Loss of CTCF-YY1-mediated regulation of human papillomavirus oncogene transcription in HPV-driven disease

BY

MEGAN BURLEY

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

November 2022

UNIVERSITY OF
BIRMINGHAM

University of Birmingham Research Archive

e-theses repository

This unpublished thesis/dissertation is copyright of the author and/or third parties.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

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

ABSTRACT

Persistent human papillomavirus (HPV) infection is a risk factor for oropharyngeal squamous cell carcinoma (OPSCC). The viral oncogenes E6 and E7 drive cell growth and delay differentiation. Expression of E6/E7 is tightly regulated during the HPV life cycle. In basal keratinocytes, the CCCTC-binding factor (CTCF) binds the HPV18 episome at the E2 open reading frame to mediate the formation of an epigenetically repressed chromatin loop through association with Yin Yang 1 (YY1) bound at the viral enhancer. Following differentiation, depletion in YY1 results in chromatin remodelling and increased transcription. As such, YY1 is a key mediator of differentiation-dependent depression of HPV oncogene transcription. Whilst CTCF-YY1-mediated chromatin looping is essential in regulating early gene expression during the HPV life cycle, it is unclear whether this mechanism of transcription control is deregulated during carcinogenesis. Using physiological models of HPV-induced disease and OPSCC cell lines we show a switch in YY1 function from transcriptional repression to transactivation of the HPV early promoter. YY1 enrichment at the viral enhancer becomes significantly increased during disease progression which is coincident with open chromatin and high E6/E7 expression. We show a robust enrichment of active histone modifications histone H3 acetylation (H3K27Ac) and histone H4 methylation (H3K4me3) when YY1 is bound. Following depletion of YY1 we see a significant reduction in E6/E7 expression as a result of reduced chromatin accessibility at the viral enhancer. Interestingly, in the context of transcriptional regulation, CTCF appears to be redundant within OPSCCs where it exerts no effect on E6/E7 transcription. Overall, we describe a mechanism in which YY1 functions to drive the overexpression of E6 and E7 during the progression of HPV-driven oropharyngeal carcinomas.

ACKNOWLEDGMENTS

I would firstly like to acknowledge my supervisor Prof Jo Parish and co-supervisor Dr Sally Roberts. Over the past five years they have constantly guided and inspired my research. They have been supportive, encouraging and amazing mentors. Not to mention, a joy to drink Prosecco with. Thank you so much for the support, I could not have done it without you both. I want to acknowledge my funding body Cancer Research UK for giving me the opportunity to undertake my research and for their continued support over the four years. Thank you to Kamini. You have made the past four years a joy. Thank you for the laughs and for always being there when I need you. Thank you for being my conference buddy, my wine pal, and my TC karaoke partner. I am very lucky to have met a friend like you. Thank you, Danny, for being a great friend over the past four years and making me laugh daily. I want to acknowledge all the amazing post-docs I have worked with. Karen, for teaching me everything you know, and for being an amazing friend. Jack, for always inspiring me to think creatively, for constantly listening about train drama and experiment stresses, and for always going to Pop world after the Christmas party. Christy, for the guidance and for also being a great friend. I would like to acknowledge everyone else I have worked with over the years including Molly, Vidushi, everyone in the office and many others. Thank you to my family for always supporting me. Especially my mum and dad for always cheering me on. Finally, I would not have been able to get to where I am today without the love and support from Will. Thank you for always being there for me. After the long, stressful days. The infuriating commute. The experiments that just wouldn't work. You always knew how to cheer me up. I am endlessly grateful for you. I dedicate this thesis to my nan and grandad who I know would have been so proud.

TABLE OF CONTENTS

TABLE OF CONTENTS 4
LIST OF FIGURES 8
LIST OF TABLES 1
ABBREVIATIONS 2
Chapter 1 Introduction 1
1.1 The human papillomavirus 1
1.1.1 Introduction 1
1.1.2 Classification and types 2
1.1.3 Epidemiology and prevalence 6
1.1.4 Cervical screening 10
1.1.5 HPV vaccination programme 11
1.1.6 Current treatments for HPV-associated cancers 14
1.1.7 HPV structure 16
1.1.8 Structure of the upstream regulatory region 16
1.2 The HPV life cycle 22
1.2.1 Initial infection and virion entry 24
1.2.2 Genome establishment and maintenance 26
1.2.3 Productive infection 29
1.3 HPV gene expression 34
1.3.1 Promoter usage 34
1.3.1 Alternative splicing 36
1.3.2 Polyadenylation 39
1.3.3 HPV oncoproteins 41
1.4 Regulation of HPV gene expression 46
1.4.1 Transcription factor binding 46
1.4.2 HPV E2 47
1.4.3 Yin-yang 1 (YY1) 50
1.4.4 Regulation of YY1 52
1.4.5 YY1 binding partners 53
1.4.6 YY1 and chromatin remodelling 55
1.4.7 YY1 mediated regulation of HPV transcription 55
1.5 Epigenetic regulation of HPV gene expression 57
1.5.1 Histone modifications 57

1.5.2 Polycomb repressive group proteins (PcG) 59
1.5.3 Three-dimensional chromatin organisation 61
1.6 The natural history of an HPV-infection 65
1.6.1 Integration 66
1.6.2 Genome instability 68
1.6.3 Current models of HPV-driven disease 68
1.7 Hypothesis and aims 71
Chapter 2 Materials and Methods 73
2.1 Cell lines and culture conditions 73
2.1.1 Culture conditions 73
2.1.2 Mycoplasma tests 73
2.1.3 Human tonsil keratinocytes 73
2.1.4 J2-3T3 Fibroblasts 74
2.1.5 Head and neck squamous cell carcinoma lines 74
2.1.6 Cryopreservation 75
2.1.7 Cell growth assay 75
2.1.8 siRNA knockdown of proteins 76
2.2 Southern blotting 76
2.2.1 DNA extraction 76
2.2.2 Restriction enzyme digestion 77
2.2.3 DNA resolution and transfer 77
2.2.4 Preparation of the radiolabelled probe 78
2.2.5 Hybridisation of radiolabelled probe to DNA 79
2.2.6 Autoradiography 79
2.3 Western blotting 80
2.3.1 Cell lysis and protein concentration 80
2.3.2 SDS-PAGE 81
2.3.3 Washes 81
2.3.4 Chemiluminescence 82
2.4 Chromatin immunoprecipitation qPCR 82
2.4.1 Formaldehyde fixation 82
2.4.2 Shearing by sonication 83
2.4.3 Shearing efficiency 83
2.4.4 Chromatin immunoprecipitation 84

2.4.5 Quantification of ChIP efficiency 84
2.5 Formaldehyde isolation of regulatory elements (FAIRE) 85
2.6 Sequencing 85
2.6.1 Library preparation and sequencing 85
2.6.2 Sequencing data analysis 87
2.7 Gene expression analysis 87
2.7.1 RNA extraction 87
2.7.2 DNase treatment 88
2.7.3 cDNA synthesis 88
2.7.4 qPCR 89
2.7.5 Quantification of gene expression 90
2.8 Statistical analysis 90
Chapter 3 Loss of CTCF-YY1-mediated HPV18 oncogene regulation in high passage primary tonsil keratinocytes 97
3.1 Introduction 97
3.2 Validating high-passage HPV18-HTKs as a model of HPV-driven disease 98
3.3 Altered expression of CTCF and YY1 in high-passage 18-HTKs 106
3.4 Enrichment of CTCF and YY1 to the HPV18 genome is significantly altered in high-passage 18-HTKs 110
3.5 YY1 binding to the HPV18 genome exerts a non-canonical regulation of the chromatin landscape in high-passage HTKs 120
3.6 Deregulation of CTCF-YY1-mediated oncogene transcription control in high-passage 18-HTKs 124
3.7 Conclusions 126
Chapter 4 Assessing the function of CTCF and YY1 in a HPV16-driven disease progression model 134
4.1 Introduction 134
4.2 Validating the high-passage 16-HTKs as a model of HPV-driven disease 134
4.3 Altered binding of CTCF and YY1 in a HPV16-driven disease progression model 140
4.4 Absence of CTCF-YY1-mediated regulation of the HPV16 early promoter in high-passage 16-HTKs 148
4.5 Conclusions 150
Chapter 5 YY1 functions as an activator of E6/E7 oncogene transcription to contribute to HPV-driven oropharyngeal squamous cell carcinoma 154

5.1 Introduction 154
5.2 Characterisation of the genome status in SCC147 and SCC104 155
5.3 Enrichment of CTCF and YY1 binding to the HPV16 genome in OPSCCs 160
5.4 Enrichment of YY1 and CTCF binding to the HPV16 genome in OPSCCs is consistent with a transcriptionally active epigenetic landscape 171
5.5 Recruitment of PRC1 component RING1B to the viral genome in HPV16 OPSCCs is independent of transcriptional repression 180
5.6 YY1 functions as a transactivator of the HPV16 early promoter to drive high expression of E6/E7 transcripts in OPSCCs 190
5.7 Conclusions 192
Chapter 6 Discussion 199
6.1 Modelling HPV-driven disease progression 199
6.2 Upregulation of CTCF and YY1 protein expression in HPV-driven disease 203
6.3 Deregulation of CTCF-YY1-mediated HPV oncogene transcription regulation in a primary human tonsil keratinocyte disease progression model 207
6.4 Loss of CTCF-mediated repression of HPV E6 and E7 transcription in HPV-driven OPSCC cell lines 213
6.5 YY1 functions as a transactivator of the HPV16 p97 promoter in HPV-driven OPSCC cell lines 214
6.6 Clinical implications 221
6.7 Limitations 223
6.8 Concluding remarks 225
References 228

LIST OF FIGURES

Figure 1.1: Evolutionary map of human papillomavirus……..……...5
Figure 1.2: Estimated age-standardised incidence rates of cervical and head and neck cancer…….……..…9
Figure 1.3: Organisation of the HPV genome…….……..18
Figure 1.4: Structure of the HPV upstream regulatory region…….…...…..19
Figure 1.5: The HPV life cycle…….22
Figure 1.6: The function of high-risk HPV oncogenes …….....30
Figure 1.7: HPV18 transcript map ……..37
Figure 1.8: The structure of Yin-Yang 1 (YY1) ……45
Figure 1.9: CTCF-YY1-mediated epigenetically repressed chromatin loop regulates E6/E7 oncogene expression during the HPV18 life cycle…….....….60
Figure 1.10: Progression of normal squamous epithelial cells into pre-cancerous lesions and cancer…….....…..….66
Figure 3.1: Schematic of primary human tonsil keratinocyte model system…….....……...…102
Figure 3.2: Increased transcription of HPV18 E6 in high-passage 18-HTKs…….....103
Figure 3.3: Increased E6 and E7 protein levels in high-passage 18-HTKs…..……..…104
Figure 3.4: High-passage 18-HTKs have an increased growth rate and reduced doubling time……...…...105
Figure 3.5: CTCF and YY1 protein expression increases in high-passage 18-HTKs…….108
Figure 3.6: CTCF and YY1 transcript expression is not significantly altered in high-passage 18-HTKs……...109
Figure 3.7: Assessment of chromatin shearing efficiency in low- and high-passage 18-HTKs. …….....……..113
Figure 3.8: Linearised HPV18 genome with forward and reverse primer binding sites……...……114
Figure 3.9: Reduced CTCF enrichment in the E2 ORF of high-passage 18-HTKs (Donor 1). ……....115
Figure 3.10: Reduction of CTCF binding to the HPV18 genome in high-passage HTKs (Donor 2). …….116

Figure 3.11: Reduced abundance of the CTCF binding site (E2 ORF) in high-passage 18-HTKs.....117
Figure 3.12: Increased YY1 enrichment at the URR of high-passage HPV18 HTKs (Donor 1). ……118
Figure 3.13: Increased YY1 enrichment at the URR in high-passage 18-HTKs (Donor 2). ……119
Figure 3.14: Increased chromatin accessibility at the viral URR in high-passage 18-HTKs (Donor 1). ……122
Figure 3.15: Increased chromatin accessibility at the viral URR in high-passage 18-HTKs (Donor 2). ……123
Figure 3.16: Depletion of YY1 in low-passage HFKs results in a significant increase in E6/E7 transcript expression.....129
Figure 3.17: Depletion of CTCF in low-passage 18-HTKs results in a significant increase in E6/E7 transcript expression.....130
Figure 3.18: Depletion of YY1 in low-passage 18-HTKs results in a significant increase in E6/E7 transcript expression.....131
Figure 3.19: Depletion of CTCF in high-passage 18-HTKs does not alter E6/E7 transcript expression. ……132
Figure 3.20: Depletion of YY1 in high-passage 18-HTKs results in a non-significant reduction in E6/E7 transcript expression……133
Figure 4.1: HPV16 genomes remain episomal in high-passage HTKs……137
Figure 4.2: Increased E6/E7 viral oncogene transcription in high-passage 16-HTKs.....138
Figure 4.3: Late passage 16-HTKs show a non-significant increase in growth rate…139
Figure 4.4: CTCF and YY1 transcription is not significantly altered in high-passage 16-HTKs……143
Figure 4.5: Increased CTCF and YY1 protein expression in UT, low- and high-passage 16-HTKs. ……144
Figure 4.6: Assessment of shearing efficiency in low- and high-passage 16-HTKs……145
Figure 4.7: Linearised HPV16 genome with forward and reverse primer binding sites……146

Figure 4.8: Reduced CTCF enrichment in the E2 ORF of high-passage 16-HTKs…..…….....147
Figure 4.9: Chromatin accessibility is unchanged at the viral URR in high-passage 16-HTKs. …….....….152
Figure 4.10: Depletion of CTCF and YY1 in high-passage 16-HTKs does not alter E6/E7 transcript expression..…….....153
Figure 5.1: Loss of episomal HPV16 genomes in SCC147 and SCC104.....158
Figure 5.2: Identification of the virus-host fusion integration events in SCC147 cell lines.....159
Figure 5.3: CTCF and YY1 protein expression is unchanged between HPV- and HPV+ OPSCC cell lines…….163
Figure 5.4: Assessment of chromatin shearing efficiency in SCC147 and SCC104…….....……..…164
Figure 5.5: CTCF binding to the human genome in SCC147 and SCC104 cell lines……...165
Figure 5.6: CTCF binding to the HPV16 genome of SCC147 cell lines……..166
Figure 5.7: YY1 binding to the HPV16 genome of SCC147 cell lines.....167
Figure 5.8: CTCF binding to the HPV16 genome of SCC104 cell lines.....168
Figure 5.9: YY1 binding to the HPV16 genome of SCC104 cell lines…….169
Figure 5.10: Enrichment of H3K27Ac at the HPV16 early promoter in SCC147…….173
Figure 5.11: Enrichment of H3K4Me3 at the HPV16 early promoter in SCC147……174
Figure 5.12: Enrichment of H3K27Ac at the HPV16 early promoter in SCC104…….175
Figure 5.13: Enrichment of H3K4Me3 at the HPV16 early promoter in SCC104……176
Figure 5.14: Enrichment of H3K27Me3 at the HPV16 late promoter in SCC147……177
Figure 5.15: Enrichment of RNAPolII at the HPV16 early promoter in SCC147…….178
Figure 5.16: Enrichment of RNAPolII at the HPV16 early promoter in SCC104…….179
Figure 5.17: Enrichment of RING1B on the HPV16 genome in SCC147 cell lines…….…...184
Figure 5.18: Enrichment of H2AK119Ub on the HPV16 genome in SCC147 cell lines……...185
Figure 5.19: Enrichment of RING1B on the HPV16 genome in SCC104 cell lines..186
Figure 5.20: Enrichment of H2AK119Ub on the HPV16 genome in SCC104 cell lines……...187

Figure 5.21: Depletion of RING1B in SCC147 cell lines does not alter chromatin accessibility at the HPV16 URR……..…..188
Figure 5.22: Depletion of RNF2 in SCC147 does not alter E6/E7 transcript expression……..….189
Figure 5.23: Depletion of CTCF in SCC147 results in unchanged E6/E7 transcript expression…….…..194
Figure 5.24: Depletion of CTCF in SCC104 results in a non-significant increase in E6/E7 transcript expression…….……..…195
Figure 5.25: Depletion of YY1 in SCC147 results in significantly reduced E6/E7 transcript expression..……...…..196
Figure 5.26: Depletion of YY1 in SCC104 results in significantly reduced E6/E7 transcript expression……..197
Figure 5.27: Depletion of YY1 in SCC147 cell lines results in reduced chromatin accessibility at the URR. ……198
Figure 6.1: Proposed model for switch in function of YY1 from a transcriptional repressor to transactivator of the HPV16 early promoter in OPSCCs……..226

LIST OF TABLES

Table 2.1: Details of oropharyngeal squamous cell carcinoma lines (OPSCC)…..…..87
Table 2.2: Dharmacon ON-TARGET plus SMARTpool siRNA target sequences..…..88
Table 2.3: Primary antibodies……..…..…89
Table 2.4: Primer sequences for HPV16……..…..90
Table 2.5: Primer sequences for HPV18……..…91
Table 2.6: Primer sequences for the human genome……...…92
Table 5.1 Summary and comparison of the genome status, CTCF enrichment, YY1 enrichment, E6/E7 expression and CTCF/YY1 protein expression in HPV16 high-passage HTKs, HPV18 high-passage HTKs and HPV16 positive OPSCC cell lines……170

ABBREVIATIONS

Abbreviation	Definition
AP-1	Activator protein 1
APOT	Amplification of Papillomavirus Oncogene Transcripts
ATM	Ataxia telangiectasia mutated
ATP	Adenosine 5'-triphosphate
ATR	ATM- and Rad3 related
BCL-2	B-cell lymphoma 2
BRCA1	Breast cancer type 1 susceptibility protein
BRD4	Bromodomain containing protein 4
BSA	Bovine serum albumin
CBX	Chromobox protein
CDK1	Cyclin dependent kinase 1
C/EBP-β	CCAT-enhancer binding protein beta
CF I/II	Cleavage factor I/II
ChIP	Chromatin immunoprecipitation
Chk1/2	Checkpoint kinase 1/2
CIN	Cervical intraepithelial neoplasia
CPSF	Cleavage and polyadenylation specificity factor
Ct	Cycle threshold
CTCF	CCCTC-binding factor
CtsF	Cleavage stimulation factor
DHFR	Dihydrofolate reductase

DLG1	Discs large 1
DMEM	Dulbecco's modified Eagle Medium
DIPS-PCR	Detection of integrated papillomavirus sequences-PCR
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
DTT	Dithiothreitol
EBV	Epstein-Barr virus
ECM	Extracellular matrix
EDTA	Ethylenediaminetetraacetic acid
EED	Embryonic ectoderm development
EGF	Epidermal growth factor
EGFR	Epidermal growth factor receptor
EV	Epidermodysplasia verruciformis
EZH2	Enhancer of zeste homolog 2
FAIRE	Formaldehyde assisted isolation of regulatory elements
FANCL2	Fanconi anaemia group L protein 2
FBS	Foetal bovine serum
FIGO	Federation of Gynaecology and Obstetrics
FISH	Fluorescent in situ hybridisation
FOXA1	Forkhead box protein A1
GAG	Glycosaminoglycan
GATA3	GATA-binding protein-3
GEO	Gene Expression Omnibus
HAT	Histone acetyltransferase

HDAC	Histone deacetylase
HLA	Human leukocyte antigen
HIV	Human immunodeficiency virus
HNSCC	Head and neck squamous cell carcinoma
HPV	Human papillomavirus
hSCRIB	Human scribble
HSPG	Heparin sulphate proteoglycan
HTK-16	Human tonsil keratinocytes maintaining HPV16 genomes
HTK-18	Human tonsil keratinocytes maintaining HPV18 genomes
H3K4Me3	Histone H3 trimethyl lysine 4
H2AK119Ub	Histone 2A lysine 119 monoubiquitination
H3K27Ac	Histone H3 acetyl lysine 27
H3K27Me3	Histone H3 trimethyl lysine 27
IMRT	Intensity-modulated radiation therapy
Kb	Kilobase
KMT	Lysine methyltransferase
KSHV	Kaposi’s sarcoma associated herpesvirus
LCR	Long control region
LncRNAs	Long noncoding RNAs
MAGI	MAGUK with inverted domain
MED1	Mediator of RNA polymerase II transcription subunit
MHC I/II	Major histocompatibility complex I/II
miRNA	MicroRNA
mRNA	Messenger RNA

MUPP1	Multi PDZ domain protein 1
MYBL2	Myb protooncogene-like 2
NaCl	Sodium chloride
NaOH	Sodium hydroxide
NF-1	Nuclear factor-1
Oct-1	Octamer transcription factor 1
OPC	Oropharyngeal cancer
ORF	Open reading frame
PAP	Poly(A) polymerase
PAS	Poly(A) signal
PBM	PDZ binding motif
PBS	Phosphate buffered saline
PCAF	p300-CBP-associated factor
PcG	Polycomb group proteins
PDZ	PSD95/DLG/ZO-1
PMSF	Phenylmethylsulfonyl fluoride
pRb	Protein retinoblastoma
PRMT1	Protein arginine methyltransferase 1
PsV	Pseudovirus
PTM	Post-translational modification
PV	Papillomavirus
PVDF	Polyvinylidene difluoride
qRT-PCR	Quantitative reverse transcription PCR
RNA Pol II	RNA polymerase II

RNF2	Ring finger protein 2
RNA	Ribonucleic acid
RPA	Replication protein A
RPM	Revolutions per minute
RYBP	RING1 and YY1 binding protein
SCJ	Squamocolumnar junction
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
SIN	Squamous intraepithelial neoplasia
siRNA	Small interfering RNA
SMURF2	SMAD specific E3 ubiquitin protein ligase 2
Sp1	Specificity protein 1
STAT5	Signal transducer and activator of transcription 5
SUZ	Suppressor of zeste
TBE	Tris-buffered saline
TBP	TATA binding protein
TBS-T	Tris-buffered saline containing 0.1% Tween 20
TCGA	The Cancer Genome Atlas
TE	Tris-EDTA
TEF-1	Transcription enhancer factor-1
TEM	Tetraspanin enriched microdomains
TFAP2B	Transcription factor AP-2 beta
TFCP2	Transcription factor CP2
TFIIB	Transcription factor II B
TFIID	Transcription factor II D

TPA	Tissue plasminogen activator
Tris-HCL	Tris-hydrochloride
URR	Upstream regulatory region
UT	untransfected
UTR	Untranslated region
UV	Ultraviolet
VLP	Virus like particle
YAF2	YY1 associated factor 2
YY1	Yin yang 1

Chapter 1 Introduction

1.1 The human papillomavirus

1.1.1 Introduction

The human papillomavirus (HPV) is a small, non-enveloped double stranded DNA virus with a tropism for mucosal and cutaneous epithelia at multiple anatomical sites. HPV is the most common sexually transmitted infection and viral spread occurs through skin-skin contacts via microabrasions, fissures and wound sites (Forcier and Musacchio, 2010). Although common, the majority of HPV infections are successfully cleared by the host immune response. However, a subset of HPV infections are able to persist which results in growth of benign lesions such as warts or malignant lesions such as carcinomas.

The etiological link between HPV and cervical cancer was first described over thirty years ago following isolation of HPV DNA from cervical biopsies (Dürst et al., 1983, Boshart et al., 1984). It is now well established that infection with oncogenic HPV is accountable for 99.7% of cervical cancers worldwide (Walboomers et al., 1999). Infection with oncogenic HPV is also a significant contributing factor to the development of head and neck squamous cell carcinoma (HNSCC) (Syrjänen et al., 1983, Syrjänen et al., 1987, Andl et al., 1998, Park et al., 1991). There is evidence that HPV-associated carcinomas at this anatomical site arise from the mucosal epithelium of the oral cavity, nasopharynx, oropharynx, hypopharynx and larynx (Johnson et al., 2020, Niedobitek et al., 1990). Recent estimations suggest that ~4%

of cancers are attributable to oncogenic HPV infection, highlighting the continuing burden of HPV infections worldwide (de Martel et al., 2020).

1.1.2 Classification and types

Human papillomavirus, classified as part of the *Papillomaviridae* (PV) family, are comprised of five distinct genera: alpha (α), beta (β), gamma (γ), mu (μ) and nu (ν) as shown in **Figure 1.1.** Each genus can be further classified into lower-order clusters termed species and type (de Villiers et al., 2004). Classification of papillomavirus is based on nucleotide differences in the L1 open reading frame (ORF) where distinct genera share less than 60% identity in the L1 ORF, distinct species share between 60-70% identity in the L1 ORF and distinct types share between 71-89% identity in the L1 ORF (de Villiers et al., 2004). To date, there are over 450 known types of human papillomavirus that have been identified (http://pave.niaid.nih.gov).

The alpha genus is the largest of the five genera and can be divided into three different classifications denoted cutaneous low-risk, mucosal low-risk and mucosal high-risk – based on their oncogenic potential (Doorbar et al., 2015). There are thirteen HPV types that are classified as high-risk, that is, they are carcinogenic to humans. These are HPV16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59 and 66, as recognised by the International Agency for Research on Cancer (IARC) which is the specialised cancer agency of the World Health Organisation (IARC, 2007). HPV16 is reported in the highest number of these cases with approximately 60% of global cervical carcinomas containing HPV16 DNA in comparison to 10-16% containing HPV18 (de Sanjose et al., 2010, Li et al., 2011a). In cancers of the head and neck

HPV16 is also the most prevalent virus type with over 80% of cancers at this site being driven by HPV16 (Kreimer et al., 2005, Castellsagué et al., 2016, Ndiaye et al., 2014). A subset of oropharyngeal carcinomas are positive for HPV18, HPV33 and HPV31 (Castellsagué et al., 2016). In comparison, the low-risk HPV types cause little to no disease. Instead, low-risk HPV types are those associated with the development of benign growth such as genital warts. This includes HPV6 and HPV11 which are part of the alpha genus (low-risk mucosal) and have been classified as a class 2B carcinogen by the IARC and can therefore be defined as possibly carcinogenic to humans (IARC, 2007). These types have been identified in human genital warts (*condyloma acuminatum*) (Gissmann and zur Hausen, 1980, Gissmann et al., 1983). HPV11 was also detected in laryngeal and tracheal papillomatoses, flat condylomatous and a subset of malignant cervical tumours (Gissmann et al., 1983). HPV42 is classified as a low-risk mucosal HPV which is ordinarily associated with the induction of benign warts. However, HPV42 has recently been linked to the development of digital papillary adenocarcinoma (DPA) – a skin cancer often found in fingers and toes (Leiendecker et al., 2023). Through performing a screen of 214 samples from 19 tumours, the authors found that HPV42 was present in 100% of DPAs. This was further confirmed by hybridisation-based sequence capture assay. Interestingly, expression of HPV42 episomes in primary human keratinocytes can extend the lifespan of the cell (Leiendecker et al., 2023).

The beta papillomavirus group is the second largest genus that exclusively infect cutaneous epithelia. Beta HPVs are responsible for the formation of benign cutaneous lesions such as flat warts, however, have also been implicated in the formation of cutaneous squamous cell carcinoma in immunocompromised patients

and in patients who have a rare genetic disorder known as epidermodysplasia verruciformis (EV). Epidermodysplasia verruciformis patients are highly susceptible to HPV infection and HPV5 and HPV8 have both been isolated from cutaneous squamous cell carcinomas derived from these patients (Orth, 1986). As such, HPV5 and HPV8 are classified as type 2B carcinogens and are defined as carcinogenic in patients who have epidermodysplasia verruciformis (IARC, 2007).

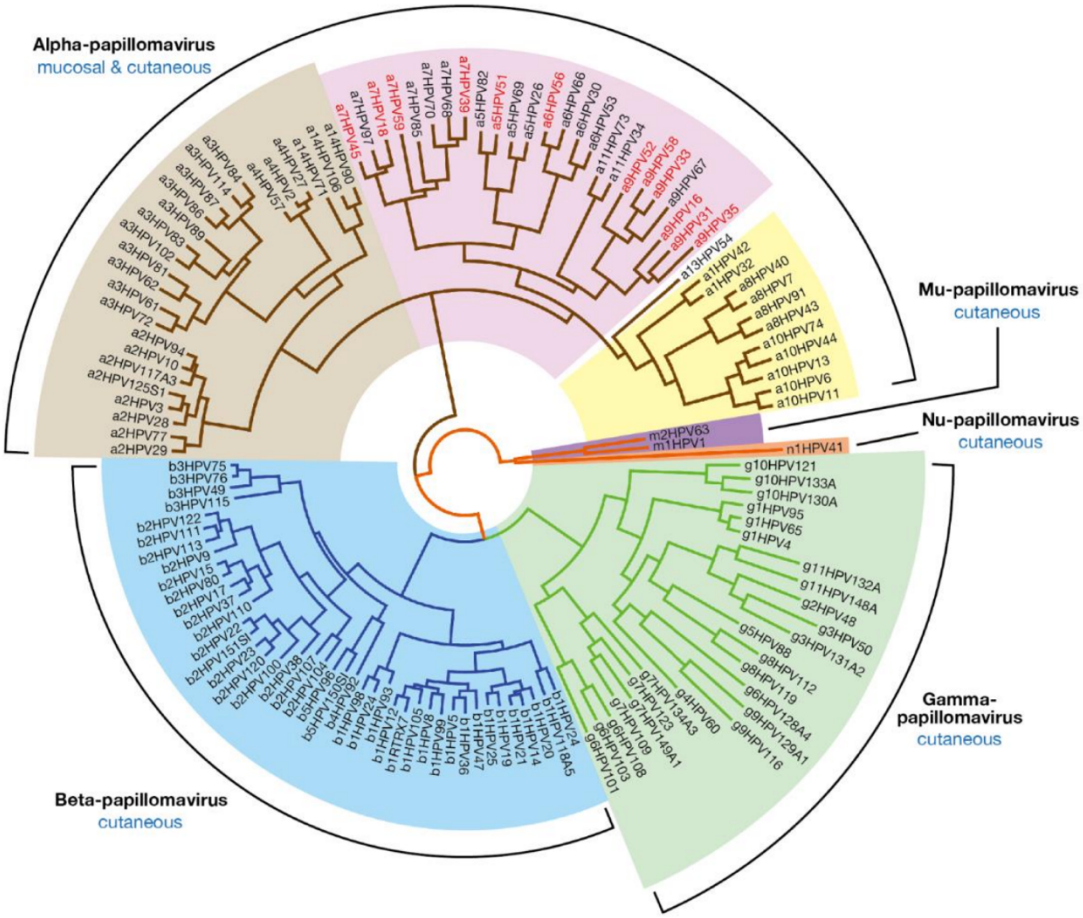


Figure 1.1 Evolutionary map of human papillomavirus. The HPV family are classified into five distinct genera alpha (α), beta (β), gamma (γ), mu (μ) and nu (ν). HPV-types from the alpha genus comprise the largest of the five groups and can be separated into low-risk cutaneous (brown), low-risk mucosal (yellow) and high-risk mucosal (pink). The high-risk papillomaviruses that are carcinogenic to humans are highlighted in red. The beta-papillomavirus (blue) is the second largest of the five groups. Beta-papillomavirus exclusively infect cutaneous epithelia. The remaining genera all infect cutaneous epithelia and are considered as low-risk. These include the Gamma-papillomavirus (green), Mu-papillomavirus (purple) and Nu-papillomavirus (orange). Figure taken from (Egawa et al., 2015).

1.1.3 Epidemiology and prevalence

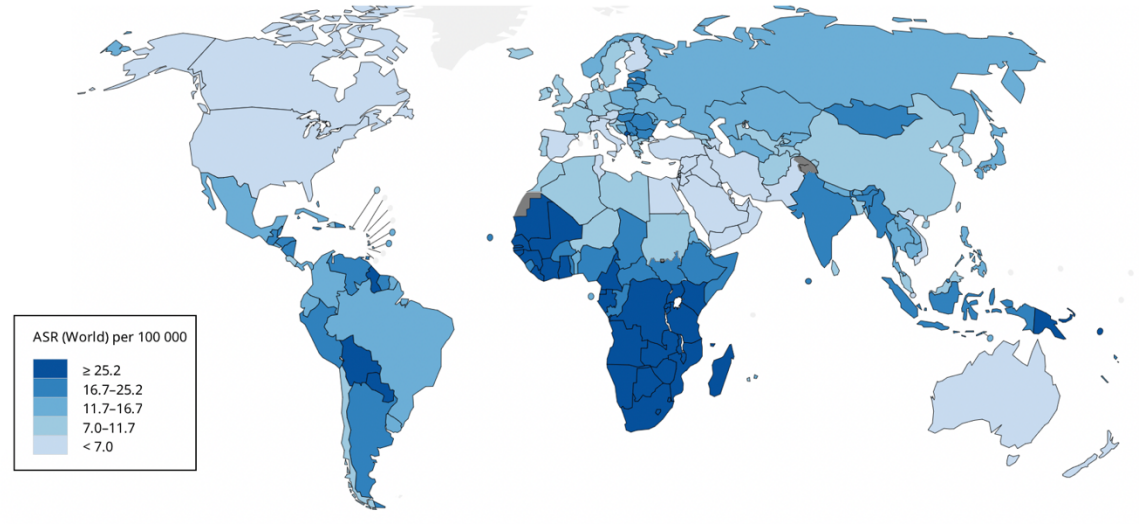
The worldwide burden of cancer incidence attributable to HPV was estimated at 690,000 (~4%) cases globally when assessed by the Global Cancer Observatory (GLOBCAN) in 2018. Over half of these cases corresponded to cancer of the uterine cervix (570,000 cases) (de Martel et al., 2020). More recent estimates suggest that cervical cancer is the fourth most frequently diagnosed cancer in women with an estimated 604,000 cases worldwide in 2020 (Sung et al., 2021). The global prevalence of HPV in women with normal cervical cytology was estimated to be at 10% in a previous meta-analysis study. The distribution of these cases was varied and the highest incidence of HPV infection was found in Africa, Central America and Mexico (de Sanjosé et al., 2007). HPV16 and HPV18 were the most common types detected and the global prevalence was highest in women under the age of 35 (de Sanjosé et al., 2007). The global estimated age standardised incidence rate of cervical cancer was 13 per 100,000 (Arbyn et al., 2020). There is a correlation with lower-income countries and increased incidence of HPV-associated cancer (de Martel et al., 2020). A breakdown of the global age-standardised rates of cervical cancer is shown in **Figure 1.2A** which demonstrates that the lowest incidence rates are found in North America, Australia and parts of the Middle East with the age standardised rates in 2020 being less than 7 per 100,000. The highest incidence rates were in Southern Africa and parts of West Africa where there is over 25 cases per 100,000 **(Fig 1.2A).** Cervical cancer is the fourth leading cause of death in women with an estimated ~300,000 deaths globally per year. A significant proportion of these deaths (>85%) occur in low- and middle-income countries (Arbyn et al., 2020).

