regulation of MHC class I is a property common to papillomavirus E5 proteins. *Virus Res,* 120**,** 208-11.

ATCHISON, L., GHIAS, A., WILKINSON, F., BONINI, N. & ATCHISON, M. L. 2003. Transcription factor YY1 functions as a PcG protein in vivo. *Embo Journal,* 22**,** 1347-1358.

BAKER, C. C., PHELPS, W. C., LINDGREN, V., BRAUN, M. J., GONDA, M. A. & HOWLEY, P. M. 1987. Structural and Transcriptional Analysis of Human Papillomavirus Type-16 Sequences in Cervical-Carcinoma Cell-Lines. *Journal of Virology,* 61**,** 962-971.

BALDWIN, A., HYPES, M. K., PIRISI, L. & CREEK, K. E. 2007. NFI is an essential positive transcription factor for human papillomavirus type 16 early gene expression. *Open Virol J,* 1**,** 33-8.

BANIAHMAD, A., STEINER, C., KOHNE, A. C. & RENKAWITZ, R. 1990. Modular structure of a chicken lysozyme silencer: involvement of an unusual thyroid hormone receptor binding site. *Cell,* 61**,** 505-14.

BAO, L., ZHOU, M. & CUI, Y. 2008. CTCFBSDB: a CTCF-binding site database for characterization of vertebrate genomic insulators. *Nucleic Acids Res,* 36**,** D83-7.

BARBOSA, M. S., EDMONDS, C., FISHER, C., SCHILLER, J. T., LOWY, D. R. & VOUSDEN, K. H. 1990. 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. *EMBO J,* 9**,** 153-60.

BARROW-LAING, L., CHEN, W. & ROMAN, A. 2010. Low- and high-risk human papillomavirus E7 proteins regulate p130 differently. *Virology,* 400**,** 233-9.

BARSKI, A., CUDDAPAH, S., CUI, K., ROH, T. Y., SCHONES, D. E., WANG, Z., WEI, G., CHEPELEV, I. & ZHAO, K. 2007a. High-resolution profiling of histone methylations in the human genome. *Cell,* 129**,** 823-37.

BARSKI, A., CUDDAPAH, S., CUI, K. R., ROH, T. Y., SCHONES, D. E., WANG, Z. B., WEI, G., CHEPELEV, I. & ZHAO, K. J. 2007b. High-resolution profiling of histone methylations in the human genome. *Cell,* 129**,** 823-837.

BARSOUM, J., PRAKASH, S. S., HAN, P. & ANDROPHY, E. J. 1992. Mechanism of action of the papillomavirus E2 repressor: repression in the absence of DNA binding. *J Virol,* 66**,** 3941-5.

BASU, A., WILKINSON, F. H., COLAVITA, K., FENNELLY, C. & ATCHISON, M. L. 2014a. YY1 DNA binding and interaction with YAF2 is essential for Polycomb recruitment. *Nucleic Acids Research,* 42**,** 2208-2223.

BASU, A., WILKINSON, F. H., COLAVITA, K., FENNELLY, C. & ATCHISON, M. L. 2014b. YY1 DNA binding and interaction with YAF2 is essential for Polycomb recruitment. *Nucleic Acids Res,* 42**,** 2208-23.

BAUKNECHT, T., ANGEL, P., ROYER, H. D. & HAUSEN, H. Z. 1992. Identification of a Negative Regulatory Domain in the Human Papillomavirus Type-18 Promoter - Interaction with the Transcriptional Repressor-Yy1. *Embo Journal,* 11**,** 4607-4617.

BEAGAN, J. A., DUONG, M. T., TITUS, K. R., ZHOU, L., CAO, Z., MA, J., LACHANSKI, C. V., GILLIS, D. R. & PHILLIPS-CREMINS, J. E. 2017. YY1 and CTCF orchestrate a 3D chromatin looping switch during early neural lineage commitment. *Genome Res*.

BELL, A. C. & FELSENFELD, G. 2000. Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene. *Nature,* 405**,** 482-5.

BELL, A. C., WEST, A. G. & FELSENFELD, G. 1999. The protein CTCF is required for the enhancer blocking activity of vertebrate insulators. *Cell,* 98**,** 387-96.

BELL, I., MARTIN, A. & ROBERTS, S. 2007. The E1circumflexE4 protein of human papillomavirus interacts with the serine-arginine-specific protein kinase SRPK1. *J Virol,* 81**,** 5437-48.

BERMUDEZ-MORALES, V. H., GUTIERREZ, L. X., ALCOCER-GONZALEZ, J. M., BURGUETE, A. & MADRID-MARINA, V. 2008. Correlation Between IL-10 Gene Expression and HPV Infection in Cervical Cancer: A Mechanism for Immune Response Escape. *Cancer Investigation,* 26**,** 1037-1043.

BERNARD, B. A., BAILLY, C., LENOIR, M. C., DARMON, M., THIERRY, F. & YANIV, M. 1989. The Human Papillomavirus Type-18 (Hpv18) E2 Gene-Product Is a Repressor of the Hpv18 Regulatory Region in Human Keratinocytes. *Journal of Virology,* 63**,** 4317-4324.

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

BERNSTEIN, B. E., HUMPHREY, E. L., ERLICH, R. L., SCHNEIDER, R., BOUMAN, P., LIU, J. S., KOUZARIDES, T. & SCHREIBER, S. L. 2002. Methylation of histone H3 Lys 4 in coding regions of active genes. *Proceedings of the National Academy of Sciences of the United States of America,* 99**,** 8695-8700.
BERUMEN, J., CASAS, L., SEGURA, E., AMEZCUA, J. L. & GARCIA-CARRANCA, A. 1994. Genome amplification of human papillomavirus types 16 and 18 in cervical carcinomas is related to the retention of E1/E2 genes. *Int J Cancer,* 56**,** 640-5.
BHATTACHARJEE, B. & SENGUPTA, S. 2006. CpG methylation of HPV 16 LCR at E2 binding site proximal to P97 is associated with cervical cancer in presence of intact E2. *Virology,* 354**,** 280-285.
BIAN, C., XU, C., RUAN, J., LEE, K. K., BURKE, T. L., TEMPEL, W., BARSYTE, D., LI, J., WU, M., ZHOU, B. O., FLEHARTY, B. E., PAULSON, A., ALLALI-HASSANI, A., ZHOU, J. Q., MER, G., GRANT, P. A., WORKMAN, J. L., ZANG, J. & MIN, J. 2011. Sgf29 binds histone H3K4me2/3 and is required for SAGA complex recruitment and histone H3 acetylation. *EMBO J,* 30**,** 2829-42.
BLACKLEDGE, N. P., FARCAS, A. M., KONDO, T., KING, H. W., MCGOURAN, J. F., HANSSEN, L. L., ITO, S., COOPER, S., KONDO, K., KOSEKI, Y., ISHIKURA, T., LONG, H. K., SHEAHAN, T. W., BROCKDORFF, N., KESSLER, B. M., KOSEKI, H. & KLOSE, R. J. 2014. Variant PRC1 complex-dependent H2A ubiquitylation drives PRC2 recruitment and polycomb domain formation. *Cell,* 157**,** 1445-59.
BODILY, J. & LAIMINS, L. A. 2011. Persistence of human papillomavirus infection: keys to malignant progression. *Trends Microbiol,* 19**,** 33-9.
BRAATEN, K. P. & LAUFER, M. R. 2008. Human Papillomavirus (HPV), HPV-Related Disease, and the HPV Vaccine. *Rev Obstet Gynecol,* 1**,** 2-10.
BRAVO, I. G. & FELEZ-SANCHEZ, M. 2015. Papillomaviruses: Viral evolution, cancer and evolutionary medicine. *Evol Med Public Health,* 2015**,** 32-51.
BROWN, J. L., MUCCI, D., WHITELEY, M., DIRKSEN, M. L. & KASSIS, J. A. 1998. The Drosophila polycomb group gene pleiohomeotic encodes a DNA binding protein with homology to the transcription factor YY1. *Molecular Cell,* 1**,** 1057-1064.
BRYAN, J. T. & BROWN, D. R. 2000. Association of the human papillomavirus type 11 E1()E4 protein with cornified cell envelopes derived from infected genital epithelium. *Virology,* 277**,** 262-9.
BUCK, C. B., CHENG, N., THOMPSON, C. D., LOWY, D. R., STEVEN, A. C., SCHILLER, J. T. & TRUS, B. L. 2008. Arrangement of L2 within the papillomavirus capsid. *J Virol,* 82**,** 5190-7.
BUCK, C. B., DAY, P. M. & TRUS, B. L. 2013. The papillomavirus major capsid protein L1. *Virology,* 445**,** 169-74.
BUENROSTRO, J. D., GIRESI, P. G., ZABA, L. C., CHANG, H. Y. & GREENLEAF, W. J. 2013. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat Methods,* 10**,** 1213-8.
BUENROSTRO, J. D., WU, B., CHANG, H. Y. & GREENLEAF, W. J. 2015. ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide. *Curr Protoc Mol Biol,* 109**,** 21 29 1-9.
BULK, S., BERKHOF, J., BULKMANS, N. W., ZIELINSKI, G. D., ROZENDAAL, L., VAN KEMENADE, F. J., SNIJDERS, P. J. & MEIJER, C. J. 2006. Preferential risk of

HPV16 for squamous cell carcinoma and of HPV18 for adenocarcinoma of the cervix compared to women with normal cytology in The Netherlands. *Br J Cancer,* 94**,** 171-5.

BURCIN, M., ARNOLD, R., LUTZ, M., KAISER, B., RUNGE, D., LOTTSPEICH, F., FILIPPOVA, G. N., LOBANENKOV, V. V. & RENKAWITZ, R. 1997. Negative protein 1, which is required for function of the chicken lysozyme gene silencer in conjunction with hormone receptors, is identical to the multivalent zinc finger repressor CTCF. *Mol Cell Biol,* 17**,** 1281-8.

CALDEIRA, S., ZEHBE, I., ACCARDI, R., MALANCHI, I., DONG, W., GIARRE, M., DE VILLIERS, E. M., FILOTICO, R., BOUKAMP, P. & TOMMASINO, M. 2003. The E6 and E7 proteins of the cutaneous human papillomavirus type 38 display transforming properties. *Journal of Virology,* 77**,** 2195-2206.

CAO, Q., WANG, X., ZHAO, M., YANG, R., MALIK, R., QIAO, Y., POLIAKOV, A., YOCUM, A. K., LI, Y., CHEN, W., CAO, X., JIANG, X., DAHIYA, A., HARRIS, C., FENG, F. Y., KALANTRY, S., QIN, Z. S., DHANASEKARAN, S. M. & CHINNAIYAN, A. M. 2014. The central role of EED in the orchestration of polycomb group complexes. *Nat Commun,* 5**,** 3127.

CAO, R., WANG, L., WANG, H., XIA, L., ERDJUMENT-BROMAGE, H., TEMPST, P., JONES, R. S. & ZHANG, Y. 2002. Role of histone H3 lysine 27 methylation in Polycomb-group silencing. *Science,* 298**,** 1039-43.

CARROZZA, M. J., LI, B., FLORENS, L., SUGANUMA, T., SWANSON, S. K., LEE, K. K., SHIA, W. J., ANDERSON, S., YATES, J., WASHBURN, M. P. & WORKMAN, J. L. 2005. Histone H3 methylation by Set2 directs deacetylation of coding regions by Rpd3S to suppress spurious intragenic transcription. *Cell,* 123**,** 581-92.

CARSON, A. & KHAN, S. A. 2006. Characterization of transcription factor binding to human papillomavirus type 16 DNA during cellular differentiation. *J Virol,* 80**,** 4356-62.

CHA, T. L., ZHOU, B. P., XIA, W., WU, Y., YANG, C. C., CHEN, C. T., PING, B., OTTE, A. P. & HUNG, M. C. 2005. Akt-mediated phosphorylation of EZH2 suppresses methylation of lysine 27 in histone H3. *Science,* 310**,** 306-10.

CHAN, W. K., CHONG, T., BERNARD, H. U. & KLOCK, G. 1990. Transcription of the Transforming Genes of the Oncogenic Human Papillomavirus-16 Is Stimulated by Tumor Promotors through Ap1 Binding-Sites. *Nucleic Acids Research,* 18**,** 763-769.

CHATURVEDI, A. K., ENGELS, E. A., ANDERSON, W. F. & GILLISON, M. L. 2008. Incidence trends for human papillomavirus-related and -unrelated oral squamous cell carcinomas in the United States. *J Clin Oncol,* 26**,** 612-9.

CHAU, C. M., ZHANG, X. Y., MCMAHON, S. B. & LIEBERMAN, P. M. 2006. Regulation of Epstein-Barr virus latency type by the chromatin boundary factor CTCF. *Journal of virology,* 80**,** 5723-32.

CHELLAPPAN, S., KRAUS, V. B., KROGER, B., MUNGER, K., HOWLEY, P. M., PHELPS, W. C. & NEVINS, J. R. 1992. Adenovirus E1A, simian virus 40 tumor antigen, and human papillomavirus E7 protein share the capacity to disrupt the interaction between transcription factor E2F and the retinoblastoma gene product. *Proc Natl Acad Sci U S A,* 89**,** 4549-53.

CHEN, H. S., MARTIN, K. A., LU, F., LUPEY, L. N., MUELLER, J. M., LIEBERMAN, P. M. & TEMPERA, I. 2013a. Epigenetic Deregulation of the LMP1/LMP2 Locus of Epstein-Barr Virus by Mutation of a Single CTCF-Cohesin Binding Site. *J Virol*.

CHEN, S. A., BOHRER, L. R., RAI, A. N., PAN, Y. Q., GAN, L., ZHOU, X. Z., BAGCHI, A., SIMON, J. A. & HUANG, H. J. 2010. Cyclin-dependent kinases regulate epigenetic gene silencing through phosphorylation of EZH2. *Nature Cell Biology,* 12**,** 1108-U118.

CHEN, Z., SCHIFFMAN, M., HERRERO, R., DESALLE, R., ANASTOS, K., SEGONDY, M., SAHASRABUDDHE, V. V., GRAVITT, P. E., HSING, A. W. & BURK, R. D. 2013b. Evolution and taxonomic classification of alphapapillomavirus 7 complete genomes: HPV18, HPV39, HPV45, HPV59, HPV68 and HPV70. *PLoS One,* 8**,** e72565.

CHERNUKHIN, I., SHAMSUDDIN, S., KANG, S. Y., BERGSTROM, R., KWON, Y. W., YU, W., WHITEHEAD, J., MUKHOPADHYAY, R., DOCQUIER, F., FARRAR, D., MORRISON, I., VIGNERON, M., WU, S. Y., CHIANG, C. M., LOUKINOV, D., LOBANENKOV, V., OHLSSON, R. & KLENOVA, E. 2007. CTCF interacts with and recruits the largest subunit of RNA polymerase II to CTCF target sites genome-wide. *Mol Cell Biol,* 27**,** 1631-48.