The estimates for the number of cases of head and neck cancer were at 600,000 worldwide in 2018 (de Martel et al., 2020). The global age-standardised rate of HNSCC are shown in **Figure 1.2B** which demonstrates that the lowest incidence rates occurred in Mexico, regions of Western South America and Central and Western Africa with the age-standardised incidence rates in 2020 less than 3.5 per 100,000. In contrast, the highest incidence rates were found in Western Europe, parts of Southern Europe, Central and Eastern Europe, Russia, parts of South Asia and Australia with an age-standardised incidence rate >10 per 100,000 **(Fig 1.2B).** The International Agency for Research on Cancer describes multiple risk factors associated with development of HNSCC including alcohol consumption, tobacco consumption and exposure to Epstein-Barr Virus (EBV) and HPV (predominantly type 16) (IARC, 2020). HPV-associated HNSCC cases were estimated at 52,000 globally in 2018 representing a ~24% increase in the age-standardised incidence rate since previous reports from 2012 (38,000 cases globally) (de Martel et al., 2020, de Martel et al., 2017). The majority of these cases correspond to oropharyngeal carcinoma – comprising the tonsil and base of tongue sites (de Martel et al., 2017). In the US, deaths attributable to cancer of the oral cavity and pharynx were 11,230 in 2022 which represents a 0.4% increase per year from 2010 to 2019 (Siegel et al., 2022). Interestingly, whilst HPV-driven cancers of this site are increasing, there appears to be a decline in HNSCC related to other external factors, such as smoking (Siegel et al., 2022). There are several lines of evidence suggesting that HPV is a prognostic marker for overall survival and progression-free survival in oropharyngeal carcinomas. Patients with HPV-positive carcinoma when compared to negative demonstrate a significantly increased survival rate and progression-free

survival. This was consistent at four different subsites (oropharynx, hypopharynx, larynx subsites, oral cavity) (Ang et al., 2010, Li et al., 2018, Zhou et al., 2021). Patients with HPV-positive HNSCC also show a better response to radiotherapy and following treatment have an increased overall survival when compared to HPV-negative patients (Zhou et al., 2021).

The remaining cancer sites attributable to HPV include the anus, penis, vagina and vulva which make up approximately 12% of HPV associated cancers (de Martel et al., 2017). HPV infections (predominantly HPV16) can cause anus squamous cell carcinoma. The prevalence of HPV found in anal cancer is 90%. This is however higher in patients who are also positive for human immunodeficiency virus (HIV) with 100% of these cases being HPV-positive (Lin et al., 2018). Recent estimations suggest that 29,000 cases of anal cancer are attributable to HPV and over half of these cases arise in women (de Martel et al., 2020). Vagina and vulva cancer attributable to HPV were estimated at 62,000 in 2020. Vagina cancer is considered relatively rare and only accounts for 2.5% of HPV driven cancers. They usually comprise squamous cell carcinoma although adenocarcinoma also occurs at this site (Rajaram et al., 2015). Often, age is a risk factor for vaginal cancer and a large proportion occur in post-menopausal or elderly women. However, increasing incidence in young women is attributable to sexually transmitted HPV infections (Rajaram et al., 2015). Vulval cancers are predominantly squamous cell carcinoma (90%) but also include melanoma, basal cell carcinoma and adenocarcinoma (Morrison et al., 2020). The incidence of vulval cancer has increased over the past twenty years and more cases are occurring in women aged 50-59 which likely represents an increase in HPV associated vulval cancer (Morrison et al., 2020).

A



B

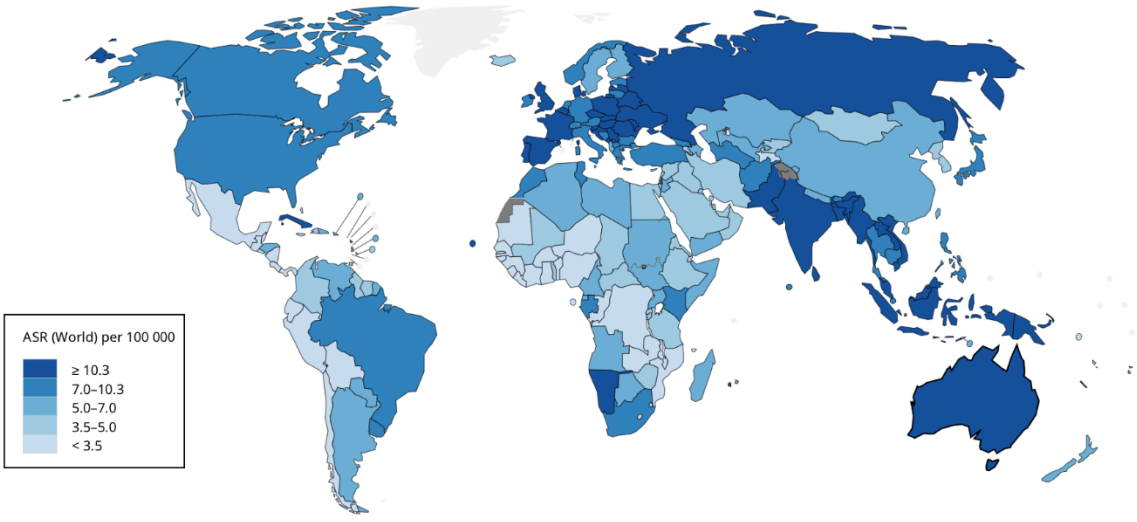


Figure 1.2 Estimated age-standardised incidence rates of cervical and head and neck cancer. The age standardised incidence rates of **(A)** cervical and **(B)** head and neck cancer (GLOBOCAN 2020). Maps were generated using the IARC cancer today visualisation tool (https://gco.iarc.fr/today/home).

1.1.4 Cervical screening

Cervical screening was introduced with the aim to diagnose intraepithelial precursor lesions of the cervix in order to reduce both the morbidity and mortality of cervical cancer. There are three modalities for cervical screening, as published by the world health organisation in 2013 and 2014, which can either be performed independently or in combination (Jeronimo et al., 2017). The Papanicolaou test (also referred to as pap test) assesses cervical cytology. The practising pathologist will diagnose biopsied tissues based on the thickness of the transformed epithelium which will be classified as negative or as a cervical intraepithelial neoplasia (CIN). Depending on the thickness of the neoplasia the CIN can be graded as CIN1 (grade 1; one third thickness of epithelium), CIN2 (grade 2; up to two thirds thickness of epithelium) or CIN3 (grade 3; two thirds or greater thickness of epithelium) (Jeronimo et al., 2017). If cytology results are abnormal then further confirmation is required in the form of colposcopy (Jeronimo et al., 2017). As an alternative, visual inspection with acetic acid (VIA) is a method to visualise the cervix to detect precancerous lesions. This technique is an affordable alternative to pap tests and HPV screening, especially in low income countries (Jeronimo et al., 2017, Azene, 2021). In addition to assessing cervical cytology, The World Health Organisation now recommend detection of HPV DNA as the primary screening method. This is more sensitive than cytology for detecting precancerous lesions. HPV screening methods include target- and signal-amplification assays of HPV DNA and HPV mRNA, and these can either be performed in conjunction with cervical cytology or alone (Perkins et al., 2020).

Regular cervical screening is associated with reduced incidence of cancer. Detection of cervical cancer diagnosed by cervical cytology results in a better clinical outcome than for women diagnosed due to symptoms. The five-year relative survival is 26% higher for those detected via screening compared to symptoms (69% vs 95%) (Andrae et al., 2012). In England, regular screening of women of all ages has been shown to prevent 70% of all cervical cancer deaths, this is estimated to be increased to 83% if everyone attended regular screening (Landy et al., 2016). This is likely because of effective treatment of the precursor lesions (Jeronimo et al., 2017). Treatments for pre-cancerous lesions include ablation with cryotherapy or thermal ablation or loop electrosurgical excision of the transformation zone.

1.1.5 HPV vaccination programme

The past 15 years has seen advances in the development of a prophylactic vaccine against HPV. In the UK, routine vaccination for HPV was introduced in 2008 for girls aged 12-13 years. This was subsequently expanded in 2019 when the HPV vaccine rollout was extended to boys aged 12-13 years. An additional programme is also available for gay and bisexual males up to 45 years of age in efforts to protect against HPV-driven cancers of the anus, penis, mouth and throat. The vaccination programme is offered as part of the NHS – previously it was recommended that three doses of the vaccine were required, however, standard guidelines set out by the centre for disease control and prevention (CDC) currently recommend two doses in a 0, 6–12-month schedule. Those who are immunocompromised, or HIV positive are however recommended three doses of the vaccine.

There are currently three licensed HPV vaccines that are approved for use: Gardasil, Cervarix and Gardasil-9. These are adjuvant vaccines produced using 360 copies of the HPV structural protein L1 which self-assemble into virus-like particles (VLPs) resembling the viral capsid (Schiller et al., 2012, Cuzick, 2015). They contain no viral genome and are therefore non-oncogenic and non-infectious (Yousefi et al., 2021). Gardasil® (Merck) is a quadrivalent vaccine composed of VLPs containing HPV6/11/16/18. Whilst HPV6 and HPV11 are classified as low-risk types, given that HPV is the most commonly sexually transmitted infection, genital warts are considered as a significant social and economic burden. Vaccination against these types is therefore important for reducing costs to the health service whilst also positively impacting quality of life (Woodhall et al., 2011). Cervarix® (GlaxoSmithKline) is a bivalent vaccine which protects against the most common high-risk types HPV16 and HPV18. In 2014, a second generation prophylactic HPV vaccine was approved known as Garsasil-9® (Merck) which is a 9-valent vaccine that offers protection against five additional high-risk types of HPV (HPV31/33/45/52/58) which successfully prevents cervical, vulvar and vaginal disease associated with these HPV-types (Joura et al., 2015, Yousefi et al., 2021). Both Gardasil vaccines are produced using *Saccharomyces cerevisiae*, a yeast expression system, whilst the Cervarix vaccine is produced using *Trichoplusia ni*, an insect cell line containing recombinant baculovirus (Pinto et al., 2018).

The quadrivalent vaccine has high prophylactic efficacy against HPV and infection with HPV6/11/16/18 was reduced by over 90% following administration (Kjaer et al., 2009). Several randomised, double-blind, placebo-controlled clinical trials found the quadrivalent vaccine to prevent HPV-associated disease including anogenital warts,

CIN2 or worse, high-grade cervical, vaginal and vulval intraepithelial neoplasia in women aged 15-26 (Villa et al., 2005, Garland et al., 2007, Future, 2007, Dillner et al., 2010). Vaccine efficacy was found to be 96% for CIN1, 100% for vulval and vaginal intraepithelial neoplasia grade I and 99% for condyloma (Dillner et al., 2010). Introduction of HPV vaccination has resulted in a significant reduction in the prevalence of HPV16 and HPV18. After 5-8 years post-vaccine the incidence rates in women aged 13-19 had reduced by 83%, women aged 20-24 had reduced by 66% and women aged 25-29 had reduced by 37% (Drolet et al., 2019). Administration of the quadrivalent vaccine substantially reduced risk of invasive cervical cancer (Lei et al., 2020). The risk of getting cervical cancer was 88% lower in women who were vaccinated before age 17 when compared to non-vaccinated women (Lei et al., 2020).

Immunisation of girls aged 10-14 with the bivalent vaccine resulted in 100% seroconversion (Pedersen et al., 2007). The bivalent vaccine has been shown to have 90% efficacy against CIN2+ lesions containing either HPV16 or HPV18 (Paavonen et al., 2007). Detection of high-grade cervical intraepithelial neoplasia (CIN2) associated with HPV16/18 decreased from 3% to 1% in women ages 24-25 years old. This suggests an estimated vaccine effectiveness of 87% (Rebolj et al., 2022). Vaccination with Gardasil-9® resulted in a 98.4% reduction in cervical disease of any grade and a 98.2% reduction in high grade cervical neoplasia. There was a reduction in disease related to all nine HPV types. There was high efficacy against vulval and vaginal disease including condylomas and an overall efficacy of between 94.9-100% against HPV6, 11, 16 and 18 (Giuliano et al., 2019).

The contribution of the HPV vaccination programme to preventing disease of the head and neck has not been as well explored. Pre-clinical trials using murine models have assessed the efficacy of the Gardasil® vaccine on preventing oral disease. This trial utilised HPV16 E6/E7 pseudovirus (PsV) and found that following immunisation, all oral HPV16 PsV infection was prevented in mice (Ahn et al., 2018). A recent phase III double-blind, placebo-controlled clinical trial has begun to explore both the efficacy, immunogenicity and safety of the use of the 9-valent vaccine for preventing HPV oral persistent infection in men. This study is for men between the age of 20-45 years of age and should provide an insight into whether the vaccine is able to prevent HPV-driven associated head and neck disease (Giuliano et al., 2022).

1.1.6 Current treatments for HPV-associated cancers

There is a widely accepted method for the staging of cervical cancer known as the Federation of Gynaecology and Obstetrics (FIGO) staging system. This ranges from stage I to stage IV and within each stage there are also further sub-categorisations (Singg, 2018). The treatment approaches for cervical cancer are dependent on the FIGO stage of the patient. Treatment options include surgery, radiotherapy, and chemotherapy, either used alone or in combination with one another. Surgery for cervical cancer is varied and can include a total or radical hysterectomy, a radical trachelectomy, a bilateral salpingo-oophorectomy or a total pelvic exenteration. The use of external-beam radiation therapy and internal radiation therapy can be used for cervical cancer treatment. Intensity-modulated radiation therapy (IMRT) is a common technique used to prevent damage to surrounding healthy tissue.

Chemotherapy is often in the form of Cisplatin treatment, this is often in combination with radiotherapy, but this can be used alone.

Treatment for head and neck cancer is often multidisciplinary and can be dependent on different factors including the stage of the cancer, HPV-status and risk factors. HPV-driven cancers can be classified as low-risk or high-risk depending on pack years of tobacco smoking, tumour stage and nodal stage (Ang et al., 2010). Treatment options can include surgery, radiotherapy and chemotherapy either as a single modality or in combination. Patients that present with early stage disease (stage I or II) are primarily treated with either surgery or radiotherapy (Pfister et al., 2020). The current standard of care for patients with locally advanced head and neck cancer is combined modality therapy including radiotherapy in combination with three weekly high-dose cisplatin treatment (100mg/m^2) (Pfister et al., 2020, Muzaffar et al., 2021).

HPV has been identified as a stratification factor for prognosis of head and neck cancer and these patients often present with high radiosensitivity. In combination with the high toxicity attributable to Cisplatin treatment, efforts are being made to de-intensify therapy based on HPV status and risk categories. Clinical trials such as the De-ESCALaTE open-label randomised controlled phase III trial and the RTOG 1016 randomised multicentre, non-inferiority trial both explored the use of Cetuximab (an EGFR inhibitor) for the treatment of HPV-positive, low-risk (<10 pack-years) oropharyngeal cancers. Both trials found that Cetuximab showed no benefit in terms of reduced toxicity and was detrimental to tumour control (Mehanna et al., 2019, Gillison et al., 2019).

1.1.7 HPV structure

Human papillomaviruses are non-enveloped, icosahedral DNA viruses with a diameter of 55 nm (Hagensee et al., 1993). The viral capsid is formed from the major and minor late (L) structural proteins L1 and L2 (Hagensee et al., 1993). The outer surface is composed of 360 monomers of the 55 kDa polypeptide L1 which are arranged in an icosahedral lattice composed of 72 capsomers (Baker et al., 1991, Joyce et al., 1999). The minor capsid protein L2 is hidden within the centre of these capsomere structures (Trus et al., 1997). Enclosed beneath the capsid is a double-stranded DNA genome of approximately 8000 base pairs (bp) (Van Doorslaer et al., 2018). Within infected keratinocytes, the genome exists in the form of extrachromosomal DNA – these HPV episomes reside within the nucleus of the infected cell and have been shown to associate with cellular histones to form nucleosome structures (Favre et al., 1977, Day et al., 2004, Graham, 2017). The viral genome consists of 7-9 open reading frames (ORFs) which encode the early genes (E1, E2, E5, E6, E7 and E8^E2) and the late genes (E1^E4, L1 and L2) **(Fig 1.3).** The expression of E5 and E8^E2 is not conserved amongst all HPV types – the spliced isoform E8^E2 is only expressed in high-risk HPV types including HPV16, HPV18 and HPV31 and E5 is only expressed by alpha-papillomaviruses (Straub et al., 2015, Basukala and Banks, 2021).

1.1.8 Structure of the upstream regulatory region

Downstream of the L1 open reading frame is a non-coding region termed the upstream regulatory region (URR; also called the long-control region (LCR)). The URR contains binding sites for both viral and cellular DNA binding proteins that act

on the downstream promoters to regulate transcription of HPV genes **(Fig 1.4).** Viral proteins known to bind the viral URR include the core proteins E1 and E2. E2 binds in the form of a dimer to the consensus palindromic motif ACCG$(N)_4$CGGT which can be found in the viral URR of all HPVs (Androphy et al., 1987, Hirochika et al., 1988, Hegde et al., 1992). The two major high-risk types HPV16 and HPV18 contain four conserved binding sites for E2, however, depending on virus type the number and location of these binding sites can vary. The 5' segment of the URR is flanked by the late polyadenylation site and an E2 binding site (E2 #4) which is situated within the enhancer. The epithelial-specific enhancer sits within the central region of the URR and just downstream from the this is the origin of replication (*ori).* Spanning an 80bp segment in the replication of origin is an E1 and an E2 binding site (E2 #3) (Mohr et al., 1990, Chiang et al., 1992, Liu et al., 1995). The proximal promoter is positioned within the 3' region of the URR and contains two E2 binding sites (E2 #1 #2) which are flanked by the TATA box downstream and a specificity protein 1 (Sp1) binding site upstream (Thierry and Yaniv, 1987, Demeret et al., 1994).

Several cellular transcription factors have been shown to bind to the viral URR and regulate transcription from the early promoter. In the HPV18 URR there are two binding motifs for Sp1. When bound to the consensus motif 5'-GGGAGT-3' within the HPV18 proximal promoter, Sp1 positively regulates E6/E7 transcription (Hoppe-Seyler and Butz, 1992). Because of the overlap between the Sp1 and E2 #2 binding site, E2 can counteract the activator function of Sp1 through direct interference with Sp1 binding (Gloss and Bernard, 1990). Transcriptional enhancer factor-1 (TEF-1) binds to the HPV16 URR and activates the activity of the early promoter (Ishiji et al.,

1992). The HPV16 URR contains ten binding motifs which contain the consensus sequences 5'-CAT-3', 5'-ACAT-3' and 5'-TCAT-3' that are all bound by the zinc finger protein Yin Yang 1 (YY1) – with binding sites #1 and #10 having the highest affinity for YY1 binding (O'Connor et al., 1996). In contrast HPV18 contains only seven binding sites for YY1. In both HPV16 and HPV18, YY1 binding sites are located within the proximal promoter, flanking the TATA box, within the origin of replication and in the enhancer. Binding of YY1 to these sites within the URR results in repression of HPV early gene transcription. Other cellular factors that bind the URR include nuclear factor-1 (NF1) which binds the consensus sequence 5'-TTGGC-3' in multiple HPV types including HPV16 and HPV18 (Gloss et al., 1989b, Gloss et al., 1989a, Chong et al., 1990). NF1 functions to positively regulate transcription from the HPV16 early promoter (Baldwin et al., 2007). Interestingly the function of NF1 as a transcriptional activator is dependent on binding of the octamer transcription factor-1 (Oct-1) which has known binding sites within the HPV16 and HPV18 enhancer region (Hoppe-Seyler et al., 1991, O'Connor and Bernard, 1995). Oct-1 is able to stabilise binding of NF1 resulting in increased enhancer activity (O'Connor and Bernard, 1995). Activator protein-1 (AP1) has been shown to bind to two sites within the HPV16 URR (Chan et al., 1990). In addition, several other transcription factors have been shown to bind to the HPV16 URR including C-Myb, Pax5, NFATx and WT1 (Carson and Khan, 2006). More recent studies into cellular transcription factor binding identified eight novel transcription factors that bind the URR to regulate transcription from the early promoter. GATA binding protein-3 (GATA3) and transcription factor AP-2 beta (TFAP2B) were shown to alter HPV18 transcription (Sichero et al., 2012). Binding of *c-myc* and Forkhead box protein A1

(FOXA1) were able to positively regulate transcription through direct binding to the URR (Sichero et al., 2012).

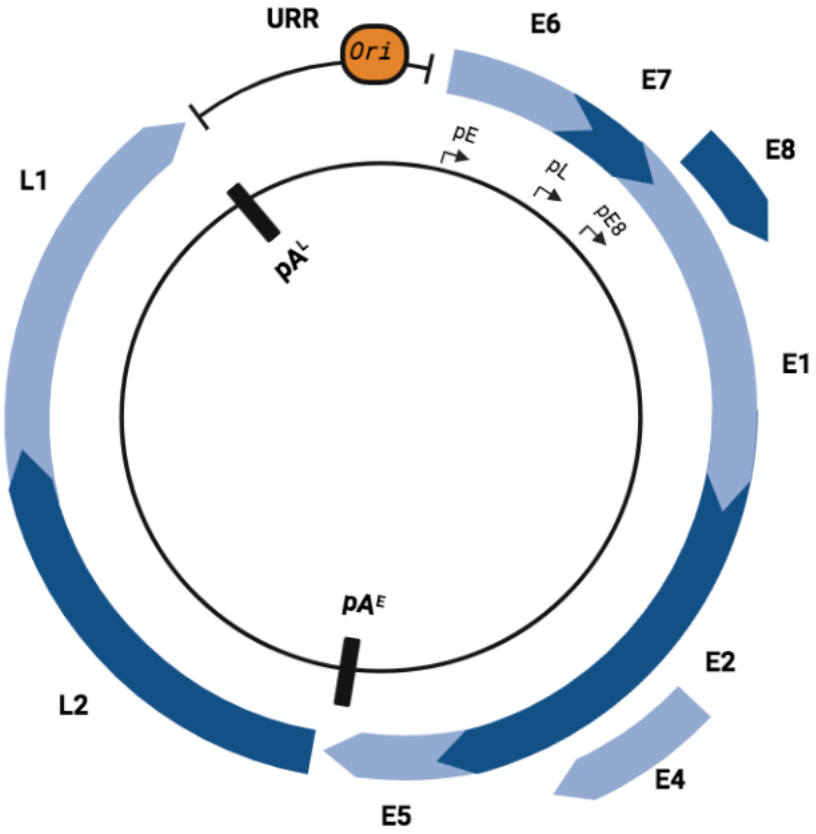


Figure 1.3 Organisation of the HPV genome. HPVs contain a double stranded, circular genome of approximately 8000 bp in length. The genome is organised into three major regions: the early gene region, the late gene region and a non-coding upstream regulatory region (URR). Papillomaviruses from the alpha genus encode nine open reading frames which can broadly be classified as regulatory (E1, E2, E4, E5, E6, E7, E8) and structural (L1, L2). The three major viral promoters are the early promoter (pE), late promoter (pL) and E8 promoter (pE8). Viral transcripts terminate at the early polyadenylation site (pAE) or late polyadenylation site (pAL). The replication origin (ori) is situated within the 5' region of the URR and is indicated in orange. Figure adapted from (Burley et al., 2020).

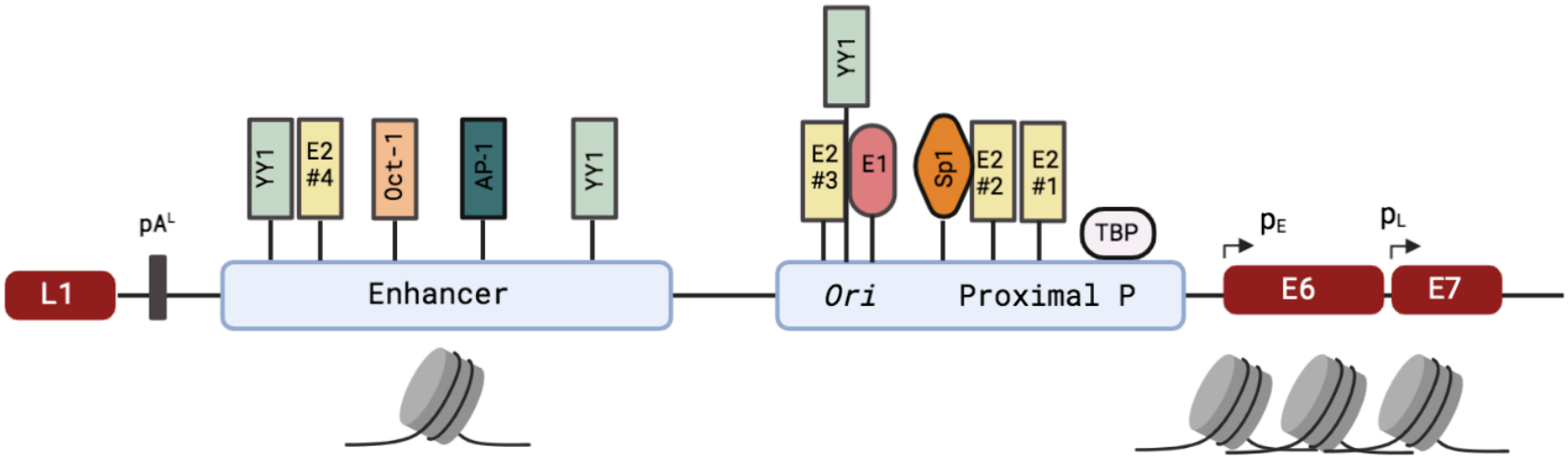


Figure 1.4 Structure of the HPV upstream regulatory region. The HPV URR spans across approximately 850 bp and contains multiple *cis*-regulatory elements that regulate transcription from the early promoter. The URR contains a central enhancer, origin of replication and proximal promoter which contains a TATA motif. There are four binding sites for HPV E2 and a single binding site for HPV E1 in the origin of replication. The URR is also bound by multiple cellular transcription factors including YY1, Oct-1, AP-1, and Sp1. Nucleosome occupancy around the URR is important for facilitating the epigenetic regulation of HPV gene expression. There are known nucleosome structures at the enhancer, across the E6 and E7 open reading frames. Figure adapted from (Burley et al., 2020).

1.2 The HPV life cycle

The life cycle of HPV is closely linked to the programme of keratinocyte differentiation. Details of the HPV life cycle are shown in **Figure 1.5**. Infection is targeted to basal keratinocytes in the epithelium, which maintain inherent proliferative capacity. During the maintenance phase of the life cycle HPV DNA is replicated synchronously with the host cell DNA and E2 tethers the virus to mitotic chromosomes to enable episome persistence and lateral expansion of the infected cells. E6 and E7 oncoproteins produce a replication competent environment by delaying the natural programme of keratinocyte differentiation through degradation of key cell cycle regulatory proteins pRb and p53. They also function to increase host-cell factors that are essential for facilitating genome replication. Following differentiation, in the intermediate layers of the stratified epithelium, in a process termed amplification, HPV genomes are replicated to thousands of copies per cell. A switch in promoter usage to the late promoter enables increased expression of late transcripts including E1^E4, L1 and L2. Finally L1 and L2 proteins are assembled into complete virus capsids and virions are released. Each stage of the HPV life cycle is detailed below.

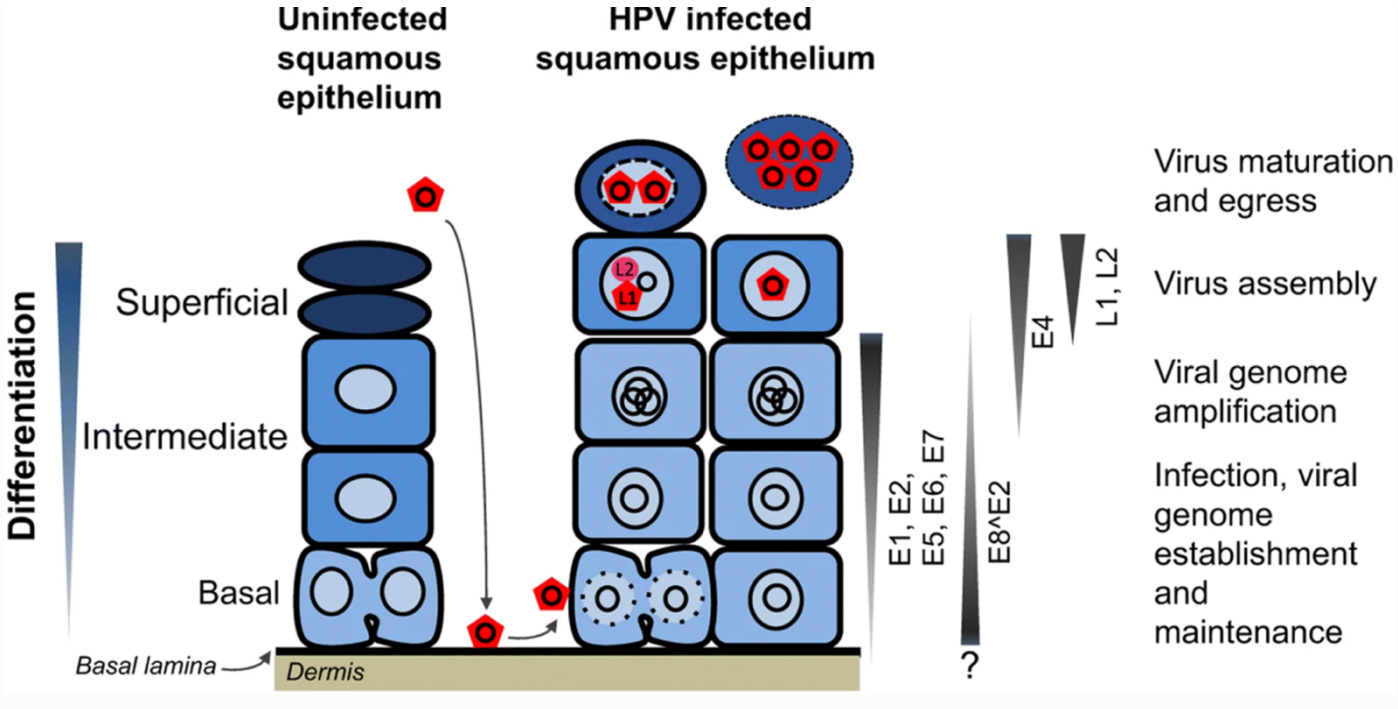


Figure 1.5 The HPV life cycle. HPV infects basal keratinocytes through wound sites and microabrasions and attaches to components of the extracellular matrix (ECM) such as heparin sulphate glycoproteins (HSGPs). Following internalisation the virus is trafficked to the nucleus and undergoes a transient round of replication to establish maintenance episomes. Expression of E1 and E2 initiate viral replication. E2 also tethers to mitotic chromosomes and coordinates equal viral genome segregation to expand the population of HPV-infected cells. Within the intermediate layers of the epithelium, differentiation is delayed through the effects of E6 and E7 which also function to produce a replication competent environment. Viral genomes are amplified to thousands of copies per cell. Within the superficial layers of the epithelium E1^E4, L1 and L2 are expressed at high levels. Capsid assembly and virion release is subsequently initiated. Figure taken from (Burley et al., 2020).

1.2.1 Initial infection and virion entry

Attachment and entry of HPV into the host cell is initiated through binding to the basement membrane – virion access is generally enabled following damage to tissue integrity, such that would be found at wound sites or abrasions. PVs interact with multiple components of the extracellular matrix (ECM) including glycoproteins such as heparin sulphate glycoproteins (HSPGs) that are expressed on both the basement membrane and the cell surface. The major capsid protein L1 is primarily involved in attachment and entry into host keratinocytes. The carboxyl terminal of L1 has been shown to directly interact with HSPGs (Joyce et al., 1999, Giroglou et al., 2001, Rommel et al., 2005, Selinka et al., 2007). This interaction is essential to facilitate attachment to the epithelial cell surface membrane and internalisation (Joyce et al., 1999). It should be noted that whilst some HPV types require direct binding to HSPG to facilitate infection, interaction with glycosaminoglycan (GAGs) is thought to be dispensable for HPV16 internalisation. HPV16 binding and entry has been shown to occur via a GAG independent pathway, suggesting the presence of a non-HSPG binding partner on the basement membrane (Cruz and Meyers, 2013). Alternative extracellular matrix membrane components, such as laminin-5, which is secreted by migrating and basal keratinocytes, have also been implicated in initial binding of certain HPV types. Laminin-5 has been show to colocalise with and facilitate the internalisation of HPV11 and HPV16 (Culp et al., 2006).

Following engagement with cell surface heparin sulphate receptors, the viral capsid proteins undergo a conformational change. The L2 conformational change is

facilitated by the cell surface protein cyclophilin B which is known to associate with HSPGs and has been found in multiple virus types (HPV16, 18, 45, 58) to expose the amino acid residues 17-36 on the L2 capsid protein (Bienkowska-Haba et al., 2009). The L2 minor capsid protein contains a consensus furin cleavage motif (R-X-K/R-R) which is conserved amongst all HPV types (Richards et al., 2006). Following exposure of the N-terminal domain of L2, this site undergoes proteolytic cleavage by furin, and these events are essential for a PV infection (Richards et al., 2006). Whilst the interaction with HSPG receptors is important for virion uptake, it is understood that this is due to the conformational changes induced by HSPG binding and not through facilitating cellular entry (Day et al., 2008). Viral entry is initiated by binding to a host-cell receptor; whilst the exact receptors used are currently unknown, there are suggestions that it could be the laminin binding integrins such as α3β1 or α6β1/β4 (Scheffer et al., 2013).