CHUNG, J. H., WHITELEY, M. & FELSENFELD, G. 1993. A 5' element of the chicken beta-globin domain serves as an insulator in human erythroid cells and protects against position effect in Drosophila. *Cell,* 74**,** 505-14.

CLIFFORD, G. M., SMITH, J. S., AGUADO, T. & FRANCESCHI, S. 2003. Comparison of HPV type distribution in high-grade cervical lesions and cervical cancer: a meta-analysis. *Br J Cancer,* 89**,** 101-5.

COGLIANO, V., BAAN, R., STRAIF, K., GROSSE, Y., SECRETAN, B. & EL GHISSASSI, F. 2005. Carcinogenicity of human papillomaviruses. *Lancet Oncology,* 6**,** 204-204.

COLLINS, A. S., NAKAHARA, T., DO, A. & LAMBERT, P. F. 2005. Interactions with pocket proteins contribute to the role of human papillomavirus type 16 E7 in the papillomavirus life cycle. *J Virol,* 79**,** 14769-80.

COMBITA, A. L., TOUZE, A., BOUSARGHIN, L., SIZARET, P. Y., MUNOZ, N. & COURSAGET, P. 2001. Gene transfer using human papillomavirus pseudovirions varies according to virus genotype and requires cell surface heparan sulfate. *Fems Microbiology Letters,* 204**,** 183-188.

CONGER, K. L., LIU, J. S., KUO, S. R., CHOW, L. T. & WANG, T. S. F. 1999. Human papillomavirus DNA replication - Interactions between the viral E1 protein and two subunits of human DNA polymerase alpha/primase. *Journal of Biological Chemistry,* 274**,** 2696-2705.

CONRAD, M., BUBB, V. J. & SCHLEGEL, R. 1993. The human papillomavirus type 6 and 16 E5 proteins are membrane-associated proteins which associate with the 16-kilodalton pore-forming protein. *J Virol,* 67**,** 6170-8.

CORNET, I., BOUVARD, V., CAMPO, M. S., THOMAS, M., BANKS, L., GISSMANN, L., LAMARTINE, J., SYLLA, B. S., ACCARDI, R. & TOMMASINO, M. 2012. Comparative Analysis of Transforming Properties of E6 and E7 from Different Beta Human Papillomavirus Types. *Journal of Virology,* 86**,** 2366-2370.

CRAMER, P., BUSHNELL, D. A. & KORNBERG, R. D. 2001. Structural basis of transcription: RNA polymerase II at 2.8 angstrom ngstrom resolution. *Science,* 292**,** 1863-1876.
CRUM, C., JONES, C. & KIRKPATRICK, P. 2006. Quadrivalent human papillomavirus recombinant vaccine. *Nat Rev Drug Discov,* 5**,** 629-30.
CUDDAPAH, S., JOTHI, R., SCHONES, D. E., ROH, T. Y., CUI, K. & ZHAO, K. 2009. Global analysis of the insulator binding protein CTCF in chromatin barrier regions reveals demarcation of active and repressive domains. *Genome Res,* 19**,** 24-32.
CULLEN, A. P., REID, R., CAMPION, M. & LORINCZ, A. T. 1991. Analysis of the physical state of different human papillomavirus DNAs in intraepithelial and invasive cervical neoplasm. *J Virol,* 65**,** 606-12.
CUZICK, J., ARBYN, M., SANKARANARAYANAN, R., TSU, V., RONCO, G., MAYRAND, M. H., DILLNER, J. & MEIJER, C. J. 2008. Overview of human papillomavirus-based and other novel options for cervical cancer screening in developed and developing countries. *Vaccine,* 26 Suppl 10**,** K29-41.
CZERMIN, B., MELFI, R., MCCABE, D., SEITZ, V., IMHOF, A. & PIRROTTA, V. 2002. Drosophila enhancer of Zeste/ESC complexes have a histone H3 methyltransferase activity that marks chromosomal Polycomb sites. *Cell,* 111**,** 185-96.
DANIEL, B., RANGARAJAN, A., MUKHERJEE, G., VALLIKAD, E. & KRISHNA, S. 1997. The link between integration and expression of human papillomavirus type 16 genomes and cellular changes in the evolution of cervical intraepithelial neoplastic lesions. *Journal of General Virology,* 78**,** 1095-1101.
DE ALMEIDA, S. F. & CARMO-FONSECA, M. 2008. The CTD role in cotranscriptional RNA processing and surveillance. *Febs Letters,* 582**,** 1971-1976.
DE MARTEL, C., FERLAY, J., FRANCESCHI, S., VIGNAT, J., BRAY, F., FORMAN, D. & PLUMMER, M. 2012. Global burden of cancers attributable to infections in 2008: a review and synthetic analysis. *Lancet Oncol,* 13**,** 607-15.
DE VILLIERS, E. M., FAUQUET, C., BROKER, T. R., BERNARD, H. U. & ZUR HAUSEN, H. 2004. Classification of papillomaviruses. *Virology,* 324**,** 17-27.
DEGNER, S. C., WONG, T. P., JANKEVICIUS, G. & FEENEY, A. J. 2009. Cutting edge: developmental stage-specific recruitment of cohesin to CTCF sites throughout immunoglobulin loci during B lymphocyte development. *J Immunol,* 182**,** 44-8.
DEKKER, J., RIPPE, K., DEKKER, M. & KLECKNER, N. 2002. Capturing chromosome conformation. *Science,* 295**,** 1306-1311.
DELGADO, M. D., CHERNUKHIN, I. V., BIGAS, A., KLENOVA, E. M. & LEON, J. 1999. Differential expression and phosphorylation of CTCF, a c-myc transcriptional regulator, during differentiation of human myeloid cells. *Febs Letters,* 444**,** 5-10.
DEMERET, C., YANIV, M. & THIERRY, F. 1994. The E2 transcriptional repressor can compensate for Sp1 activation of the human papillomavirus type 18 early promoter. *J Virol,* 68**,** 7075-82.
DENG, W., LIN, B. Y., JIN, G., WHEELER, C. G., MA, T., HARPER, J. W., BROKER, T. R. & CHOW, L. T. 2004. Cyclin/CDK regulates the nucleocytoplasmic localization of the human papillomavirus E1 DNA helicase. *J Virol,* 78**,** 13954-65.
DERKAY, C. S. 1995. Task force on recurrent respiratory papillomas. A preliminary report. *Arch Otolaryngol Head Neck Surg,* 121**,** 1386-91.

DIMAIO, D. & MATTOON, D. 2001. Mechanisms of cell transformation by papillomavirus E5 proteins. *Oncogene,* 20**,** 7866-73.
DIXON, J. R., SELVARAJ, S., YUE, F., KIM, A., LI, Y., SHEN, Y., HU, M., LIU, J. S. & REN, B. 2012. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature,* 485**,** 376-80.
DONG, X. P., STUBENRAUCH, F., BEYERFINKLER, E. & PFISTER, H. 1994. Prevalence of Deletions of Yy1-Binding Sites in Episomal Hpv-16 DNA from Cervical Cancers. *International Journal of Cancer,* 58**,** 803-808.
DONOHOE, M. E., ZHANG, L. F., XU, N., SHI, Y. & LEE, J. T. 2007. Identification of a Ctcf cofactor, Yy1, for the X chromosome binary switch. *Mol Cell,* 25**,** 43-56.
DOORBAR, J. 2013. The E4 protein; structure, function and patterns of expression. *Virology,* 445**,** 80-98.
DOORBAR, J., EGAWA, N., GRIFFIN, H., KRANJEC, C. & MURAKAMI, I. 2015. Human papillomavirus molecular biology and disease association. *Reviews in Medical Virology,* 25**,** 2-23.
DOORBAR, J., ELY, S., STERLING, J., MCLEAN, C. & CRAWFORD, L. 1991. Specific interaction between HPV-16 E1-E4 and cytokeratins results in collapse of the epithelial cell intermediate filament network. *Nature,* 352**,** 824-7.
DOORBAR, J., FOO, C., COLEMAN, N., MEDCALF, L., HARTLEY, O., PROSPERO, T., NAPTHINE, S., STERLING, J., WINTER, G. & GRIFFIN, H. 1997. Characterization of events during the late stages of HPV16 infection in vivo using high-affinity synthetic Fabs to E4. *Virology,* 238**,** 40-52.
DOORBAR, J., QUINT, W., BANKS, L., BRAVO, I. G., STOLER, M., BROKER, T. R. & STANLEY, M. A. 2012. The biology and life-cycle of human papillomaviruses. *Vaccine,* 30 Suppl 5**,** F55-70.
DOWHANICK, J. J., MCBRIDE, A. A. & HOWLEY, P. M. 1995. Suppression of cellular proliferation by the papillomavirus E2 protein. *J Virol,* 69**,** 7791-9.
DOYON, Y., CAYROU, C., ULLAH, M., LANDRY, A. J., COTE, V., SELLECK, W., LANE, W. S., TAN, S., YANG, X. J. & COTE, J. 2006. ING tumor suppressor proteins are critical regulators of chromatin acetylation required for genome expression and perpetuation. *Molecular Cell,* 21**,** 51-64.
DROLET, M., BENARD, E., BOILY, M. C., ALI, H., BAANDRUP, L., BAUER, H., BEDDOWS, S., BRISSON, J., BROTHERTON, J. M., CUMMINGS, T., DONOVAN, B., FAIRLEY, C. K., FLAGG, E. W., JOHNSON, A. M., KAHN, J. A., KAVANAGH, K., KJAER, S. K., KLIEWER, E. V., LEMIEUX-MELLOUKI, P., MARKOWITZ, L., MBOUP, A., MESHER, D., NICCOLAI, L., OLIPHANT, J., POLLOCK, K. G., SOLDAN, K., SONNENBERG, P., TABRIZI, S. N., TANTON, C. & BRISSON, M. 2015. Population-level impact and herd effects following human papillomavirus vaccination programmes: a systematic review and meta-analysis. *Lancet Infect Dis,* 15**,** 565-80.
DUENSING, S., LEE, L. Y., DUENSING, A., BASILE, J., PIBOONNIYOM, S., GONZALEZ, S., CRUM, C. P. & MUNGER, K. 2000. The human papillomavirus type 16 E6 and E7 oncoproteins cooperate to induce mitotic defects and genomic instability by uncoupling centrosome duplication from the cell division cycle. *Proc Natl Acad Sci U S A,* 97**,** 10002-7.

DURST, M., CROCE, C. M., GISSMANN, L., SCHWARZ, E. & HUEBNER, K. 1987. Papillomavirus sequences integrate near cellular oncogenes in some cervical carcinomas. *Proc Natl Acad Sci U S A,* 84**,** 1070-4.

DYSON, N., HOWLEY, P. M., MUNGER, K. & HARLOW, E. 1989. The human papilloma virus-16 E7 oncoprotein is able to bind to the retinoblastoma gene product. *Science,* 243**,** 934-7.

EGAWA, N., NAKAHARA, T., OHNO, S., NARISAWA-SAITO, M., YUGAWA, T., FUJITA, M., YAMATO, K., NATORI, Y. & KIYONO, T. 2012. The E1 protein of human papillomavirus type 16 is dispensable for maintenance replication of the viral genome. *J Virol,* 86**,** 3276-83.

EL AWADY, M. K., KAPLAN, J. B., O'BRIEN, S. J. & BURK, R. D. 1987. Molecular analysis of integrated human papillomavirus 16 sequences in the cervical cancer cell line SiHa. *Virology,* 159**,** 389-98.

ESSIEN, K., VIGNEAU, S., APRELEVA, S., SINGH, L. N., BARTOLOMEI, M. S. & HANNENHALLI, S. 2009. CTCF binding site classes exhibit distinct evolutionary, genomic, epigenomic and transcriptomic features. *Genome Biol,* 10**,** R131.

FAKHRY, C., WESTRA, W. H., LI, S., CMELAK, A., RIDGE, J. A., PINTO, H., FORASTIERE, A. & GILLISON, M. L. 2008. Improved survival of patients with human papillomavirus-positive head and neck squamous cell carcinoma in a prospective clinical trial. *J Natl Cancer Inst,* 100**,** 261-9.

FANELLI, M., AMATORI, S., BAROZZI, I. & MINUCCI, S. 2011. Chromatin immunoprecipitation and high-throughput sequencing from paraffin-embedded pathology tissue. *Nat Protoc,* 6**,** 1905-19.

FARRELL, C. M., WEST, A. G. & FELSENFELD, G. 2002. Conserved CTCF insulator elements flank the mouse and human beta-globin loci. *Molecular and Cellular Biology,* 22**,** 3820-3831.

FERLAY, J., FORMAN, D., MATHERS, C. D. & BRAY, F. 2012. Breast and cervical cancer in 187 countries between 1980 and 2010. *Lancet,* 379**,** 1390-1.

FERLAY, J., SHIN, H. R., BRAY, F., FORMAN, D., MATHERS, C. & PARKIN, D. M. 2010. Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. *Int J Cancer,* 127**,** 2893-917.

FILIPPOVA, G. N., QI, C. F., ULMER, J. E., MOORE, J. M., WARD, M. D., HU, Y. J., LOUKINOV, D. I., PUGACHEVA, E. M., KLENOVA, E. M., GRUNDY, P. E., FEINBERG, A. P., CLETON-JANSEN, A. M., MOERLAND, E. W., CORNELISSE, C. J., SUZUKI, H., KOMIYA, A., LINDBLOM, A., DORION-BONNET, F., NEIMAN, P. E., MORSE, H. C., 3RD, COLLINS, S. J. & LOBANENKOV, V. V. 2002. Tumor-associated zinc finger mutations in the CTCF transcription factor selectively alter tts DNA-binding specificity. *Cancer Res,* 62**,** 48-52.

FILIPPOVA, M., EVANS, W., ARAGON, R., FILIPPOV, V., WILLIAMS, V. M., HONG, L., REEVES, M. E. & DUERKSEN-HUGHES, P. 2014. The small splice variant of HPV16 E6, E6, reduces tumor formation in cervical carcinoma xenografts. *Virology,* 450-451**,** 153-164.

FILIPPOVA, M., JOHNSON, M. M., BAUTISTA, M., FILIPPOV, V., FODOR, N., TUNGTEAKKHUN, S. S., WILLIAMS, K. & DUERKSEN-HUGHES, P. J. 2007. The large and small isoforms of human papillomavirus type 16 E6 bind to and differentially affect procaspase 8 stability and activity. *J Virol,* 81**,** 4116-29.

FIRZLAFF, J. M., LUSCHER, B. & EISENMAN, R. N. 1991. Negative Charge at the Casein Kinase-Ii Phosphorylation Site Is Important for Transformation but Not for Rb Protein-Binding by the E7 Protein of Human Papillomavirus Type-16. *Proceedings of the National Academy of Sciences of the United States of America,* 88**,** 5187-5191.

FLAVAHAN, W. A., DRIER, Y., LIAU, B. B., GILLESPIE, S. M., VENTEICHER, A. S., STEMMER-RACHAMIMOV, A. O., SUVA, M. L. & BERNSTEIN, B. E. 2016. Insulator dysfunction and oncogene activation in IDH mutant gliomas. *Nature,* 529**,** 110-4.

FONG, N. & BENTLEY, D. L. 2001. Capping, splicing, and 3' processing are independently stimulated by RNA polymerase II: different functions for different segments of the CTD. *Genes Dev,* 15**,** 1783-95.

FRANCESCHI, S. & DE VUYST, H. 2009. Human papillomavirus vaccines and anal carcinoma. *Curr Opin HIV AIDS,* 4**,** 57-63.

FRISCH, M., HJALGRIM, H., JAEGER, A. B. & BIGGAR, R. J. 2000. Changing patterns of tonsillar squamous cell carcinoma in the United States. *Cancer Causes Control,* 11**,** 489-95.

FRITSCH, C., BROWN, J. L., KASSIS, J. A. & MULLER, J. 1999. The DNA-binding polycomb group protein pleiohomeotic mediates silencing of a Drosophila homeotic gene. *Development,* 126**,** 3905-13.

FROLOV, M. V. & DYSON, N. J. 2004. Molecular mechanisms of E2F-dependent activation and pRB-mediated repression. *J Cell Sci,* 117**,** 2173-81.

FU, Y. T., SINHA, M., PETERSON, C. L. & WENG, Z. P. 2008. The Insulator Binding Protein CTCF Positions 20 Nucleosomes around Its Binding Sites across the Human Genome. *Plos Genetics,* 4.

GARLAND, S. M., HERNANDEZ-AVILA, M., WHEELER, C. M., PEREZ, G., HARPER, D. M., LEODOLTER, S., TANG, G. W., FERRIS, D. G., STEBEN, M., BRYAN, J., TADDEO, F. J., RAILKAR, R., ESSER, M. T., SINGS, H. L., NELSON, M., BOSLEGO, J., SATTLER, C., BARR, E., KOUTSKY, L. A. & FEMALES UNITED TO UNILATERALLY REDUCE ENDO/ECTOCERVICAL DISEASE, I. I. 2007. Quadrivalent vaccine against human papillomavirus to prevent anogenital diseases. *N Engl J Med,* 356**,** 1928-43.

GENOVESE, N. J., BANERJEE, N. S., BROKER, T. R. & CHOW, L. T. 2008. Casein kinase II motif-dependent phosphorylation of human papillomavirus E7 protein promotes p130 degradation and S-phase induction in differentiated human keratinocytes. *J Virol,* 82**,** 4862-73.

GENTHER WILLIAMS, S. M., DISBROW, G. L., SCHLEGEL, R., LEE, D., THREADGILL, D. W. & LAMBERT, P. F. 2005. Requirement of epidermal growth factor receptor for hyperplasia induced by E5, a high-risk human papillomavirus oncogene. *Cancer Res,* 65**,** 6534-42.

GEORGOPOULOS, N. T., PROFFITT, J. L. & BLAIR, G. E. 2000. Transcriptional regulation of the major histocompatibility complex (MHC) class I heavy chain, TAP1 and LMP2 genes by the human papillomavirus (HPV) type 6b, 16 and 18 E7 oncoproteins. *Oncogene,* 19**,** 4930-5.

GIRESI, P. G., KIM, J., MCDANIELL, R. M., IYER, V. R. & LIEB, J. D. 2007. FAIRE (Formaldehyde-Assisted Isolation of Regulatory Elements) isolates active regulatory elements from human chromatin. *Genome Res,* 17**,** 877-85.

GIROGLOU, T., FLORIN, L., SCHAFER, F., STREECK, R. E. & SAPP, M. 2001. Human papillomavirus infection requires cell surface heparan sulfate. *J Virol,* 75**,** 1565-70.

GISSMANN, L., WOLNIK, L., IKENBERG, H., KOLDOVSKY, U., SCHNURCH, H. G. & ZUR HAUSEN, H. 1983. Human papillomavirus types 6 and 11 DNA sequences in genital and laryngeal papillomas and in some cervical cancers. *Proc Natl Acad Sci U S A,* 80**,** 560-3.

GIULIANO, A. R., PALEFSKY, J. M., GOLDSTONE, S., MOREIRA, E. D., JR., PENNY, M. E., ARANDA, C., VARDAS, E., MOI, H., JESSEN, H., HILLMAN, R., CHANG, Y. H., FERRIS, D., ROULEAU, D., BRYAN, J., MARSHALL, J. B., VUOCOLO, S., BARR, E., RADLEY, D., HAUPT, R. M. & GURIS, D. 2011. Efficacy of quadrivalent HPV vaccine against HPV Infection and disease in males. *N Engl J Med,* 364**,** 401-11.

GLAXOSMITHKLINE VACCINE, H. P. V. S. G., ROMANOWSKI, B., DE BORBA, P. C., NAUD, P. S., ROTELI-MARTINS, C. M., DE CARVALHO, N. S., TEIXEIRA, J. C., AOKI, F., RAMJATTAN, B., SHIER, R. M., SOMANI, R., BARBIER, S., BLATTER, M. M., CHAMBERS, C., FERRIS, D., GALL, S. A., GUERRA, F. A., HARPER, D. M., HEDRICK, J. A., HENRY, D. C., KORN, A. P., KROLL, R., MOSCICKI, A. B., ROSENFELD, W. D., SULLIVAN, B. J., THOMING, C. S., TYRING, S. K., WHEELER, C. M., DUBIN, G., SCHUIND, A., ZAHAF, T., GREENACRE, M. & SGRIOBHADAIR, A. 2009. Sustained efficacy and immunogenicity of the human papillomavirus (HPV)-16/18 AS04-adjuvanted vaccine: analysis of a randomised placebo-controlled trial up to 6.4 years. *Lancet,* 374**,** 1975-85.

GOH, Y. F., FULLWOOD, M. J., POH, H. M., PEH, S. Q., ONG, C. T., ZHANG, J. Y., RUAN, X. A. & RUAN, Y. J. 2012. Chromatin Interaction Analysis with Paired-End Tag Sequencing (ChIA-PET) for Mapping Chromatin Interactions and Understanding Transcription Regulation. *Jove-Journal of Visualized Experiments*.

GOLDSTROHM, A. C., ALBRECHT, T. R., SUNE, C., BEDFORD, M. T. & GARCIA-BLANCO, M. A. 2001a. The transcription elongation factor CA150 interacts with RNA polymerase II and the pre-mRNA splicing factor SF1. *Mol Cell Biol,* 21**,** 7617-28.

GOLDSTROHM, A. C., GREENLEAF, A. L. & GARCIA-BLANCO, M. A. 2001b. Co-transcriptional splicing of pre-messenger RNAs: considerations for the mechanism of alternative splicing. *Gene,* 277**,** 31-47.

GOMEZ-DIAZ, E. & CORCES, V. G. 2014. Architectural proteins: regulators of 3D genome organization in cell fate. *Trends Cell Biol*.

GORYSHIN, I. Y. & REZNIKOFF, W. S. 1998. Tn5 in vitro transposition. *J Biol Chem,* 273**,** 7367-74.

GRAHAM, S. V. 2010. Human papillomavirus: gene expression, regulation and prospects for novel diagnostic methods and antiviral therapies. *Future Microbiol,* 5**,** 1493-506.

GRASSMANN, K., RAPP, B., MASCHEK, H., PETRY, K. U. & IFTNER, T. 1996. Identification of a differentiation-inducible promoter in the E7 open reading frame of human papillomavirus type 16 (HPV-16) in raft cultures of a new cell line containing high copy numbers of episomal HPV-16 DNA. *J Virol,* 70**,** 2339-49.

GRULICH, A. E., JIN, F., CONWAY, E. L., STEIN, A. N. & HOCKING, J. 2010. Cancers attributable to human papillomavirus infection. *Sex Health,* 7**,** 244-52.

GUASTAFIERRO, T., CECCHINELLI, B., ZAMPIERI, M., REALE, A., RIGGIO, G., STHANDIER, O., ZUPI, G., CALABRESE, L. & CAIAFA, P. 2008. CCCTC-binding factor activates PARP-1 affecting DNA methylation machinery. *J Biol Chem,* 283**,** 21873-80.

GUNTHER, T., THEISS, J. M., FISCHER, N. & GRUNDHOFF, A. 2016. Investigation of Viral and Host Chromatin by ChIP-PCR or ChIP-Seq Analysis. *Curr Protoc Microbiol,* 40**,** 1E 10 1-21.

GUO, Y., XU, Q., CANZIO, D., SHOU, J., LI, J., GORKIN, D. U., JUNG, I., WU, H., ZHAI, Y., TANG, Y., LU, Y., WU, Y., JIA, Z., LI, W., ZHANG, M. Q., REN, B., KRAINER, A. R., MANIATIS, T. & WU, Q. 2015. CRISPR Inversion of CTCF Sites Alters Genome Topology and Enhancer/Promoter Function. *Cell,* 162**,** 900-10.

HAEDICKE, J. & IFTNER, T. 2013. Human papillomaviruses and cancer. *Radiotherapy and oncology : journal of the European Society for Therapeutic Radiology and Oncology,* 108**,** 397-402.

HAMID, N. A., BROWN, C. & GASTON, K. 2009. The regulation of cell proliferation by the papillomavirus early proteins. *Cell Mol Life Sci,* 66**,** 1700-17.

HAN, Z., XING, X., HU, M., ZHANG, Y., LIU, P. & CHAI, J. 2007. Structural basis of EZH2 recognition by EED. *Structure,* 15**,** 1306-15.

HANDOKO, L., XU, H., LI, G., NGAN, C. Y., CHEW, E., SCHNAPP, M., LEE, C. W., YE, C., PING, J. L., MULAWADI, F., WONG, E., SHENG, J., ZHANG, Y., POH, T., CHAN, C. S., KUNARSO, G., SHAHAB, A., BOURQUE, G., CACHEUX-RATABOUL, V., SUNG, W. K., RUAN, Y. & WEI, C. L. 2011. CTCF-mediated functional chromatin interactome in pluripotent cells. *Nat Genet,* 43**,** 630-8.

HARK, A. T., SCHOENHERR, C. J., KATZ, D. J., INGRAM, R. S., LEVORSE, J. M. & TILGHMAN, S. M. 2000. CTCF mediates methylation-sensitive enhancer-blocking activity at the H19/Igf2 locus. *Nature,* 405**,** 486-9.

HAWLEY-NELSON, P., VOUSDEN, K. H., HUBBERT, N. L., LOWY, D. R. & SCHILLER, J. T. 1989. HPV16 E6 and E7 proteins cooperate to immortalize human foreskin keratinocytes. *EMBO J,* 8**,** 3905-10.

HERFS, M., YAMAMOTO, Y., LAURY, A., WANG, X., NUCCI, M. R., MCLAUGHLIN-DRUBIN, M. E., MUNGER, K., FELDMAN, S., MCKEON, F. D., XIAN, W. & CRUM, C. P. 2012. A discrete population of squamocolumnar junction cells implicated in the pathogenesis of cervical cancer. *Proceedings of the National Academy of Sciences of the United States of America,* 109**,** 10516-10521.

HILDESHEIM, A. & HERRERO, R. 2007. Human papillomavirus vaccine should be given before sexual debut for maximum benefit. *J Infect Dis,* 196**,** 1431-2.

HILDESHEIM, A., HERRERO, R., WACHOLDER, S., RODRIGUEZ, A. C., SOLOMON, D., BRATTI, M. C., SCHILLER, J. T., GONZALEZ, P., DUBIN, G., PORRAS, C., JIMENEZ, S. E., LOWY, D. R. & COSTA RICAN, H. P. V. V. T. G. 2007. Effect of human papillomavirus 16/18 L1 viruslike particle vaccine among young women with preexisting infection: a randomized trial. *JAMA,* 298**,** 743-53.

HILLER, T., POPPELREUTHER, S., STUBENRAUCH, F. & IFTNER, T. 2006. Comparative analysis of 19 genital human papillomavirus types with regard to p53 degradation, immortalization, phylogeny, and epidemiologic risk classification. *Cancer Epidemiol Biomarkers Prev,* 15**,** 1262-7.

HNISZ, D., WEINTRAUB, A. S., DAY, D. S., VALTON, A. L., BAK, R. O., LI, C. H., GOLDMANN, J., LAJOIE, B. R., FAN, Z. P., SIGOVA, A. A., REDDY, J., BORGES-

RIVERA, D., LEE, T. I., JAENISCH, R., PORTEUS, M. H., DEKKER, J. & YOUNG, R. A. 2016. Activation of proto-oncogenes by disruption of chromosome neighborhoods. *Science,* 351**,** 1454-1458.
HOLOHAN, E. E., KWONG, C., ADRYAN, B., BARTKUHN, M., HEROLD, M., RENKAWITZ, R., RUSSELL, S. & WHITE, R. 2007. CTCF genomic binding sites in Drosophila and the organisation of the bithorax complex. *Plos Genetics,* 3**,** 1211-1222.
HOLWERDA, S. J. & DE LAAT, W. 2013. CTCF: the protein, the binding partners, the binding sites and their chromatin loops. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences,* 368**,** 20120369.
HOWE, K. J. 2002. RNA polymerase II conducts a symphony of pre-mRNA processing activities. *Biochim Biophys Acta,* 1577**,** 308-24.
HSU, S. C., GILGENAST, T. G., BARTMAN, C. R., EDWARDS, C. R., STONESTROM, A. J., HUANG, P., EMERSON, D. J., EVANS, P., WERNER, M. T., KELLER, C. A., GIARDINE, B., HARDISON, R. C., RAJ, A., PHILLIPS-CREMINS, J. E. & BLOBEL, G. A. 2017. The BET Protein BRD2 Cooperates with CTCF to Enforce Transcriptional and Architectural Boundaries. *Mol Cell,* 66**,** 102-116 e7.
HUBERT, P., HERMAN, L., RONCARATI, P., MAILLARD, C., RENOUX, V., DEMOULIN, S., ERPICUM, C., FOIDART, J. M., BONIVER, J., NOEL, A., DELVENNE, P. & HERFS, M. 2014. Altered alpha-defensin 5 expression in cervical squamocolumnar junction: implication in the formation of a viral/tumour-permissive microenvironment. *J Pathol,* 234**,** 464-77.
HUBERT, P., VAN DEN BRULE, F., GIANNINI, S. L., FRANZEN-DETROOZ, E., BONIVER, J. & DELVENNE, P. 1999. Colonization of in vitro-formed cervical human papillomavirus- associated (pre)neoplastic lesions with dendritic cells: role of granulocyte/macrophage colony-stimulating factor. *Am J Pathol,* 154**,** 775-84.
HUDELIST, G., MANAVI, M., PISCHINGER, K. I., WATKINS-RIEDEL, T., SINGER, C. F., KUBISTA, E. & CZERWENKA, K. F. 2004. Physical state and expression of HPV DNA in benign and dysplastic cervical tissue: different levels of viral integration are correlated with lesion grade. *Gynecol Oncol,* 92**,** 873-80.
HUMANS, I. W. G. O. T. E. O. C. R. T. 2007. Human papillomaviruses. *IARC Monogr Eval Carcinog Risks Hum,* 90**,** 1-636.
HUMMEL, M., HUDSON, J. B. & LAIMINS, L. A. 1992. Differentiation-induced and constitutive transcription of human papillomavirus type 31b in cell lines containing viral episomes. *J Virol,* 66**,** 6070-80.
HYLAND, P. L., MCDADE, S. S., MCCLOSKEY, R., DICKSON, G. J., ARTHUR, K., MCCANCE, D. J. & PATEL, D. 2011. Evidence for alteration of EZH2, BMI1, and KDM6A and epigenetic reprogramming in human papillomavirus type 16 E6/E7-expressing keratinocytes. *J Virol,* 85**,** 10999-1006.
ISHIJI, T., LACE, M. J., PARKKINEN, S., ANDERSON, R. D., HAUGEN, T. H., CRIPE, T. P., XIAO, J. H., DAVIDSON, I., CHAMBON, P. & TUREK, L. P. 1992a. Transcriptional enhancer factor (TEF)-1 and its cell-specific co-activator activate human papillomavirus-16 E6 and E7 oncogene transcription in keratinocytes and cervical carcinoma cells. *EMBO J,* 11**,** 2271-81.
ISHIJI, T., LACE, M. J., PARKKINEN, S., ANDERSON, R. D., HAUGEN, T. H., CRIPE, T. P., XIAO, J. H., DAVIDSON, I., CHAMBON, P. & TUREK, L. P. 1992b. Transcriptional Enhancer Factor (Tef)-1 and Its Cell-Specific Co-Activator Activate Human