There have been several proposed mechanisms for virion entry into the target cell, yet the precise internalisation pathways remain elusive. All PV types enter the cell via the process of endocytosis which occurs asynchronously over multiple hours. Different HPV-types appear to exhibit different endocytic pathways. HPV31 for example, utilises a caveolin1- and dynamin 2- dependent pathway but does not depend on clathrin expression (Smith et al., 2007). Early studies looking into HPV16 endocytosis suggested a clathrin-dependent mechanism of entry into the host cell. HPV16 virions were shown to colocalise with the adapter protein AP-2 which is present in clathrin coated pits and biochemical inhibitors of clathrin-mediated endocytosis further supported this suggesting entry via a distinct route which was independent of dynamin and caveolin (Day et al., 2003, Smith et al., 2007). In

contrast there is also evidence to suggest that HPV16 cellular internalisation occurs via a clathrin-, caveolin-, flotillin- and dynamin-2-independent pathway (Scheffer et al., 2013, Schelhaas et al., 2012). Instead HPV16 uptake requires association with tetraspanins present on the cell surface membrane. Tetraspanins can interact with one another and form tetraspanin enriched microdomains (TEMs) – HPV16 associates with multiple tetraspanins including CD9, CD81, CD82, CD63 and CD51 (Scheffer et al., 2013, Spoden et al., 2008). Tetraspanin CD151 is thought to be directly involved in the endocytosis of HPV16 and it has been found to function as a chaperone to facilitate movement to sites of entry and is also co-internalised with HPV16 into the endosome (Scheffer et al., 2013). Additionally, actin dynamics were found to be important in HPV16 internalisation, and infection was dependent on actin polymerisation (Schelhaas et al., 2012). The retrograde pathway has also been shown to be implicated during viral entry into host cells and for virion transport from the endosome to the golgi-like compartments (Lipovsky et al., 2013). Capsid proteins L1 and L2 have been shown to colocalise and directly interact with golgi apparatus markers including GM130, and L2 additionally co-localises with the trans-golgi network marker TGN46 (Lipovsky et al., 2013).

1.2.2 Genome establishment and maintenance

Following capsid uncoating, a transient round of DNA replication amplifies viral episomes to relatively low expression levels (~50-100 copies per cell). During the establishment stages of the HPV life cycle episomal DNA is replicated synchronously with the host cell DNA. Essential during this process is the expression of the core proteins E1 and E2 (Ustav and Stenlund, 1991). This has

been supported by the findings that E1 and E2 are the first of the virally encoded transcripts to be expressed immediately after infection with HPV (Ozbun, 2002, Mirkovic et al., 2015). Full-length papillomavirus E2 is a DNA-binding protein which is encoded by the full length E2 open reading frame. E2 has multiple functions including, but not exclusive to, transcriptional regulation, viral DNA replication and genome maintenance. Structurally, E2 contains two highly-conserved functional domains, the N-terminus, which serves as the transactivation domain, and, the carboxyl terminus which is required for DNA binding (McBride et al., 1988). These two domains are connected by a central linker region termed the “hinge region” which is involved in the nuclear localisation of E2 during viral replication (Giri and Yaniv, 1988, Zou et al., 2000). The function of E2 during HPV replication is to load E1 at the replication of origin – these core proteins form an E1E2-*ori* complex which binds DNA with high sequence specificity and affinity (Sanders and Stenlund, 1998). The presence of ATP results in conformational changes that results in transition to a multimeric E1-*ori* complex (Sanders and Stenlund, 1998). In the absence of E2, E1 is still able to bind DNA and initiate replication. However, this is very non-specific and inefficient (Conger et al., 1999).

HPV replication initiates from the replication of origin, it occurs via theta structures and is bidirectional in nature (Melendy et al., 1995, Flores and Lambert, 1997). E1 is part of the ATPase associated with diverse cellular activities (AAA+) family and therefore binding to the viral URR is dependent on the presence of ATP (Liu et al., 1995). E1 is essential for both the initiation of DNA replication and elongation. It functions as a DNA helicase and unwinds super helical DNA and initiates replication in a 3’-5’ direction, E1 also directly interacts with the host-cell polymerase α-primase

complex (Seo et al., 1993, Liu et al., 1995, Conger et al., 1999). Several host cell enzymes are also required for successful replication including DNA polymerase alpha-primase and delta-primase and replication protein A (RPA) (Melendy et al., 1995). Within basal keratinocytes genome replication is limited, presumably to avoid detection from cellular immune proteins. The regulation of gene expression is detailed below (see section 1.5) but one such process relevant to the regulation of E1-mediated replication is the expression of the fusion protein E8^E2 (Choe et al., 1989). E8^E2 is the result of a splicing event between E8 and the 3' exon of E2 (Wang et al., 2011). E8^E2 functions to repress DNA replication by competing for E2 binding sites in the viral URR and therefore exclude E1 from the origin of replication and directly inhibit E1 expression (Lace et al., 2008).

In order to maintain a persistent infection and productive life cycle, HPV genomes must partition into daughter cells which migrate up through the epithelium. HPV E2 is able to tether to mitotic chromosomes and coordinate equal viral genome segregation between basal cells and these newly formed daughter cells. Interactions with the viral genome occur through the E2-DNA binding domain whilst attachment to the host chromosome depends on protein-protein interactions. One protein important in this process is the DNA helicase ChIR1 – E2 binding to ChIR1 is necessary for localisation to mitotic chromosomes (Parish et al., 2006). ChIR1 is required for tethering of viral genomes onto host chromosomes during DNA replication. During early mitosis E2 and ChIR1 co-localise, as the cell cycle progresses into metaphase, ChIR1 relocates to the spindle poles, leaving E2 associated with mitotic chromatin (Parish et al., 2006).

1.2.3 Productive infection

As the keratinocytes begin to migrate up from the basal lamina, the process of cellular differentiation is delayed as a result of expression of high-risk HPV oncoproteins. HPV E5, E6 and E7 function to create a replication competent environment, to promote cell cycle re-entry and proliferation **(Fig 1.6).** This is in part mediated through the targeting of key cell cycle control pathways including pRb and p53. HPV E7 binds to and initiates ubiquitin-proteasomal degradation of the retinoblastoma protein (pRb) (Dyson et al., 1989, Boyer et al., 1996, Berezutskaya et al., 1997). In addition to degradation, E7 can also bind to hypophosphorylated pRb to inhibit binding of E2F resulting in cell cycle progression into S-phase and aberrant proliferation (Boyer et al., 1996). Ordinarily, the consequence of pRb degradation can be circumvented by expression of the tumour suppressor protein p53, which results in cell cycle arrest and apoptosis. However, expression of high-risk HPV E6 has been shown to bind to p53 and initiate subsequent proteasomal degradation (Werness et al., 1990, Scheffner et al., 1993). Degradation of p53 inactivates the p53-mediated apoptotic response, the result of which is cellular immortalisation. A more comprehensive insight into the mechanism of E5, E6 and E7 is detailed below (see section 1.4.3).

Within the intermediate layers of the stratified epithelium, following initiation of keratinocyte differentiation, late transcripts are highly expressed, and genomes are amplified to ~10,000 copies per cell. HPV genome amplification is dependent on cellular differentiation both in vivo and in vitro in raft cultures (Bedell et al., 1991). HPV DNA replication in undifferentiated keratinocytes has been classified as

bidirectional theta replication. However, upon keratinocyte differentiation there is a switch in the mode of HPV DNA replication. Lack of identification of replication initiation or termination sites following differentiation suggests a switch to a rolling circle mechanism of replication (Flores and Lambert, 1997). This switch occurs immediately following differentiation and is therefore thought to be an essential prerequisite for effective HPV DNA amplification.

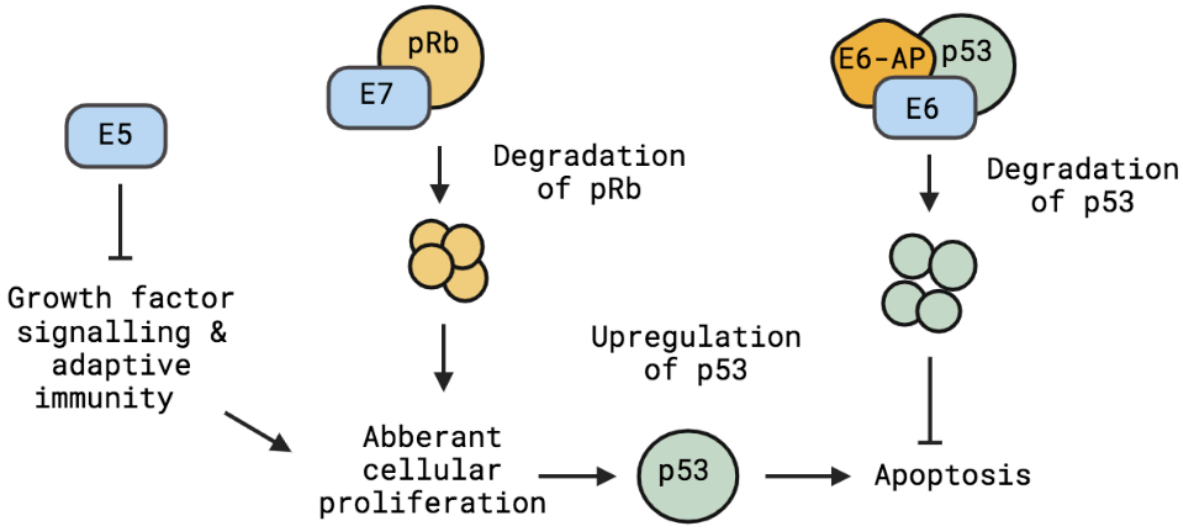


Figure 1.6 The function of high-risk HPV oncogenes. High-risk HPV oncogenes E5, E6 and E7 function synergistically to target multiple cellular proteins resulting in aberrant cellular proliferation. HPV E5 functions to target growth factor signalling and the major histocompatibility complex. HPV E7 targets the Rb proteins pRb and associated pocket proteins p130 and p107 resulting in deregulation of the cell cycle and continued proliferation.

Expression of the viral oncoproteins E6 and E7 in HPV infected keratinocytes results in altered expression of host cell proteins that enable efficient viral DNA amplification following differentiation. The viral oncoproteins can induce double strand breaks – predominantly in differentiated keratinocytes. Fluorescent in situ hybridisation demonstrated that the double strand breaks are mostly localised to host chromosomal DNA whilst viral DNA has minimal breaks (Mehta and Laimins, 2018). Induction of the cellular DNA damage response has also been shown to be essential for effective genome amplification in differentiated keratinocytes. This includes activation of both Ataxia telangiectasia mutated (ATM) and the ATM- and Rad3-Related (ATR) pathways. ATM is a key mediator of the DNA damage response pathway. Phosphorylation of ATM results in activation of a downstream signalling cascade including phosphorylation of NBS1, CHK2 and BRCA1. ATM-induced phosphorylation of CHK2 is essential for genome amplification in differentiated keratinocytes (Moody and Laimins, 2009). Recruitment of DNA damage response factors RAD51 and BRCA1 are also significantly increased in HPV infected keratinocytes. Following keratinocyte differentiation binding of RAD51 to the HPV genome is significantly increased in comparison to host DNA. This represents a scenario where DNA damage induced to host DNA induces a DNA damage response to favour rapid repair of HPV DNA (Mehta and Laimins, 2018).

Caspase activation has also been found to be important for viral genome amplification and is again increased as a result of E6 and E7 expression. Activation of caspase-3 within 3D-organotypic raft cultures is present within cells across all layers of the stratified epithelium. This is also consistent following high-calcium induced differentiation which results in increased activation of caspase-3 as well as

caspase-7 and caspase-9 (Moody et al., 2007). Caspase activation does not induce an apoptotic effect within the differentiating keratinocytes but instead results in increased levels of anti-apoptotic factors including Survivin and B-cell lymphoma 2 (BCL-2). Additionally, following differentiation, activated caspases target and cleave HPV E1 which is essential for genome amplification (Moody et al., 2007).

Following genome amplification, increased expression of E2 functions to repress the activity of the early promoter. In the superficial layers of the epithelium there is an increase in E1^E4, L1 and L2 transcription as a result of activation of the late viral promoter (Ruesch et al., 1998, Milligan et al., 2007). The presence of L1 and E1^E4 RNA products in undifferentiated keratinocytes provides evidence for a post-translational modification of late gene products – fully polyadenylated late RNAs are only present following differentiation (Milligan et al., 2007). E1^E4 is the first late mRNA to be expressed in the differentiated layers of the stratified epithelium (Milligan et al., 2007). E1^E4 protein (synonymous with E4) functions to amplify viral genomes and initiate the expression of late viral transcripts, as such E1^E4 is important in regulating late viral functions (Ruesch et al., 1998, Wilson et al., 2005). The minor and major structural proteins L2 and L1 are highly expressed in the superficial layers of the epithelium. Here, HPV genomes are encapsulated to produce mature virions. Cells infected with newly matured virus particles egress from the epithelial surface (Doorbar et al., 2015).

1.3 HPV gene expression

1.3.1 Promoter usage

There have been numerous transcriptional start sites (TSS) identified in the HPV genome. The alpha-papillomaviruses contain two well characterised promoters, first is the early promoter, which is situated upstream from the E6 open reading frame (termed p102 in HPV18 and p97 in HPV16; based on the nucleotide position within the viral genome) (Thierry et al., 1987, Smotkin et al., 1989). The early promoter is constitutively active throughout the entire HPV life cycle – it is active within undifferentiated basal keratinocytes but becomes further activated following cellular differentiation (Pentland et al., 2018, Ferguson et al., 2021). This therefore results in early transcripts being abundant in undifferentiated keratinocytes and then increasing in abundance following keratinocyte differentiation. Upstream from the early promoter is a TATA box which is important for recruitment of transcription factor II D (TFIID) and RNA polymerase II to initiate transcription (Hou et al., 2000). The late promoter is situated within the E7 open reading frame (termed p670 for HPV16 and p811 for HPV18) and is regulated by cellular differentiation (Grassmann et al., 1996, Wang et al., 2011).

Unlike the early promoter, the late promoter does not contain an associated TATA box. Within undifferentiated keratinocytes the late promoter is repressed, resulting in a very low abundance of transcripts initiating from this site within the initial stages of the viral life cycle (Ruesch et al., 1998, Wooldridge and Laimins, 2008, Ferguson et al., 2021). Following keratinocyte differentiation, the late

promoter is activated and transcripts initiating from this site are significantly increased.

More recently, using Nanopore direct RNA sequencing, several additional transcriptional start sites have also been identified (and/or validated) in both HPV16 and HPV18, although some are perhaps less well-characterised than the major early and late promoters. In HPV18 these include start sites at nucleotides 351-700 (p520), 1000-1400 (pE8) and 2800-4000 (p3000) (Ferguson et al., 2021). In HPV16 these include start sites at nucleotides 390-579 (p500), 1000-1400 (pE8) and 2800-3900 (p3500) (Evans et al., 2022, Unpublished Data). The HPV18 promoter p3000 had previously been identified (Karlen and Beard, 1993, Toots et al., 2014). The transcripts initiated from this promoter are thought to encode E5 and as with p811, p3000 is activated upon keratinocyte differentiation (Toots et al., 2014). The promoter pE8 is located within the E1 open reading frame (Milligan et al., 2007, Chen et al., 2014, Ferguson et al., 2021). In HPV18 this promoter has very low activity within both undifferentiated and differentiated keratinocytes. Two transcripts are encoded from this promoter; E8^E2 and to a lesser extent E5 (Ferguson et al., 2021).

1.3.1 Alternative splicing

HPV gene expression is polycistronic and as such multiple proteins can be transcribed from the 7-9 open reading frames as a result of alternative splicing **(Fig 1.7).** During the early stages of the HPV life cycle transcription is initiated from the early promoter. Alternatively spliced mRNAs are generated through the use of different splice donor and splice acceptor sites (also termed 5' and 3' splice sites). The E6 protein of both HPV16 and HPV18 is transcribed from an unspliced transcript which includes both the E6 and E7 ORFs. The major 5' splice site linked to the HPV16 early promoter p97, splice donor 226, is situated within the E6 open reading frame. This is spliced to one of four acceptor sites (SA409, SA526, SA3358 and SA2709) and generates the transcripts E6*I, E6*II, E6*III and E6*IV (Baker, 1997). In HPV18 the major splice donor site is SD233 which is situated within the E6 open reading frame. When spliced into one of three splice acceptor sites (SA416, SA3434, SA3465) this generates the transcripts E6*I, E6*II and E6*III (Wang et al., 2011).

Expression of late transcripts occurs as a result of a differentiation-dependent switch in promoter usage from the early promoter to the late promoter. The 5' splice sites within the E1 open reading frame of HPV16 are SD880 and SD1302 and the HPV18 genome are SD929. These donor sites can form splice junctions with four individual 3' acceptors to form mRNAs that code for the late viral proteins. E1^E4 is splice product formed as a result of splicing between the SD880 and SA3358 and splice acceptors within the E4 open reading frame, SA3358 in HPV16 and SA3434 in HPV18 (Baker, 1997, Wang et al., 2011). Although initiated from the late promoter,

E1^E4 transcripts are expressed in both early and late stages of the HPV life cycle. Recent quantitative analysis of the HPV18 transcriptome has also revealed novel transcripts that are generated from previously uncharacterised splice variants. These are both initiated from SD929 which can splice into SA3440 to code for E5 or into SA3506 which has the potential to encode for E6*I, E7 and E5 (Ferguson et al., 2021).

In addition to the E6* and E6/E7 transcripts, both HPV16 and HPV18 express an additional spliced isoform that is initiated from the E8 promoter, termed E8^E2 (Straub et al., 2015). This is generated as a result of a splice junction between SD1302 and SA3358 in HPV16 and SD1357 and SA3434 in HPV18 (Lace et al., 2008, Wang et al., 2011).

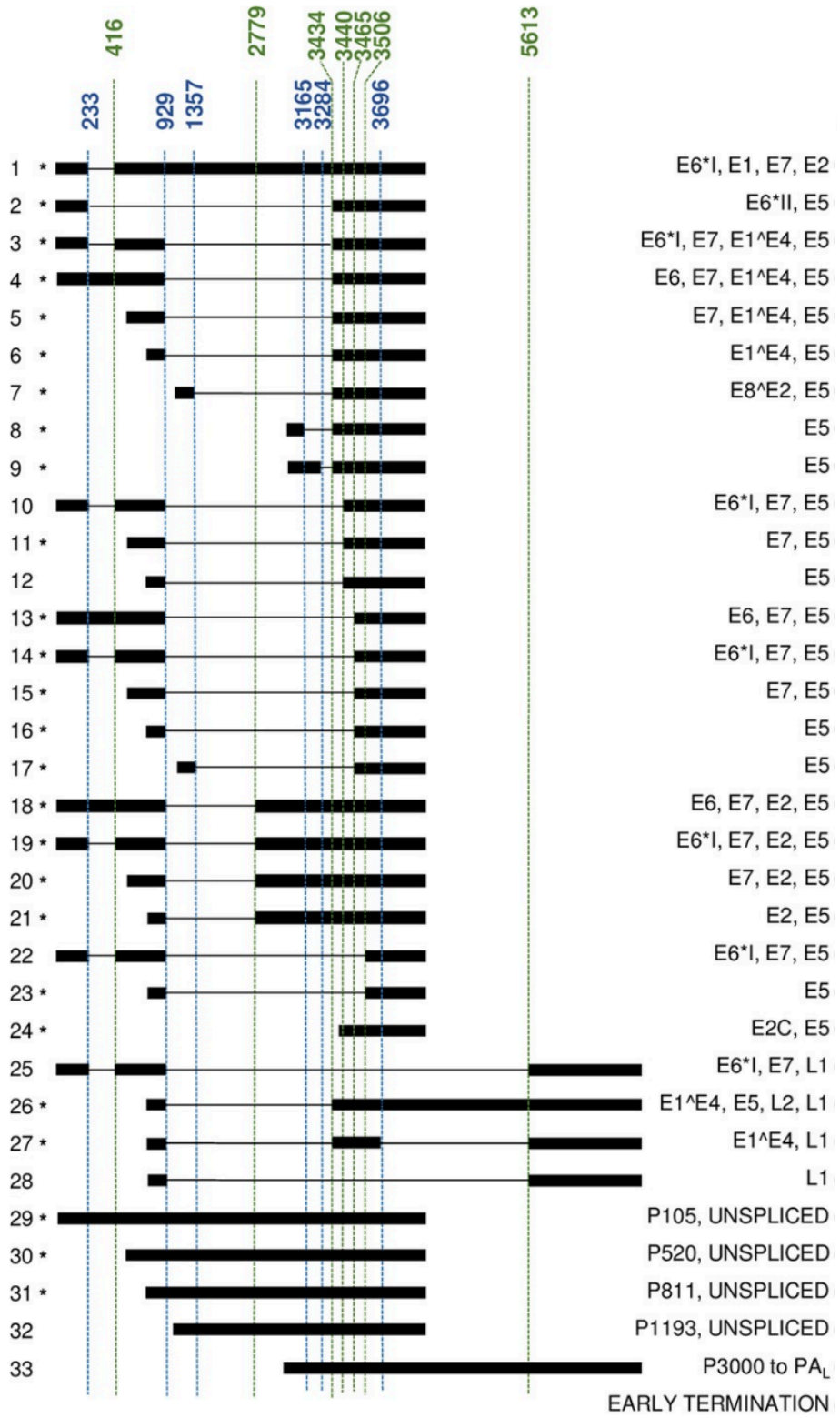


Figure 1.7 HPV18 transcript map. HPV18 transcriptome as determined by Nanopore direct RNA sequencing. Splice donors are shown in green and splice acceptors are shown in blue. * Denotes previously identified transcripts. Figure taken from (Ferguson et al., 2021).

1.3.2 Polyadenylation

RNA processing in the form of polyadenylation is a two-step process which involves the endonucleolytic cleavage of the nascent RNA and the subsequent addition of a long chain of adenine nucleotides, known as a poly(A) tail, to the 3' end of the cleaved RNA product. The process of polyadenylation is targeted to polyadenylation sites which are flanked by two *cis*-acting regulatory elements termed the poly(A) signals (PAS). The first, is an upstream A(A/U)UAAA hexamer and the second a downstream U-rich or GU-rich element (Reviewed by (Tian and Manley, 2017). The process of polyadenylation relies on multiple cellular factors including the cleavage and polyadenylation specificity factor (CPSF) which recognises and binds to the A(A/U)UAAA hexamers and the cleavage stimulation factor (CstF) which recognises and binds to the U-rich or GU-rich elements (Zhao et al., 1999). In addition, there are two cleavage factors CFI and CFII and a poly-A-polymerase (PAP) which mediates the addition of the poly(A) tails (Takagaki et al., 1989). The process of polyadenylation is thought to stabilise mRNA, aid cytoplasmic export and also translation. It is also required for transcription termination.

The polyadenylation cleavage sites of high-risk HPV genomes have been mapped using 3'RACE. HPV16 contains three identified polyadenylation sites; an early polyadenylation site (pA_E) which is situated just upstream from the L2 open reading frame at nucleotide 4215 and two late polyadenylation sites (pA_L) which both flank the 5' end of the L1 open reading frame and the 3' end of the URR. The first is situated at nucleotide 7284, with an upstream poly(A) signal at nucleotide 7620 and the second at nucleotide 7343 with an upstream poly(A) signal at nucleotide 7319

(Milligan et al., 2007). In HPV18 the early polyadenylation (pA_E) site is located at the 3' end of the L1 open reading frame at nucleotide 4270 with an upstream polyadenylation signal (AAUAAA) positioned at nucleotide 4235. Downstream of this pA_E cleavage site is a GU-rich element containing four U/GU motifs located between nucleotides 4334-4339. There are five identified late polyadenylation cleavage sites (pA_L) for HPV18; located at nucleotides 7299, 7307, 7294, 7297 and 7306. Upstream from the polyadenylation site is an AUAAA sequence containing the polyadenylation signal, downstream from the L1 open reading frame at nucleotide 7278. Between nucleotides 7341-7350 are multiple U/GU motifs (Wang et al., 2011).

Within basal keratinocytes, transcription is initiated from the early promoter and terminated at the early polyadenylation site. In the intermediate layers of the epithelium, following differentiation, transcription is initiated from the late promoter and is terminated at the late polyadenylation site. Expression of L1 and L2 genes is dependent on read-through at the early polyadenylation signal – there are several mechanisms that enable the pAE to be bypassed. Following differentiation, in the upper stratified layers of the epithelium, HPV E2 protein is upregulated. E2 has been shown to upregulate late gene transcription through inhibiting the HPV early polyadenylation signal and therefore enabling readthrough into the late gene region. This process is independent of E2 binding to the viral URR and seems to require both the hinge region and N-terminal of the E2 protein (Johansson et al., 2012). E2 functions to repress early polyadenylation by targeting the CPSF-containing polyadenylation complex through a direct interaction with CSPF30 (Johansson et al., 2012). Activation of the DNA damage response has been shown to be important

in HPV genome amplification following differentiation. Activation of the DNA damage response, including the ATM and the checkpoint kinase 1/2 (Chk1/2) pathways, is also important in inducing HPV16 late gene expression. Initiation of the DNA damage response causes altered expression of RNA processing factors resulting in decreased binding of CPSF30 and increased binding of the heterogenous ribonucleoprotein C (hnRNPC) at the early polyadenylation site. Increased binding of hnRNPC results in bypassing of the early polyadenylation site and increased polyadenylation at the pAL site – as a result late gene expression is increased (Nilsson et al., 2018).

1.3.3 HPV oncoproteins

1.3.3.1 E5

Papillomaviruses from the alpha genus encode a short transmembrane protein called E5 which has known transforming capabilities. The E5 open reading frame is situated at the 3' end of the early gene region and the 80-amino acid protein has hydrophobic properties (Bubb et al., 1988). Multiple cellular targets have been identified for E5 including proteins that function in immune evasion, cell-cell communication, and apoptosis – these can be directly or indirectly modulated by E5 expression. Early studies into the function of HPV E5 identified the transforming capabilities of this protein. The expression of HPV6 E5 in murine fibroblasts resulted in a morphologically transformed phenotype which included accelerated growth rate, anchorage independent growth and the ability to grow in soft agarose (Chen and Mounts, 1990). This phenotype was also consistent in cells expressing HPV16 E5 and this was found to be facilitated through the epidermal growth factor receptor.

Expression of HPV16 E5 results in increased sensitivity to EGF and enhanced mitogenic signals from the epidermal growth factor receptor. As a result these cells have increased proliferation and transformation efficiency (Leechanachai et al., 1992, Pim et al., 1992, Straight et al., 1993).

Evasion of host immune clearance is an essential prerequisite for establishment of a persistent papillomavirus infection. Presentation of viral antigens by the major histocompatibility complex I/II (MHC I/II) is a fundamental aspect of adaptive immunity. MHC class I and II proteins include the human leukocyte antigens (HLA) which present antigens to cytotoxic T-lymphocytes such as CD8+ cells (Wieczorek et al., 2017). HPV16 E5 can directly interact with the HLA-I heavy chain to prevent translocation from the golgi apparatus to the cell surface membrane. As a result there is reduced expression of MHC class I and II proteins HLA-A and HLA-B on the cell surface membrane. Papillomavirus antigens therefore go unrecognised by CD8+ cells (Ashrafi et al., 2005, Ashrafi et al., 2006, Campo et al., 2010).

1.3.3.2 High-risk E6 and E7

One of the most well-established functions for HPV E7 is the interaction with the retinoblastoma tumour suppressor gene product pRb (p105/RB) and the associated pocket proteins including p130/RBL2 and p107/RBL1 (Münger et al., 1989b, Dyson et al., 1989, Münger et al., 1992, Dyson et al., 1992). These interactions are facilitated through a Leu-X-Cys-X-Glu motif located in the conserved region 2 (CR2 domain) of E7 (Münger et al., 1989b, Dyson et al., 1989). The retinoblastoma protein is a tumour suppressor that is involved in regulating the G1 cell cycle checkpoints, blocking entry into S-phase and inhibiting cellular growth (Reviewed by (Giacinti and

Giordano, 2006). One mechanism for this is through an interaction with the E2F genes – a family of transcription factors which can be regulated through direct binding with Rb proteins (Chellappan et al., 1991, Bagchi et al., 1991, Cobrinik et al., 1993). Increasing phosphorylation of pRb through early G_1 towards S phase, in normal cells, enables pRb to initiate cell cycle entry. In its hypophosphorylated state, pRb complexes with E2F transcription factors, sequestering the activity of E2F and therefore repressing transcription from E2F-dependent promoters (Fehrmann and Laimins, 2003). Since E2F regulates the expression of multiple genes that are implicated in cell cycle regulation, including MYBL2 (B-*myb)*, cyclin E, cyclin-dependent kinase 1 (CDK1) and dihydrofolate reductase (DHFR), and drives cells into S-phase, pRb binding is able to negatively regulate E2F-mediated cell cycle progression (Dyson, 1998). HPV E7 sequesters the function of pRb through directly competing with E2F and therefore disrupting pRB-E2F complexes in addition to targeting pRb for proteasomal degradation (Huang et al., 1993, Boyer et al., 1996). Proteasomal degradation is thought to occur via the Cullin 2 ubiquitin ligase complex (Huh et al., 2007).

The Rb-like proteins p130 and p107 can function to regulate cell cycle gene expression through mechanisms independent of pRb. This includes the formation of a multi-protein complex composed of dimerization partner (DP 1/2), RB-like, E2F and a MuvB core consisting of four MuvB-like proteins (LIN9, LIN37, LIN52, LIN54) (DREAM) (Litovchick et al., 2007). The DREAM complex regulates cell cycle gene expression and can exist in two forms; a repressive form and an active form, also termed the MMB complex (for review see (Sadasivam and DeCaprio, 2013). HPV16 E7 can interact directly with the Rb-like proteins including p130. As a result, E7 can

disrupt the MuvB core from association with p130. Decreased levels of MuvB-like protein LIN37 bound to p130 results in a switch from the DREAM complex to the MMB complex as LIN37 consequently interacts with B-Myb (James et al., 2021). As such, HPV16 positive keratinocytes and indeed HPV-driven head and neck cancers show significant enrichment of expression of DREAM target genes (2007, James et al., 2021).

E6 expressed from high-risk alpha papillomaviruses are able to interact with and degrade p53 tumour suppressor protein. E6-mediated degradation of p53 is facilitated through a stable interaction a 100-kDa E6-associated protein (E6AP) which promotes binding to p53 (Huibregtse et al., 1991, Huibregtse et al., 1993). An 18-amino acid sequence within E6AP is required for this interaction and the deletion of this site (leading to the abrogation of E6/E6AP interactions) results in loss of p53 binding activity (Scheffner et al., 1993). E6AP is an E3 ubiquitin ligase which forms thioester bonds with ubiquitin. When in complex, E6/E6AP is able to transfer ubiquitin to p53 and initiate subsequent proteasome-mediated degradation (Scheffner et al., 1993).

Another function of HPV E6 is targeting of proteins that contain a PSD95/DLG/ZO-1 (PDZ) domain, also referred to as PDZ proteins. PDZ proteins primarily function as cell junction proteins which act to maintain cell polarity, cell-cell adhesions and cell proliferation. All E6 proteins from high-risk HPV types contain a PDZ-binding motif (PBM) which is located within the C-terminal region (Thomas et al., 2016). Most recent estimates have confirmed 19 PDZ proteins that are targeted by E6 resulting in the proteasomal degradation of these proteins (Thomas et al., 2016).

These include the human homologues of the *Drosophila* tumour suppressor proteins Discs Large (DLG1) and Scribble (hSCRIB), the MAGUK with Inverted Domain (MAGI) proteins MAGI-1, -2 and -3 and the multi-PDZ domain protein 1 (MUPP1) (Kiyono et al., 1997, Nakagawa and Huibregtse, 2000, Glaunsinger et al., 2000, Thomas et al., 2002, Lee et al., 2000).

1.4 Regulation of HPV gene expression

1.4.1 Transcription factor binding

As previously described, the HPV URR contains a number of *cis*-regulatory elements that influence the activity of the early and late viral promoters. Depending on a number of factors, including binding partners, post-translational modifications and the epigenetic landscape, these proteins can function to either activate or repress HPV transcription. Regulation is dependent on both viral and cellular factors, the expression of which is temporally regulated during the differentiation dependent life cycle of HPV.

In HPV31, binding of several transcription factors to the viral genome act to regulate the activity of the early promoter and the pattern of transcription factor binding changes between undifferentiated and differentiated keratinocytes. In undifferentiated keratinocytes there is enrichment of Sp1, YY1, C/EBP-B and OCT-1 (Wooldridge and Laimins, 2008). In HPV31 YY1 is enriched at both the early and late promoter regions, although more robustly bound at the late promoter. In HPV18, YY1 also binds across the viral URR including to the enhancer and replication of origin. Interestingly, following differentiation of HPV31-containing keratinocytes YY1 binding is significantly increased by 20-50-fold at the early promoter, whereas differentiation of HPV18-containing keratinocytes results in abrogated YY1 binding (Wooldridge and Laimins, 2008, Pentland et al., 2018). Differentiation of HPV16-containing keratinocytes results in altered expression and binding of several transcription factors. Binding of transcription factors such as C/EBPα, NF1 and Stat5 are all significantly increased following differentiation (Carson and Khan, 2006).

1.4.2 HPV E2

E2 has been found to act as both an activator and repressor of the HPV early promoter. In order to mediate transcriptional activity, E2 binds to the E2-responsive sequences in the URR. Whilst there are four conserved E2 binding sites, only three are required for HPV11 E6 promoter regulation. The promoter-distal E2 binding site is not involved in promoter repression (Dong et al., 1994a). E2 mediated transcriptional repression is in part mediated through interference with binding or function of cellular proteins that ordinarily act as *cis*-elements which activate activity of the early promoter. This includes TATA binding proteins (TBP), TFIID and Sp1 (Bernard et al., 1989, Dong et al., 1994a). One of the E2 binding sites (E2 #2) overlaps with the binding site of transcription factor Sp1 (Demeret et al., 1994, Tan et al., 1994). E2 can mediate early gene repression by the interfering with the transcription pre-initiation complex assembly, namely, the TFIIB-Pol II-TFIIF complex. In addition to displacing TATA-binding proteins and TFIID, E2 is thought to also inhibit HPV transcription at steps after TATA recognition (Hou et al., 2000).

In addition to full length E2, HPVs can also encode shorter E2 isoforms which in comparison to full-length E2, are able to act as transcriptional repressors. Full-length E2 has been shown to activate the early promoter in HPV11. However, short forms of E2 which encode the carboxyl terminal are able to competitively repress the *trans*-activating abilities of full-length E2 on the HPV11 enhancer (Chin et al., 1988). In both HPV16 and HPV18 containing human keratinocytes, full-length E2 functions as a transcriptional activator of the early promoter. However, a spliced isoform of E2 encodes a putative repressor of the early promoter even if co-expressed with full

length E2 (Bouvard et al., 1994). Whilst full-length and short-isoform E2 both have similar binding affinities, the short isoform of E2 is able to mediate transcriptional repression by interfering with binding of full-length E2 (Bouvard et al., 1994). As such, the relative amounts of the two isoforms is important in regulating viral gene expression.

E8^E2 is a splice product formed by splicing between the E8 open reading frame and the E2 open reading frame. E8^E2 functions to repress the transactivation functions of E2 and therefore regulates early transcript expression from the early promoter (Choe et al., 1989, Lace et al., 2008).

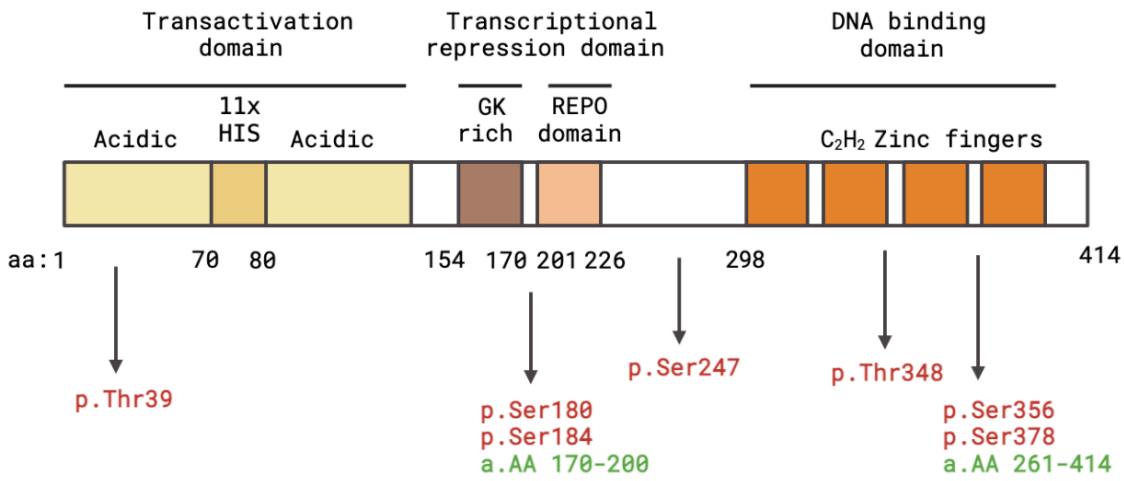


Figure 1.8 The structure of Yin-Yang 1 (YY1). YY1 is comprised of 414 amino acids. The transactivation domain is located within the N-terminal region of the protein. The REPO domain functions as a bridging molecule between YY1 and PcG proteins. Within the C-terminal domain are four C_2H_2 zinc finger proteins which are primarily involved in DNA binding function. Post-translational modifications of YY1 and the amino acid residue(s) are listed below. p, phosphorylation, a, acetylation, m, methylation, AA, amino acids. Figure adapted from (Verheul et al., 2020).

1.4.3 Yin-yang 1 (YY1)

YY1 is a ubiquitously expressed zinc-finger transcription factor that has been implicated in both negative and positive regulation of gene expression (Austen et al., 1997). Early studies first described YY1 as a transcription factor when it was found to repress transcription from the adeno-associated p5 promoter in the absence of the adenovirus E1A protein (Shi et al., 1991). The adenovirus E1A protein is an essential activator of both adenoviral genes and cellular genes. E1A induced transcriptional activation of the p5 promoter is mapped to a 20 bp region termed the p5-60 element. This same sequence was also found to contain an overlapping binding site for two cellular proteins – one of which being YY1. Through directly binding to this sequence, YY1 interferes with E1A function as a transcriptional activator and as a result can repress transcription from the p5 promoter (Shi et al., 1991). Later studies then identified that to mitigate this, E1A is able to circumvent YY1 mediated repression by interacting with YY1, and this was found to be dependent on the histone acetyltransferase p300. Through forming a direct interaction with YY1, p300 is able to activate YY1-mediated transcription by recruiting E1A to DNA-bound YY1 resulting in transactivation of the adenovirus p5 promoter (Lee et al., 1995a). YY1-mediated transcriptional repression was further supported by a second group who at the same time had independently cloned YY1 (which they termed NF-E1) and showed that it was able to bind to a negative acting region of the Igκ 3′ enhancer and repress activity (Park and Atchison, 1991).

As shown in **Figure 1.8** YY1 is comprised of 414 amino acids and these are separated into three major functional domains. Within the C-terminal domain of the

protein (between amino acids 298-414) are four C_2H_2-type zinc fingers (Hyde-DeRuyscher et al., 1995). The zinc fingers facilitate the DNA binding capabilities of YY1 which binds the consensus motif 5'-CCGCCAT(N)TT-3' (Houbaviy et al., 1996). There is also a region within the C-terminal domain which overlaps the zinc fingers that can function to repress transcription. This is located between amino acids 333-397 – although it does not appear that normal structure of the zinc fingers are required to mediate repression from this domain (Lee et al., 1994, Bushmeyer et al., 1995). The transactivation domain of YY1 resides within the N-terminus of the protein (Lee et al., 1994, Lee et al., 1995b, Bushmeyer et al., 1995). Two regions in the N-terminus are thought to act to activate transcription, amino acids 16-29 and amino acids 80-100 (Bushmeyer et al., 1995). Amino acids 1-90 are also required for YY1 activation function (Lee et al., 1994). The transcriptional repression domain of YY1 is situated within both the central and C-terminal region of the protein. To facilitate transcriptional repression YY1 can recruit several co-factors to sites of enhancers and promoters. A comprehensive list of YY1 interacting partners is detailed below (see section 1.4.5). Within the central region of the protein there is a small 25 amino acid segment situated between amino acids 201-226 which is known as the REPO domain (REcruit POlycomb) (Wilkinson et al., 2010). As the name suggests, the REPO domain acts as a bridging molecule between PcG proteins and sites of DNA. The YY1 associated factor 2 (YAF2) binds to the REPO domain of YY1 and functions to recruit PcG proteins to DNA in order to facilitate PcG mediated transcriptional repression (Wilkinson et al., 2010).

1.4.4 Regulation of YY1

Post-translational modifications (PTMs) can alter the structure and therefore function of proteins through the process of covalent modifications usually in the form of adding modifying groups such as phosphoryl, acetyl or methyl to amino acid(s) (Ramazi and Zahiri, 2021). Regulation of YY1 is closely related to a number of post-translational modifications **(Fig 8)** (Zhang et al., 2016, Rizkallah and Hurt, 2009, Rizkallah et al., 2011, Alexander and Rizkallah, 2017). Ubiquitination is the process of covalently linking a ubiquitin protein to a target protein, often resulting in the proteasomal degradation of the target protein via an E3-ubiquitin ligase complex (Ramazi and Zahiri, 2021). The stability of YY1 protein expression can be altered by ubiquitin mediated proteasomal degradation. The E3 ubiquitin ligase SMAD specific E3 ubiquitin protein ligase 2 (SMURF2) is responsible for downregulating YY1 protein expression and the expression of SMURF2 drastically reduces the half-life of YY1 (Ramkumar et al., 2013, Jeong et al., 2014, Gao et al., 2021). The histone acetyltransferase (HAT) p300 and the p300/CBP-associated factor (PCAF) influence the transcriptional activity of YY1 through acetylation. Amino acid residues 170-200 are acetylated by both PCAF and p300 and an additional site within the C-terminus at amino acids 261-414 can also be acetylated by PCAF, but not p300 (Yao et al., 2001). Acetylation of YY1 by p300 and PCAF was found to be necessary for the transcriptional repression activity of YY1 (Yao et al., 2001). YY1 is phosphorylated by Aurora B kinase at serine 184 and serine 180 during the G2/M phase of the cell cycle. As the cells exit the cell cycle and enter G1, YY1 is then de-phosphorylated at this site. Interestingly, phosphorylation of YY1 at serine 184 results in altered DNA binding activity, by enhancing binding at repressive sites and

reducing binding at activating sites. Additionally, phosphorylation at this residue interferes with p300 mediated acetylation of YY1. YY1 is also phosphorylated by Aurora kinase A and C although this did not occur at serine 184 (Kassardjian et al., 2012). Instead Aurora A kinase phosphorylates YY1 at serine 365 – this only occurs in mitotic cells. Phosphorylation with Aurora kinase A abolishes YY1 DNA binding capabilities (Alexander and Rizkallah, 2017).

MicroRNAs (miRNAs) are short, single stranded, non-coding RNA molecules that can regulate gene expression. They function primarily by interacting with the 3' untranslated regions of target messenger RNA sequences. As a result miRNAs degrade mRNA sequences and directly block translation (O'Brien et al., 2018). Multiple miRNAs have been implicated in the regulation of YY1 expression such as miR-29a. Several studies have identified a role for miR-29a in downregulating YY1 expression (Li et al., 2011b, Liu et al., 2021). The overexpression of miR-29a results in reduced YY1 expression and conversely blocking miR-29a activity results in increased YY1 expression. This is unsurprising given that there is a direct target sequence match of miR-29a in the 3' untranslated region of YY1 (Liu et al., 2021). Interestingly, the expression of miR-29a is significantly downregulated in cervical intraepithelial neoplasia and squamous cell carcinoma (Li et al., 2011b). This downregulation appears to be a direct result of expression of HPV E7 (Liu et al., 2021).

1.4.5 YY1 binding partners

YY1 has been implicated as both an activator and repressor of gene transcription. This is in part mediated through direct protein-protein interactions which in different

contexts can either relieve or activate YY1 mediated repression. YY1 can interact with multiple members of the transcription initiation complex including transcription factor II B (TFIIB), TATA binding protein (TBP), TBP-associated factors, CREB-binding protein and RNA pol II (Austen et al., 1997). TFIIB-YY1 interactions enhance the affinity for YY1 to its consensus DNA motif and in this complex, in this sense, YY1 is able to function like the TATA binding protein and recruit RNA pol II to the transcription initiation complex. This is independent of the TATA binding protein and appears to be relatively stable as YY1 remains bound at the promoter through multiple rounds of initiation (Usheva and Shenk, 1994, Usheva and Shenk, 1996). The presence of Sp1 binding sites has been shown to enhance the activity of YY1 when bound to initiator sequences. This was found to be mediated through a direct protein-protein interaction between Sp1 and YY1 (Seto et al., 1993).

Other interacting partners include chromatin modifiers. YY1 has been shown to recruit and directly interact with the protein arginine methyltransferase-1 (PRMT1) to facilitate methylation of histone H4 at a YY1 activated promoter resulting in increased transcription from this site (Rezai-Zadeh et al., 2003). YY1 can interact with the BET bromodomain-containing family proteins (Brd2/Brd4) and this interaction is thought to be significantly enhanced by acetylation of YY1 at its lysine residues (Wu et al., 2013, Ren et al., 2018, Xu et al., 2021). YY1 enrichment at gene promoters significantly overlaps with Brd4 and H3K27Ac and YY1 can function to activate gene expression via the recruitment of both Brd4 and histone acetyltransferases such as the Ada2a-containing (ATAC) HATs which accelerates gene activation via a feedforward loop (Xu et al., 2021).

1.4.6 YY1 and chromatin remodelling

The three-dimensional (3D) organisation of chromatin is important in regulating transcription. Long-range chromatin interactions and topological changes can alter gene expression, and these can occur as a result of binding of architectural proteins to sites of enhancers and promoters. YY1 can serve as an architectural protein and facilitate DNA loops, DNA interactions and long-range contacts in order to connect distant promoters and enhancers (Weintraub et al., 2017, Beagan et al., 2017). These can be both enhancer-enhancer, promoter-promoter, or enhancer-promoter contacts (Weintraub et al., 2017). In doing so, YY1 is able to both negatively and positively alter gene expression. One mechanism to which this is facilitated is via a direct protein-protein interaction between YY1 and the CCCTC-binding factor (CTCF). This interaction is directed through the zinc finger domain of YY1 and the N-terminus of CTCF – often, long-range interactions that are anchored by YY1 are connected by constitutively bound CTCF (Donohoe et al., 2007, Beagan et al., 2017).

1.4.7 YY1 mediated regulation of HPV transcription

YY1 binding to the HPV18 genome was first shown when a negative regulatory element was identified within the URR. The transcription factor AP-1 was unable to be activated by tissue plasmogen activator (TPA) raising the possibility that this was counteracted by a negative regulator (Bauknecht et al., 1992). This regulatory motif situated within the URR was found to have sequence homology to the consensus YY1 binding site. When bound to this motif, YY1 was found to repress transcription from the HPV18 early promoter (Bauknecht et al., 1992). Interestingly, the same

authors went on to show that YY1 is able to act as an activator or repressor of HPV18 transcription depending on a sequence termed the switch region, which is situated upstream of the proximal promoter YY1 binding site. The switch region enables YY1 to act as a transcriptional activator and if absent, YY1 instead functions to repress HPV18 transcription from the URR (Bauknecht et al., 1995).

YY1 has also been shown to regulate the HPV16 early promoter (p97), both in vitro and in vivo. Isolated cervical biopsies containing HPV16 DNA were found to contain multiple mutated or deleted YY1 binding sites which resulted in increased activity of the early promoter (May et al., 1994). When bound to the early transcription start site YY1 represses the expression of both E6 and E7 from the early promoter which can be attenuated following the introduction of simultaneous mutations across four YY1 binding sites in the URR (Lace et al., 2009b, Lace et al., 2009a). Interestingly, out of the ten YY1 binding sites present in the HPV16 URR, only five are required for regulation of transcription from the early promoter (binding sites #4, #5, #7, #8, #9) which are all located proximal to the origin of replication and the early promoter (O'Connor et al., 1996). YY1 was thought to function as a repressor of E6/E7 transcription by inactivating AP-1 activity through recruitment and direct interaction with the CREB-binding protein (O'Connor et al., 1996).

YY1 functions to regulate transcription from the HPV18 early promoter (p102) during the differentiation dependent life cycle. YY1 is recruited to the viral genome through a long-range chromatin interaction with CTCF bound at the HPV18 E2 open reading frame. Enrichment of YY1 binding at the HPV18 URR is most profound within undifferentiated keratinocytes. Here, YY1 functions to recruit histone modifying

proteins that function to condense chromatin. As such, in undifferentiated keratinocytes, YY1 functions to repress E6/E7 transcription. Following differentiation, YY1 protein is significantly reduced, resulting in increasing chromatin accessibility at the viral URR and recruitment of RNA PolII which functions to initiate transcription from the early promoter. This mechanism of YY1 mediated transcription repression is thought to be important in maintaining relatively low levels of oncogene transcription during the initial stages of the HPV life cycle to reduce the likelihood of immune clearance and as such enable a persistent HPV infection.

1.5 Epigenetic regulation of HPV gene expression

1.5.1 Histone modifications

The nucleosome is the unit of chromatin that is composed of repeated stretches of ~146 bp of DNA wrapped around a histone octamer. Histone octamers are composed of the core histone proteins H2, H3 and H4 – they can be defined by two H2A-H2B dimers and a H3-H4 tetramer (Luger et al., 1997). The N-terminal region of histone proteins is referred to as the histone tail and these can be covalently modified. These modifications include, but are not exclusive to, acetylation, methylation, and ubiquitination. Enzymes known as writers mediate these modifications and can include histone acetyl transferases (HATs) or lysine methyltransferases (KMTs). Nucleosome occupancy across the URR is an important regulator of HPV transcription with known nucleoprotein structures forming over the enhancer, upstream of E6 including the early promoter, across the origin replication and within the E7 open reading frame (Stünkel and Bernard, 1999).

Excluding *cis*-regulatory control of HPV gene expression the epigenetic landscape of the viral upstream regulatory region is also fundamental in regulating transcription. Using DNase I footprinting the chromatin changes that occur upon keratinocyte differentiation have been observed in HPV31. Following differentiation, chromatin accessibility at the E7 open reading frame is increased. In contrast chromatin around the early promoter becomes less accessible following differentiation (del Mar Peña and Laimins, 2001). Histone modifications have been shown to regulate differentiation-dependent transcription in HPV31 containing CIN612 cells. In undifferentiated keratinocytes the early and late promoter contained acetylated histones H3, H4 and H3K4Me2 indicative of active chromatin. These regions also showed low levels of lysine 9 and lysine 27 dimethylation. Following differentiation, acetylation of histone H3 and H4 increased and H3K4Me2 increased at the early promoter, and this was also consistent at the late promoter, although less drastically. H3K9Me2 and H3K27Me2 was unchanged at the early promoter but decreased at the late promoter (Wooldridge and Laimins, 2008). Expression of the viral oncogenes E6 and E7 is dependent on the chromatin topology at the URR – with open or euchromatin, associated with active histone modifications including H3 lysine 4 trimethylation or H2 lysine 27 acetylation, favouring high viral gene expression (Groves et al., 2016). Alternatively, chromatin enriched with repressive modifying marks including H3 lysine 9 di- and trimethylation and H3 lysine 27 di- and tri-methylation were associated with repressed gene expression (Groves et al., 2016).

1.5.2 Polycomb repressive group proteins (PcG)

Polycomb-group (PcG) proteins are chromatin modifiers that function to regulate transcription. Following their original identification in *Drosophila melanogaster* they have since been shown to be conserved in mammals. There are numerous PcG proteins which form multimeric complexes best known as PRC1 and PRC2 (for review see (Piunti and Shilatifard, 2022). There are several distinct groups of PRC1 complexes all of which contain a core PCGF subunit and a RING1A/B ubiquitin ligase which is responsible for the catalytic activity of the complex (Gao et al., 2012). RING1B (gene name RNF2) is an E3 ubiquitin ligase which monoubiquitinates histone H2A at lysine 119. This modification is important in gene silencing (Wang et al., 2004). The canonical PRC1 complexes also contain a chromobox protein (CBX). Non-canonical PRC1 complexes do not contain CBX subunits. Instead, these complexes contain either recruitment of the RING1 and YY1 binding protein (RYBP) or the YY1 associated factor 2 (YAF2). RYBP or YAF2 can both interact with RING1A/B and PCGF2/4. The RYBP plays an essential role in facilitating the enzymatic activity of RING1B and is required for the monoubiquitination of H2A lysine 119 (Gao et al., 2012). In addition to functioning to catalytically regulate gene expression there is also evidence to suggest that RING1B can function to compact chromatin in order to repress the expression of genes. This is independent of RING1B mediated ubiquitination (Eskeland et al., 2010).

The PRC2 complex is composed of three subunits, EED, SUZ12 and EZH1/2. EZH2 is a member of the SET-domain family of proteins and as such has methyltransferase activity. The SET domain of EZH2 is essential for its ability to

methylate histones. In particular, EZH2 is responsible for both di- and tri-methylation of lysine 9 and lysine 27 on histone H3 (Cao et al., 2002, Kuzmichev et al., 2002). EZH2 is able to directly interact with EED via the WD-40 domain on EED (Sewalt et al., 1998). SUZ12 interacts directly with EZH2 or EED-EZH2 and is thought to be required for the enzymatic activity of these complexes. SUZ12 is therefore directly involved in the transcriptional repression of the PRC2 complex (Cao and Zhang, 2004, Yamamoto et al., 2004).

There are several mechanisms which are involved in the recruitment of PcG proteins to sites of DNA. Deposition of the histone mark H3K27Me3 can facilitate recruitment of both PRC1 and PRC2 complexes. In a feedback loop, PRC2 can bind directly to the methylated lysine 27 (Hansen et al., 2008). The PRC1 components CBX2 and PCGF2/4 bind to H3K27Me3 rich regions and RING1B can also recognise and bind specifically to this mark, suggesting that PRC2 functions to facilitate PRC1 mediated transcription repression (Gao et al., 2012, Kuzmichev et al., 2002). However, there is also H3K27Me3-independent recruitment of PRC1 components. In EED null cells which were also devoid of H3K27Me3 there was still enrichment of RING1B. This was also consistent with deposition of H2AK119Ub (Tavares et al., 2012). The H3K27Me3-independent recruitment was shown to be mediated by RYBP. The YY1 associated factor 2 (YAF2) has been shown to directly interact with YY1 via the first and second zinc fingers of YY1 (Kalenik et al., 1997). The REPO domain of YY1 interacts with the C-terminal region of YAF2 and provides a link between YY1 and RING1B. This is mediated by the recruitment RYBP to DNA via a direct interaction with YY1. As RYBP can directly interact with both RING1A/RING1B, this highlights a mechanism for YY1 mediated recruitment of PRC1 components to DNA (García

et al., 1999, Satijn et al., 2001, Basu et al., 2014). YY1 has also been shown to interact with the PRC2 component EED where it functions to direct EED-EZH2 to target genes (Satijn et al., 2001).

Recruitment of PcG proteins to the HPV18 genome regulates differentiation dependent transcription from the early promoter. In undifferentiated keratinocytes YY1 binds to the HPV18 URR and recruits PcG proteins to the viral enhancer and early promoter. Enrichment of RING1B, EED and EZH2 at the HPV18 genome has been confirmed by ChIP-qPCR in undifferentiated human foreskin keratinocytes. As a result, deposition of H3K27Me3 and H2AK119Ub results in repression of E6 and E7 transcription (Pentland et al., 2018). Following HPV18 keratinocyte differentiation the expression of YY1 protein is significantly abrogated, relieving PcG mediated repression of the early promoter (Pentland et al., 2018).

1.5.3 Three-dimensional chromatin organisation

Architectural proteins such as the zinc-finger binding protein, CTCF, are able to regulate gene expression through chromatin rearrangements which result in blocking of *cis*-regulatory elements. CTCF is a ubiquitously expressed DNA binding protein with an affinity for ~55,000 sites in the mammalian genome and functions as a chromatin insulator, transcriptional activator and repressor (Chen et al., 2012, Bell et al., 1999). Through reorganising chromatin into loop formations, CTCF can mediate long-range interactions and regulate transcriptional control in a number of DNA viruses. CTCF was first identified when a novel DNA-binding protein was found to bind regularly spaced (12-13bp interval) repeated motifs on the chicken *c-myc* gene. The protein was named CTCF based on the consensus binding motif 5'-

CCCTC-3' (Lobanenkov et al., 1990). Structurally, CTCF is comprised of 727 amino acids, it contains a central zinc finger domain consisting of ten C_2H_2-type and one C_2HC-type zinc fingers (Filippova et al., 1996). The zinc fingers function to facilitate DNA binding and nine of the zinc fingers are required for CTCF binding to the human *c-myc* promoter. Within the carboxyl terminal are three highly acidic regions and the zinc finger region is flanked by two positive domains (Filippova et al., 1996).

CTCF has been shown to regulate transcription in several DNA viruses (For Review see (Pentland and Parish, 2015). Within the genome of the Kaposi's sarcoma-associated herpesvirus (KSHV) CTCF can associate with multiple regions. Through co-localising and dimerising with a ring-like cohesion complex and binding to these regions in the KSHV genome, CTCF functions to repress expression of lytic genes (Chen et al., 2012). In the Epstein-Barr virus (EBV) genome, there are different sites for CTCF binding, located at the type I latency promoter and the type III latency promoter. Binding at these sites results in long-range interactions with the OriP enhancer (Tempera et al., 2011). There are two conserved binding sites for CTCF in the hepatitis B virus (HBV) genome. Binding of CTCF to these sites results in repression of the HBV enhancer (EnhI) and therefore repression of HBV gene transcription (D'Arienzo et al., 2021).

CTCF is important in regulating HPV early transcripts during the differentiation dependent life cycle of HPV. High-risk HPV types contain a conserved binding site for CTCF situated within the E2 open reading frame. In HPV18 this is positioned around nucleotide ~2926. Although located approximately 3000bp downstream from the viral upstream regulatory region, CTCF is able to repress transcription of

E6 and E7 when bound to this site (Paris et al., 2015, Pentland et al., 2018). This is because CTCF forms a long-range interaction with the transcription factor YY1, bound at the viral enhancer. This epigenetically repressed chromatin loop is formed within undifferentiated keratinocytes where the abundance of YY1 at the upstream regulatory region is enriched. During this process, recruitment of PcG proteins to the chromatin results in histone modification that favour transcriptional repression using mechanisms as detailed above. The resulting heterochromatin configuration represses E6/E7 transcription from the early promoter until a differentiation-dependent reduction in YY1 alleviates the loop formation and enables a productive infection through increased E6/E7 oncogene expression (Pentland et al., 2018). The mechanism of CTCF-YY1 mediated transcriptional repression is detailed in **Figure 1.9.**

In the context of HPV16 binding of CTCF in W12 cells has been shown to be important for the regulation of early transcripts. In monolayer W12 cells CTCF is enriched at the E2 ORF and following differentiation there is a significant increase in binding both at the E2 ORF and also at the L2 CTCF binding site. In monolayer W12 cells depletion of CTCF results in a significant increase in early transcripts (E6all, E6full, E6*I, E6*II, E6*III, E6*X, E2) except from E6*IV which is significantly reduced following depletion of CTCF (Groves et al., 2021).

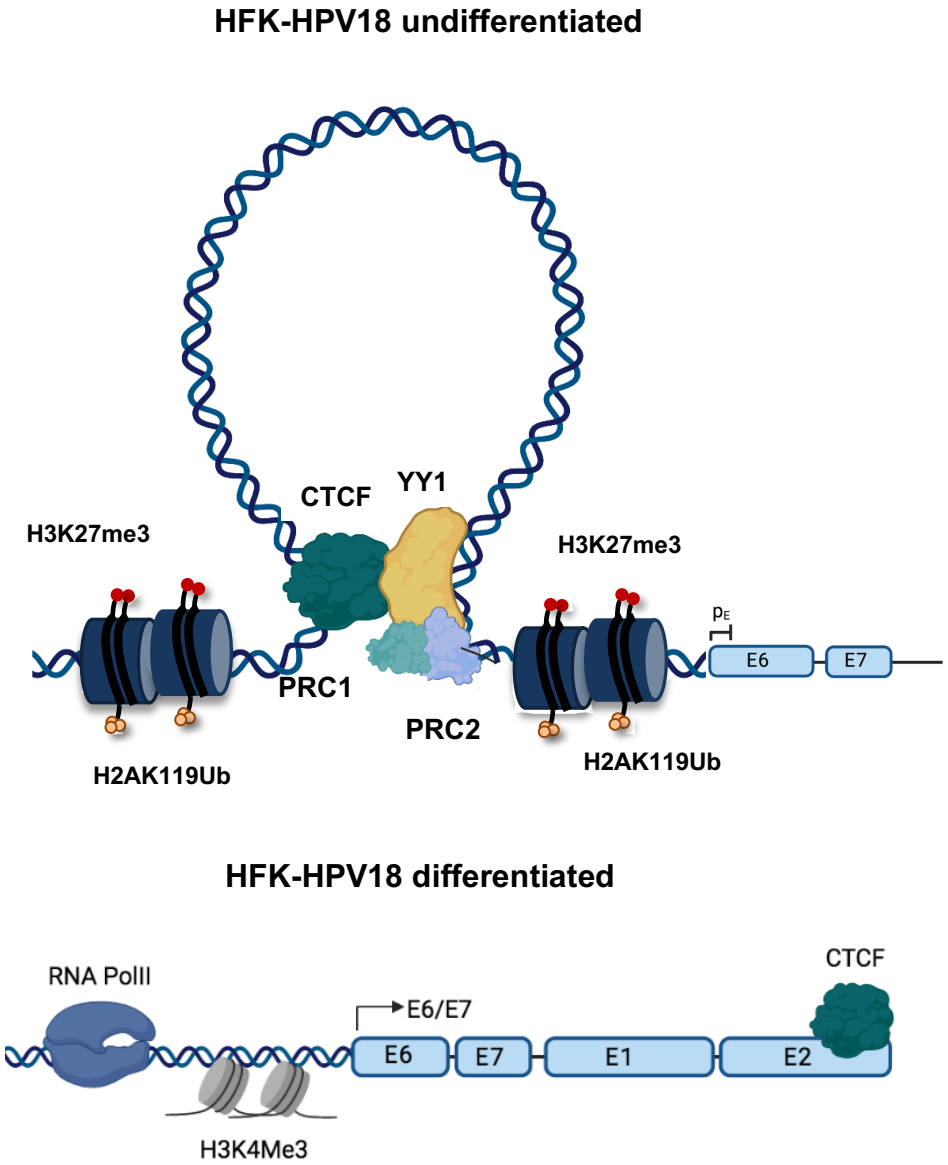


Figure 1.9 CTCF-YY1-mediated epigenetically repressed chromatin loop regulates E6/E7 oncogene expression during the HPV18 life cycle. Within an undifferentiated keratinocyte (top) CTCF and YY1 bound to the HPV18 episome form a long-range epigenetically repressive chromatin loop. YY1 functions as a repressor of HPV transcription by recruiting PcG proteins to the URR which results in deposition of H3k27Me3 and H2AK119Ub. The resulting chromatin is compacted and as such transcription of E6/E7 oncogenes is repressed. Differentiation (bottom) results in a significant depletion of YY1 protein expression and abrogation of binding to the URR. CTCF binding is unaffected. This results in a loss of chromatin looping and therefore no recruitment of the PcG complex. The resulting chromatin accessibility increases and host-cell factors such as RNA Pol II are recruited to the viral URR to initiate transcription from the early promoter. Figure adapted from (Pentland et al., 2018).

1.6 The natural history of an HPV-infection

HPV infection and replication is targeted to keratinocytes at multiple anatomical sites including the anogenital and oropharyngeal epithelium. However, the risk of HPV-associated cancer varies depending on the anatomical site. Whilst 99.7% of cervical cancers are associated with oncogenic HPV this is estimated to be at only ~70% in the head and neck sites. This might represent differences in the predominant virus type, or perhaps differences in the anatomy of each site. The majority of cervical cancers arise in a region termed the transformation zone at the squamo-columnar junction. This site contains a population of layered cuboid cells that sit between the interface of the squamous and columnar epithelium. The cells that comprise this site are highly susceptible to HPV infection and as such are commonly the origin of squamous and adeno-carcinomas (Mirkovic et al., 2015, Herfs et al., 2012).

The transition from normal cervical epithelial cells into invasive cervical cancer is a well-defined process **(Fig 1.10).** Persistent infection with oncogenic HPV (predominantly HPV16) is a prerequisite for cervical cancer development. This results in the continued expression of high-risk oncogenes. High-risk E6 and E7 work synergistically to target multiple cellular factors that are involved in regulating cell growth and arrest and as such are essential for the immortalisation of human keratinocytes (Münger et al., 1989a). Long-term infection with high-risk HPV results in pre-malignant changes known as dysplasia. Dysplasia is defined as abnormal cellular growth – the presence of dysplastic areas within the epithelium is associated with an increased risk of developing cancer. Dysplastic

lesions in the cervix are defined as cervical intraepithelial neoplasia (CIN) which can be graded depending on the severity (mild, moderate and severe dysplasia) (Woodman et al., 2007). If not detected early through cervical screening, these dysplastic lesions can eventually culminate in invasive cervical cancer.

The transformation zone for oropharyngeal cancer has not been elucidated. The oropharynx is comprised of the palatine tonsils, base of tongue, soft palate, tonsillar pillars and uvula (Haeggblom et al., 2017). Although this site is relatively heterogenous the onset of HPV-induced cancer predominantly resides within the tonsils and base of tongue with over 80% of cancers arising in this site. In contrast, only a subset of HPV-driven cancers are found within the soft palate and oropharyngeal walls (Haeggblom et al., 2017).

1.6.1 Integration

A hallmark of HPV-driven cancers is integration of the viral genome into the host cell chromatin. Precancerous cervical lesions including CIN2 and CIN3 harbour a subset of integration events, the frequency of which appears to increase during the progression to invasive cervical cancer which predominantly contains integrated HPV **(Fig 1.10)** (Klaes et al., 1999, Bodelon et al., 2016). Over half of all integration events involve HPV16 and only a handful contain integrated HPV18. In OSPCCs integration of HPV also occurs albeit at a lower rate (Hafkamp et al., 2003, Deng et al., 2013, Parfenov et al., 2014). At this anatomical site integration of HPV18 is most prevalent (Roberts et al., 2019).

There have been two well characterised integration events; type 1, where all copies of viral DNA are disrupted, and type 2, which consists of multiple lengths

of HPV in the host genome (Jeon et al., 1995). Mapping of integration has revealed that the majority of these events occur either within genes or within 20kb of a gene, often within common fragile sites (Parfenov et al., 2014, Bodelon et al., 2016). This can result in disruption of cellular oncogenes or tumour suppressor genes. Integration events can also occur close to enhancers and transcription start sites. As such they have been shown to be close to sites of open chromatin as indicated by the presence of active histone marks such as H3K4Me3 and H3K27Ac or the presence of writers such as EZH2 (Bodelon et al., 2016). An example of this is the W12 clone 20861 which contains 30 copies of tandemly integrated HPV16 at a single locus. In this clone, the integration locus is enriched with Brd4, H3k27Ac and mediator of RNA polymerase II transcription subunit 1 (MED1). This suggests that Brd4 can form a super-enhancer-like element to drive E6/E7 oncogene expression within these cells (Dooley et al., 2016).

Integration of HPV16 in cervical epithelial cells can result in a selective growth advantage, one mechanism of which is thought to be increased E7 oncogene expression (Jeon et al., 1995). In the head and neck integration does not always correlate with increased E6E7 oncogene expression, suggesting that HPV-integration might function to alter the host genome at this anatomical site. This includes DNA amplification, disruption of tumour suppressor genes, intrachromosomal rearrangements and altered transcript expression (Parfenov et al., 2014). Although HPV-driven cancers often contain multiple integration sites, usually only one of these is transcriptionally active (Van Tine et al., 2004, Xu et al., 2013, Warburton et al., 2021).

1.6.2 Genome instability

Several integration hotspots have been identified within HPV driven cancers including chromosome 3q28, 8q24.21 and 13q22.1 (Bodelon et al., 2016). The result of these events is often enhanced chromosomal instability. As an example, chromosome 8q24.21 harbours the oncogene *c-myc*, therefore integration at this site reflects possible growth advantages. High grade cervical intraepithelial lesions often display greater DNA copy number imbalance when compared to normal or even low-grade lesions. Often, integration events that result in loss of HPV E2 expression appear to result in greater genomic instability (Alazawi et al., 2004). Examples of genes that HPV integrates into within HNSCC: RAD51 homolog 2 (RAD51B), ETS2 gene, NR4A2 – HPV integration can drive cancer by altering the host genome (Parfenov et al., 2014).

1.6.3 Current models of HPV-driven disease

A well-established system to model the HPV life cycle is the use of primary human keratinocytes. These can be transfected with recircularised high-risk HPV genomes and selected for using a drug resistant marker. Establishment of HPV-containing keratinocytes can take several months but the end result is cells that are immortalised by the expression of oncogenic E6 and E7. These models, whilst physiologically relevant, are mostly used to study the events which occur during the HPV life cycle.