Papillomavirus-16 E6 and E7 Oncogene Transcription in Keratinocytes and Cervical-Carcinoma Cells. *Embo Journal,* 11**,** 2271-2281.
JACKSON, S., HARWOOD, C., THOMAS, M., BANKS, L. & STOREY, A. 2000. Role of Bak in UV-induced apoptosis in skin cancer and abrogation by HPV E6 proteins. *Genes Dev,* 14**,** 3065-73.
JEON, S., ALLENHOFFMANN, B. L. & LAMBERT, P. F. 1995. Integration of Human Papillomavirus Type-16 into the Human Genome Correlates with a Selective Growth Advantage of Cells. *Journal of Virology,* 69**,** 2989-2997.
JIA, R., LIU, X., TAO, M., KRUHLAK, M., GUO, M., MEYERS, C., BAKER, C. C. & ZHENG, Z. M. 2009. Control of the papillomavirus early-to-late switch by differentially expressed SRp20. *J Virol,* 83**,** 167-80.
JOHANNSEN, E. & LAMBERT, P. F. 2013. Epigenetics of human papillomaviruses. *Virology,* 445**,** 205-212.
JOHANSSON, C. & SCHWARTZ, S. 2013. Regulation of human papillomavirus gene expression by splicing and polyadenylation. *Nat Rev Microbiol,* 11**,** 239-51.
JOHANSSON, C., SOMBERG, M., LI, X., BACKSTROM WINQUIST, E., FAY, J., RYAN, F., PIM, D., BANKS, L. & SCHWARTZ, S. 2012. HPV-16 E2 contributes to induction of HPV-16 late gene expression by inhibiting early polyadenylation. *EMBO J,* 31**,** 3212-27.
JONES, D. L., ALANI, R. M. & MUNGER, K. 1997. The human papillomavirus E7 oncoprotein can uncouple cellular differentiation and proliferation in human keratinocytes by abrogating p21Cip1-mediated inhibition of cdk2. *Genes Dev,* 11**,** 2101-11.
JOSHI, A. A. & STRUHL, K. 2005. Eaf3 chromodomain interaction with methylated H3-K36 links histone deacetylation to Pol II elongation. *Molecular Cell,* 20**,** 971-978.
JOYCE, J. G., TUNG, J. S., PRZYSIECKI, C. T., COOK, J. C., LEHMAN, E. D., SANDS, J. A., JANSENS, K. U. & KELLER, P. M. 1999. The L1 major capsid protein of human papillomavirus type 11 recombinant virus-like particles interacts with heparin and cell-surface glycosaminoglycans on human keratinocytes. *Journal of Biological Chemistry,* 274**,** 5810-5822.
KANDURI, C., PANT, V., LOUKINOV, D., PUGACHEVA, E., QI, C. F., WOLFFE, A., OHLSSON, R. & LOBANENKOV, V. V. 2000. Functional association of CTCF with the insulator upstream of the H19 gene is parent of origin-specific and methylation-sensitive. *Curr Biol,* 10**,** 853-6.
KANG, H., CHO, H., SUNG, G. H. & LIEBERMAN, P. M. 2013. CTCF regulates Kaposi's sarcoma-associated herpesvirus latency transcription by nucleosome displacement and RNA polymerase programming. *Journal of virology,* 87**,** 1789-99.
KANG, H. & LIEBERMAN, P. M. 2009. Cell Cycle Control of Kaposi's Sarcoma-Associated Herpesvirus Latency Transcription by CTCF-Cohesin Interactions. *Journal of Virology,* 83**,** 6199-6210.
KANG, H., WIEDMER, A., YUAN, Y., ROBERTSON, E. & LIEBERMAN, P. M. 2011. Coordination of KSHV latent and lytic gene control by CTCF-cohesin mediated chromosome conformation. *PLoS pathogens,* 7**,** e1002140.
KATAINEN, R., DAVE, K., PITKANEN, E., PALIN, K., KIVIOJA, T., VALIMAKI, N., GYLFE, A. E., RISTOLAINEN, H., HANNINEN, U. A., CAJUSO, T., KONDELIN, J.,

TANSKANEN, T., MECKLIN, J. P., JARVINEN, H., RENKONEN-SINISALO, L., LEPISTO, A., KAASINEN, E., KILPIVAARA, O., TUUPANEN, S., ENGE, M., TAIPALE, J. & AALTONEN, L. A. 2015. CTCF/cohesin-binding sites are frequently mutated in cancer. *Nat Genet,* 47**,** 818-21.

KEOGH, M. C., KURDISTANI, S. K., MORRIS, S. A., AHN, S. H., PODOLNY, V., COLLINS, S. R., SCHULDINER, M., CHIN, K. Y., PUNNA, T., THOMPSON, N. J., BOONE, C., EMILI, A., WEISSMAN, J. S., HUGHES, T. R., STRAHL, B. D., GRUNSTEIN, M., GREENBLATT, J. F., BURATOWSKI, S. & KROGAN, N. J. 2005. Cotranscriptional Set2 methylation of histone H3 lysine 36 recruits a repressive Rpd3 complex. *Cell,* 123**,** 593-605.

KIM, K., GARNER-HAMRICK, P. A., FISHER, C., LEE, D. & LAMBERT, P. F. 2003. Methylation patterns of papillomavirus DNA, its influence on E2 function, and implications in viral infection. *Journal of Virology,* 77**,** 12450-12459.

KIM, T. H., ABDULLAEV, Z. K., SMITH, A. D., CHING, K. A., LOUKINOV, D. I., GREEN, R. D., ZHANG, M. Q., LOBANENKOV, V. V. & REN, B. 2007. Analysis of the vertebrate insulator protein CTCF-binding sites in the human genome. *Cell,* 128**,** 1231-45.

KIM, Y. J., CECCHINI, K. R. & KIM, T. H. 2011. Conserved, developmentally regulated mechanism couples chromosomal looping and heterochromatin barrier activity at the homeobox gene A locus (vol 108, pg 7391, 2011). *Proceedings of the National Academy of Sciences of the United States of America,* 108**,** 19096-19096.

KIYONO, T., HIRAIWA, A., FUJITA, M., HAYASHI, Y., AKIYAMA, T. & ISHIBASHI, M. 1997. Binding of high-risk human papillomavirus E6 oncoproteins to the human homologue of the Drosophila discs large tumor suppressor protein. *Proceedings of the National Academy of Sciences of the United States of America,* 94**,** 11612-11616.

KLINGELHUTZ, A. J., FOSTER, S. A. & MCDOUGALL, J. K. 1996. Telomerase activation by the E6 gene product of human papillomavirus type 16. *Nature,* 380**,** 79-82.

KLINGELHUTZ, A. J. & ROMAN, A. 2012. Cellular transformation by human papillomaviruses: lessons learned by comparing high- and low-risk viruses. *Virology,* 424**,** 77-98.

KOMARNITSKY, P., CHO, E. J. & BURATOWSKI, S. 2000. Different phosphorylated forms of RNA polymerase II and associated mRNA processing factors during transcription. *Genes Dev,* 14**,** 2452-60.

KOMATSU, T., SEKIYA, T. & NAGATA, K. 2013. DNA replication-dependent binding of CTCF plays a critical role in adenovirus genome functions. *Sci Rep,* 3**,** 2187.

KOUTSKY, L. A., AULT, K. A., WHEELER, C. M., BROWN, D. R., BARR, E., ALVAREZ, F. B., CHIACCHIERINI, L. M., JANSEN, K. U. & PROOF OF PRINCIPLE STUDY, I. 2002. A controlled trial of a human papillomavirus type 16 vaccine. *N Engl J Med,* 347**,** 1645-51.

KOUZARIDES, T. 2007. Chromatin modifications and their function. *Cell,* 128**,** 693-705.

KRAWCZYK, E., SUPRYNOWICZ, F. A., SUDARSHAN, S. R. & SCHLEGEL, R. 2010. Membrane orientation of the human papillomavirus type 16 E5 oncoprotein. *J Virol,* 84**,** 1696-703.

KREIMER, A. R., CLIFFORD, G. M., BOYLE, P. & FRANCESCHI, S. 2005. Human papillomavirus types in head and neck squamous cell carcinomas worldwide: a systematic review. *Cancer Epidemiol Biomarkers Prev,* 14**,** 467-75.

KROGAN, N. J., KIM, M., TONG, A., GOLSHANI, A., CAGNEY, G., CANADIEN, V., RICHARDS, D. P., BEATTIE, B. K., EMILI, A., BOONE, C., SHILATIFARD, A., BURATOWSKI, S. & GREENBLATT, J. 2003. Methylation of histone H3 by Set2 in Saccharomyces cerevisiae is linked to transcriptional elongation by RNA polymerase II. *Mol Cell Biol,* 23**,** 4207-18.

KURUKUTI, S., TIWARI, V. K., TAVOOSIDANA, G., PUGACHEVA, E., MURRELL, A., ZHAO, Z. H., LOBANENKOV, V., REIK, W. & OHLSSON, R. 2006. CTCF binding at the H19 imprinting control region mediates maternally inherited higher-order chromatin conformation to restrict enhancer access to Igf2. *Proceedings of the National Academy of Sciences of the United States of America,* 103**,** 10684-10689.

KUZMICHEV, A., NISHIOKA, K., ERDJUMENT-BROMAGE, H., TEMPST, P. & REINBERG, D. 2002. Histone methyltransferase activity associated with a human multiprotein complex containing the Enhancer of Zeste protein. *Genes & Development,* 16**,** 2893-2905.

LACE, M. J., ANSON, J. R., THOMAS, G. S., TUREK, L. P. & HAUGEN, T. H. 2008. The E8 boolean AND E2 Gene Product of Human Papillomavirus Type 16 Represses Early Transcription and Replication but Is Dispensable for Viral Plasmid Persistence in Keratinocytes. *Journal of Virology,* 82**,** 10841-10853.

LACE, M. J., YAMAKAWA, Y., USHIKAI, M., ANSON, J. R., HAUGEN, T. H. & TUREK, L. P. 2009. Cellular factor YY1 downregulates the human papillomavirus 16 E6/E7 promoter, P97, in vivo and in vitro from a negative element overlapping the transcription-initiation site. *J Gen Virol,* 90**,** 2402-12.

LAMBERT, P. F., SPALHOLZ, B. A. & HOWLEY, P. M. 1987. A transcriptional repressor encoded by BPV-1 shares a common carboxy-terminal domain with the E2 transactivator. *Cell,* 50**,** 69-78.

LANG, F., LI, X., VLADIMIROVA, O., HU, B., CHEN, G., XIAO, Y., SINGH, V., LU, D., LI, L., HAN, H., WICKRAMASINGHE, J. M., SMITH, S. T., ZHENG, C., LI, Q., LIEBERMAN, P. M., FRASER, N. W. & ZHOU, J. 2017. CTCF interacts with the lytic HSV-1 genome to promote viral transcription. *Sci Rep,* 7**,** 39861.

LECHNER, M. S. & LAIMINS, L. A. 1994. Inhibition of p53 DNA binding by human papillomavirus E6 proteins. *J Virol,* 68**,** 4262-73.

LEE, C. K., SHIBATA, Y., RAO, B., STRAHL, B. D. & LIEB, J. D. 2004. Evidence for nucleosome depletion at active regulatory regions genome-wide. *Nature Genetics,* 36**,** 900-905.

LEE, S. S., WEISS, R. S. & JAVIER, R. T. 1997. Binding of human virus oncoproteins to hDlg/SAP97, a mammalian homolog of the Drosophila discs large tumor suppressor protein. *Proc Natl Acad Sci U S A,* 94**,** 6670-5.

LI, B., CAREY, M. & WORKMAN, J. L. 2007. The role of chromatin during transcription. *Cell,* 128**,** 707-19.

LI, H. T., ILIN, S., WANG, W. K., DUNCAN, E. M., WYSOCKA, J., ALLIS, C. D. & PATEL, D. J. 2006. Molecular basis for site-specific read-out of histone H3K4me3 by the BPTF PHD finger of NURF. *Nature,* 442**,** 91-95.