Using cell lines established from cervical intraepithelial lesions enables insight into cellular changes that occur in in premalignant lesions. Examples of this are CIN162 which was established from a CIN1 lesion. This cell line contains

episomal HPV16 DNA (De Geest et al., 1993). Another example is W12 cells which again are derived from a CIN1 lesion and contain HPV16 DNA (Stanley et al., 1989). Interestingly, multiple subclones of W12 cells exist which contain either episomal or integrated DNA. These subclones have been important in elucidating the function of HPV integration in driving high E6/E7 oncogene expression (Jeon et al., 1995, Jeon and Lambert, 1995, Warburton et al., 2018). However, there are currently no available cell lines established from dysplastic lesions from the oropharynx (Roberts et al., 2019).

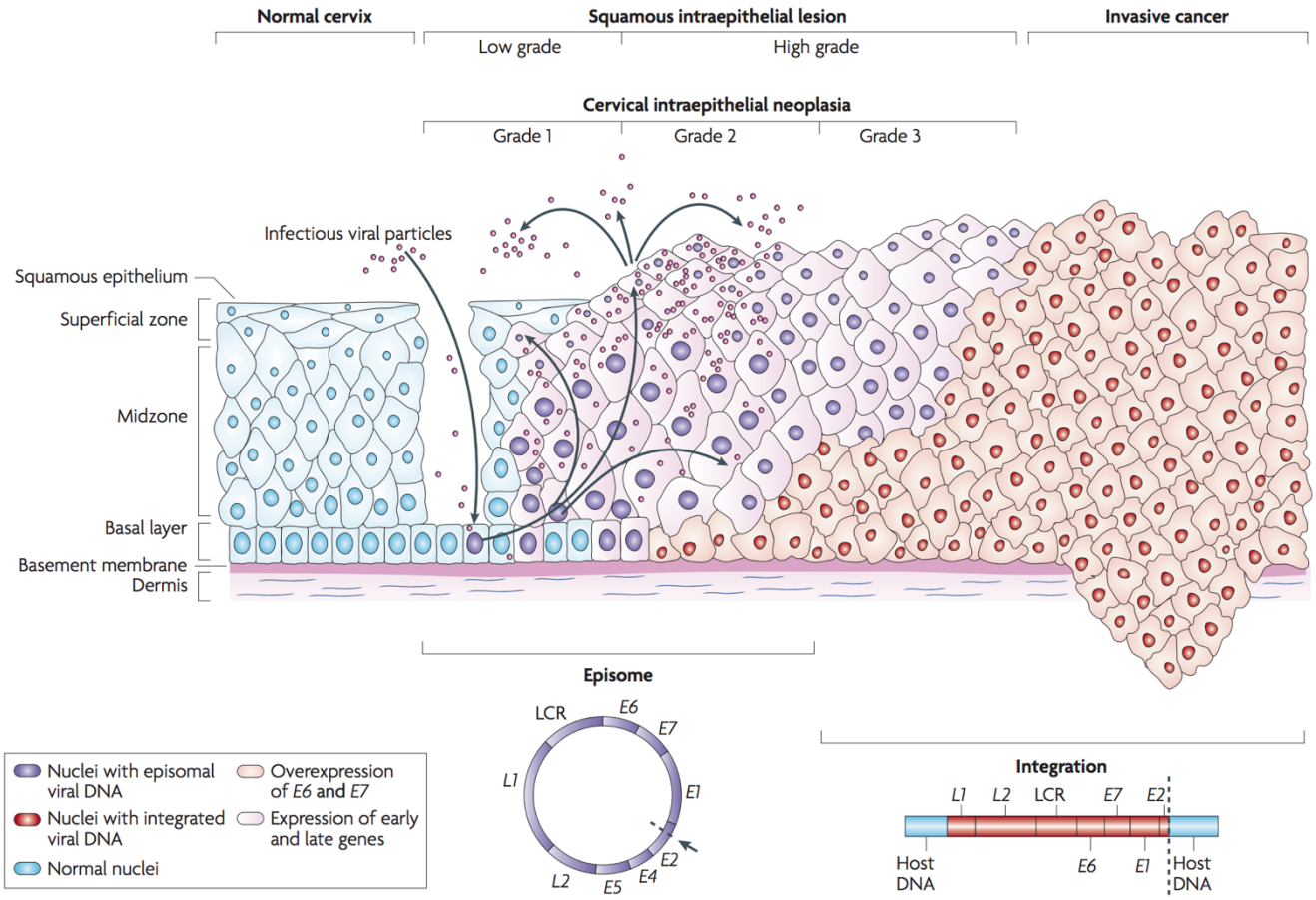


Figure 1.10 Progression of HPV-mediated cervical cancer. Persistent infection with oncogenic HPV can progress to invasive cervical cancer. This is a well characterised process which involves the development of low grade cervical intraepithelial neoplasia (CIN) and eventually culminates in invasive cervical cancer. This is associated with high expression of the viral oncogenes E6 and E7 which contribute to aberrant cellular proliferation. Constitutively high viral oncogene expression occurs as a result of viral integration into host cell chromatin which primarily disrupts the E2 open reading frame. Progression from a normal cervical epithelium into invasive cancer can take many years. Figure taken from (Woodman et al., 2007).

1.7 Hypothesis and aims

In order to reduce host immune clearance and increase the likelihood of a persistent infection, HPV gene expression is tightly controlled throughout the viral life cycle. This is in part regulated via transcription factor binding to the viral upstream regulatory region, but also integral to this process is epigenetic modification. We have shown that during the HPV life cycle, E6/E7 transcription is repressed by the host cell DNA binding proteins CTCF and YY1 through an epigenetically repressive, long-range, chromatin loop. Whilst we have demonstrated that CTCF-YY1-dependent looping is essential in regulating HPV early gene expression within the normal HPV life cycle, it is unclear whether this mechanism of transcriptional control is deregulated during carcinogenesis. Mutational analysis of HPV-driven oropharyngeal cancers using both TCGA analysis and whole exome sequencing of tumour samples revealed that chromatin regulators are highly mutated in these cancers. This included the lysine methyltransferase (KMT2C), the E1A binding protein p300, CBP and importantly, CTCF (Haft et al., 2019). CTCF is often prone to missense mutations at this anatomical site. Interestingly, YY1 has been shown to be overexpressed in cervical squamous cell carcinoma and this was positively correlated with high HPV16 E6 expression (Wang et al., 2018). We have previously shown that CTCF and YY1 are significantly more abundant in HPV+ tonsil tumour sections when compared to HPV- tonsil tumour sections (data not published). Prompted by this, we hypothesised that the CTCF-YY1-mediated regulation of HPV oncogenes would become deregulated during a persistent infection with oncogenic HPV to drive carcinogenesis of the head and neck.

Therefore, the aims of this thesis are:

1. To validate the use of low-passage and high-passage HPV16 and HPV18 containing human tonsil keratinocytes as a model of HPV-driven disease progression.
2. To determine whether CTCF and YY1 function to repress HPV early gene transcription within the normal HPV life cycle and to characterise the changes to CTCF-YY1-mediated regulation of E6/E7 oncogene transcription following the long-term culture of HPV16 and HPV18 genome containing tonsil keratinocytes.
3. To assess whether CTCF-YY1-mediated viral oncogene repression is disrupted during HPV-driven cancer using patient derived HPV16-driven oropharyngeal squamous cell carcinoma cell lines.

Chapter 2 Materials and Methods

2.1 Cell lines and culture conditions

2.1.1 Culture conditions

All cell culture work was carried out in a class II biosafety cabinet (Holton Laminar S200 1.2) using standard aseptic techniques. Routine cleaning and servicing of the safety cabinets took place.

2.1.2 Mycoplasma tests

Routine mycoplasma tests were carried out using the MycoAlert® Mycoplasma detection kit (Lonza) as per manufacturer's instructions. Briefly, 1 ml of media was taken from confluent cell culture plates and centrifuged for 5 minutes at 2.5 xg. The lysed test sample was incubated for five minutes with 75 µl of the MycoAlert® reagent (1:1 ratio). Using the GloMax® Explorer Microplate Reader (Promega) the levels of ATP were measured, and a first reading (read A) was taken. The sample was subsequently incubated for 10 minutes with 75 µl of MycoAlert® substrate which catalyses the conversion of ADP to ATP. A second reading (read B) was taken again using the GloMax® plate reader. To determine the presence of Mycoplasma, a ratio between read A and read B was calculated.

2.1.3 Human tonsil keratinocytes

Human tonsil keratinocytes (hereafter HTKs) were isolated from tonsil epithelia by Dr S. Roberts (HSE ethical approval number 06/Q1702/45) and transfected with re-circularised HPV16 or HPV18 genomes. HTKs that had been established from two

independent donors (#donor1, #donor2) were used throughout the study. HTK-16 and HTK-18 lines were cultured in E-medium (DMEM 60% v/v (Sigma; D6171), Ham's F12 32% v/v (Gibco), Pen-strep 2% v/v (Gibco), Hydrocortisone 0.1% v/v (Sigma), Hyclone FBS 5% v/v (Sigma), mouse epidermal growth factor 0.5% v/v (BD Biosciences), Cholera toxin A 0.1% v/v (Corning), Insulin 0.1% v/v (Sigma), Transferrin 0.1% v/v (Sigma), Tri-idio- 1L-thyronine, 2 nM (Sigma) and Adenine, 1.8 mM (Sigma)). HTK-16 and HTK-18 lines were grown on irradiated J2-3T3 fibroblasts in 10 cm Falcon® tissue culture dishes (Corning). Irradiated J2-3T3s were seeded at a density of $2x10^6$ and left to adhere for 24 hours before HTKs were seeded. Low-passage HTKs were maintained between passage 5 and passage 10 whilst high-passage HTKs were maintained between passage 20 and passage 30.

2.1.4 J2-3T3 Fibroblasts

J2-3T3 fibroblasts were cultured and maintained in DMEM (Sigma; D1671) supplemented with bovine serum (10% v/v, Gibco) and 2 mM L-glutamine (2% v/v). At 90% confluency, J2-3T3s were trypsinized and resuspended in complete E-media to a final density of $2x10^6$/ml. J2-3T3s were exposed to 30 Gray (Gy) irradiation using the CellRad x-ray source irradiator and were stored at 4°C for no longer than one week.

2.1.5 Head and neck squamous cell carcinoma lines

Human oropharyngeal squamous cell carcinoma (OPSCC) lines were donated by Professor Hisham Mehanna, University of Birmingham. The details of the cell lines used are listed in **Table 2.1.** UM-SCC-47, VU-SCC-147, UPCI-SCC-040 and

CAL27 were cultured in Dulbecco's Modified Eagles medium (DMEM) (Sigma; D1671) supplemented with fetal bovine serum (FBS) (10% v/v, Sigma), pen-strep (1% v/v), non- essential amino acids (1% v/v) and 1 mM L-glutamine (1% v/v, Gibco). UM-SCC-104 were cultured and maintained in DMEM (Sigma; D1671) supplemented with 10% FBS, 1% pen-strep and 1 mM GlutaMAX ™. All OPSCC cell lines were assessed for HPV using HPV16 PCR, p16 protein expression and p21 protein expression.

2.1.6 Cryopreservation

Cell stocks were generated by resuspending cell pellets containing $1x10^6$ cells into 1 ml of freezing media. HTKs were preserved in E-media containing 20% FBS (v/v) and 10% (v/v) glycerol and OPSCCs and J2-3T3s were preserved in DMEM with 20% FBS (v/v) and 10% (v/v) DMSO. Cell stocks were placed in cryovials (Greiner Bio-One, Stroudwater, UK) and stored at -80°C in a Nalgene Mr Frosty system. Cell stocks were preserved for long-term in liquid nitrogen.

2.1.7 Cell growth assay

HTK-16 and HTK-18 were seeded at a density of $3x10^4$ on to a layer of irradiated J2-3T3s ($2x10^5$) in six-well dishes. Cell counts were taken 48 hours, 72 hours, 144 hours and 168 hours post-seeding. Irradiated J2-3T3s were removed and cells were washed twice in PBS. Cells were trypsinized and transferred into 1.5 ml tubes before centrifugation at 5000 xg for 5 minutes. Cell pellets were resuspended into 1 ml of Trypan Blue 0.4% (v/v) solution (Lonza) and viable cells were counted. Cell doubling times were calculated as described by Delury et al (Delury et al., 2013).

2.1.8 siRNA knockdown of proteins

Cells were seeded onto a 6-well dish and cultured until they had reached 70% confluency. HTK-16 and HTK-18 were seeded onto $2x10^5$ irradiated 3T3-J2s. CTCF-specific, YY1-specific, RNF2-specific or scrambled TARGET*plus* Smartpool siRNAs were purchased from Horizon and were prepared in Opti-MEM reduced serum media (ThermoFisher Scientific). Cells were transfected with 10 nM or 25 nM siRNA using Lipofectamine™ RNAiMAX for 48 hours or 72 hours. Details of the ON-TARGET plus SMARTpool siRNA sequences can be found in **table 2.2**.

2.2 Southern blotting

2.2.1 DNA extraction

Cell pellets were resuspended in lysis buffer (400 mM NaCl, 10 mM Tris-HCl (pH 7.4), 10 mM EDTA) and vortexed for 30 seconds. RNase A was added to a final concentration of 400 µg/ml and the sample was incubated at 37°C for 30 minutes. Proteinase K was added to a final concentration of 50 µg/ml in addition to 0.2% (v/v) SDS. The sample was incubated at 37°C overnight. The cell lysate was passed through a blunt 19-gauge needle to shear cellular DNA. DNA was isolated using 6 ml of phenol: chloroform: isoamyl alcohol (25:24:1) followed by a centrifugation step at 537 *x*g for 5 minutes at 4°C. The aqueous phase was retained, and this step was repeated. Chloroform: isoamyl alcohol (24:1) was then added to the sample followed by a final centrifugation step at 537 xg for 5 minutes at 4°C. The aqueous phase was transferred to a new tube and precipitated overnight with 2-volumes of absolute ethanol and 0.25 M sodium acetate at -20°C. The samples were centrifuged at 537

xg for 30 minutes at 4°C and the pellet was washed with 70% ethanol. This wash step was repeated, and the sample was centrifuged at 537 xg for 10 minutes at 4°C. The DNA pellet was air dried for 5 minutes and resuspended in 50-100 µl Tris-EDTA buffer (TE buffer). DNA was incubated at 60°C for 10 minutes and the concentration determined using NanoDrop ND-1000 Spectrophotometer (Thermo Fisher Scientific).

2.2.2 Restriction enzyme digestion

Purified DNA was restriction enzyme digested overnight at 37°C. The digestion reaction was set up as follows: 5 µg of genomic DNA, 2 Units *Dpn*I, 10 Units restriction enzyme and 1X restriction enzyme buffer in a total volume of 20 µl. *Bam*HI was used to linearise the HPV16 genome and *Hind*III was used as a non-cutter. *Eco*RI was used to linearise the HPV18 genome and *Bgl*II was used as a non-cutter.

2.2.3 DNA resolution and transfer

Digested DNA was separated on a 0.8% (w/v) agarose gel prepared in Tris-Borate-EDTA buffer (TBE; 45 mM Tris-HCl, 45 mM boric acid and 1 mM EDTA) containing 0.2 µg/ml ethidium bromide. Digested samples were resolved alongside a HyperLadder 1kB (Bioline) overnight at 50 V. Assessment of DNA loading and digestion was carried out under ultraviolet light (UV) using the GeneFlash transluminometer (Syngene, UK). The gel was washed twice in 0.25 M HCL for 40 minutes in total followed by a wash step in 0.4 M NaOH for 30 minutes before being transferred onto a to Gene Screen plus^{TM} nylon membrane (Perkin Elmer, UK) via capillary action.

2.2.4 Preparation of the radiolabelled probe

HPV16 and HPV18 genomes were isolated from pUC-HPV16 and pGEMII-HPV18 plasmid using restriction enzyme digestion. The pUC-HPV16 plasmid was digested with *Bam*HI and the pGEMII-18 plasmid was digested with *Eco*RI and the released genome was resolved on a 1% (w/v) agarose gel. The DNA was visualised using a UV transilluminator and an 8 kbp band was excised using a scalpel ensuring to trim any excess agarose from the perimeter of the band. The DNA was purified from the gel using the Monarch® DNA gel extraction kit following manufacturers guidelines (New England Biolabs). Briefly, the excised band was dissolved in 4X dissolving buffer for 10 minutes at 50°C. The dissolved gel was placed into a spin column with a collection tube and this was centrifuged at 537 xg for 1 minute. The flow-through was discarded and the gel was washed twice using 200 µl of DNA wash buffer followed by a subsequent centrifugation step at 537 xg for 1 minute. The sample was transferred to a new microcentrifuge tube followed by the addition of 6 µl of DNA elution buffer which had been pre-heated to 50°C. The sample was incubated for 1 minute at room temperature and the DNA eluted by centrifugation at 537 xg for 1 minute.

To prepare the probe, 50 ng of purified DNA was diluted with TE buffer to a final volume of 13 µl. This was boiled at 100°C for 5 minutes and subsequently placed into an ice-water slurry for a further 2 minutes. The following components were then added to the purified DNA sample; 5 µl of oligonucleotide labelling buffer (OLB; 1 mg/ml random deoxynucleotide primers, 1 M HEPES, pH 6.6), 1.5 µl molecular grade BSA, 1.5 µl Klenow polymerase and 30 µCi of radiolabelled deoxynucleotide

(^{32}P-dCTP, EasyTides, PerkinElmer). The reaction was incubated for three hours at room temperature in a lead container. Unincorporated deoxynucleotide was then removed using Illustra Probe Quant G-50 microcolumn following manufacturer's instructions. The incorporation was assessed using a Geiger counter and the probe was stored at -20°C.

2.2.5 Hybridisation of radiolabelled probe to DNA

Hybridisation buffer was prepared at a 2X concentration (5X SSC: 750 mM sodium chloride and 75 mM trisodium citrate (pH 7.0)), 10X Denhardt's solution (0.2% (w/v) Ficoll 400, 0.2% (w/v) polyvinylpyrrolidone), 0.2% (w/v) BSA fraction V and 20% (w/v) dextran sulphate). This was then diluted 1:1 with formamide and SDS was added to a final concentration of 0.1% (v/v). Two micrograms of salmon sperm DNA was boiled at 100ºC for 10 minutes before cooling on an ice/water slurry for two minutes. This was then added to 10 ml of hybridisation buffer to form the pre-hybridisation buffer. The nylon membrane was rolled into a gauze and placed into a hybridisation cannister with 10 ml of pre-hybridisation buffer. This was gently rotated for 1 hour at 42ºC before being replaced with 10 ml hybridisation buffer containing 2 µg of boiled salmon sperm DNA and the radiolabelled DNA probe. The hybridisation cannister was rotated overnight at 42ºC.

2.2.6 Autoradiography

The membrane was washed twice in 2X SCC 0.1% SDS (600 mM sodium citrate, 300 mM sodium chloride and 0.1% v/v SDS, pH 7.0) for a total of 30 minutes. This was followed by two washes in 0.5X SCC 0.1% SDS (150 mM sodium citrate, 75

mM sodium chloride and 0.1% v/v SDS, pH 7.0) and two subsequent washes in 0.1X SCC 0.1% SDS (30 mM sodium citrate, 15 mM sodium chloride and 0.1% v/v SDS, pH 7.0) again for a total of 30 minutes. Following a final wash step in 0.1X SCC 1% SDS (150 mM sodium citrate, 75 mM sodium chloride and 1% (v/v) SDS, pH 7.0) at 55°C for 30 minutes, the membrane was then exposed to x-ray film at -80°C and developed using an Xograph imaging system compact X4.

2.3 Western blotting

2.3.1 Cell lysis and protein concentration

Cells were grown to 80% confluency on a 10 cm culture dish and were harvested by trypsinisation. Cell pellets were lysed in 125 µl of urea lysis buffer (8M Urea, 100mM Tris-HCL, pH 7.4) which had been supplemented with 14 mM β-mercaptoethanol and 14 mM protease inhibitors. Cell lysates were sonicated for 10 seconds at 30% amplitude using the VibraCell™ high-intensity ultrasonic processor fitted with a micro probe. Lysed samples were centrifuged at 16,000 xg for 15 minutes and the supernatant transferred to a new 1.5 ml tube.

To determine protein concentration a standard curve was generated using six bovine serum albumin (BSA) (Sigma-Aldrich) standards (2 mg/ml – 0.0625 mg/ml). Protein lysates were diluted 1:5 and the protein concentration was determined using 1ml of Bio-Rad protein assay reagent (Hemel Hempstead, UK). The absorbance (595 nm) of the BSA standards and protein lysates was measured using a Biophotometer (Eppendorf, Hamburg, Germany). A standard curve was generated

from the absorbance readings using the equation x=(y-c)/m where x = unknown concentration, y = absorbance reading, c = y-intercept and m = gradient.

2.3.2 SDS-PAGE

SDS-PAGE was carried out using 10 µg – 30 µg of total protein. Protein lysates were prepared with 1X SDS-gel loading buffer (50 mM Tris pH 6.8, 2% (w/v) SDS, Bromophenol Blue, 10% glycerol, 100 mM DTT) and denatured at 95°C for 5 minutes. Protein samples were resolved on a 4-20% Mini-PROTEAN® precast polyacrylamide gel alongside a PageRuler™ Plus Prestained Protein Ladder (10-250 kDa). Electrophoresis was performed using the Bio-Rad Mini-Protean™ system at 120V for 1-2 hours. Protein samples were transferred onto a 0.45 µm PVDF membrane (Thermo Scientific™). Four sheets of Whatman™ paper were pre-soaked in semi-dry transfer buffer (48 mM Tris, 39 mM glycine, 0.037% (v/v) SDS, 20% (v/v) methanol). The PVDF membrane was pre-soaked in methanol and placed onto the Whatman™ paper. The gel was then placed on to the membrane followed by four more pre-soaked sheets of Whatman™ paper. Any surface bubbles were removed by gentle rolling and the protein samples were transferred using the Trans-blot® Turbo™ Transfer System (Bio-Rad) at 25V 1A for 35 minutes.

2.3.3 Washes

The membrane was blocked in 5% (w/v) skimmed milk powder in TBS-T or 3% (w/v) BSA in TBS-T (HPV18E6) for 1 hour at room temperature and subsequently incubated in primary antibody overnight at 4°C under gentle agitation. Details of the primary antibodies can be found in **Table 2.3.** The membrane was washed in TBS-

T for 10 minutes at room temperature for three repetitions before being incubated in species-specific HRP-conjugated secondary antibody for 1 hour at room temperature. This was followed by three further washes in TBS-T for 10 minutes each.

2.3.4 Chemiluminescence

Proteins were visualised using Pierce™ enhanced chemiluminescence (ECL) western blotting substrate (Thermo Scientific™) and signals were captured using the Fusion FX digital chemiluminescence detection system (Vilber Lourmat, Collegien, France). Densitometry was performed using the Fusion FX software.

2.4 Chromatin immunoprecipitation qPCR

2.4.1 Formaldehyde fixation

Five 10 cm dishes of cells which had reached 80% confluency were trypsinized and harvested to generate a single pellet containing approximately $1x10^7$ cells. Cell pellets were resuspended in 5 ml of minimal media (D6171) and fixed in 2% (v/v) formaldehyde**.** The fixation time was limited to 5 minutes in total; pellets were inverted at room temperature for 3 minutes and subsequently centrifuged at 244 xg for 2 minutes (4°C). The supernatant was removed, and pellets were washed in ice-cold PBS and then centrifuged at 244 xg for 2 minutes (4°C). The formaldehyde fixation was quenched using 0.25 M glycine (Active Motif) by gentle inversion at room temperature for 3 minutes. Samples were subsequently centrifuged at 244 xg for 2 minutes (4°C) and washed again in ice-cold PBS. Fixed pellets were

transferred to a 15 ml Falcon and subsequently washed in 5 ml ice-cold PBS supplemented with 0.5 mM PMSF and a centrifuged for 10 minutes at 720 xg (4°C). Fixed pellets were stored at -80°C.

2.4.2 Shearing by sonication

Formaldehyde fixed pellets were lysed on ice for 30 minutes in 1 ml Lysis Buffer (Active Motif) supplemented with 0.5 mM PMSF and protease inhibitor cocktail (1X). Samples were Dounce homogenised for 30 repetitions using a tight pestle in order to aid release of intact nuclei. Samples were centrifuged for 10 minutes at 2500 xg (4°C) and the nuclei pellet was resuspended in 350 µl of shearing buffer supplemented with 0.5 mM PMSF and 1.75 µl PIC. The samples were sonicated on ice at 25% amplitude (30 seconds on/30 seconds off) for a total duration of 15 minutes and centrifuged at 17,949 xg for 10 minutes (4°C). Chromatin was stored at -80°C.

2.4.3 Shearing efficiency

Aliquots of chromatin were taken to assess shearing efficiency and to quantify samples. To 30 µl aliquots of chromatin 150 µl of water and 0.25 M of NaCl was added. Samples were heated to 65°C for 4 hours to reverse the cross links followed by the addition of 10 µg RNaseA and another incubation at 37°C for 15 minutes. Proteinase K (5 µg) was added, and the sample was incubated at 42°C for 1.5 hours. DNA was extracted with 200 µl phenol: chloroform: isoamylalcohol (25:24:1) and centrifuged at 17,949 xg for 5 minutes. The aqueous phase was retained, and DNA was precipitated overnight with 500 µl 100% ethanol, 0.12 M sodium acetate (pH

5.2) and 1 µl glycogen for visualisation of the DNA pellet. Samples were centrifuged at 17,949 xg for 10 minutes at 4°C and the pellet was subsequently air dried for 10 minutes at room temperature. The DNA pellet was eluted in 30 µl of water. A total of 10 µg of DNA was separated on a 1.5% (w/v) agarose gel alongside a HyperLadder 1 kb (Bioline) to assess shearing efficiency.

2.4.4 Chromatin immunoprecipitation

The immunoprecipitation reaction was set up as follows: 25 µg chromatin, 25 µl protein G magnetic beads, 1 µl PIC, 10 µl ChIP buffer-1 (Active Motif) and 5-10 µg of target antibody in siliconized 1.5 ml tubes **(table 2.3).** Input chromatin (10 µl of chromatin for 10% input) was set aside. Chromatin was immunoprecipitated overnight on an end-to-end rotator at 4°C and then washed once with ChIP buffer-1 and twice with ChIP buffer-2 using a magnetic tube rack. Chromatin was eluted in 50 µl elution buffer AM2 on an end-to-end rotator for 15 minutes at room temperature. Cross-linking was reversed in all immunoprecipitated samples using 50 µl of reverse cross-linking buffer. Both immunoprecipitated sample and the input control were incubated at 65°C for 2.5 hours followed by the addition of 1 µg proteinase K. Samples were incubated at 37°C for 1 hour and proteinase K stop solution was added to stop the reaction.

2.4.5 Quantification of ChIP efficiency

Both the input and immunoprecipitated samples were diluted 1:5 with nuclease free water and the amount of immunoprecipitated DNA was assessed using qPCR. The input cycle threshold (Ct) value was normalised based on the dilution factor (DF) of

the IP. The input fraction was 10% of the IP (DF 10) or 3.322 cycles (log2 of 10). This was subtracted from the Ct value of the diluted input to generate an adjusted input. Percent input was calculated using 100*2^ (Adjusted Input – Ct (IP)).

2.5 Formaldehyde isolation of regulatory elements (FAIRE)

A total of $4x10^6$ cells were pooled together and fixed in 2% (v/v) formaldehyde and sheared by sonication as per ChIP-qPCR. Chromatin was separated into two conditions: FAIRE or Input. To FAIRE DNA 150 µl of water was added. To the input sample, 150 µl of water and 0.2 M NaCl was added and this was then heated to 95°C for 15 minutes for reversal of cross-links. RNaseA (10 µg) was added and the sample incubated at 37°C for 15 minutes followed by the addition of 10 µg of Proteinase K and incubation at 67°C for 15 minutes. DNA was extracted from both the FAIRE and Input sample with 200 µl of phenol: chloroform: isoamylalcohol (25:24:1) as previously detailed and resuspended in 150 µl of TE buffer. The ratio between FAIRE:Input DNA was determined by qPCR using 2^-∆Ct.

2.6 Sequencing

2.6.1 Library preparation and sequencing

Sequencing libraries were prepared using the Next Gen DNA library Kit (#53216) and the Next Gen Indexing Kit (#53264) by Active Motif following manufacturer's instructions. Briefly, 10 pg of sheared DNA was adjusted to a final volume of 40 µl using low EDTA TE buffer. The fragmented DNA was repaired in two steps: repair I and repair II. The repair reaction I (13 µl low EDTA TE buffer, 6 µl Buffer W1 and 1

µl enzyme W2) was added to the sample and this was incubated at 37 ºC for 10 minutes. Post-repair clean-up was carried out using 54 µl SPRI beads (ratio: 0.9). The sample was incubated for five minutes at room temperature and the beads were subsequently separated using a magnetic rack. The beads were washed twice in 80% ethanol before being air dried. The repair reaction II (30 µl low EDTA TE buffer, 5 µl buffer G1, 13 µl reagent G2, 1 µl enzyme G3 and 1 µl enzyme G4) was added to the sample and this was incubated at 20ºC for 20 minutes. A further clean-up was carried out as previously described using 37.5 µl (ration: 0.75) of PEG NaCl solution.

Adapter fragments were ligated to the samples using a two-step ligation. The ligation I reaction (20 µl low EDTA TE buffer, 3 µl buffer Y1 and 2 µl of enzyme Y3) was added to the sample and incubated at 25ºC for 15 minutes. A further post-ligation I clean-up step was carried out using 36 µl (ration 1:2) PEG NaCl solution. The ligation II reaction (30 µl low EDTA TE buffer, 5 µl buffer B1, 2 µl reagent B2-MID, 9 µl reagent B3, 1 µl enzyme B4, 2 µl enzyme B5 and 1 µl of enzyme B6) was subsequently added to the sample followed by the addition of 5µl of the Y2 reagent (indexed adapter). This was incubated at 40ºC for 10 minutes followed by two washes in in 60 µl (ration 1:2) of PEG NaCl and 80% ethanol. The beads were then resuspended in 20 µl low EDTA TE buffer.

Using a magnetic rack, the supernatant containing the adapted library was transferred to a new 1.5 ml low retention microcentrifuge tube and this was amplified using PCR. The PCR reaction was set up as follows: 10 µl low EDTA TE buffer, 5 µl reagent R1, 4 µl reagent R2, 10 µl buffer R3 and 1 µl enzyme R4. This was added to the eluted library mix and placed in a thermal cycler. The thermal cycle was as

follows: 98ºC for 30 seconds; x6 (98ºC for 10 seconds, 60ºC for 30 seconds and 68ºC for 60 seconds). The library was cleaned up using 70 µl (ratio 1:4) of PEG NaCl and 80% ethanol. Libraries were sequenced using the NextSeq 500 (Illumina) using single read (1x75) mid-flow cells.

2.6.2 Sequencing data analysis

Sequencing data was analysed by Dr M. Wiench and Dr S. Clokie. Quality filtered single end reads were aligned to the HPV16 genome (NC_001526.2) using HISAT2. The sequencing reads were soft-clipped – this masks portions of reads that do not align to the reference genome from either the 3' or the 5' end. To identify potential integration events, reads which contained only one soft clipping were identified. In this instance, the soft-clipped reads would be those that do not align to the HPV16 reference sequence. The soft-clipped reads were recovered and aligned to the human genome reference sequence (Hg19) using HISAT. The reads were filtered using Ensembl and they were subsequently merged and overlapped to generate a bed file (.bed). Visualisation of potential integration events (soft-clipped fragments that aligned to the human genome) was carried out using IGV (2007, Thorvaldsdóttir et al., 2013).

2.7 Gene expression analysis

2.7.1 RNA extraction

RNA was extracted using the RNeasy Mini Kit (Qiagen, Hilden, Germany) following manufacturer's instructions. Cell pellets were lysed in buffer RLT (350 µl <5 x 10^6

cells or 600 µl per 5 x 10^6 – 1 x 10^7 cells) which had been supplemented with 14mM β-mercaptoethanol. Samples were vortexed for 30 seconds followed by the addition of 1-volume of 70% ethanol to the homogenised lysate. The sample was transferred to a RNeasy spin column and centrifuged at 8000 xg for 15 seconds. The flow-through was discarded and the column was washed with 700 µl of Buffer RW1 followed by two further wash steps with Buffer RPE. RNA was eluted in 20-35 µl of nuclease free water and quantified using the NanoDrop Nucleic Acid Quantification system.

2.7.2 DNase treatment

Genomic DNA was degraded using the Invitrogen ™ DNA-*free* ™ DNA removal kit (ThermoFisher Scientific). The reaction was set up as follows: 0.1 volume of 10X DNase I Buffer, 2 Units of rDNase I and up to 10 µg of RNA. The sample was incubated at 37°C for 30 minutes followed by the addition of 0.1 volume of DNase inactivation reagent for 2 minutes at room temperature. The sample was centrifuged at 10,000 xg for 1.5 minutes and the RNA was transferred to a fresh 1.5 ml tube.

2.7.3 cDNA synthesis

cDNA was transcribed using the Tetro cDNA synthesis kit (meridian Bioscience). RNA (200 ng – 500 ng) was added to a reaction tube with 40 µM of Random Hexamer, 10 nM dNTP, 1X RT Buffer, 1 µl RiboSafe RNase Inhibitor, 1 Unit of Tetro Reverse Transcriptase and nuclease free water in a final reaction volume of 20 µl. The samples were incubated at 25°C for 10 minutes, followed by 45°C for 30

minutes. The reaction was terminated by a final incubation step at 85°C for 5 minutes. All transcribed cDNA was diluted 1:5 before analysis by qRT-PCR.

2.7.4 qPCR

Quantitative PCR (qPCR) was set up with either cDNA or genomic DNA to assess gene expression analysis and DNA target detection between cells. The reaction utilises a SensiFAST™ probe which enables quantification of targets by binding to double-stranded DNA and emitting a fluorescent signal. Increasing DNA amplification results in greater intercalation and therefore a more intense fluorescent signal. The qPCR reaction was set up to a total volume of 20 µl as follows: 2 µl DNA/cDNA, 10 µl SensiFAST™ (No-ROX) (Meridian Bioscience®) and 0.125 nM forward and reverse primer **(listed in tables 2.4, 2.5 and 2.6).** Each sample was loaded in triplicate into a 96-well plate (Starlab) and the qPCR was performed using the Stratagene Mx3005P system and software. To ensure correct primer efficiency standard curves were also set up on the plate using serially diluted DNA/cDNA. Primer efficiency was considered acceptable if it fell between the range of 90-110%. The PCR cycle was a 3-step 40 cycle PCR, and the thermal profile was as follows: 95ºC for 10 minutes, x40 (95ºC for 30 seconds, T_a 30 seconds, 72 ºC 45 seconds), 72 ºC for 10 minutes. Assessment of primer efficiency, amplification curves, dissociation curves and expression changes were all carried out using the Mx3005P software.

2.7.5 Quantification of gene expression

The cycle threshold (Ct) is the number at which the fluorescent signal emitted from the SensiFAST^{TM} reaches levels above the fluorescence threshold (or background fluorescence). The Ct value is inversely proportional to the original relative expression level of the gene of interest. To assess gene expression changes between samples relative fold increase was calculated using the delta-delta Ct method (ΔΔCt). Firstly, the Ct values for each target are normalised to the Ct values of the housekeeping gene (β-actin) (ΔCt). Next, the normalised tested sample (e.g. low passage) is subtracted from the experimental sample (e.g. high passage) to generate the ΔΔCt. Finally, fold change is calculated using the following equation: 2^-ΔΔCt.

2.8 Statistical analysis

Statistical analysis was carried out using GraphPad Prism (version 9.5.0). The statistical test used was either a Student's *t test*, two-tailed, or a repeated measures one-way ANOVA using the Geisser-Greenhouse correction and comparing the mean of each column with the mean of every other column.

Table 2.1 Details of oropharyngeal squamous cell carcinoma lines (OPSCC)

OPSCC line	*HPV-status*	*Integration status*	*Origin*	*p53 protein expression*	*Tumour*
FaDu	Negative	N/A	Pharynx	-	Primary
CAL27	Negative	N/A	Tongue	Yes	Primary
UPCI-SCC-040	Negative	N/A	Tongue	No	Primary
VU-SCC-147	Positive	Integrated	Floor of mouth	Yes	Primary
UM-SCC-47	Positive	Integrated	Lateral tongue	-	Primary
UM-SCC-104	Positive	Integrated	Floor of mouth	-	Recurrent

Table 2.2 Dharmacon ON-TARGET plus SMARTpool siRNA target sequences

SMARTpool	*Target sequence*
YY1 (J-011796-08)	CAUAAAGGCUGCACAAAGA
YY1 (J-011796-09)	UAACACAUGCUAAGGCCAA
YY1 (J-011796-10)	GAACUCACCUCCUGAUUAU
YY1 (J-011796-11)	CCAAACAACUGGCAGAUU
CTCF (J-020165-07)	GAUGAAGACUGAAGUAAUG
CTCF (J-020165-08)	GGAGAAACGAAGAAGAGUA
CTCF (J-020165-09)	GAAGAUGCCUGCCACUUAC
CTCF (J-020165-10)	GAACAGCCCAUAAACAUAG
RNF2 (J-006556-05)	CGAGAUACAUAAGACUUC
RNF2 (J-006556-06)	GUAUCUGGCUGUGAGGUUA
RNF2 (J-006556-07)	GGAAUUGAUCCAGUAAUG
RNF2 (J-006556-08)	ACAAAGGAGUGUUUACAUC

Table 2.3 Primary antibodies

Antibody	*Clone/ID*	*Manufacturer*	*Catalogue number*	*Species*	*ChIP*	*WB dilution*
Anti-FLAG® (mAb)	M2	Sigma Aldrich	F1804	Mouse	5 µg	-
β-actin	AC-15	Sigma Aldrich	A5441	Mouse	-	1 in 1000
CTCF (pAb)	AB_2614975	Active Motif	61311	Rabbit	8 µg	1 in 1000
GAPDH	6C5	Santa Cruz	sc-32233	Mouse	-	1 in 1000
HPV18 E6 (mA)	(G-7)	Santa Cruz	sc-365089	Mouse	-	1 in 50
HPV18 E7 (mAb)	8E2	Abcam	ab100953	Mouse	-	1 in 1000
H3K4Me3 (pAb)	AB_2615077	Active Motif	39159	Rabbit	5 µg	-
H3K27Me3 (pAb)	AB_2561020	Active Motif	39155	Rabbit	5 µg	-
H3K27Ac (pAb)	AB_2614979	Active Motif	39133	Rabbit	5 µg	-
Ring1b (mAb)	3-3	Active Motif	39663	Mouse	5 µg	1 in 1000
RNA pol II (mAb)	4F8	Active Motif	61081	Rat	5 µg	-
Ubiquityl-Histone H2A (Lys119) (mAb)	D27C4 XP®	Cell Signalling	8240	Rabbit	5 µg	1 in 1000
YY1 (pAb)	AB_2793763	Active Motif	61779	Rabbit	10 µg	1 in 1000

Table 2.4 Primer sequences for HPV16. Nt, nucleotide. Ta, annealing temperature. Based on HPV16 sequence NC_001526.2

Primer pair	*Midpoint (Nt)*	*Fw (5'-3')*	*Rv (5'-3')*	*Ta (℃)*
L2 (4419- 4542)	4483	CAGGGTCGGGTACAGGCGGA	GGATCGGAAGGGCCCACAGGA	50
L2 CTCF (5123-5309)	5216	AGGCGTACTGGCATTAGGTACAGT	AGGTAAGGCTGCATGTGAAGTGGT	50
L1 CTCF (6039-6157)	6098	TGCAGCAAATGCAGGTGTGGAT	TGGGGATCCTTTGCCCCAGTGT	50
L1 (7045-7121)	7083	ACAAGCAGGATTGAAGGCCAAACCA	AGAGGTAGATGAGGTGGTGGGTGT	50
5'URR (7419-7552)	7485	TTTGTAGCGCCAGCGGCCATTT	GCATGGCAAGCAGGAAACGTACAA	50
Enh (7555-7681)	7618	CCAAATCCCTGTTTTCCTGA	CGTTGGCGCATAGTGATTTA	50
Early promoter (7854-7865)	7859	GCAAACCGATTTTGGGTTACA	ACTAACCGGTTTCGGTTCAA	50
E6 unspliced (111-223)	167	AGGACCCACAGGAGCGACCC	ACGTCGCAGTAACTGTTGCTTGCA	50
E6*1 (427-506)	466	GCCACTGTGTCCTGAAGAAAAGCA	GACCGGTCCACCGACCCCTT	50
Late promoter (649-765)	707	GACAGCTCAGAGGAGGAGGA	GCACAACCGAAGCGTAGAGT	50
E1 (2158-2316)	2237	AGGGTAGATGATGGAGGTGATTGG	GATTTACCTGTGTTAGCTGCACCA	50
E2 CTCF (2853-2950)	2901	GGAAACACATGCGCCTAGAATGTGC	TGATACAGCCAGTGTTGGCACC	50
E5 (3936-3958)	3947	ACGTCCGCTGCTTTTGTCTGTGT	ACCTAAACGCAGAGGCTGCTGT	50

Table 2.5 Primer sequences for HPV18. Nt, nucleotide, Ta, annealing temperature. Based on HPV18 sequence AY262282.1

Primer pair	*Midpoint (Nt)*	*Fw (5'-3')*	*Rv (5'-3')*	*Ta (°C)*
5'URR (7367-7433)	7367	GTGTGTTATGTGGTTGCGCC	GGATGCTGTAAGGTGTGCAG	58
ORI (7746-7845)	7795	ACTTTCATGTCCAACATTCTGTCT	ATGTGCTGCCCAACCTATTT	56
E6 (121-295)	208	ATCCAACACGGCGACCCTAC	GCAGCATGCGGTATACTGTCTCTA	56
Early (154-294)	224	TGTGCACGGAACTGAACACT	CAGCATGCGGTATACTGTCTC	58
Late (751-890)	820	CGAACCACAACGTCACACAAT	ACGGACACACAAAGGACAGG	58
E1 (1383-1460)	1460	GCAATGTATGTAGTGGCGGC	TACACTGCTGTTGTTGCCCT	58
E2 (2818-2949)	2883	TGCAGACACCGAAGGAAACC	CATTTCCCAACGTATTAGTTCC	58
E2 CTCF (2926-3118)	3022	GGCAACTAATACGTTGGGAAAA	TGTCTTGCAGTGTCCAATCC	56
E4 (3557-3575)	3556	TGGGAAGTACATTTTGGGAATAA	TCCACAGTGTCCAGGTCGT	58
E5 (3978-4080)	4029	TATGTGTGCTGCCATGTCCC	TCCACAGTGTCCAGGTCGT	56

Table 2.6 Primer sequences for the human genome. Ta, annealing temperature.

Primer pair	*Spanning exon*	*Fw*	*Rv*	*Ta (ºC)*
CTCF set 1	Exon 5	CAGTGGAGAATTGGTTCGGCA	CTGGCGTAATCGCACATGGA	60
CTCF set 2	Exon 4	ATGTGCGATTACGCCAGTGTA	TGAAACGGACGCTCTCCAGTA	60
CTCF set 3	Exon 6-7 junction	TCAGTGCAGTTGTGCAGTTA	TTCCCCTGAATGGGTTCTCAT	60
YY1 set 1	Exon 1	ACGGCTTCGAGGATCAGATTC	TGACCAGCGTTTGTTCAATGT	60
YY1 set 2	Exon 1	AAGAGCGGCAAGAAGAGTTAC	CAACCACTGTCTCATGGTCAATA	60
YY1 set 3	Exon 2	CCTCTCAGATCCCAAACAACTG	GCCTTTATGAGGGCAAGCTATT	60
RNF1 set 1	-	GGCCAGCGGAGCCCT	AGCCATCTGTTATTGCCTCCTGA	60
RNF1 set 2	-	AGTATTAGCCAGGATCAACAAGCAC	TGCCTCGCTGCAGTCTGTT	60
RNF1 set 3	-	GGCCAGACCCAAACTTTGATGC	CCTCAATGCTGTGACTGAGTGC	60
Positive Ctrl (CTCF)	-	TTGGGTAATGAGGAGGCGGGT	TCTTGGTCCTTGGGGATGGC	60
Negative Ctrl (CTCF)	-	AGAGGTCAAGTCTCTGCCCAG	AGAAATCCCACCGGGTTACTC	60

Chapter 3 Loss of CTCF-YY1-mediated HPV18 oncogene regulation in high passage primary tonsil keratinocytes

3.1 Introduction

HPV gene expression is tightly controlled throughout the virus life cycle. Expression of early proteins (including E6 and E7) in undifferentiated basal keratinocytes are maintained at low levels, presumably to avoid immune activation. Transcription regulation of early genes is in part mediated via transcription factor binding to the viral URR but also integral to this process is epigenetic modification through chromatin remodelling. We have previously identified the role of host cell binding proteins CTCF and YY1 in regulating E6/E7 oncogene expression through stabilisation of an epigenetically repressed chromatin loop. CTCF binds to the HPV18 episome within the E2 open reading frame and mediates the formation a chromatin loop through association with YY1 which occupies the viral URR. Repression of viral gene expression by chromatin looping is most profound in the undifferentiated basal cells. As the cells enter terminal differentiation, reduced YY1 expression results in a loss of chromatin looping in the viral genome and enhanced activation of viral transcription (Pentland et al., 2018). Given that uncontrolled E6/E7 expression drives carcinogenesis, we hypothesised that the CTCF-YY1-dependent epigenetic repression of virus transcription in the normal HPV life cycle would become deregulated during HPV pathogenesis to uncouple viral oncogene expression from cellular differentiation.

3.2 Validating high-passage HPV18-HTKs as a model of HPV-driven disease

One approach to model HPV pathogenesis includes establishing HPV genome-containing primary human keratinocytes. Primary tonsil keratinocytes were isolated from tonsil tissue and subsequently transfected with re-circularised HPV18 genomes and a neomycin resistance plasmid by Dr S. Roberts (HSE ethical approval number 06/Q1702/45). HPV containing keratinocytes were selected for using G418 antibiotic selection. The result of which being a population of cells that contain episomal HPV18 genomes that are maintained at approximately 50-100 copies per cell. The expression of high-risk HPV oncoproteins extends the lifespan of the keratinocytes by targeting tumour suppressor proteins to enable cell cycle re-entry and continuous proliferation. This model provides a physiologically relevant system to investigate HPV infections in the tonsil. Extended culture of these HTKs can be used to highlight events which occur over the natural course of an HPV infection, including cellular changes that are consistent with HPV-driven disease progression **(Fig. 3.1)** (Roberts et al., 2019). This is achieved through serially passaging the HPV containing tonsil keratinocytes resulting in multiple population doublings. In the current study, low-passage HTKs can be defined as those maintained between passage 5 and passage 10 whilst high-passage HTKs can be defined as those maintained between passage 20 and passage 30. The use of this system has not yet been fully characterised and therefore our initial aim was to validate the effect of long-term culture as model of HPV-disease progression

In a large number of cervical cancers HPV genomes integrate into host cell chromatin (Boshart et al., 1984, Schwarz et al., 1985). This is also true for oropharyngeal cancers albeit in a smaller proportion (Hafkamp et al., 2003, Parfenov et al., 2014). Disruption of any of the 7-9 open reading frames means that complete viral episomes cannot be synthesised, as such integration prevents completion of the HPV life cycle. However, in regard to carcinogenesis, integration can often disrupt the E1 and E2 open reading frames and transcription regulation therefore becomes disrupted. This results in constitutively high expression of E6 and E7 oncoproteins and aberrant cellular growth. Integration is therefore considered as a crucial event in the progression to cancer and as a consequence is often associated with poorer clinical outcomes. Often, the severity of disease is associated with integration. As such, precancerous lesions show a smaller frequency of integration events which increase during cancer progression (Klaes et al., 1999). The viral genome status in the low- and high-passage 18-HTKs has previously been assessed using Southern blotting in order to assess whether prolonged culture results in changes to the viral genome status that would mimic the events which occur during the natural history of an HPV infection. Genomic DNA was extracted, and restriction enzyme digested with *Bgl*II, a non-cutter of HPV DNA. The digested DNA was resolved on a 0.8% agarose gel, transferred to a GeneScreen Hybridisation transfer membrane and cross linked. HPV18 DNA was then visualised using a ^{32}P radiolabelled probe made from a purified HPV18 plasmid (Roberts et al., 2019). Southern blotting confirmed that all tested low-passage HPV18-HTKs contain episomal HPV DNA. In comparison, following long-term

culture, episomal DNA is drastically reduced, highlighting a likely integration event in the high-passage 18-HTKs (Roberts et al., 2019).

We wanted to assess whether integration had resulted in altered E6 transcription within the high-passage 18-HTKs to mimic the events which occur during in vivo disease progression. To assess this, the expression of full length E6 relative to β-actin was analysed by qRT-PCR. In both tested donors there was an increase in E6 transcript expression within the high-passage HTKs (#donor 1 p <0.05, #donor 2 n.s.) **(Fig. 3.2**). Given that it is the expression of E6 and E7 oncoproteins that is essential for perturbing host cell proteins involved in cell cycle regulation, including p53 and pRb, it was important to determine whether increased transcription was biologically significant. The expression E6 and E7 protein was therefore assessed by western blotting in the low- and high-passage 18-HTKs and normalised to β-actin expression. As shown in **Figure 3.3 (top panel)** there is an increase in E6 and E7 protein expression in the high-passage 18-HTKs for both tested donors in comparison to the low-passage cultures. In donor 1 there is a 1.5-fold increase in the expression of E6 and E7 protein within the high-passage keratinocytes. In donor 2 there is a 2-fold increase in E6 protein expression and a 3-fold increase in E7 protein expression within the high-passage keratinocytes **(Fig 3.3. bottom panel).**

A contributing factor to HPV driven premalignant lesions is increased cellular growth as a result of uncontrolled oncogene expression. In order to explore the biological significance of the increased oncoprotein expression in the high-passage 18-HTKs we assessed the cellular growth rate of the low-passage and high-passage 18-

HTKs. Cells were seeded at an equal density onto irradiated J2-3T3 feeder cells and were subsequently counted at 48, 72, 144 and 168 hours post seeding. Growth curves were generated using Trypan blue as a means of measuring viable cells. As expected, increased E6 and E7 oncoprotein expression is coincident with an increased growth rate and reduced doubling time within the high-passage 18-HTKs in both assessed donors – although this did not reach statistical significance **(Fig. 3.4).** Taken together, this data suggests that the prolonged culture of HPV18-HTKs can be used as a model to recapitulate the early events that occur in HPV-driven disease.

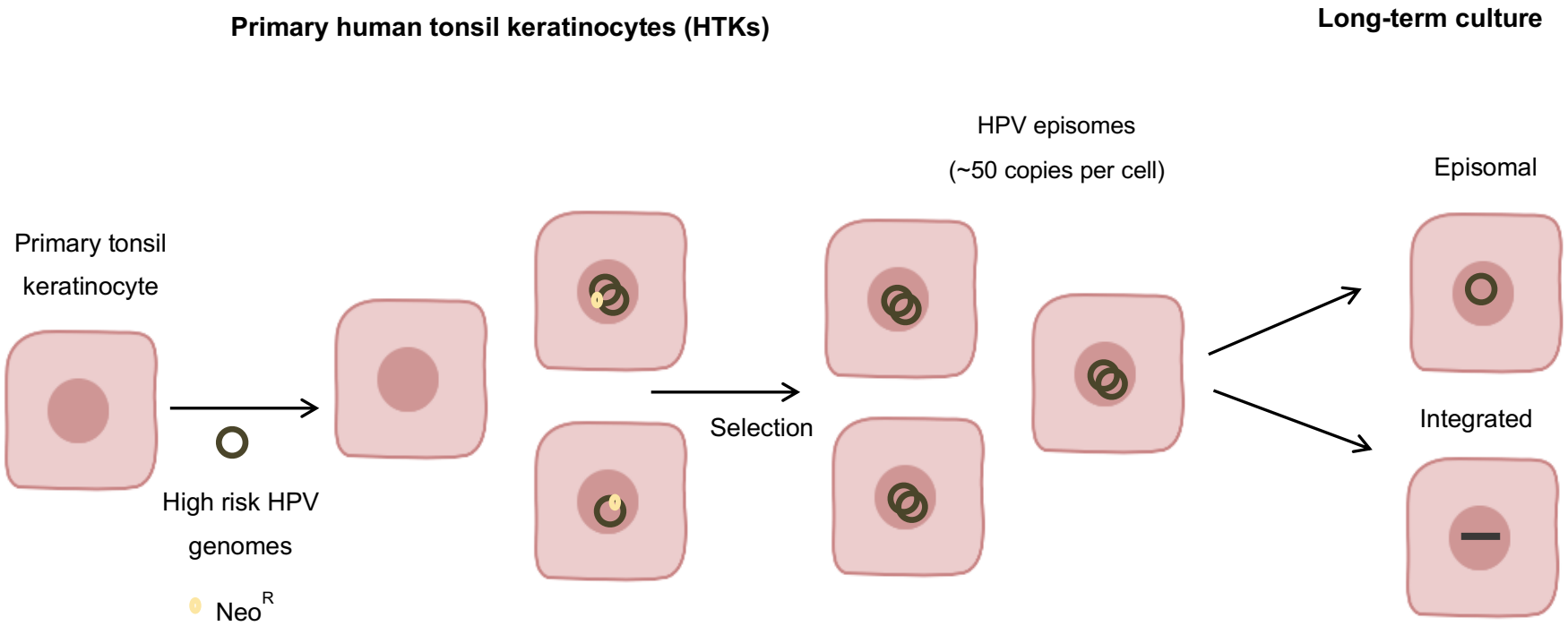


Figure 3.1 Schematic of primary human tonsil keratinocyte model system. Primary human tonsil keratinocytes (HTKs) were established from tonsillectomies. HTKs were then transfected with re-circularised high-risk HPV genomes alongside a neomycin resistance marker for selection. This results in a population of cells containing episomal HPV genomes. Long-term culture of these HTKs results in changes that are consistent with HPV-driven disease progression. Figure adapted from (Roberts et al., 2019).

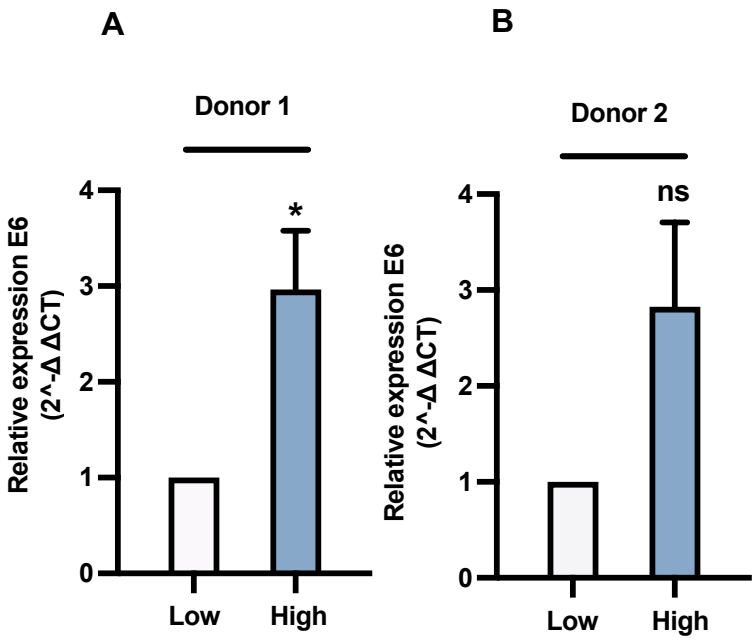


Figure 3.2 Increased transcription of HPV18 E6 in high-passage 18-HTKs. Relative fold change of E6 transcripts in high-passage HPV18 containing HTKs (A; donor 1, B; donor 2) normalised to β-actin (2^-ΔΔCt). Data show the mean ± SEM of three independent repeats. * $p < 0.05$ as determined by an unpaired *t* test.

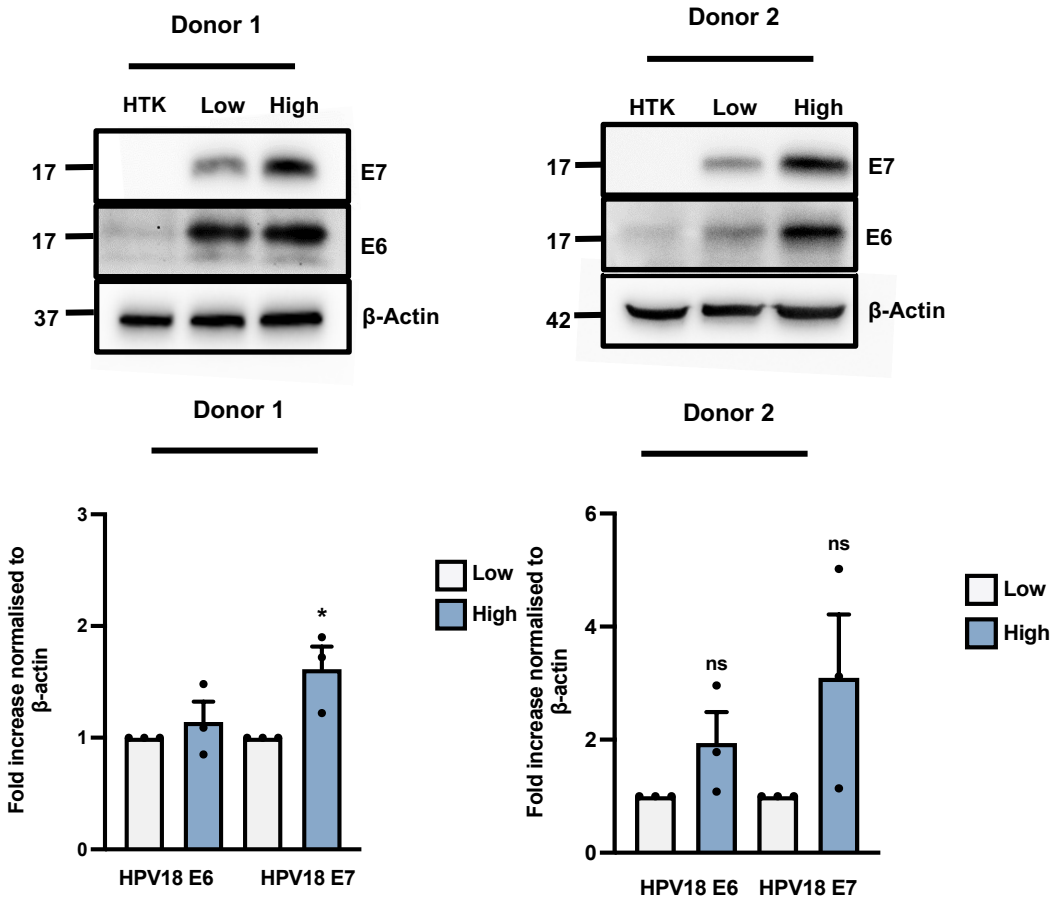


Figure 3.3 Increased E6 and E7 protein levels in high-passage 18-HTKs. Top panel: E6 and E7 protein levels were assessed in untransfected HTKs, low-passage and high-passage 18-HTKs normalised to β-actin. A total of 30 µg of protein was separated using a 4-12% polyacrylamide gel. Protein was transferred onto a PDVF membrane and imaged using a Fusion FX7 digital detection system. Data shown are one representative image from three independent repeats. Bottom panel: Protein densitometry was carried out using the Fusion FX7 digital detection system and relative fold increase normalised to β-actin was calculated. Data show the mean ± SEM from ≥ two independent repeats. * p <0.05 as determined by an unpaired t test.

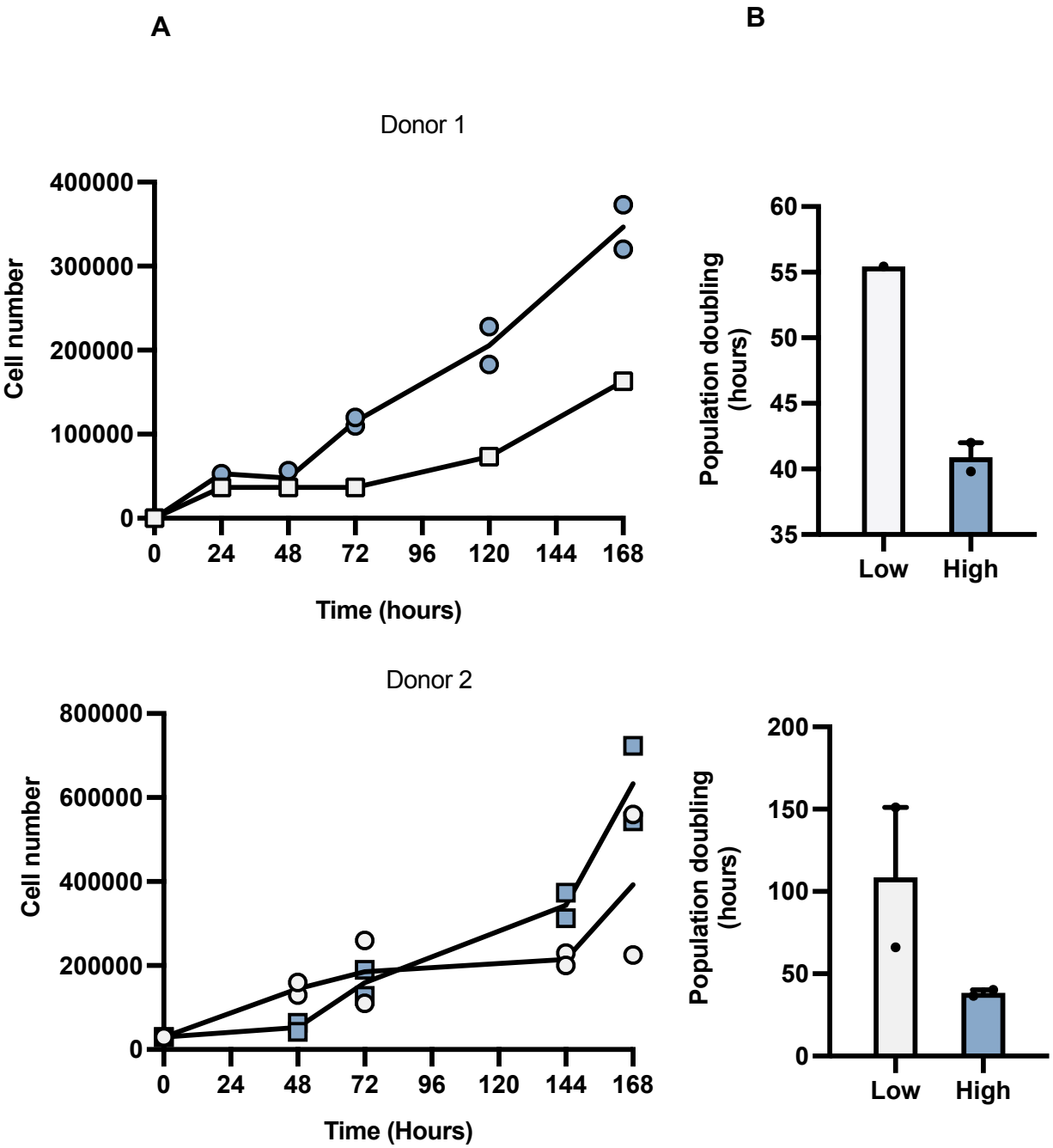


Figure 3.4 High-passage 18-HTKs show no change in growth rate or doubling time. (A) Cell growth assays were carried out using Trypan Blue staining as an indicator of viable cells. Cells were seeded at a density of 2 x 10^4 and viable cells were counted at from 24 hours up to 168 hours postseeding. (B) Doubling time of low-passage and high-passage 18-HTKs. Data show the mean ± SEM from ≥ one independent repeats.

3.3 Altered expression of CTCF and YY1 in high-passage 18-HTKs

Having established that the long-term culture of HPV18 containing HTKs results in changes consistent with the phenotype of HPV-driven disease, we next sought to determine whether CTCF and YY1 function to regulate HPV oncogene expression within our progression model. Our previous studies showed that following establishment of HPV18 episomes in primary foreskin keratinocytes CTCF protein expression is significantly increased (Pentland et al., 2018). To confirm whether this was consistent within the tonsil and to elucidate whether serially passaging the 18-HTKs alters CTCF and YY1 protein expression, we used western blotting. As shown in **Figure 3.5 (top panel)** western blot analysis of lysates harvested from untransfected HTKs, low-passage 18-HTKs and high-passage 18-HTKs confirmed that following establishment of HPV18 episomes in tonsil keratinocytes, CTCF and YY1 protein expression is increased.

Given that in the context of the HPV life cycle CTCF and YY1 function to repress transcription from the early promoter, we hypothesised that following long-term culture of the HPV18-HTKs we would see reduced expression of both proteins. This would be consistent with the detailed increase in E6 transcription in the high-passage 18-HTKs (see section 3.2). However, following long-term culture of the HPV18 HTKs we observed a further 2-fold increase in CTCF and YY1 protein which was consistent in two independent donors (#donor 1 $p < 0.01$, $p < 0.001$, #donor 2 $p < 0.05$, $p < 0.01$) **(Fig 3.5. bottom panel).** To assess whether this increase was as a result of altered transcription, we used qRT-PCR and analysed the expression of CTCF and YY1 transcripts using three different primer sets. Transcription of CTCF

was quantified relative to β-actin using primers that anneal to exon 4, exon 5 and the exon-exon junction between 5 and 6 of CTCF. Transcription of YY1 was quantified relative to β-actin using primers that anneal to exon 1 and exon 2 of YY1. Interestingly, the increase in CTCF and YY1 protein expression is the result of a post-transcriptional modification since there was no observed significant differences in either CTCF or YY1 transcript levels between the low-passage and high-passage HPV18-HTKs. This was consistent in both tested donors **(Fig 3.6).**

This data is in agreement with our previous studies where RNA-Seq and qRT-PCR analysis of CTCF transcripts revealed no significant differences in transcript levels following HPV18 establishment (Pentland et al., 2018). Taken together, this data shows that during the long-term culture of HPV18-HTKs there is a significant increase in the protein expression of CTCF and YY1. This is independent of any changes to transcription and is therefore likely the result of a post-transcriptional modification. Importantly, the significant increase in both proteins is coincident with high expression of E6 and E7 oncoproteins. This suggests that during HPV-driven disease progression the expression of CTCF and YY1 becomes uncoupled from E6/E7 transcription repression.

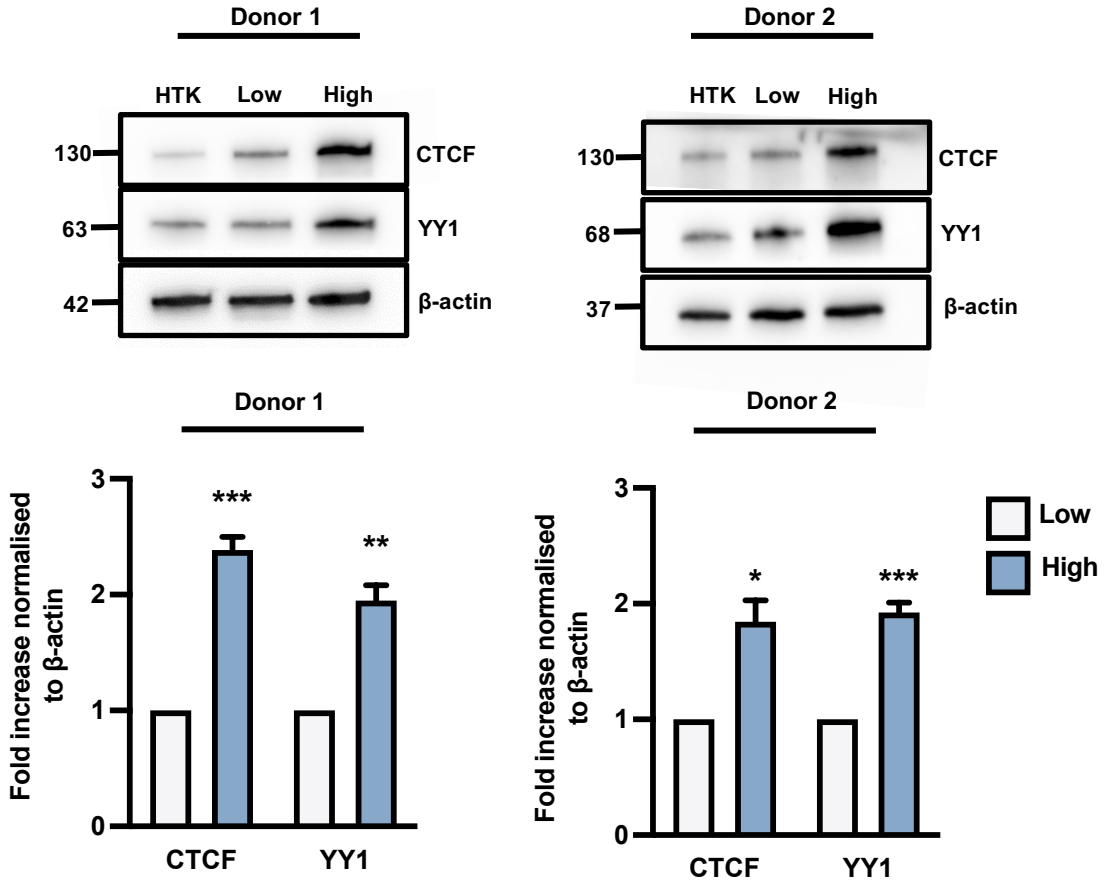


Figure 3.5 CTCF and YY1 protein expression increases in high-passage 18-HTKs. Top panel: CTCF and YY1 protein expression was assessed in untransfected HTKs, low-passage, and high-passage HPV18 containing HTKs and normalised to β-actin. A total of 10 µg of protein was separated using a 4-12% polyacrylamide gel. Protein was transferred onto a PDVF membrane and imaged using a Fusion FX7 digital detection system. Data shown are one representative image from three independent repeats. Bottom panel: Protein densitometry was carried out using the Fusion FX7 digital detection system and relative fold increase normalised to β-actin was calculated. Data show the mean ± SEM from three independent repeats. * $p < 0.05$ ** $p < 0.01$, *** $p < 0.001$ as determined by an unpaired *t* test.