LI, J., MOAZED, D. & GYGI, S. P. 2002. Association of the histone methyltransferase Set2 with RNA polymerase II plays a role in transcription elongation. *J Biol Chem,* 277**,** 49383-8.
LI, T., HU, J. F., QIU, X., LING, J., CHEN, H., WANG, S., HOU, A., VU, T. H. & HOFFMAN, A. R. 2008. CTCF regulates allelic expression of Igf2 by orchestrating a promoter-polycomb repressive complex 2 intrachromosomal loop. *Mol Cell Biol,* 28**,** 6473-82.
LI, X. & COFFINO, P. 1996. High-risk human papillomavirus E6 protein has two distinct binding sites within p53, of which only one determines degradation. *J Virol,* 70**,** 4509-16.
LIM, D. A., GOSSEN, M., LEHMAN, C. W. & BOTCHAN, M. R. 1998. 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,* 72**,** 1931-1940.
LING, J. Q., LI, T., HU, J. F., VU, T. H., CHEN, H. L., QIU, X. W., CHERRY, A. M. & HOFFMAN, A. R. 2006. CTCF mediates interchromosomal colocalization between Igf2/H19 and Wsb1/Nf1. *Science,* 312**,** 269-272.
LOBANENKOV, V. V., NICOLAS, R. H., ADLER, V. V., PATERSON, H., KLENOVA, E. M., POLOTSKAJA, A. V. & GOODWIN, G. H. 1990. 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,* 5**,** 1743-53.
LONGWORTH, M. S., WILSON, R. & LAIMINS, L. A. 2005. HPV31 E7 facilitates replication by activating E2F2 transcription through its interaction with HDACs. *Embo Journal,* 24**,** 1821-1830.
LOWY, D. R. & SCHILLER, J. T. 2006. Prophylactic human papillomavirus vaccines. *J Clin Invest,* 116**,** 1167-73.
LUCO, R. F., PAN, Q., TOMINAGA, K., BLENCOWE, B. J., PEREIRA-SMITH, O. M. & MISTELI, T. 2010. Regulation of alternative splicing by histone modifications. *Science,* 327**,** 996-1000.
LUSKY, M., HURWITZ, J. & SEO, Y. S. 1994. The Bovine Papillomavirus E2 Protein Modulates the Assembly of but Is Not Stably Maintained in a Replication-Competent Multimeric E1-Replication Origin Complex. *Proceedings of the National Academy of Sciences of the United States of America,* 91**,** 8895-8899.
MA, T., ZOU, N., LIN, B. Y., CHOW, L. T. & HARPER, J. W. 1999. Interaction between cyclin-dependent kinases and human papillomavirus replication-initiation protein E1 is required for efficient viral replication. *Proc Natl Acad Sci U S A,* 96**,** 382-7.
MACALUSO, M., MONTANARI, M. & GIORDANO, A. 2006. Rb family proteins as modulators of gene expression and new aspects regarding the interaction with chromatin remodeling enzymes. *Oncogene,* 25**,** 5263-7.
MACLEAN, J., KOEKEMOER, M., OLIVIER, A. J., STEWART, D., HITZEROTH, II, RADEMACHER, T., FISCHER, R., WILLIAMSON, A. L. & RYBICKI, E. P. 2007. Optimization of human papillomavirus type 16 (HPV-16) L1 expression in plants: comparison of the suitability of different HPV-16 L1 gene variants and different cell-compartment localization. *J Gen Virol,* 88**,** 1460-9.

MAJUMDER, P., GOMEZ, J. A., CHADWICK, B. P. & BOSS, J. M. 2008. The insulator factor CTCF controls MHC class II gene expression and is required for the formation of long-distance chromatin interactions. *J Exp Med,* 205**,** 785-98.
MARGUERON, R., JUSTIN, N., OHNO, K., SHARPE, M. L., SON, J., DRURY, W. J., 3RD, VOIGT, P., MARTIN, S. R., TAYLOR, W. R., DE MARCO, V., PIRROTTA, V., REINBERG, D. & GAMBLIN, S. J. 2009. Role of the polycomb protein EED in the propagation of repressive histone marks. *Nature,* 461**,** 762-7.
MARIANI, L. & VENUTI, A. 2010. HPV vaccine: an overview of immune response, clinical protection, and new approaches for the future. *J Transl Med,* 8**,** 105.
MARTINEZ, F. P., CRUZ, R., LU, F., PLASSCHAERT, R., DENG, Z., RIVERA-MOLINA, Y. A., BARTOLOMEI, M. S., LIEBERMAN, P. M. & TANG, Q. 2014. CTCF Binding to the First Intron of the Major Immediate-Early (MIE) Gene of Human Cytomegalovirus (HCMV) Negatively Regulates MIE Gene Expression and HCMV Replication. *J Virol*.
MARTINEZ-ZAPIEN, D., RUIZ, F. X., POIRSON, J., MITSCHLER, A., RAMIREZ, J., FORSTER, A., COUSIDO-SIAH, A., MASSON, M., VANDE POL, S., PODJARNY, A., TRAVE, G. & ZANIER, K. 2016. Structure of the E6/E6AP/p53 complex required for HPV-mediated degradation of p53. *Nature,* 529**,** 541-5.
MARUR, S., D'SOUZA, G., WESTRA, W. H. & FORASTIERE, A. A. 2010. HPV-associated head and neck cancer: a virus-related cancer epidemic. *Lancet Oncol,* 11**,** 781-9.
MAURANO, M. T., WANG, H., JOHN, S., SHAFER, A., CANFIELD, T., LEE, K. & STAMATOYANNOPOULOS, J. A. 2015. Role of DNA Methylation in Modulating Transcription Factor Occupancy. *Cell Reports,* 12**,** 1184-1195.
MAY, M., DONG, X. P., BEYER-FINKLER, E., STUBENRAUCH, F., FUCHS, P. G. & PFISTER, H. 1994. The E6/E7 promoter of extrachromosomal HPV16 DNA in cervical cancers escapes from cellular repression by mutation of target sequences for YY1. *EMBO J,* 13**,** 1460-6.
MCBRIDE, A. A. 2008. Replication and partitioning of papillomavirus genomes. *Adv Virus Res,* 72**,** 155-205.
MCBRIDE, A. A. 2013. The papillomavirus E2 proteins. *Virology,* 445**,** 57-79.
MCINTOSH, P. B., LASKEY, P., SULLIVAN, K., DAVY, C., WANG, Q., JACKSON, D. J., GRIFFIN, H. M. & DOORBAR, J. 2010. E1--E4-mediated keratin phosphorylation and ubiquitylation: a mechanism for keratin depletion in HPV16-infected epithelium. *J Cell Sci,* 123**,** 2810-22.
MCLAUGHLIN-DRUBIN, M. E., CRUM, C. P. & MUNGER, K. 2011. Human papillomavirus E7 oncoprotein induces KDM6A and KDM6B histone demethylase expression and causes epigenetic reprogramming. *Proc Natl Acad Sci U S A,* 108**,** 2130-5.
MCLAUGHLIN-DRUBIN, M. E., HUH, K. W. & MUNGER, K. 2008. Human papillomavirus type 16 E7 oncoprotein associates with E2F6. *Journal of Virology,* 82**,** 8695-8705.
MCLAUGHLIN-DRUBIN, M. E. & MUNGER, K. 2009. The human papillomavirus E7 oncoprotein. *Virology,* 384**,** 335-44.
MEHTA, K., GUNASEKHARAN, V., SATSUKA, A. & LAIMINS, L. A. 2015. Human papillomaviruses activate and recruit SMC1 cohesin proteins for the

differentiation-dependent life cycle through association with CTCF insulators. *PLoS Pathog,* 11**,** e1004763.
MERCK 2014. HPV Vaccine Works against Nine Viral Type. *Cancer Discov,* 4.
MIDDLETON, K., PEH, W., SOUTHERN, S., GRIFFIN, H., SOTLAR, K., NAKAHARA, T., EL-SHERIF, A., MORRIS, L., SETH, R., HIBMA, M., JENKINS, D., LAMBERT, P., COLEMAN, N. & DOORBAR, J. 2003. Organization of human papillomavirus productive cycle during neoplastic progression provides a basis for selection of diagnostic markers. *J Virol,* 77**,** 10186-201.
MIKKELSEN, T. S., KU, M., JAFFE, D. B., ISSAC, B., LIEBERMAN, E., GIANNOUKOS, G., ALVAREZ, P., BROCKMAN, W., KIM, T. K., KOCHE, R. P., LEE, W., MENDENHALL, E., O'DONOVAN, A., PRESSER, A., RUSS, C., XIE, X., MEISSNER, A., WERNIG, M., JAENISCH, R., NUSBAUM, C., LANDER, E. S. & BERNSTEIN, B. E. 2007. Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature,* 448**,** 553-60.
MOHR, I. J., CLARK, R., SUN, S., ANDROPHY, E. J., MACPHERSON, P. & BOTCHAN, M. R. 1990. Targeting the E1 Replication Protein to the Papillomavirus Origin of Replication by Complex-Formation with the E2 Transactivator. *Science,* 250**,** 1694-1699.
MONINI, P., BLITZ, I. L. & CASSAI, E. 1993. Cooperative DNA-Binding of the Bovine Papillomavirus-E2 Transcriptional Activator Is Antagonized by Truncated E2-Polypeptides. *Journal of Virology,* 67**,** 5668-5676.
MONTALVO, E. A., COTTAM, M., HILL, S. & WANG, Y. J. 1995. YY1 binds to and regulates cis-acting negative elements in the Epstein-Barr virus BZLF1 promoter. *J Virol,* 69**,** 4158-65.
MOODY, C. A., FRADET-TURCOTTE, A., ARCHAMBAULT, J. & LAIMINS, L. A. 2007. Human papillomaviruses activate caspases upon epithelial differentiation to induce viral genome amplification. *Proc Natl Acad Sci U S A,* 104**,** 19541-6.
MOODY, C. A. & LAIMINS, L. A. 2009a. Human Papillomaviruses Activate the ATM DNA Damage Pathway for Viral Genome Amplification upon Differentiation. *Plos Pathogens,* 5.
MOODY, C. A. & LAIMINS, L. A. 2009b. Human papillomaviruses activate the ATM DNA damage pathway for viral genome amplification upon differentiation. *PLoS Pathog,* 5**,** e1000605.
MOODY, C. A. & LAIMINS, L. A. 2010. Human papillomavirus oncoproteins: pathways to transformation. *Nat Rev Cancer,* 10**,** 550-60.
MOORE, B. L., AITKEN, S. & SEMPLE, C. A. 2015. Integrative modeling reveals the principles of multi-scale chromatin boundary formation in human nuclear organization. *Genome Biol,* 16**,** 110.
MORI, S., TAKEUCHI, T., ISHII, Y., YUGAWA, T., KIYONO, T., NISHINA, H. & KUKIMOTO, I. 2017. Human Papillomavirus 16 E6 Upregulates APOBEC3B via the TEAD Transcription Factor. *J Virol,* 91.
MULLER, J., HART, C. M., FRANCIS, N. J., VARGAS, M. L., SENGUPTA, A., WILD, B., MILLER, E. L., O'CONNOR, M. B., KINGSTON, R. E. & SIMON, J. A. 2002. Histone methyltransferase activity of a Drosophila Polycomb group repressor complex. *Cell,* 111**,** 197-208.
MULLER, S., FILIPPAKOPOULOS, P. & KNAPP, S. 2011. Bromodomains as therapeutic targets. *Expert Reviews in Molecular Medicine,* 13**,** 1-21.

MUNGER, K., BALDWIN, A., EDWARDS, K. M., HAYAKAWA, H., NGUYEN, C. L., OWENS, M., GRACE, M. & HUH, K. 2004. Mechanisms of human papillomavirus-induced oncogenesis. *J Virol,* 78**,** 11451-60.

MUNGER, K., PHELPS, W. C., BUBB, V., HOWLEY, P. M. & SCHLEGEL, R. 1989. The E6 and E7 genes of the human papillomavirus type 16 together are necessary and sufficient for transformation of primary human keratinocytes. *J Virol,* 63**,** 4417-21.

MURTON, B. L., CHIN, W. L., PONTING, C. P. & ITZHAKI, L. S. 2010. Characterising the Binding Specificities of the Subunits Associated with the KMT2/Set1 Histone Lysine Methyltransferase. *Journal of Molecular Biology,* 398**,** 481-488.

NAKAGAWA, S. & HUIBREGTSE, J. M. 2000. Human scribble (Vartul) is targeted for ubiquitin-mediated degradation by the high-risk papillomavirus E6 proteins and the E6AP ubiquitin-protein ligase. *Molecular and Cellular Biology,* 20**,** 8244-8253.

NAKAMURA, M., BODILY, J. M., BEGLIN, M., KYO, S., INOUE, M. & LAIMINS, L. A. 2009. Hypoxia-specific stabilization of HIF-1alpha by human papillomaviruses. *Virology,* 387**,** 442-8.

NASSERI, M., HIROCHIKA, R., BROKER, T. R. & CHOW, L. T. 1987. A Human Papilloma-Virus Type-11 Transcript Encoding an E1 or E4 Protein. *Virology,* 159**,** 433-439.

NATESAN, S. & GILMAN, M. Z. 1993. DNA bending and orientation-dependent function of YY1 in the c-fos promoter. *Genes Dev,* 7**,** 2497-509.

NEES, M., GEOGHEGAN, J. M., HYMAN, T., FRANK, S., MILLER, L. & WOODWORTH, C. D. 2001. Papillomavirus type 16 oncogenes downregulate expression of interferon-responsive genes and upregulate proliferation-associated and NF-kappaB-responsive genes in cervical keratinocytes. *J Virol,* 75**,** 4283-96.

NG, H. H., ROBERT, F., YOUNG, R. A. & STRUHL, K. 2003. Targeted recruitment of set1 histone methylase by elongating pol II provides a localized mark and memory of recent transcriptional activity. *Molecular Cell,* 11**,** 709-719.

NUOVO, J., MELNIKOW, J., WILLAN, A. R. & CHAN, B. K. S. 2000. Treatment outcomes for squamous intraepithelial lesions. *International Journal of Gynecology & Obstetrics,* 68**,** 25-33.

O'BRIEN, P. M. & SAVERIA CAMPO, M. 2002. Evasion of host immunity directed by papillomavirus-encoded proteins. *Virus Res,* 88**,** 103-17.

O'CONNOR, M. J., STUNKEL, W., ZIMMERMANN, H., KOH, C. H. & BERNARD, H. U. 1998. A novel YY1-independent silencer represses the activity of the human papillomavirus type 16 enhancer. *Journal of Virology,* 72**,** 10083-10092.

O'CONNOR, M. J., TAN, S. H., TAN, C. H. & BERNARD, H. U. 1996. YY1 represses human papillomavirus type 16 transcription by quenching AP-1 activity. *J Virol,* 70**,** 6529-39.

OBERG, D., FAY, J., LAMBKIN, H. & SCHWARTZ, S. 2005. A downstream polyadenylation element in human papillomavirus type 16 L2 encodes multiple GGG motifs and interacts with hnRNP H. *Journal of Virology,* 79**,** 9254-9269.