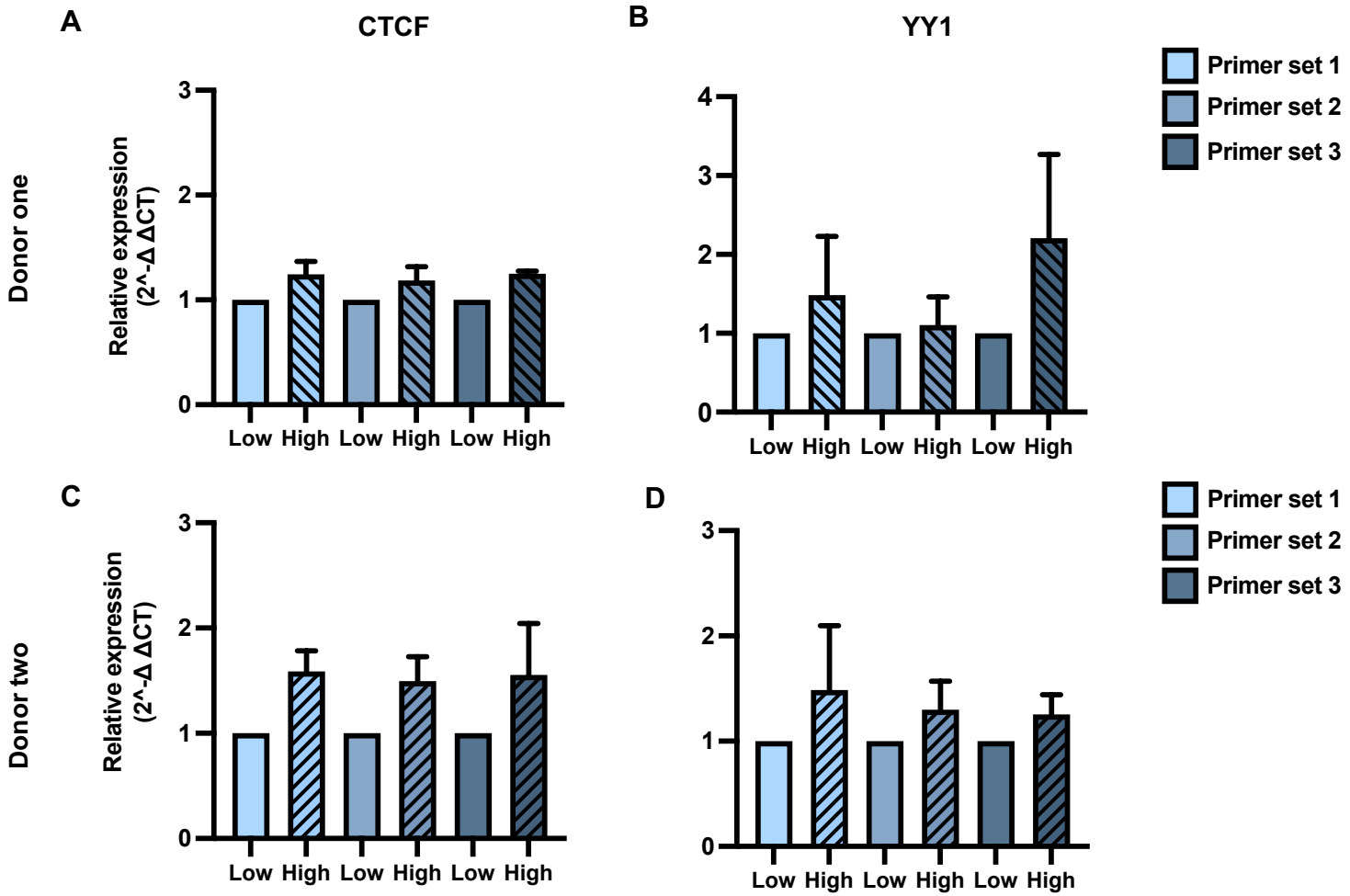


Figure 3.6 CTCF and YY1 transcript expression is not significantly altered in high-passage 18-HTKs. Relative fold change of CTCF (A, C) and YY1 (B, D) transcripts were assessed in low-passage and high-passage 18-HTKs by qRT-PCR. Data were normalised to β-actin (2^-ΔΔCt) and show the mean ± SEM from three independent repeats. *p* values (n.s.) were determined using an unpaired *t* test.

3.4 Enrichment of CTCF and YY1 to the HPV18 genome is significantly altered in high-passage 18-HTKs

In low-passage human foreskin keratinocytes CTCF and YY1 regulate viral oncogene expression by binding directly to the HPV18 genome (Pentland et al., 2018). This mechanism enables tight regulation of E6 and E7 expression during the initial stages of an HPV infection. Although we have shown a significant increase in CTCF and YY1 protein expression, we had not yet assessed whether there is enrichment of these proteins at the HPV18 genome within the tonsil keratinocytes. We therefore aimed to determine whether CTCF and YY1 binding to the HPV18 genome is altered as the cells are serially passaged to uncouple increased protein expression from oncogene regulation. We firstly assessed CTCF binding to the HPV18 genome in low- and high-passage 18-HTKs using ChIP-qPCR. Cells were fixed in 2% formaldehyde, Dounce homogenised and sheared using sonication. DNA was extracted from a small volume of sheared chromatin and resolved on a 1.5% agarose gel to ensure efficient shearing conditions had been met. Optimal shearing conditions for ChIP-qPCR should result in chromatin fragments between 100-600 bp with anything larger than 600 bp thought to disrupt ChIP quality. As shown in **Figure 3.7,** the resulting sheared DNA forms a smear which runs between approximately 200-400 bp indicating efficient shearing.

The sheared chromatin was immunoprecipitated overnight with either 8 µg of CTCF polyclonal antibody or 5 µg of anti-Flag M2 antibody and protein G magnetic beads. Eluted chromatin was then analysed by qPCR with primers that anneal at multiple

sites along the HPV18 genome. Enrichment was determined by percentage of binding over the input chromatin. The position of the primer pairs used for ChIP-qPCR are shown in **Figure 3.8** In both assessed tonsil donors binding of CTCF was seen across the HPV18 genome in low-passage 18-HTKs. There was a robust peak of enrichment of CTCF within in the E2 region of the genome (nucleotides 2926-3118) where the conserved CTCF binding site is situated **(Fig 3.9, Fig 3.10).** Chromatin from high-passage HTKs was also assessed for enrichment of CTCF and interestingly, it was demonstrated that at this same consensus binding site there is a significant depletion of CTCF enrichment in both assessed donors (#donor 1 p <0.05, #donor 2 p <0.01) **(Fig 3.9, Fig 3.10).**

Given the integration status of the high-passage 18-HTKs, we wanted to assess whether the E2 open reading frame was still present, specifically the region spanning the CTCF binding site. To assess this, DNA was extracted from low-passage and high-passage 18-HTKs and adjusted to equal concentrations. The abundance of the E2 CTCF binding site was then quantified by qPCR with primers that span nucleotides 2926-3118 of the HPV18 genome. This demonstrated a reduction (#donor 1 p<0.001, #donor 2 n.s.) in the abundance of the E2-CTCF binding site within the high-passage 18-HTKs **(Fig 3.11).** However, although this was reduced, there was still a detectable level of amplification of DNA by qPCR within the high-passage 18-HTKs, suggesting that some viral genomes retain the CTCF binding site within the E2 open reading frame.

CTCF mediates the repression of E6 and E7 oncogenes through a long-range repressive chromatin looping interaction with YY1. Therefore, it was essential to

assess whether there were changes to YY1 binding within the high-passage 18-HTKs consistent with reduced binding of CTCF. Again, sheared chromatin was immunoprecipitated overnight with either 10 µg of YY1 polyclonal antibody or anti-Flag M2 antibody as previously described. It was previously shown that YY1 enrichment was mostly situated at the 5' region of the viral URR (Pentland et al., 2018). This was also true for low-passage tonsil keratinocytes **(Fig 3.12 Fig 3.13).** Intriguingly, in the high-passage HTKs there was a significant increase (#donor 1 p <0.01, #donor 2 p <0.01) in YY1 enrichment at the viral URR in both tested donors, as shown in **Figure 3.12** and **Figure 3.13**. This finding was somewhat surprising given that the canonical function of YY1 is a transcriptional repressor of E6 and E7 through recruitment of a group of proteins known as the polycomb repressive group (PcG) which mediate the catalysis of H3 lysine 27-tri-methylation and H2A lysine 119 ubiquitination, resulting in heterochromatin formation around the viral URR. Taken together, this data indicates that consistent with the foreskin model, CTCF and YY1 both bind the HPV18 genome in low-passage HTKs. Interestingly, as the tonsil keratinocytes are serially passaged there appears to be differences in the binding patterns of both proteins. CTCF binding is significantly reduced, likely as a result of an integration event which disrupts the presence of the E2 open reading frame. Alternatively, there is a significant increase in the enrichment of YY1 at the viral URR of the high-passage 18-HTKs.

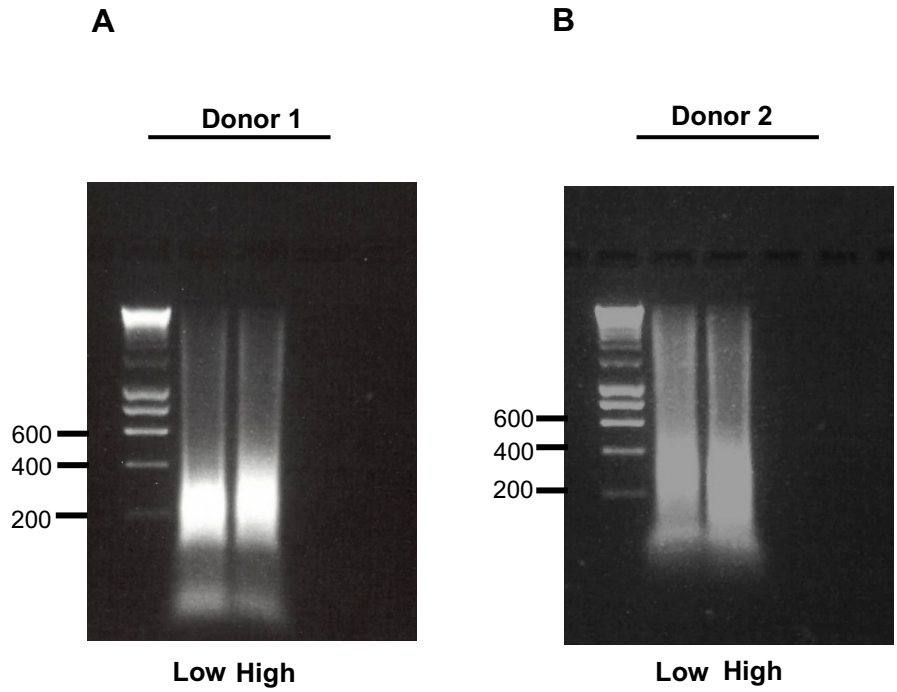


Figure 3.7 Assessment of chromatin shearing efficiency in low- and high-passage 18-HTKs. Cell pellets were fixed in 2% formaldehyde, lysed and Dounce homogenized for the preparation of intact nuclei. Chromatin was fragmented via sonication for 15 minutes at 25% amplitude (30 seconds on, 30 seconds off). DNA was extracted from sheared chromatin using phenol: chloroform: isoamylalcohol (25:24:1). A total of 10 µg of sheared DNA and a 1 kbp DNA ladder and was separated in a 1.5% agarose gel and visualised using a Gene Flash UV lightbox to assess sheared chromatin (Donor 1; A, Donor 2; B). Data shown are representative images taken from three independent repeats.

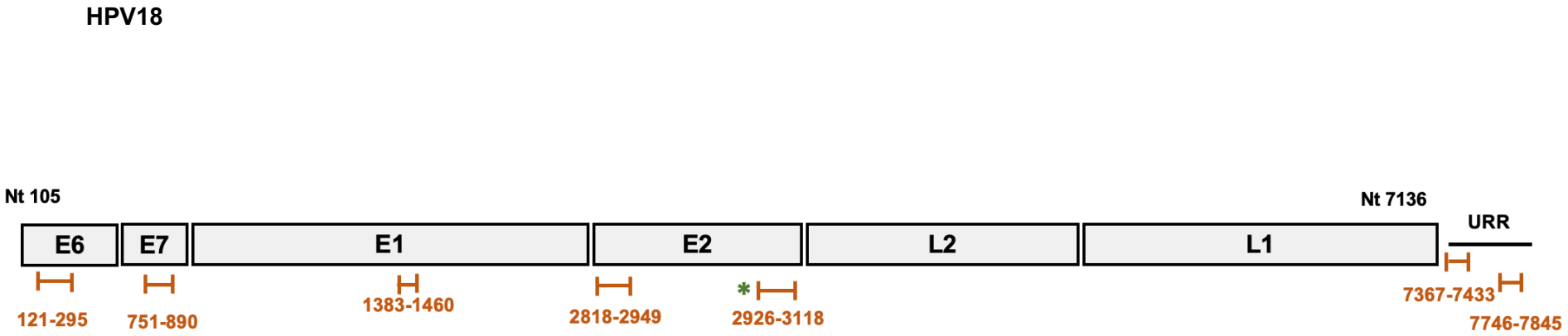


Figure 3.8 Linearised HPV18 genome with forward and reverse primer binding sites. HPV18 linearised genome with marked primer binding sites used for ChIP-qPCR displayed in orange. Primer details are the forward start site and reverse start site. * denotes previously validated binding site for CTCF.

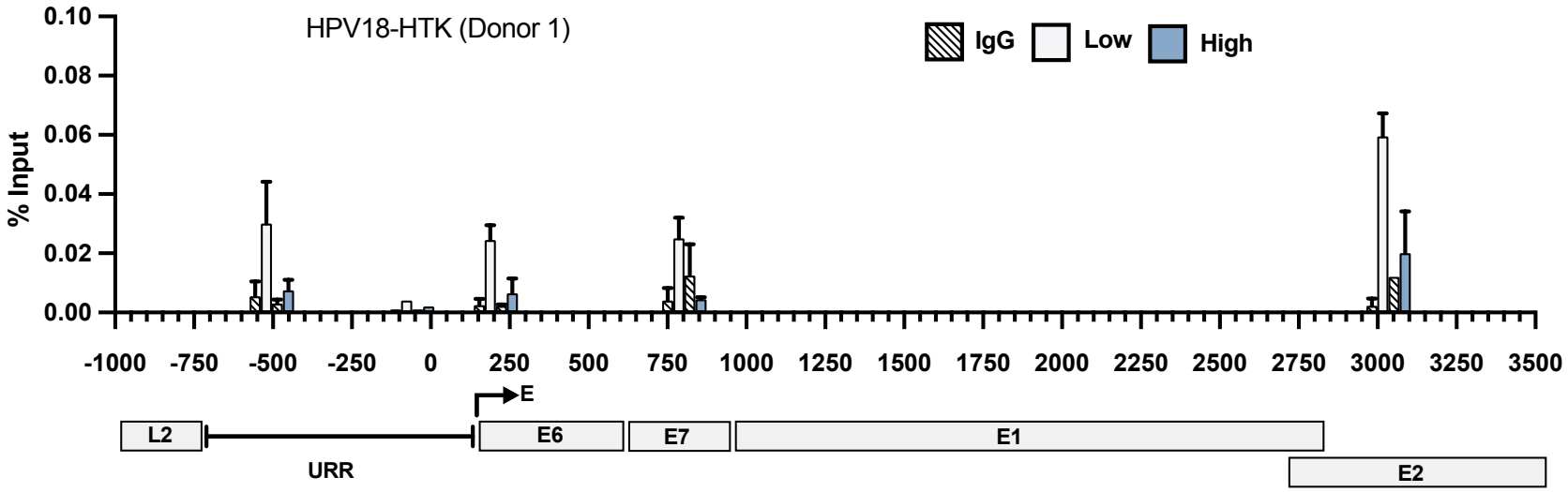


Figure 3.9 Reduction of CTCF binding to the HPV18 genome in high-passage HTKs (Donor 1). CTCF association to the HPV18 genome was assessed by ChIP-qPCR in low-passage and high-passage 18-HTKs. Cell pellets were fixed in 2% formaldehyde, lysed and Dounce homogenized for the preparation of intact nuclei. Chromatin was fragmented via sonication and 25 µg was immunoprecipitated with 8 µg of CTCF polyclonal antibody or 5 µg of anti-Flag M2 antibody. Eluted DNA was amplified using qPCR with primers that anneal at multiple sites along the HPV18 genome. The linearized HPV18 genome is displayed on the x-axis based on the 1/7857bp sequence (AY262282.1). Peaks represent the midpoint of the PCR amplicon. Data show the mean ±SD (% Input) from two independent repeats.

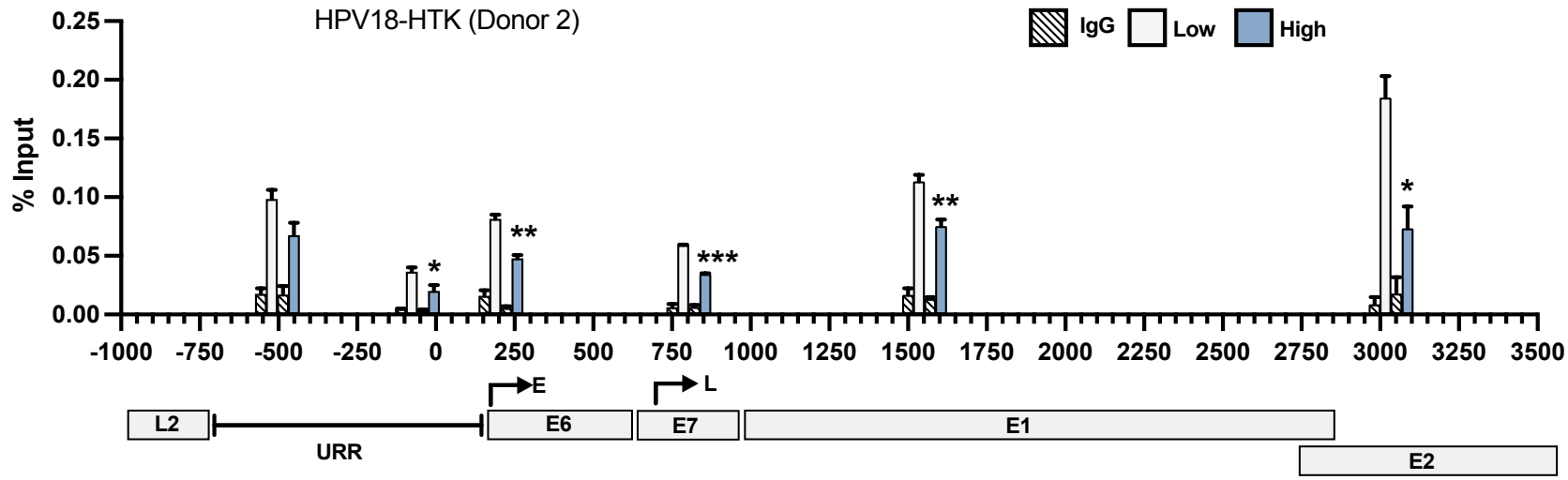


Figure 3.10 Reduction of CTCF binding to the HPV18 genome in high-passage HTKs (Donor 2). CTCF binding to the HPV18 genome was assessed by ChIP-qPCR in low-passage and high-passage 18-HTKs. Cell pellets were fixed in 2% formaldehyde lysed and Dounce homogenized for the preparation of intact nuclei. Chromatin was fragmented via sonication and 25 µg was immunoprecipitated with 8 µg of CTCF polyclonal antibody or 5 µg anti-Flag M2 antibody. Eluted DNA was amplified using qPCR with primers that anneal at multiple sites along the HPV18 genome. The linearized HPV18 genome is displayed on the x-axis based on the 1/7857bp sequence (AY262282.1). Peaks represent the midpoint of the PCR amplicon. Data shown are the mean ±SD (% Input) from three technical replicates and are representative of three independent repeats. *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$ as determined using an unpaired *t* test comparing low-passage and high-passage 18-HTKs.

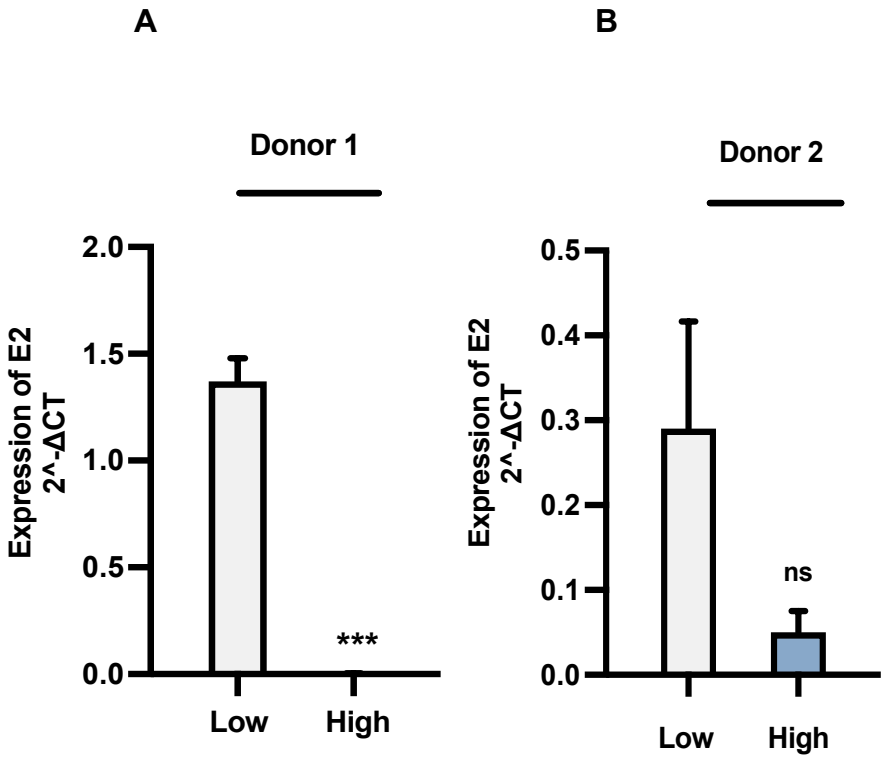


Figure 3.11 Reduced abundance of the CTCF binding site (E2 ORF) in high-passage 18-HTKs. DNA was extracted from low-passage and high-passage 18-HTKs using phenol: chloroform: isoamylalcohol (25:24:1). Samples were adjusted to the same concentration and the expression of the CTCF binding site (E2 ORF) was assessed using qPCR and normalised to β-actin (2^-ΔCt) (A; donor 1, B; donor 2). Data shown are mean ± SEM from three independent repeats. *** $p < 0.001$ as determined by an unpaired *t* test.

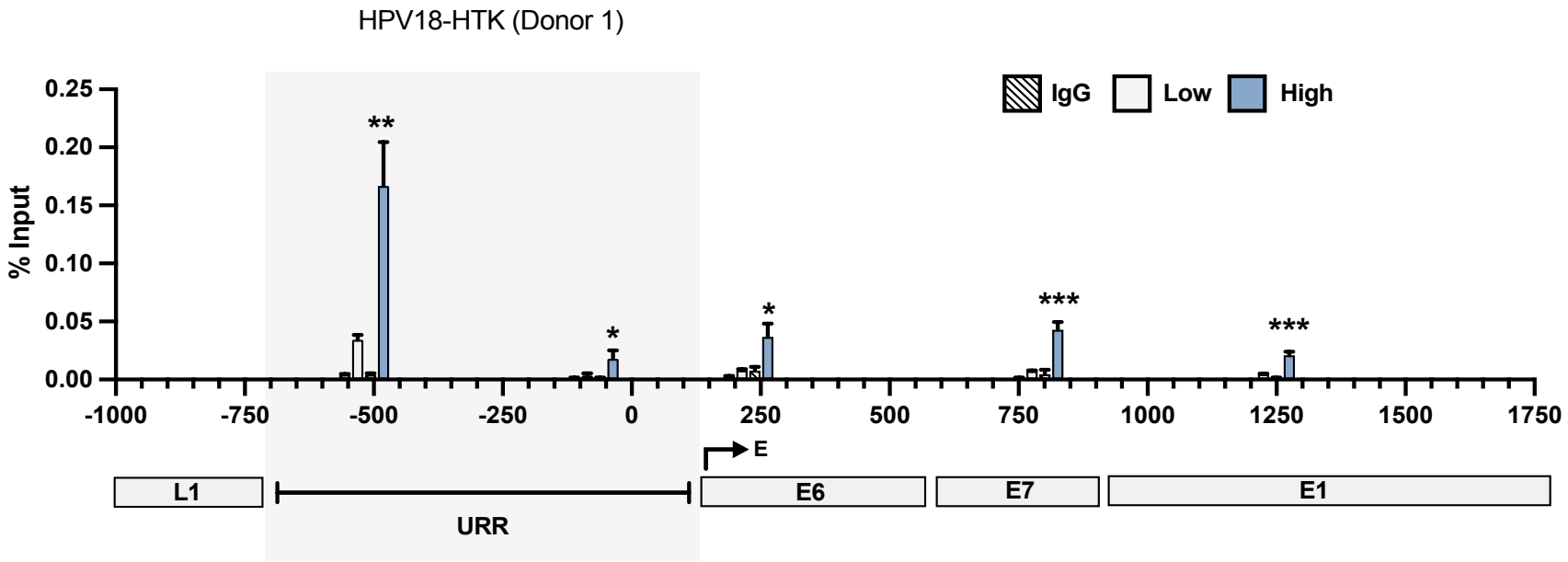


Figure 3.12 Increased YY1 enrichment at the URR of high-passage HPV18 HTKs (Donor 1). YY1 association to the HPV18 genome was assessed by ChIP-qPCR in low-passage and high-passage HPV18 HTKs. Cell pellets were fixed in 2% formaldehyde, lysed and Dounce homogenized for the preparation of intact nuclei. Chromatin was fragmented via sonication and 25 µg was immunoprecipitated with 10 µg of YY1 polyclonal antibody or 10 µg of anti-Flag M2 antibody. Eluted DNA was amplified using qPCR with primers that anneal at multiple sites along the HPV18 genome. The linearized HPV18 genome is displayed on the x-axis based on the 1/7857bp sequence (AY262282.1). Peaks represent the midpoint of the PCR amplicon. Data shown are the mean ±SD (% Input) from three technical replicates and are representative of three independent repeats. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$ as determined by an unpaired *t* test comparing low-passage and high-passage 18-HTKs.

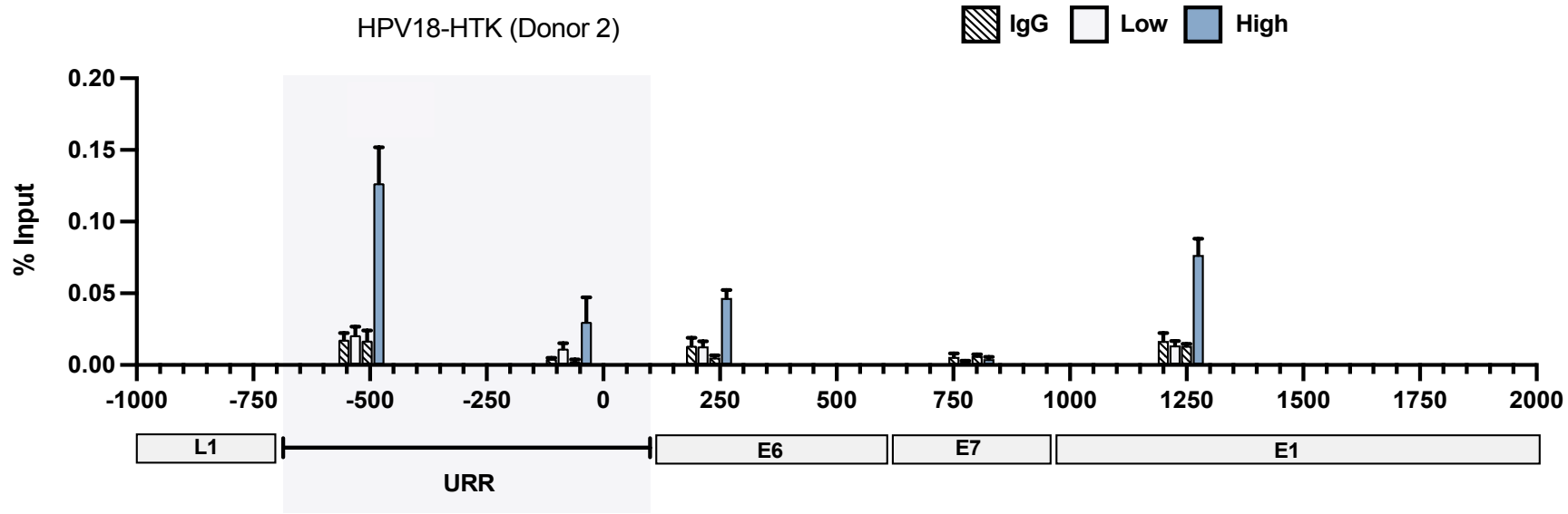


Figure 3.13 Increased YY1 enrichment at the URR in high-passage 18-HTKs (Donor 2). YY1 association to the HPV18 genome was assessed using ChIP-qPCR in low-passage and high-passage 18-HTKs. Cell pellets were fixed in 2% formaldehyde lysed and Dounce homogenized for the preparation of intact nuclei. Chromatin was fragmented via sonication and 25 µg was immunoprecipitated with 10 µg of YY1 polyclonal antibody or 5 µg anti-Flag M2 antibody. Eluted DNA was amplified using qPCR with primers that anneal at multiple sites along the HPV18 genome. The linearized HPV18 genome is displayed on the x-axis based on the 1/7857bp sequence (AY262282.1). Peaks represent the midpoint of the PCR amplicon. Data shown are the mean ±SD (% Input) from three technical replicates and are representative of two independent repeats.

3.5 YY1 binding to the HPV18 genome exerts a non-canonical regulation of the chromatin landscape in high-passage HTKs

Transcription of E6 and E7 is initiated from the HPV18 early promoter, p102, situated upstream from the E6 open reading frame. HPV genomes are known to interact with cellular histones and nucleosome positioning across the URR is an important regulator of HPV transcription. Nucleoprotein formation has been shown over the viral enhancer, at the early promoter, across the origin of replication and within the E7 open reading frame (Stünkel and Bernard, 1999). Changes to chromatin accessibility at these sites are in part mediated by YY1 binding. When functioning as a transcriptional repressor, YY1 recruits PcG proteins to the viral URR – this includes EED and RING1B. As a result of this, there is deposition of H3K27me3 and H2AK119Ub, which results in the formation of heterochromatin. Given that E6 transcription is significantly increased in the late passage 18-HTKs, despite YY1 binding, it is possible that YY1 binding has become uncoupled from transcriptional repression.

Areas of nucleosome depleted DNA are associated with active transcription to allow for transcription factor binding. To identify whether increased oncogene expression was consistent with changes to chromatin topology across the HPV18 genome of low and high-passage HTKs, the chromatin accessibility was assessed using FAIRE. FAIRE distinguishes areas of open and closed chromatin due to the difference in crosslinking efficiency of nucleosome bound and nucleosome depleted DNA (Simon et al., 2012). The ability for nucleosome depleted DNA to be

crosslinked is far less than nucleosome enriched DNA and therefore following phenol extraction nucleosome depleted and enriched DNA are separated in different phases (Simon et al., 2012).

As per ChIP-qPCR, protein-DNA interactions were crosslinked using 2% formaldehyde followed by subsequent shearing. The sample was then split into two conditions: FAIRE or Input. Crosslinking was reversed in the input sample and both samples then had DNA extracted using phenol-chloroform. The chromatin accessibility was then assessed using qPCR and quantified based on a ratio between FAIRE:Input. Across the early and late gene regions of the HPV18 genome there is very little difference between the chromatin accessibility between the low- and high-passage HTKs. These regions show very closed chromatin conformation (**Fig 3.14 Fig 3.15**). However, the upstream regulatory region shows an increase in open chromatin in the low-passage HTKs as compared to the remainder of the genome, which is expected as this region contains the majority of *cis*-regulatory elements that act on the downstream promoters. This is also consistent with the expression of E6/E7 transcripts, albeit at low levels, in the low-passage 18-HTKs. Within this same site, the chromatin accessibility is significantly increased in the high-passage HTKs (#donor 1 $p<0.05$, #donor 2 p n.s.), again correlating with a significant increase in oncogene expression in the high-passage 18-HTKs (**Fig 3.14, Fig 3.15**). This data suggests that YY1 binding to the HPV18 genome of high-passage HTKs is no longer functioning to repress E6/E7 oncogene expression through chromatin remodelling to favour transcriptional repression.