OCONNOR M., Chan S., Bernard H.U.Transcription factor binding sites in the long control region of genital HPVs Human Papillomaviruses, 1995

Compendium, Part III, Los Alamos National Laboratory, Los Alamos, N.Mex (1995), pp. 21-40
OCONNOR, M. J., TAN, S. H., TAN, C. H. & BERNARD, H. U. 1996. YY1 represses human papillomavirus type 16 transcription by quenching AP-1 activity. *Journal of Virology,* 70**,** 6529-6539.
OHLSSON, R., RENKAWITZ, R. & LOBANENKOV, V. 2001. CTCF is a uniquely versatile transcription regulator linked to epigenetics and disease. *Trends in genetics : TIG,* 17**,** 520-7.
ONG, C. T. & CORCES, V. G. 2014. CTCF: an architectural protein bridging genome topology and function. *Nat Rev Genet,* 15**,** 234-46.
ORTH, G. 1986. Epidermodysplasia verruciformis: a model for understanding the oncogenicity of human papillomaviruses. *Ciba Found Symp,* 120**,** 157-74.
ORTH, G. 2005. Human papillomaviruses associated with epidermodysplasia verruciformis in non-melanoma skin cancers: guilty or innocent? *J Invest Dermatol,* 125**,** xii-xiii.
PALEFSKY, J. M., GIULIANO, A. R., GOLDSTONE, S., MOREIRA, E. D., JR., ARANDA, C., JESSEN, H., HILLMAN, R., FERRIS, D., COUTLEE, F., STOLER, M. H., MARSHALL, J. B., RADLEY, D., VUOCOLO, S., HAUPT, R. M., GURIS, D. & GARNER, E. I. 2011. HPV vaccine against anal HPV infection and anal intraepithelial neoplasia. *N Engl J Med,* 365**,** 1576-85.
PALSTRA, R. J., TOLHUIS, B., SPLINTER, E., NIJMEIJER, R., GROSVELD, F. & DE LAAT, W. 2003. The beta-globin nuclear compartment in development and erythroid differentiation. *Nature Genetics,* 35**,** 190-194.
PAN, G., TIAN, S., NIE, J., YANG, C., RUOTTI, V., WEI, H., JONSDOTTIR, G. A., STEWART, R. & THOMSON, J. A. 2007. Whole-genome analysis of histone H3 lysine 4 and lysine 27 methylation in human embryonic stem cells. *Cell Stem Cell,* 1**,** 299-312.
PARAMIO, J. M., LAIN, S., SEGRELLES, C., LANE, E. B. & JORCANO, J. L. 1998. Differential expression and functionally co-operative roles for the retinoblastoma family of proteins in epidermal differentiation. *Oncogene,* 17**,** 949-57.
PARELHO, V., HADJUR, S., SPIVAKOV, M., LELEU, M., SAUER, S., GREGSON, H. C., JARMUZ, A., CANZONETTA, C., WEBSTER, Z., NESTEROVA, T., COBB, B. S., YOKOMORI, K., DILLON, N., ARAGON, L., FISHER, A. G. & MERKENSCHIAGER, M. 2008. Cohesins functionally associate with CTCF on mammalian chromosome arms. *Cell,* 132**,** 422-433.
PARIS, C., PENTLAND, I., GROVES, I., ROBERTS, D. C., POWIS, S. J., COLEMAN, N., ROBERTS, S. & PARISH, J. L. 2015. CCCTC-Binding Factor Recruitment to the Early Region of the Human Papillomavirus 18 Genome Regulates Viral Oncogene Expression. *Journal of Virology,* 89**,** 4770-4785.
PARISH, J. L., BEAN, A. M., PARK, R. B. & ANDROPHY, E. J. 2006a. ChlR1 is required for loading papillomavirus E2 onto mitotic chromosomes and viral genome maintenance. *Mol Cell,* 24**,** 867-76.
PARISH, J. L., ROSA, J., WANG, X., LAHTI, J. M., DOXSEY, S. J. & ANDROPHY, E. J. 2006b. The DNA helicase ChlR1 is required for sister chromatid cohesion in mammalian cells. *J Cell Sci,* 119**,** 4857-65.

PARK, J. S., KIM, E. J., KWON, H. J., HWANG, E. S., NAMKOONG, S. E. & UM, S. J. 2000. Inactivation of interferon regulatory factor-1 tumor suppressor protein by HPV E7 oncoprotein. Implication for the E7-mediated immune evasion mechanism in cervical carcinogenesis. *J Biol Chem,* 275**,** 6764-9.
PARKIN, D. M. 2011. 11. Cancers attributable to infection in the UK in 2010. *Br J Cancer,* 105 Suppl 2**,** S49-56.
PARKIN, D. M. & BRAY, F. 2006. Chapter 2: The burden of HPV-related cancers. *Vaccine,* 24 Suppl 3**,** S3/11-25.
PARKIN, D. M., BRAY, F., FERLAY, J. & PISANI, P. 2005. Global cancer statistics, 2002. *Ca-a Cancer Journal for Clinicians,* 55**,** 74-108.
PARO, R. 1995. Propagating memory of transcriptional states. *Trends Genet,* 11**,** 295-7.
PASINI, D., BRACKEN, A. P., JENSEN, M. R., DENCHI, E. L. & HELIN, K. 2004. Suz12 is essential for mouse development and for EZH2 histone methyltransferase activity. *Embo Journal,* 23**,** 4061-4071.
PATEL, D., HUANG, S. M., BAGLIA, L. A. & MCCANCE, D. J. 1999. The E6 protein of human papillomavirus type 16 binds to and inhibits co-activation by CBP and p300. *EMBO J,* 18**,** 5061-72.
PENA-HERNANDEZ, R., MARQUES, M., HILMI, K., ZHAO, T., SAAD, A., ALAOUI-JAMALI, M. A., DEL RINCON, S. V., ASHWORTH, T., ROY, A. L., EMERSON, B. M. & WITCHER, M. 2015. Genome-wide targeting of the epigenetic regulatory protein CTCF to gene promoters by the transcription factor TFII-I. *Proc Natl Acad Sci U S A,* 112**,** E677-86.
PENG, J. C., VALOUEV, A., SWIGUT, T., ZHANG, J., ZHAO, Y., SIDOW, A. & WYSOCKA, J. 2009. Jarid2/Jumonji coordinates control of PRC2 enzymatic activity and target gene occupancy in pluripotent cells. *Cell,* 139**,** 1290-302.
PETT, M. & COLEMAN, N. 2007. Integration of high-risk human papillomavirus: a key event in cervical carcinogenesis? *J Pathol,* 212**,** 356-67.
PHILLIPS, J. E. & CORCES, V. G. 2009. CTCF: master weaver of the genome. *Cell,* 137**,** 1194-211.
PIM, D., MASSIMI, P. & BANKS, L. 1997. Alternatively spliced HPV-18 E6* protein inhibits E6 mediated degradation of p53 and suppresses transformed cell growth. *Oncogene,* 15**,** 257-64.
PIRROTTA, V. 1998. Polycombing the genome: PcG, trxG, and chromatin silencing. *Cell,* 93**,** 333-336.
POKHOLOK, D. K., HARBISON, C. T., LEVINE, S., COLE, M., HANNETT, N. M., LEE, T. I., BELL, G. W., WALKER, K., ROLFE, P. A., HERBOLSHEIMER, E., ZEITLINGER, J., LEWITTER, F., GIFFORD, D. K. & YOUNG, R. A. 2005. Genome-wide map of nucleosorne acetylation and methylation in yeast. *Cell,* 122**,** 517-527.
PRESCOTT, E. L., BRIMACOMBE, C. L., HARTLEY, M., BELL, I., GRAHAM, S. & ROBERTS, S. 2014. Human papillomavirus type 1 E1^E4 protein is a potent inhibitor of the serine-arginine (SR) protein kinase SRPK1 and inhibits phosphorylation of host SR proteins and of the viral transcription and replication regulator E2. *J Virol,* 88**,** 12599-611.
PUTHANAKIT, T., HUANG, L. M., CHIU, C. H., TANG, R. B., SCHWARZ, T. F., ESPOSITO, S., FRENETTE, L., GIAQUINTO, C., MCNEIL, S., RHEAULT, P., DURANDO, P., HORN, M., KLAR, M., PONCELET, S., DE SIMONI, S., FRIEL, D., DE MUYNCK, B.,

SURYAKIRAN, P. V., HEZAREH, M., DESCAMPS, D., THOMAS, F. & STRUYF, F. 2016. Randomized Open Trial Comparing 2-Dose Regimens of the Human Papillomavirus 16/18 AS04-Adjuvanted Vaccine in Girls Aged 9-14 Years Versus a 3-Dose Regimen in Women Aged 15-25 Years. *J Infect Dis,* 214**,** 525-36.

REINSON, T., TOOTS, M., KADAJA, M., PIPITCH, R., ALLIK, M., USTAV, E. & USTAV, M. 2013. Engagement of the ATR-Dependent DNA Damage Response at the Human Papillomavirus 18 Replication Centers during the Initial Amplification. *Journal of Virology,* 87**,** 951-964.

REMM, M., BRAIN, R. & JENKINS, J. R. 1992. The E2 binding sites determine the efficiency of replication for the origin of human papillomavirus type 18. *Nucleic Acids Res,* 20**,** 6015-21.

RENDA, M., BAGLIVO, I., BURGESS-BEUSSE, B., ESPOSITO, S., FATTORUSSO, R., FELSENFELD, G. & PEDONE, P. V. 2007. Critical DNA binding interactions of the insulator protein CTCF: a small number of zinc fingers mediate strong binding, and a single finger-DNA interaction controls binding at imprinted loci. *J Biol Chem,* 282**,** 33336-45.

RHEE, H. S. & PUGH, B. F. 2011. Comprehensive genome-wide protein-DNA interactions detected at single-nucleotide resolution. *Cell,* 147**,** 1408-19.

RICHARDS, R. M., LOWY, D. R., SCHILLER, J. T. & DAY, P. M. 2006. Cleavage of the papillomavirus minor capsid protein, L2, at a furin consensus site is necessary for infection. *Proc Natl Acad Sci U S A,* 103**,** 1522-7.

ROBERTS, S. Papillomaviruses; John Wiley & Sons, Ltd: Chichester, UK, 2001; pp. 1–11.

ROBERTS, S., ASHMOLE, I., JOHNSON, G. D., KREIDER, J. W. & GALLIMORE, P. H. 1993. Cutaneous and mucosal human papillomavirus E4 proteins form intermediate filament-like structures in epithelial cells. *Virology,* 197**,** 176-87.

ROMANOWSKI, B., SCHWARZ, T. F., FERGUSON, L. M., PETERS, K., DIONNE, M., SCHULZE, K., RAMJATTAN, B., HILLEMANNS, P., CATTEAU, G., DOBBELAERE, K., SCHUIND, A. & DESCAMPS, D. 2011. Immunogenicity and safety of the HPV-16/18 AS04-adjuvanted vaccine administered as a 2-dose schedule compared with the licensed 3-dose schedule Results from a randomized study. *Human Vaccines,* 7**,** 1374-1386.

ROSENBERGER, S., DE-CASTRO ARCE, J., LANGBEIN, L., STEENBERGEN, R. D. & ROSL, F. 2010. Alternative splicing of human papillomavirus type-16 E6/E6* early mRNA is coupled to EGF signaling via Erk1/2 activation. *Proc Natl Acad Sci U S A,* 107**,** 7006-11.

ROUS, P. & BEARD, J. W. 1935. The Progression to Carcinoma of Virus-Induced Rabbit Papillomas (Shope). *J Exp Med,* 62**,** 523-48.

RYERSON, A. B., PETERS, E. S., COUGHLIN, S. S., CHEN, V. W., GILLISON, M. L., REICHMAN, M. E., WU, X., CHATURVEDI, A. K. & KAWAOKA, K. 2008. Burden of potentially human papillomavirus-associated cancers of the oropharynx and oral cavity in the US, 1998-2003. *Cancer,* 113**,** 2901-9.

SAKAKIBARA, N., MITRA, R. & MCBRIDE, A. A. 2011. The Papillomavirus E1 Helicase Activates a Cellular DNA Damage Response in Viral Replication Foci. *Journal of Virology,* 85**,** 8981-8995.

SANDERS, C. M. & STENLUND, A. 1998. Recruitment and loading of the E1 initiator protein: an ATP-dependent process catalysed by a transcription factor. *EMBO J,* 17**,** 7044-55.
SANDERS, C. M. & STENLUND, A. 2000. Transcription factor-dependent loading of the E1 initiator reveals modular assembly of the papillomavirus origin melting complex. *J Biol Chem,* 275**,** 3522-34.
SANKARANARAYANAN, R., NENE, B. M., SHASTRI, S. S., JAYANT, K., MUWONGE, R., BUDUKH, A. M., HINGMIRE, S., MALVI, S. G., THORAT, R., KOTHARI, A., CHINOY, R., KELKAR, R., KANE, S., DESAI, S., KESKAR, V. R., RAJESHWARKAR, R., PANSE, N. & DINSHAW, K. A. 2009. HPV screening for cervical cancer in rural India. *N Engl J Med,* 360**,** 1385-94.
SANTOS-ROSA, H., SCHNEIDER, R., BANNISTER, A. J., SHERRIFF, J., BERNSTEIN, B. E., EMRE, N. C. T., SCHREIBER, S. L., MELLOR, J. & KOUZARIDES, T. 2002. Active genes are tri-methylated at K4 of histone H3. *Nature,* 419**,** 407-411.
SANYAL, A., LAJOIE, B. R., JAIN, G. & DEKKER, J. 2012. The long-range interaction landscape of gene promoters. *Nature,* 489**,** 109-13.
SCHAFER, A., LENGENFELDER, D., GRILLHOSL, C., WIESER, C., FLECKENSTEIN, B. & ENSSER, A. 2003. The latency-associated nuclear antigen homolog of herpesvirus saimiri inhibits lytic virus replication. *J Virol,* 77**,** 5911-25.
SCHEFFNER, M., WERNESS, B. A., HUIBREGTSE, J. M., LEVINE, A. J. & HOWLEY, P. M. 1990. The E6 oncoprotein encoded by human papillomavirus types 16 and 18 promotes the degradation of p53. *Cell,* 63**,** 1129-36.
SCHIFFMAN, M., CASTLE, P. E., JERONIMO, J., RODRIGUEZ, A. C. & WACHOLDER, S. 2007. Human papillomavirus and cervical cancer. *Lancet,* 370**,** 890-907.
SCHMIDT, D., SCHWALIE, P. C., WILSON, M. D., BALLESTER, B., GONCALVES, A., KUTTER, C., BROWN, G. D., MARSHALL, A., FLICEK, P. & ODOM, D. T. 2012. Waves of retrotransposon expansion remodel genome organization and CTCF binding in multiple mammalian lineages. *Cell,* 148**,** 335-48.
SCHNEIDER, R., BANNISTER, A. J., MYERS, F. A., THORNE, A. W., CRANE-ROBINSON, C. & KOUZARIDES, T. 2004. Histone H3 lysine 4 methylation patterns in higher eukaryotic genes. *Nat Cell Biol,* 6**,** 73-7.
SCHOEFTNER, S., SENGUPTA, A. K., KUBICEK, S., MECHTLER, K., SPAHN, L., KOSEKI, H. H., JENUWEIN, T. & WUTZ, A. 2006. Recruitment of PRC1 function at the initiation of X inactivation independent of PRC2 and silencing. *Embo Journal,* 25**,** 3110-3122.
SCHWARTZ, S. & AST, G. 2010. Chromatin density and splicing destiny: on the cross-talk between chromatin structure and splicing. *EMBO J,* 29**,** 1629-36.
SCHWARTZ, S., MESHORER, E. & AST, G. 2009. Chromatin organization marks exon-intron structure. *Nature Structural & Molecular Biology,* 16**,** 990-U117.
SCHWARZ, T. F. & LEO, O. 2008. Immune response to human papillomavirus after prophylactic vaccination with AS04-adjuvanted HPV-16/18 vaccine: improving upon nature. *Gynecol Oncol,* 110**,** S1-10.
SHI, X., HONG, T., WALTER, K. L., EWALT, M., MICHISHITA, E., HUNG, T., CARNEY, D., PENA, P., LAN, F., KAADIGE, M. R., LACOSTE, N., CAYROU, C., DAVRAZOU, F., SAHA, A., CAIRNS, B. R., AYER, D. E., KUTATELADZE, T. G., SHI, Y., COTE, J., CHUA, K. F. & GOZANI, O. 2006. ING2 PHD domain links histone H3 lysine 4 methylation to active gene repression. *Nature,* 442**,** 96-9.

SHI, Y., SETO, E., CHANG, L. S. & SHENK, T. 1991. Transcriptional Repression by Yy1, a Human Gli-Kruppel-Related Protein, and Relief of Repression by Adenovirus E1a Protein. *Cell,* 67**,** 377-388.
SHI, Y. & WHETSTINE, J. R. 2007. Dynamic regulation of histone lysine methylation by demethylases. *Molecular Cell,* 25**,** 1-14.
SHIBOSKI, C. H., SCHMIDT, B. L. & JORDAN, R. C. 2005. Tongue and tonsil carcinoma: increasing trends in the U.S. population ages 20-44 years. *Cancer,* 103**,** 1843-9.
SHIRASAWA, H., TOMITA, Y., SEKIYA, S., TAKAMIZAWA, H. & SIMIZU, B. 1987. Integration and Transcription of Human Papillomavirus Type-16 and Type-18 Sequences in Cell-Lines Derived from Cervical Carcinomas. *Journal of General Virology,* 68**,** 583-591.
SHUKLA, S., KAVAK, E., GREGORY, M., IMASHIMIZU, M., SHUTINOSKI, B., KASHLEV, M., OBERDOERFFER, P., SANDBERG, R. & OBERDOERFFER, S. 2011. CTCF-promoted RNA polymerase II pausing links DNA methylation to splicing. *Nature,* 479**,** 74-9.
SICHERO, L., SOBRINHO, J. S. & VILLA, L. L. 2012. Identification of novel cellular transcription factors that regulate early promoters of human papillomavirus types 18 and 16. *J Infect Dis,* 206**,** 867-74.
SIMON, J. M., GIRESI, P. G., DAVIS, I. J. & LIEB, J. D. 2014. Using formaldehyde-assisted isolation of regulatory elements (FAIRE) to isolate active regulatory DNA (vol 7, pg 256, 2012). *Nature Protocols,* 9**,** 501-503.
SMOTKIN, D. & WETTSTEIN, F. O. 1986. Transcription of human papillomavirus type 16 early genes in a cervical cancer and a cancer-derived cell line and identification of the E7 protein. *Proc Natl Acad Sci U S A,* 83**,** 4680-4.
SMOTKIN, D. & WETTSTEIN, F. O. 1987. The major human papillomavirus protein in cervical cancers is a cytoplasmic phosphoprotein. *J Virol,* 61**,** 1686-9.
SOKOLOWSKI, M., TAN, W., JELLNE, M. & SCHWARTZ, S. 1998. mRNA instability elements in the human papillomavirus type 16 L2 coding region. *J Virol,* 72**,** 1504-15.
SOMBERG, M. & SCHWARTZ, S. 2010. Multiple ASF/SF2 Sites in the Human Papillomavirus Type 16 (HPV-16) E4-Coding Region Promote Splicing to the Most Commonly Used 3 '-Splice Site on the HPV-16 Genome. *Journal of Virology,* 84**,** 8219-8230.
SONG, S., LIEM, A., MILLER, J. A. & LAMBERT, P. F. 2000. Human papillomavirus types 16 E6 and E7 contribute differently to carcinogenesis. *Virology,* 267**,** 141-50.
SPLINTER, E., HEATH, H., KOOREN, J., PALSTRA, R. J., KLOUS, P., GROSVELD, F., GALJART, N. & DE LAAT, W. 2006. CTCF mediates long-range chromatin looping and local histone modification in the beta-globin locus. *Genes & Development,* 20**,** 2349-2354.
SRINIVASAN, L. & ATCHISON, M. L. 2004. YY1 DNA binding and PcG recruitment requires CtBP. *Genes Dev,* 18**,** 2596-601.
STEBEN, M. & DUARTE-FRANCO, E. 2007. Human papillomavirus infection: epidemiology and pathophysiology. *Gynecol Oncol,* 107**,** S2-5.
STEDMAN, W., KANG, H., LIN, S., KISSIL, J. L., BARTOLOMEI, M. S. & LIEBERMAN, P. M. 2008. Cohesins localize with CTCF at the KSHV latency control region and at cellular c-myc and H19/Igf2 insulators. *EMBO J,* 27**,** 654-66.

STENTELLA, P., FREGA, A., CICCARONE, M., CIPRIANO, L., TINARI, A., TZANTZOGLOU, S. & PACHI, A. 1998. HPV and intraepithelial neoplasia recurrent lesions of the lower genital tract: assessment of the immune system. *Eur J Gynaecol Oncol,* 19**,** 466-9.

STRAIGHT, S. W., HERMAN, B. & MCCANCE, D. J. 1995. The E5 oncoprotein of human papillomavirus type 16 inhibits the acidification of endosomes in human keratinocytes. *J Virol,* 69**,** 3185-92.

STUNKEL, W. & BERNARD, H. U. 1999. The chromatin structure of the long control region of human papillomavirus type 16 represses viral oncoprotein expression. *J Virol,* 73**,** 1918-30.

SUN, L., ZHANG, G., LEI, T., HUANG, C., SONG, T. & SI, L. 2008. Two different HPV-11E6 fusion proteins trap p53 in the cytoplasm and induce apoptosis. *Cancer Biol Ther,* 7**,** 1909-15.

SUPRYNOWICZ, F. A., KRAWCZYK, E., HEBERT, J. D., SUDARSHAN, S. R., SIMIC, V., KAMONJOH, C. M. & SCHLEGEL, R. 2010. The human papillomavirus type 16 E5 oncoprotein inhibits epidermal growth factor trafficking independently of endosome acidification. *J Virol,* 84**,** 10619-29.

SWINDLE, C. S. & ENGLER, J. A. 1998. Association of the human papillomavirus type 11 E1 protein with histone H1. *Journal of Virology,* 72**,** 1994-2001.

SYVERTON, J. T. 1952. The pathogenesis of the rabbit papilloma-to-carcinoma sequence. *Ann N Y Acad Sci,* 54**,** 1126-40.

SZABO, P., TANG, S. H., RENTSENDORJ, A., PFEIFER, G. P. & MANN, J. R. 2000. Maternal-specific footprints at putative CTCF sites in the H19 imprinting control region give evidence for insulator function. *Curr Biol,* 10**,** 607-10.

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

TAVARES, L., DIMITROVA, E., OXLEY, D., WEBSTER, J., POOT, R., DEMMERS, J., BEZSTAROSTI, K., TAYLOR, S., URA, H., KOIDE, H., WUTZ, A., VIDAL, M., ELDERKIN, S. & BROCKDORFF, N. 2012. RYBP-PRC1 Complexes Mediate H2A Ubiquitylation at Polycomb Target Sites Independently of PRC2 and H3K27me3 (vol 148, pg 664, 2012). *Cell,* 149**,** 1647-1648.

TEMPERA, I., KLICHINSKY, M. & LIEBERMAN, P. M. 2011. EBV latency types adopt alternative chromatin conformations. *PLoS Pathog,* 7**,** e1002180.

TEMPERA, I., WIEDMER, A., DHEEKOLLU, J. & LIEBERMAN, P. M. 2010. CTCF prevents the epigenetic drift of EBV latency promoter Qp. *PLoS pathogens,* 6**,** e1001048.

TERHUNE, S. S., HUBERT, W. G., THOMAS, J. T. & LAIMINS, L. A. 2001. Early polyadenylation signals of human papillomavirus type 31 negatively regulate capsid gene expression. *J Virol,* 75**,** 8147-57.

TERHUNE, S. S., MILCAREK, C. & LAIMINS, L. A. 1999. Regulation of human papillomavirus type 31 polyadenylation during the differentiation-dependent life cycle. *J Virol,* 73**,** 7185-92.

THIERRY, F. & HOWLEY, P. M. 1991. Functional analysis of E2-mediated repression of the HPV18 P105 promoter. *New Biol,* 3**,** 90-100.

THIERRY, F., SPYROU, G., YANIV, M. & HOWLEY, P. 1992. Two AP1 sites binding JunB are essential for human papillomavirus type 18 transcription in keratinocytes. *J Virol,* 66**,** 3740-8.

TOMAIC, V. 2016. Functional Roles of E6 and E7 Oncoproteins in HPV-Induced Malignancies at Diverse Anatomical Sites. *Cancers (Basel),* 8.
TOMLINS, C. & STOREY, A. 2010. Cutaneous HPV5 E6 causes increased expression of Osteoprotegerin and Interleukin 6 which contribute to evasion of UV-induced apoptosis. *Carcinogenesis,* 31**,** 2155-64.
TOUSSAINT-SMITH, E., DONNER, D. B. & ROMAN, A. 2004. Expression of human papillomavirus type 16 E6 and E7 oncoproteins in primary foreskin keratinocytes is sufficient to alter the expression of angiogenic factors. *Oncogene,* 23**,** 2988-2995.
TSOMPANA, M. & BUCK, M. J. 2014. Chromatin accessibility: a window into the genome. *Epigenetics Chromatin,* 7**,** 33.
TSUI, S., WANG, J., WANG, L., DAI, W. & LU, L. 2016. CTCF-Mediated and Pax6-Associated Gene Expression in Corneal Epithelial Cell-Specific Differentiation. *PLoS One,* 11**,** e0162071.
UNDERBRINK, M. P., HOWIE, H. L., BEDARD, K. M., KOOP, J. I. & GALLOWAY, D. A. 2008. E6 proteins from multiple human betapapillomavirus types degrade Bak and protect keratinocytes from apoptosis after UVB irradiation. *J Virol,* 82**,** 10408-17.
USTAV, M. & STENLUND, A. 1991. Transient replication of BPV-1 requires two viral polypeptides encoded by the E1 and E2 open reading frames. *EMBO J,* 10**,** 449-57.
VENUTI, A., PAOLINI, F., NASIR, L., CORTEGGIO, A., ROPERTO, S., CAMPO, M. S. & BORZACCHIELLO, G. 2011. Papillomavirus E5: the smallest oncoprotein with many functions. *Molecular Cancer,* 10.
VESIKARI, T., BRODSZKI, N., VAN DAMME, P., DIEZ-DOMINGO, J., ICARDI, G., PETERSEN, L. K., TRAN, C., THOMAS, S., LUXEMBOURG, A. & BAUDIN, M. 2015. A Randomized, Double-Blind, Phase III Study of the Immunogenicity and Safety of a 9-Valent Human Papillomavirus L1 Virus-Like Particle Vaccine (V503) Versus Gardasil((R)) in 9-15-Year-Old Girls. *Pediatric Infectious Disease Journal,* 34**,** 992-998.
VIARISIO, D., MUELLER-DECKER, K., KLOZ, U., AENGENEYNDT, B., KOPP-SCHNEIDER, A., GRONE, H. J., GHEIT, T., FLECHTENMACHER, C., GISSMANN, L. & TOMMASINO, M. 2011. E6 and E7 from beta HPV38 cooperate with ultraviolet light in the development of actinic keratosis-like lesions and squamous cell carcinoma in mice. *PLoS Pathog,* 7**,** e1002125.
VILLA, L. L., AULT, K. A., GIULIANO, A. R., COSTA, R. L., PETTA, C. A., ANDRADE, R. P., BROWN, D. R., FERENCZY, A., HARPER, D. M., KOUTSKY, L. A., KURMAN, R. J., LEHTINEN, M., MALM, C., OLSSON, S. E., RONNETT, B. M., SKJELDESTAD, F. E., STEINWALL, M., STOLER, M. H., WHEELER, C. M., TADDEO, F. J., YU, J., LUPINACCI, L., RAILKAR, R., MARCHESE, R., ESSER, M. T., BRYAN, J., JANSEN, K. U., SINGS, H. L., TAMMS, G. M., SAAH, A. J. & BARR, E. 2006. Immunologic responses following administration of a vaccine targeting human papillomavirus Types 6, 11, 16, and 18. *Vaccine,* 24**,** 5571-83.
VINOKUROVA, S., WENTZENSEN, N., KRAUS, I., KLAES, R., DRIESCH, C., MELSHEIMER, P., KISSELJOV, F., DURST, M., SCHNEIDER, A. & VON KNEBEL DOEBERITZ, M. 2008. Type-dependent integration frequency of human papillomavirus genomes in cervical lesions. *Cancer Res,* 68**,** 307-13.

VOSTROV, A. A. & QUITSCHKE, W. W. 1997. The zinc finger protein CTCF binds to the APBbeta domain of the amyloid beta-protein precursor promoter. Evidence for a role in transcriptional activation. *J Biol Chem,* 272**,** 33353-9.

WADA, Y., OHTA, Y., XU, M., TSUTSUMI, S., MINAMI, T., INOUE, K., KOMURA, D., KITAKAMI, J., OSHIDA, N., PAPANTONIS, A., IZUMI, A., KOBAYASHI, M., MEGURO, H., KANKI, Y., MIMURA, I., YAMAMOTO, K., MATAKI, C., HAMAKUBO, T., SHIRAHIGE, K., ABURATANI, H., KIMURA, H., KODAMA, T., COOK, P. R. & IHARA, S. 2009. A wave of nascent transcription on activated human genes. *Proceedings of the National Academy of Sciences of the United States of America,* 106**,** 18357-18361.

WALBOOMERS, J. M. M., JACOBS, M. V., MANOS, M. M., BOSCH, F. X., KUMMER, J. A., SHAH, K. V., SNIJDERS, P. J. F., PETO, J., MEIJER, C. J. L. M. & MUNOZ, N. 1999. Human papillomavirus is a necessary cause of invasive cervical cancer worldwide. *Journal of Pathology,* 189**,** 12-19.

WALKER, C. J., MIRANDA, M. A., O'HERN, M. J., MCELROY, J. P., COOMBES, K. R., BUNDSCHUH, R., COHN, D. E., MUTCH, D. G. & GOODFELLOW, P. J. 2015. Patterns of CTCF and ZFHX3 Mutation and Associated Outcomes in Endometrial Cancer. *Jnci-Journal of the National Cancer Institute,* 107.

WANG, H., MAURANO, M. T., QU, H., VARLEY, K. E., GERTZ, J., PAULI, F., LEE, K., CANFIELD, T., WEAVER, M., SANDSTROM, R., THURMAN, R. E., KAUL, R., MYERS, R. M. & STAMATOYANNOPOULOS, J. A. 2012. Widespread plasticity in CTCF occupancy linked to DNA methylation. *Genome research,* 22**,** 1680-8.

WANG, H., WANG, L., ERDJUMENT-BROMAGE, H., VIDAL, M., TEMPST, P., JONES, R. S. & ZHANG, Y. 2004a. Role of histone H2A ubiquitination in Polycomb silencing. *Nature,* 431**,** 873-8.

WANG, Q., GRIFFIN, H., SOUTHERN, S., JACKSON, D., MARTIN, A., MCINTOSH, P., DAVY, C., MASTERSON, P. J., WALKER, P. A., LASKEY, P., OMARY, M. B. & DOORBAR, J. 2004b. Functional analysis of the human papillomavirus type 16 E1=E4 protein provides a mechanism for in vivo and in vitro keratin filament reorganization. *J Virol,* 78**,** 821-33.

WANG, X. H., MEYERS, C., WANG, H. K., CHOW, L. T. & ZHENG, Z. M. 2011. Construction of a Full Transcription Map of Human Papillomavirus Type 18 during Productive Viral Infection. *Journal of Virology,* 85**,** 8080-8092.

WANG, Z. & BURGE, C. B. 2008. Splicing regulation: from a parts list of regulatory elements to an integrated splicing code. *RNA,* 14**,** 802-13.

WARZECHA, H., MASON, H. S., LANE, C., TRYGGVESSON, A., RYBICKI, E., WILLIAMSON, A. L., CLEMENTS, J. D. & ROSE, R. C. 2003. Oral immunogenicity of human papillomavirus-like particles expressed in potato. *J Virol,* 77**,** 8702-11.

WEI, Y. K., CHEN, Y. H., LI, L. Y., LANG, J. Y., YEH, S. P., SHI, B., YANG, C. C., YANG, J. Y., LIN, C. Y., LAI, C. C. & HUNG, M. C. 2011. CDK1-dependent phosphorylation of EZH2 suppresses methylation of H3K27 and promotes osteogenic differentiation of human mesenchymal stem cells. *Nature Cell Biology,* 13**,** 87-U211.

WEN, K. W. & DAMANIA, B. 2010. Kaposi sarcoma-associated herpesvirus (KSHV): molecular biology and oncogenesis. *Cancer Lett,* 289**,** 140-50.

WENDT, K. S., YOSHIDA, K., ITOH, T., BANDO, M., KOCH, B., SCHIRGHUBER, E., TSUTSUMI, S., NAGAE, G., ISHIHARA, K., MISHIRO, T., YAHATA, K., IMAMOTO,

F., ABURATANI, H., NAKAO, M., IMAMOTO, N., MAESHIMA, K., SHIRAHIGE, K. & PETERS, J. M. 2008. Cohesin mediates transcriptional insulation by CCCTC-binding factor. *Nature,* 451**,** 796-801.

WERNESS, B. A., LEVINE, A. J. & HOWLEY, P. M. 1990. Association of human papillomavirus types 16 and 18 E6 proteins with p53. *Science,* 248**,** 76-9.

WETHERILL, L. F., HOLMES, K. K., VEROW, M., MULLER, M., HOWELL, G., HARRIS, M., FISHWICK, C., STONEHOUSE, N., FOSTER, R., BLAIR, G. E., GRIFFIN, S. & MACDONALD, A. 2012. High-risk human papillomavirus E5 oncoprotein displays channel-forming activity sensitive to small-molecule inhibitors. *J Virol,* 86**,** 5341-51.

WHEELER, C. M., CASTELLSAGUE, X., GARLAND, S. M., SZAREWSKI, A., PAAVONEN, J., NAUD, P., SALMERON, J., CHOW, S. N., APTER, D., KITCHENER, H., TEIXEIRA, J. C., SKINNER, S. R., JAISAMRARN, U., LIMSON, G., ROMANOWSKI, B., AOKI, F. Y., SCHWARZ, T. F., POPPE, W. A. J., BOSCH, F. X., HARPER, D. M., HUH, W., HARDT, K., ZAHAF, T. K., DESCAMPS, D., STRUYF, F., DUBIN, G., LEHTINEN, M. & GRP, H. P. S. 2012. Cross-protective efficacy of HPV-16/18 AS04-adjuvanted vaccine against cervical infection and precancer caused by non-vaccine oncogenic HPV types: 4-year end-of-study analysis of the randomised, double-blind PATRICIA trial. *Lancet Oncology,* 13**,** 100-110.

WIELAND, U., KREUTER, A. & PFISTER, H. 2014. Human papillomavirus and immunosuppression. *Curr Probl Dermatol,* 45**,** 154-65.

WILKINSON, F. H., PARK, K. & ATCHISON, M. L. 2006. Polycomb recruitment to DNA in vivo by the YY1 REPO domain. *Proc Natl Acad Sci U S A,* 103**,** 19296-301.

WINER, R. L., LEE, S. K., HUGHES, J. P., ADAM, D. E., KIVIAT, N. B. & KOUTSKY, L. A. 2003. Genital human papillomavirus infection: incidence and risk factors in a cohort of female university students. *Am J Epidemiol,* 157**,** 218-26.

WOODMAN, C. B., COLLINS, S. I. & YOUNG, L. S. 2007. The natural history of cervical HPV infection: unresolved issues. *Nature reviews. Cancer,* 7**,** 11-22.

WOOLDRIDGE, T. R. & LAIMINS, L. A. 2008. Regulation of human papillomavirus type 31 gene expression during the differentiation-dependent life cycle through histone modifications and transcription factor binding. *Virology,* 374**,** 371-380.

WU, S. Y., LEE, A. Y., HOU, S. Y., KEMPER, J. K., ERDJUMENT-BROMAGE, H., TEMPST, P. & CHIANG, C. M. 2006. Brd4 links chromatin targeting to HPV transcriptional silencing. *Genes Dev,* 20**,** 2383-96.

XI, H., YU, Y., FU, Y., FOLEY, J., HALEES, A. & WENG, Z. 2007. Analysis of overrepresented motifs in human core promoters reveals dual regulatory roles of YY1. *Genome Res,* 17**,** 798-806.

XIAO, T., WALLACE, J. & FELSENFELD, G. 2011. Specific sites in the C terminus of CTCF interact with the SA2 subunit of the cohesin complex and are required for cohesin-dependent insulation activity. *Mol Cell Biol,* 31**,** 2174-83.

XIAO, T. J., WONGTRAKOONGATE, P., TRAINOR, C. & FELSENFELD, G. 2015. CTCF Recruits Centromeric Protein CENP-E to the Pericentromeric/Centromeric Regions of Chromosomes through Unusual CTCF-Binding Sites. *Cell Reports,* 12**,** 1704-1714.

XIE, X. H., MIKKELSEN, T. S., GNIRKE, A., LINDBLAD-TOH, K., KELLIS, M. & LANDER, E. S. 2007. Systematic discovery of regulatory motifs in conserved regions of the

human genome, including thousands of CTCF insulator sites. *Proceedings of the National Academy of Sciences of the United States of America,* 104**,** 7145-7150.
XU, C., BIAN, C., YANG, W., GALKA, M., OUYANG, H., CHEN, C., QIU, W., LIU, H., JONES, A. E., MACKENZIE, F., PAN, P., LI, S. S., WANG, H. & MIN, J. 2010. Binding of different histone marks differentially regulates the activity and specificity of polycomb repressive complex 2 (PRC2). *Proc Natl Acad Sci U S A,* 107**,** 19266-71.
XU, C. & CORCES, V. G. 2016. Towards a predictive model of chromatin 3D organization. *Semin Cell Dev Biol,* 57**,** 24-30.
XUE, Y. Z., BELLANGER, S., ZHANG, W. Y., LIM, D., LOW, J., LUNNY, D. & THIERRY, F. 2010. HPV16 E2 Is an Immediate Early Marker of Viral Infection, Preceding E7 Expression in Precursor Structures of Cervical Carcinoma. *Cancer Research,* 70**,** 5316-5325.
YAN, J. L., LI, B. H., LIN, B. H., LEE, P. T., CHUNG, T. H., TAN, J., BI, C. L., LEE, X. T., SELVARAJAN, V., NG, S. B., YANG, H., YU, Q. & CHNG, W. J. 2016. EZH2 phosphorylation by JAK3 mediates a switch to noncanonical function in natural killer/T-cell lymphoma. *Blood,* 128**,** 948-958.
YING, H., JING, F., FANGHUI, Z., YOULIN, Q. & YALI, H. 2014. High-risk HPV nucleic acid detection kit-the careHPV test -a new detection method for screening. *Sci Rep,* 4**,** 4704.
YOON, C. S., KIM, K. D., PARK, S. N. & CHEONG, S. W. 2001. alpha(6) Integrin is the main receptor of human papillomavirus type 16 VLP. *Biochem Biophys Res Commun,* 283**,** 668-73.
YOU, J., CROYLE, J. L., NISHIMURA, A., OZATO, K. & HOWLEY, P. M. 2004. Interaction of the bovine papillomavirus E2 protein with Brd4 tethers the viral DNA to host mitotic chromosomes. *Cell,* 117**,** 349-60.
YU, J. H., LIN, B. Y., DENG, W., BROKER, T. R. & CHOW, L. T. 2007. Mitogen-activated protein kinases activate the nuclear localization sequence of human papillomavirus type 11 E1 DNA helicase to promote efficient nuclear import. *J Virol,* 81**,** 5066-78.
YU, W., GINJALA, V., PANT, V., CHERNUKHIN, I., WHITEHEAD, J., DOCQUIER, F., FARRAR, D., TAVOOSIDANA, G., MUKHOPADHYAY, R., KANDURI, C., OSHIMURA, M., FEINBERG, A. P., LOBANENKOV, V., KLENOVA, E. & OHLSSON, R. 2004. Poly(ADP-ribosyl)ation regulates CTCF-dependent chromatin insulation. *Nat Genet,* 36**,** 1105-10.
ZAMPIERI, M., GUASTAFIERRO, T., CALABRESE, R., CICCARONE, F., BACALINI, M. G., REALE, A., PERILLI, M., PASSANANTI, C. & CAIAFA, P. 2012. ADP-ribose polymers localized on Ctcf-Parp1-Dnmt1 complex prevent methylation of Ctcf target sites. *Biochem J,* 441**,** 645-52.
ZANDBERG, D. P., BHARGAVA, R., BADIN, S. & CULLEN, K. J. 2013. The role of human papillomavirus in nongenital cancers. *CA Cancer J Clin,* 63**,** 57-81.
ZANIER, K., CHARBONNIER, S., SIDI, A. O., MCEWEN, A. G., FERRARIO, M. G., POUSSIN-COURMONTAGNE, P., CURA, V., BRIMER, N., BABAH, K. O., ANSARI, T., MULLER, I., STOTE, R. H., CAVARELLI, J., VANDE POL, S. & TRAVE, G. 2013. Structural basis for hijacking of cellular LxxLL motifs by papillomavirus E6 oncoproteins. *Science,* 339**,** 694-8.

ZHANG, T. Y., COOPER, S. & BROCKDORFF, N. 2015. The interplay of histone modifications - writers that read. *Embo Reports,* 16**,** 1467-1481.
ZHAO, K. N., HENGST, K., LIU, W. J., LIU, Y. H., LIU, X. S., MCMILLAN, N. A. J. & FRAZER, I. H. 2000. BPV1 E2 protein enhances packaging of full-length plasmid DNA in BPV1 pseudovirions. *Virology,* 272**,** 382-393.
ZHAO, X. M., OBERG, D., RUSH, M., FAY, J., LAMBKIN, H. & SCHWARTZ, S. 2005. A 57-nucleotide upstream early polyadenylation element in human papillomavirus type 16 interacts with hFip1, CstF-64, hnRNP C1/C2, and polypyrimidine tract binding protein. *Journal of Virology,* 79**,** 4270-4288.
ZHENG, Z. M. & BAKER, C. C. 2006. Papillomavirus genome structure, expression, and post-transcriptional regulation. *Front Biosci,* 11**,** 2286-302.
ZIEBARTH, J. D., BHATTACHARYA, A. & CUI, Y. 2013. CTCFBSDB 2.0: a database for CTCF-binding sites and genome organization. *Nucleic Acids Res,* 41**,** D188-94.
ZIELKE, K., FULL, F., TEUFERT, N., SCHMIDT, M., MULLER-FLECKENSTEIN, I., ALBERTER, B. & ENSSER, A. 2012. The insulator protein CTCF binding sites in the orf73/LANA promoter region of herpesvirus saimiri are involved in conferring episomal stability in latently infected human T cells. *J Virol,* 86**,** 1862-73.
ZIGHELBOIM, I., MUTCH, D. G., KNAPP, A., DING, L., XIE, M., COHN, D. E. & GOODFELLOW, P. J. 2014. High frequency strand slippage mutations in CTCF in MSI-positive endometrial cancers. *Hum Mutat,* 35**,** 63-5.
ZLATANOVA, J. & CAIAFA, P. 2009. CTCF and its protein partners: divide and rule? *Journal of cell science,* 122**,** 1275-84.
ZUIN, J., DIXON, J. R., VAN DER REIJDEN, M. I., YE, Z., KOLOVOS, P., BROUWER, R. W., VAN DE CORPUT, M. P., VAN DE WERKEN, H. J., KNOCH, T. A., VAN, I. W. F., GROSVELD, F. G., REN, B. & WENDT, K. S. 2014. Cohesin and CTCF differentially affect chromatin architecture and gene expression in human cells. *Proc Natl Acad Sci U S A,* 111**,** 996-1001.
ZUR HAUSEN, H. 2002. Papillomaviruses and cancer: from basic studies to clinical application. *Nat Rev Cancer,* 2**,** 342-50.
ZUR HAUSEN, H. 2009. Papillomaviruses in the causation of human cancers - a brief historical account. *Virology,* 384**,** 260-5.
ZUR HAUSEN, H., GISSMANN, L., STEINER, W., DIPPOLD, W. & DREGER, I. 1975. Human papilloma viruses and cancer. *Bibl Haematol***,** 569-71.