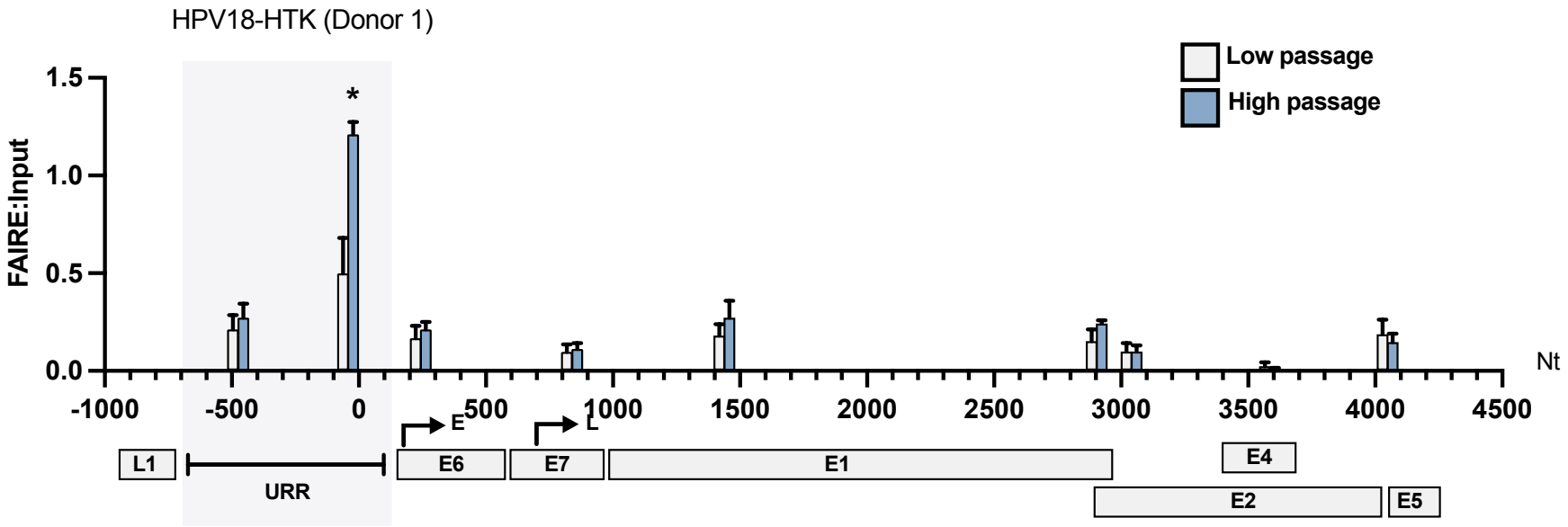


Figure 3.14 Increased chromatin accessibility at the viral URR in high-passage 18-HTKs (Donor 1). Chromatin topology was assessed using FAIRE in low- and high-passage 18-HTKs. Cell pellets were fixed in 2% formaldehyde, lysed and Dounce homogenized for the preparation of intact nuclei. Chromatin was fragmented via sonication and samples were split into FAIRE or Input. Crosslinks were reversed in the Input sample and DNA was extracted from both FAIRE and Input using phenol: chloroform: isoamylalcohol (25:24:1). Isolated DNA was amplified using qPCR with primers that anneal at different points on the HPV18 genome. Peaks represent the midpoint of the PCR amplicon. The linearized HPV18 genome is displayed on the x-axis based on the 1/7857 bp sequence (AY262282.1). Data show the mean ± SEM of FAIRE: Input (2^-ΔCt) from three independent repeats. * $p < 0.05$ as determined by an unpaired *t* test.

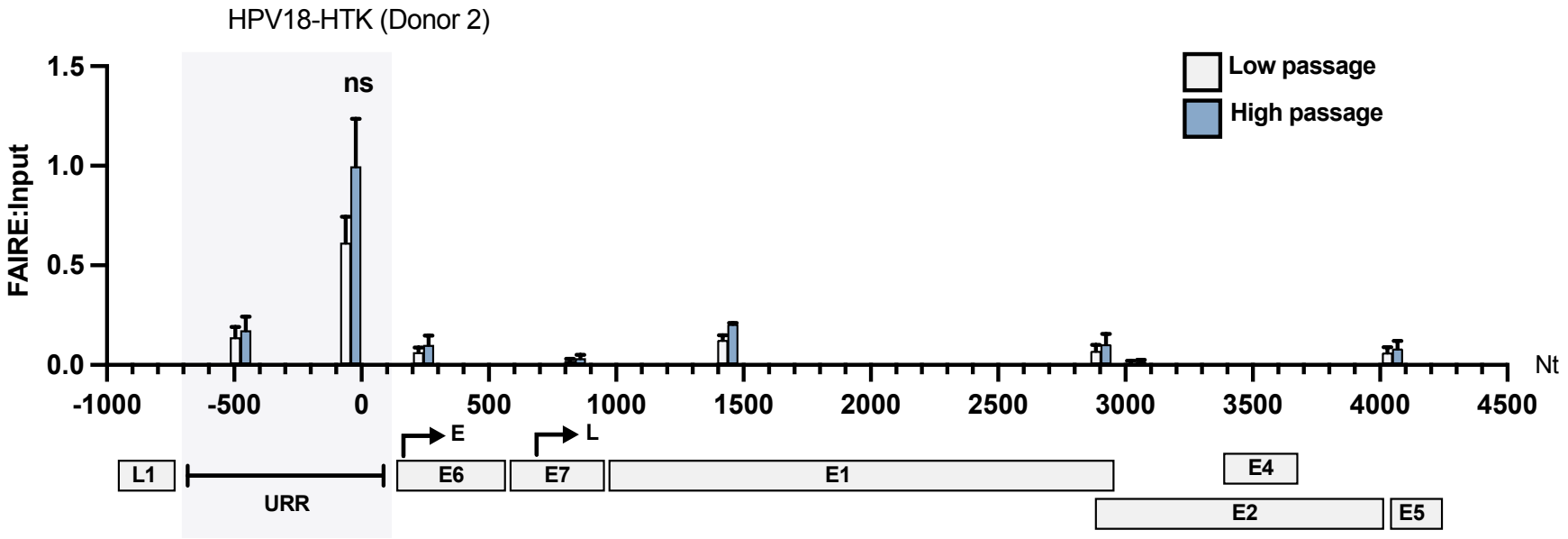


Figure 3.15 Chromatin accessibility is unchanged at the viral URR in high-passage 18-HTKs (Donor 2). Chromatin topology was assessed using FAIRE in low- and high-passage HPV18 HTKs. Cell pellets were fixed in 2% formaldehyde, lysed and Dounce homogenized for the preparation of intact nuclei. Chromatin was then fragmented via sonication and samples were split into FAIRE or Input. Crosslinks were reversed in the Input sample and DNA was extracted from both FAIRE and Input using phenol-chloroform. Isolated DNA was amplified using qPCR with primers that anneal at different sites across the entire HPV18 genome. Peaks represent the midpoint of the PCR amplicon. The linearized HPV18 genome is displayed on the x-axis based on the 1/7857 bp sequence (AY262282.1). Data show the mean ± SEM of FAIRE: Input (2^-ΔCt) from three independent repeats. *p* values (n.s.) were determined using an unpaired *t* test.

3.6 Deregulation of CTCF-YY1-mediated oncogene transcription control in high-passage 18-HTKs

The results so far suggest a loss of CTCF-YY1-mediated oncogene repression in high-passage 18-HTKs. However, it is unclear whether increased protein expression and changes to enrichment are directly involved in creating an open chromatin landscape and driving high expression of E6 and E7. Previously, we have used a doxycycline inducible shRNA system to deplete YY1 in low-passage HFKs. We found that knockdown of YY1 results in a 6-fold increase in the expression of E6 (p<0.01) **(Fig 3.16).**

In order to corroborate these findings in the tonsil, we opted to use an siRNA based approach to deplete both CTCF and YY1 in the low-passage and high-passage 18-HTKs. Low-passage 18-HTKs (#donor 2) were either left untransfected (UT) or transfected with scrambled (Scr), CTCF-specific (CTCF) or YY1-specific (YY1) siRNA at a final concentration of 10 nM for 48 hours. These conditions were based on results determined from optimisation experiments assessing siRNA concentration and length of treatment (data not shown). After 48 hours the cells were harvested, and the knockdown efficiency was assessed using qRT-PCR. As shown in **Figure 3.17** treatment with 10 nM of siCTCF for 48 hours resulted in reduced expression of CTCF transcripts across three tested primer sets – although this did not reach significance. We then assessed whether targeting CTCF for knockdown would alter expression of HPV oncogene transcription. As expected, depletion of CTCF for 48 hours resulted in a significant

increase transcript expression of E6 (p<0.01) **(Fig 3.18B).** Next we assessed the expression of YY1 transcripts following depletion of YY1 with 10 nM of siYY1 for 48 hours. Using three different primer sets we show a significant reduction in YY1 transcript expression in the low-passage 18-HTKs following siRNA treatment **(Fig 3.18A).** Consistent with our previous work, we show that following depletion of YY1, there is a significant increase in the expression of E6 transcription (p <0.05) **(Fig 3.18B).** Taken together, this data confirms that both CTCF and YY1 function to repress the expression of E6/E7 transcription from the HPV18 early promoter within low-passage HPV18 containing tonsil keratinocytes.

Given the changes to both CTCF and YY1 binding and increased open chromatin in the high-passage HTKs, we were interested to see whether depleting either protein would exert the same effect in the high-passage 18-HTKs. We hypothesised that CTCF and YY1 would no longer function to repress transcription from the early promoter. We therefore used the same siRNA based system to assess this in the high-passage keratinocytes. HTKs (#donor 2) were either left untransfected (UT) or transfected with scrambled (Scr), CTCF-specific (CTCF) or YY1-specific (YY1) siRNA at a final concentration of 25 nM for 72 hours. After 72 hours the cells were harvested, and the knockdown efficiency was assessed using qRT-PCR and western blotting. As shown in **Figure 3.19** treatment with 25 nM siCTCF for 72 hours resulted in reduced expression of CTCF transcripts with all three tested primer sets when compared to the scrambled control. Treatment with 25 nM siCTCF also resulted in a 3-fold reduction in CTCF protein expression, demonstrating that the siRNA is targeting CTCF **(Fig 3.19A).** The effect that siCTCF had on E6 transcription was assessed

using qRT-PCR. Interestingly, depletion of CTCF in the high-passage 18-HTKs resulted in no significant difference in the expression of E6 transcripts **(Fig 3.19C).** Following treatment with 25 nM siYY1 for 72 hours, there was a significant reduction in the transcript expression of YY1 in all three tested primer sets **(Fig 3.20B).** Treatment with siYY1 also resulted in a 2-fold reduction in the protein expression of YY1 **(Fig 3.20A).** Interestingly, assessment of E6 transcript expression following siYY1 treatment revealed that depletion of YY1 results in a reduction in the expression of E6, although this did not reach statistical significance **(Fig 3.20C).** Taken together, this data suggest that within the high-passage 18-HTKs, CTCF and YY1 are no longer functioning to repress the transcription of the HPV oncogenes.

3.7 Conclusions

Regulation of high-risk HPV oncogenes during the virus life cycle is a tightly controlled process to enable temporal expression of E6 and E7. Uncontrolled expression of the viral oncogenes has no favourable outcome in terms of the HPV life cycle and consequently results in uncontrolled cellular growth. The activity of the HPV18 early promoter is in part regulated by the chromatin loop formation, stabilised by interaction between CTCF and YY1. The differentiation dependent loss of YY1 binding is associated with increased transcription from the early promoter. Therefore, we hypothesised that during the natural history of an HPV infection, and indeed in HPV-driven disease, this mechanism of transcriptional regulation would become deregulated to drive increased expression of E6 and E7.

Using a physiologically relevant model of HPV18-driven disease progression within the tonsil we have confirmed this hypothesis. The long-term culture of HPV18 containing HTKs results in changes to the expression of CTCF and YY1. We have previously shown that establishment of HPV18 within human foreskin keratinocytes results in the upregulation of CTCF protein. We can now conclude that this is also the case in human tonsil keratinocytes that are established with HPV18 and we have also shown that as the cells are serially passaged this protein expression continues to significantly increase. Interestingly, this mechanism appears to be independent of transcriptional changes.

Using ChIP-qPCR we also showed that binding of these cellular proteins is altered as the cells are passaged. Despite the increased protein expression, CTCF enrichment at the E2 open reading frame is significantly abrogated – this is likely as a result of integration within the E2 open reading frame. However, YY1 enrichment is significantly increased at the viral URR, consistent with increased protein. In contrast to the canonical function of YY1 as a transcriptional repressor, using formaldehyde assisted isolation of regulatory elements, we have shown that increased enrichment of YY1 at the URR is coincident with increased open chromatin. This was also supported by increased E6 and E7 oncoprotein expression.

We have previously shown that depletion of CTCF and YY1 in low-passage HFKs results in a significant increase in HPV oncogene transcription. We have now corroborated these findings in the tonsil where we show that depletion of CTCF and YY1 using an siRNA-based system, results in significantly increased

expression of E6/E7 transcription in low-passage 18-HTKs. In contrast, we have shown that in a the high-passage 18-HTKs depletion of YY1 and CTCF by siRNA results in no significant change in E6/E7 transcription. Overall, we conclude that the function of CTCF and YY1 becomes altered following long-term infection with oncogenic HPV to favour an active transcriptional landscape and an increase in oncogene expression. This appears to represent a shift in function when compared to the life cycle of HPV.

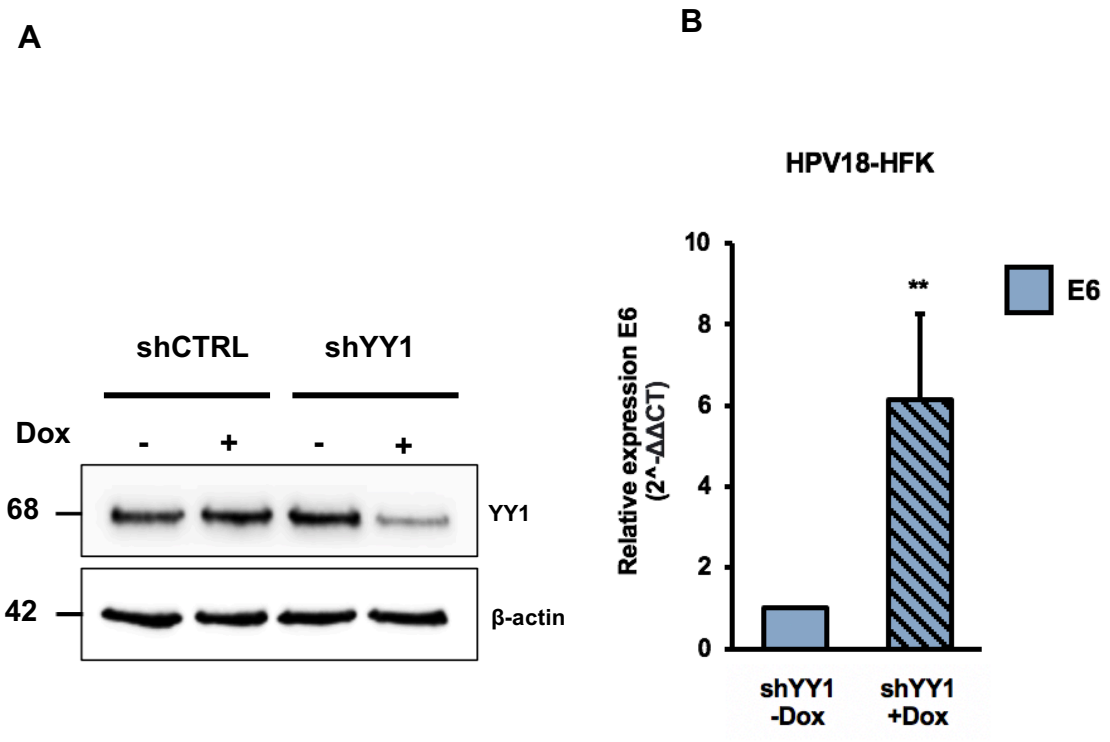


Figure 3.16 Depletion of YY1 in low-passage HFKs results in a significant increase in E6 transcript expression. (A) HFKs were transduced with shCTRL or shYY1 lentiviruses and induced with 1.5 µg of doxycycline for 48 hours and YY1 and β-actin protein expression was analysed by western blotting. (B) E6 unspliced transcripts were assessed via qRT-PCR. Data show the mean ± SEM fold change normalised to β-actin (2^-ΔΔCt) from three independent repeats. ** $p < 0.01$ as determined by an unpaired *t* test. Burley, M (2018) MRes thesis.

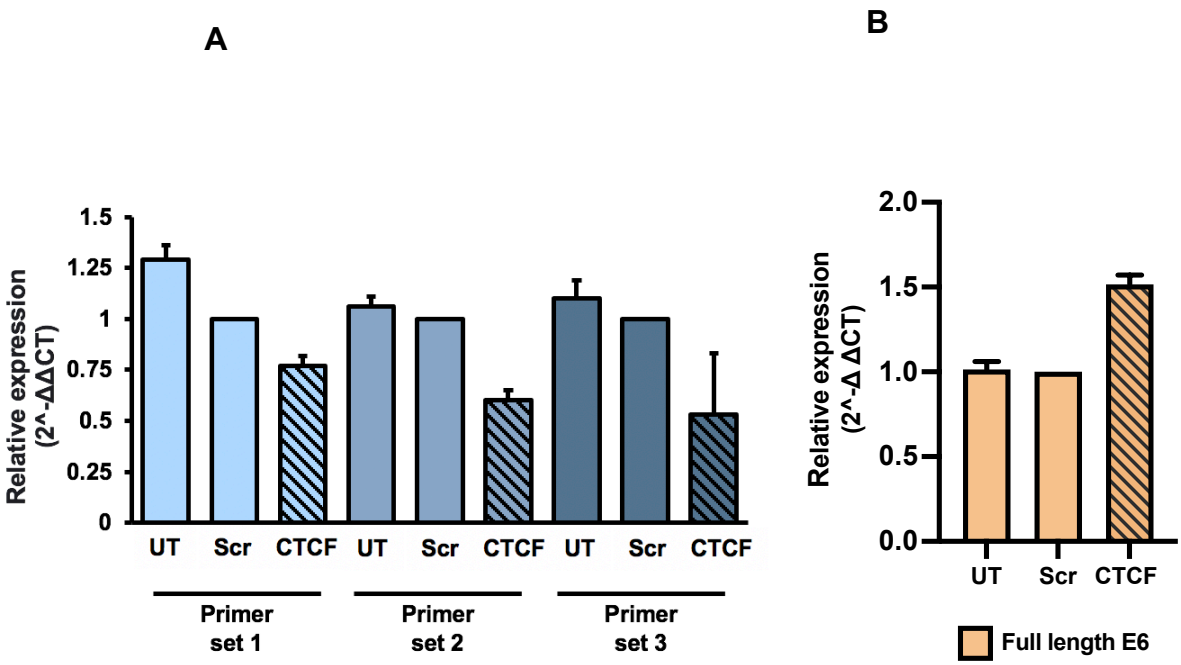


Figure 3.17 Depletion of CTCF in low-passage 18-HTKs results in a significant increase in E6 transcript expression. (A) Low-passage 18-HTKs were left untransfected (UT) or transfected with scrambled (Scr) or CTCF-specific siRNA (CTCF) at 25 nM for 72 hours and CTCF and β-actin protein expression analysed by western blotting. Relative protein expression normalised to β-actin was assessed using densitometry. (B) CTCF and (C) E6 unspliced transcripts were assessed via qRT-PCR. Data show the mean ± SEM fold change normalised to β-actin (2^-ΔΔCt) from one independent repeat.

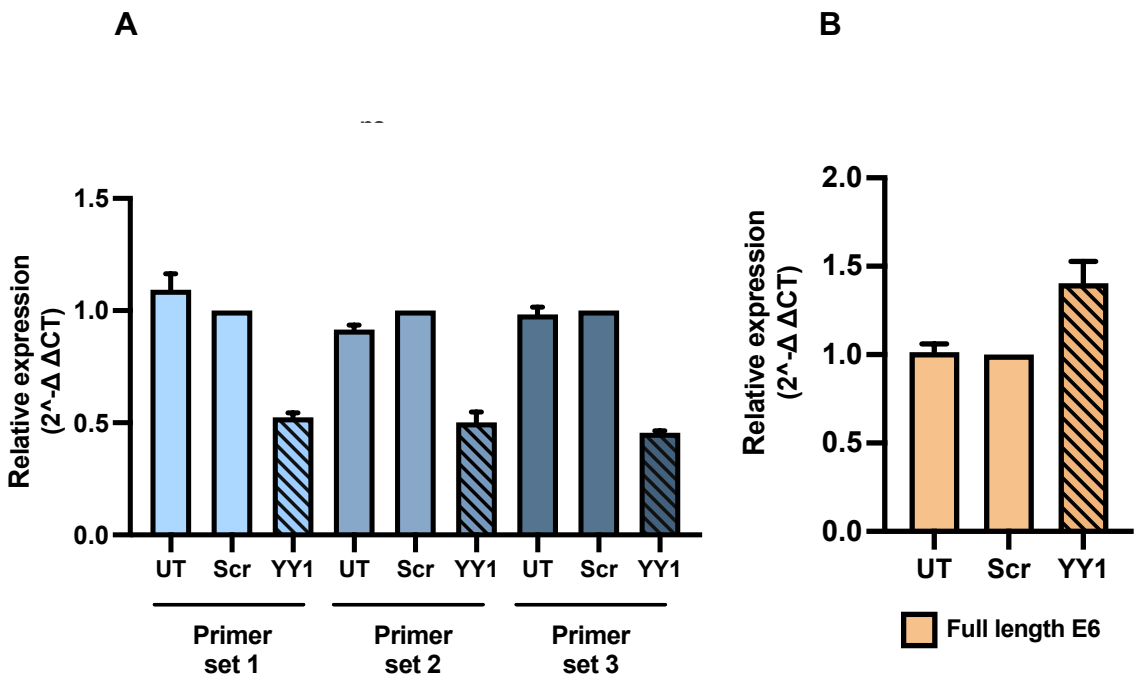


Figure 3.18 Depletion of YY1 in low-passage 18-HTKs results in a significant increase in E6 transcript expression. (A) Low-passage 18-HTKs were left untransfected (UT) or transfected with scrambled (Scr) or YY1-specific siRNA (YY1) at 10 nM for 48 hours and YY1 and β-actin protein expression analysed by western blotting. Relative protein expression normalised to β-actin was assessed using densitometry. (B) YY1 and (C) E6 unspliced transcripts were assessed via qRT-PCR. Data show the mean ± SD fold change normalised to β-actin (2^-ΔΔCt) from one independent repeat.

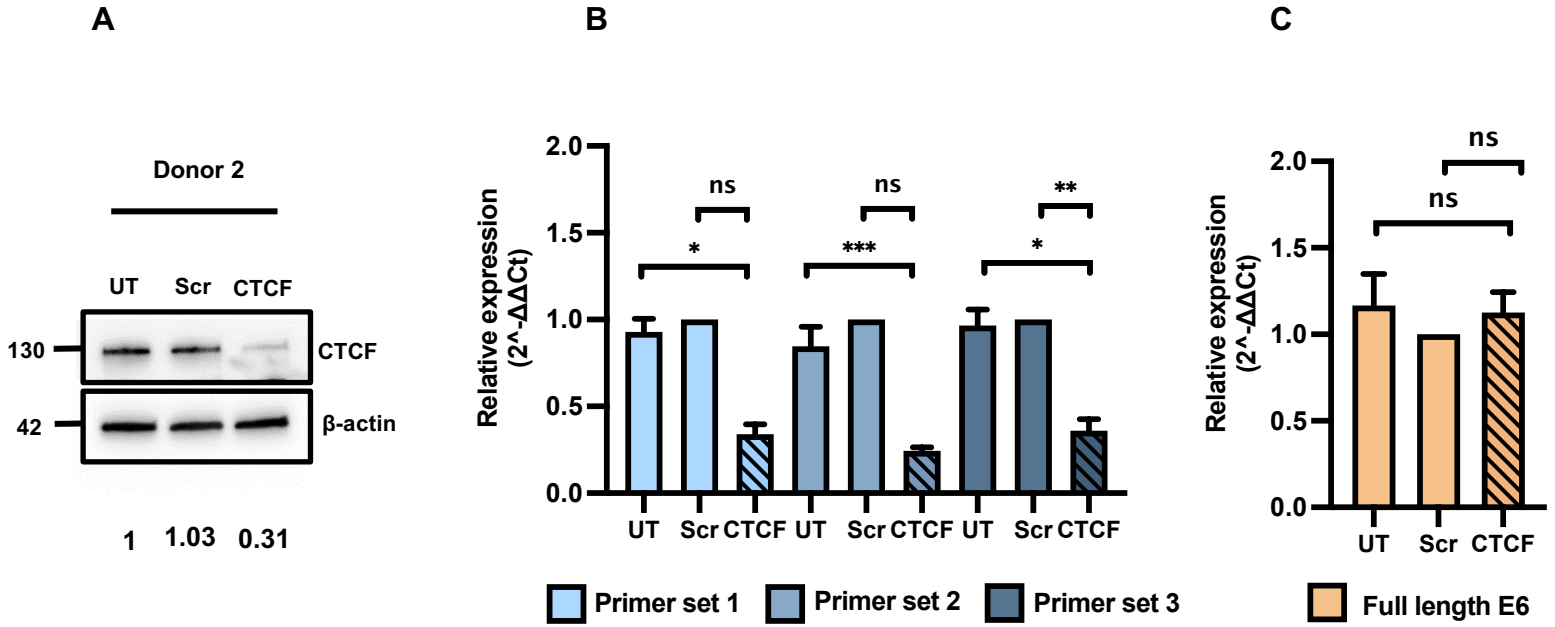


Figure 3.19 Depletion of CTCF in high-passage 18-HTKs does not alter E6 transcript expression. (A) High-passage 18-HTKs were left untransfected (UT) or transfected with scrambled (Scr) or CTCF-specific siRNA (CTCF) at 25 nM for 72 hours and CTCF and β-actin protein expression analysed by western blotting. Relative protein expression normalised to β-actin was assessed using densitometry. (B) CTCF and (C) E6 unspliced transcripts were assessed via qRT-PCR. Data show the mean ± SEM fold change normalised to β-actin (2^-ΔΔCt) from three independent repeats. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$ as determined by a one-way ANOVA.

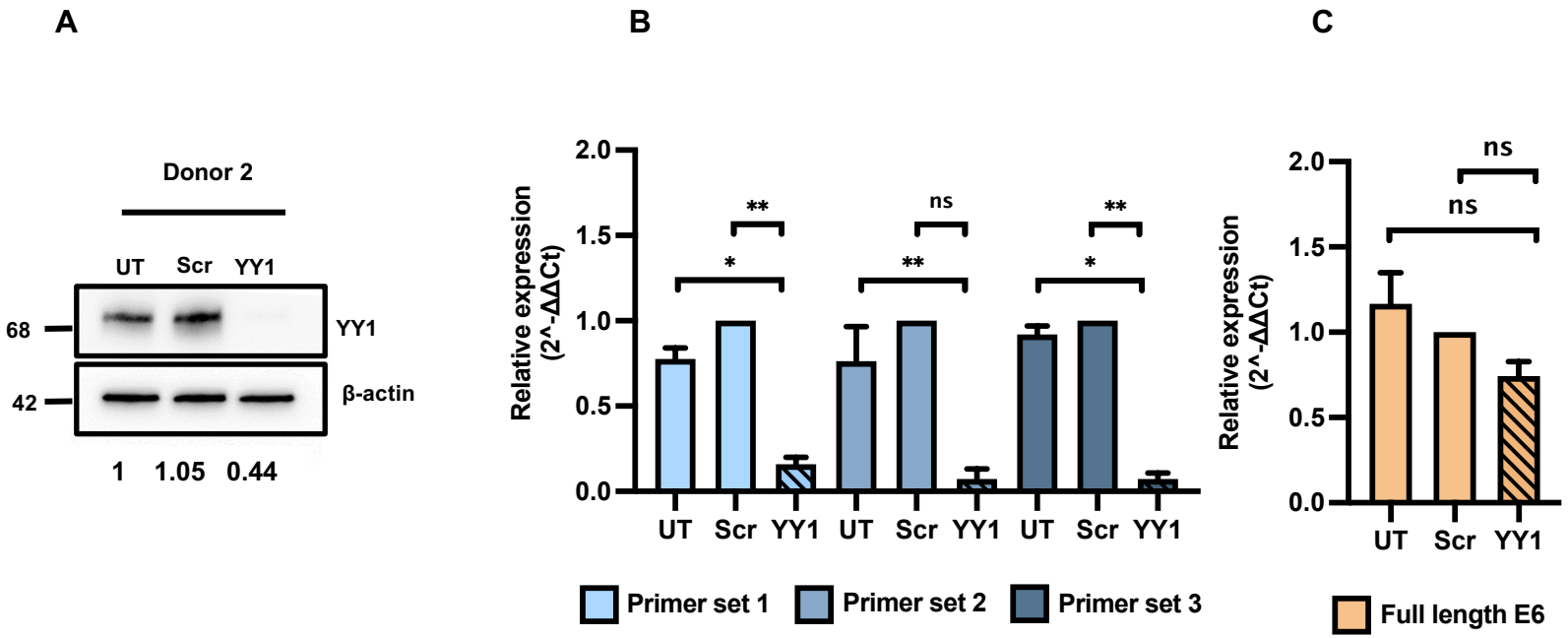


Figure 3.20 Depletion of YY1 in high-passage 18-HTKs results in a non-significant reduction in E6 transcript expression. (A) High-passage 18-HTKs were left untransfected (UT) or transfected with scrambled (Scr) or YY1-specific siRNA (YY1) at 25 nM for 72 hours and YY1 and β-actin protein expression analysed by western blotting. Relative protein expression normalised to β-actin was assessed using densitometry. (B) YY1 and (C) E6 unspliced transcripts were assessed via qRT-PCR. Data show the mean ± SEM fold change normalised to β-actin (2^-ΔΔCt) from three independent repeats. * $p < 0.05$, ** $p < 0.01$ as determined by a one-way ANOVA.

Chapter 4 Assessing the function of CTCF and YY1 in a HPV16-driven disease progression model

4.1 Introduction

The majority of the work so far in exploring the function of CTCF and YY1 in the context of the HPV life cycle has been assessed in HPV18. However, in both cervical and oropharyngeal anatomical sites HPV16 is the prevailing virus type – specifically, in the oropharynx, HPV16 accounts for over ~90% of HPV-driven cancers. There is evidence that both CTCF and YY1 bind to the HPV16 genome and that CTCF functions to regulate both oncogene splicing and expression during the viral life cycle (Paris et al., 2015, Groves et al., 2021). We therefore wanted to assess whether CTCF and YY1 are important in regulating HPV16 oncogene expression within a HPV16 disease progression model. Because HPV16 and HPV18 often behave differently in terms of pathogenesis, we also wanted to explore the differences between virus types. To do this, we used the same donor (#donor 1) as in the previous chapter but transfected with recircularised HPV16 – therefore, we are able to compare virus type specific differences as the cells are serially passaged.

4.2 Validating the high-passage 16-HTKs as a model of HPV-driven disease

Integration of the viral genome into the host cell chromatin is a characteristic associated with many HPV-driven cancers. However, in HPV-driven cancers of the head and neck a large proportion of viral genomes remain extrachromosomal. The

genome status of HPV18 in donor 1 was previously assessed by Southern blotting (See section 3.2). We therefore wanted to compare whether the same HTK donor transfected with HPV16 would exhibit a similar loss of viral episomes as the cells are serially passaged. To assess this, DNA was extracted from serially passaged 16-HTKs (#donor 1) and restriction enzyme digested with either *Bam*HI*,* to linearise the viral genome*,* or *Hind*III*,* a non-cutter of HPV16. Digested DNA was then resolved using a 0.8% agarose gel, transferred onto a GeneScreen nitrocellulose membrane and visualised using a ^{32}P radiolabelled HPV16 probe. As shown in **Figure 4.1** the three low-passage 16-HTK samples that have been digested with *Bam*HI show a linear band of 8 kbp in size which represent episomal DNA. Digestion with *Hind*III results in two bands; one which runs above 10 kbp and one which runs at 5 kbp, representing open circular and supercoiled DNA, respectively **(Fig 4.1).** Interestingly, as the cells are passaged, the genome status remains consistently episomal. Again, the three high-passage 16-HTKs show a linear 8 kbp band on digestion with *Bam*HI and a >10 kbp and 5 kbp band following digestion with *Hind*III **(Fig 4.1).** Importantly, this data highlights that unlike HPV18 containing HTKs, serially passaging HPV16-HTKs does not result in integration of the viral genome into host cell chromatin.

Often the result of viral integration is disruption of the E2 open reading frame resulting in loss of negative regulation of E6/E7 oncogene expression from the early promoter. However, in cancers that maintain HPV episomes, which presumably express full length E2, alternative mechanisms must exist to drive increased E6/E7 oncogene expression. Given that the viral DNA in high-passage 16-HTKs remains episomal, we next wanted to assess whether there was any change to the

expression of full length E6 and spliced E6 (which codes for E7). To assess this, the levels of E6 spliced and unspliced transcripts from the low-passage and high-passage 16-HTKs were quantified using qRT-PCR. As shown in **Figure 4.2** the high-passage 16-HTKs have significantly higher expression of E6 spliced and unspliced transcripts compared to the low-passage 16-HTKs ($p<0.05$, $p<0.001$). Due to lack of availability of reliable antibodies for HPV16 E6 and E7 oncoproteins we were unable to determine whether increased transcription resulted in increased protein expression. Instead, to corroborate our results and explore the significance of increased oncogene expression we assessed the growth rate of the low-passage and high-passage 16-HTKs. Cells were seeded at equal densities onto irradiated J2-3T3 fibroblasts. Cells were subsequently counted at 48-, 72-, 144- and 168-hours post-seeding using Trypan blue to measure viability. Consistent with an increase in oncogene transcript expression, we show that following prolonged culture of HPV16-HTKs there is an increase in growth rate – although this did not reach significance **(Fig 4.3).** We show that as a result of prolonged culture there is a 1.2 fold reduction in doubling time between the low-passage and high-passage 16-HTKs from 44 hours to 35 hours, respectively **(Fig 4.3B).** Taken together, this data highlights that prolonged culture of HPV16-HTKs does not result in integration into the host cell chromatin. However, importantly, these keratinocytes do still display significantly increased expression of both E6 spliced and unspliced transcripts and a non-significant increase in growth rate. This suggests that serially passaging the 16-HTKs might mimic the events that occur during the early stages of HPV16-driven cancer.

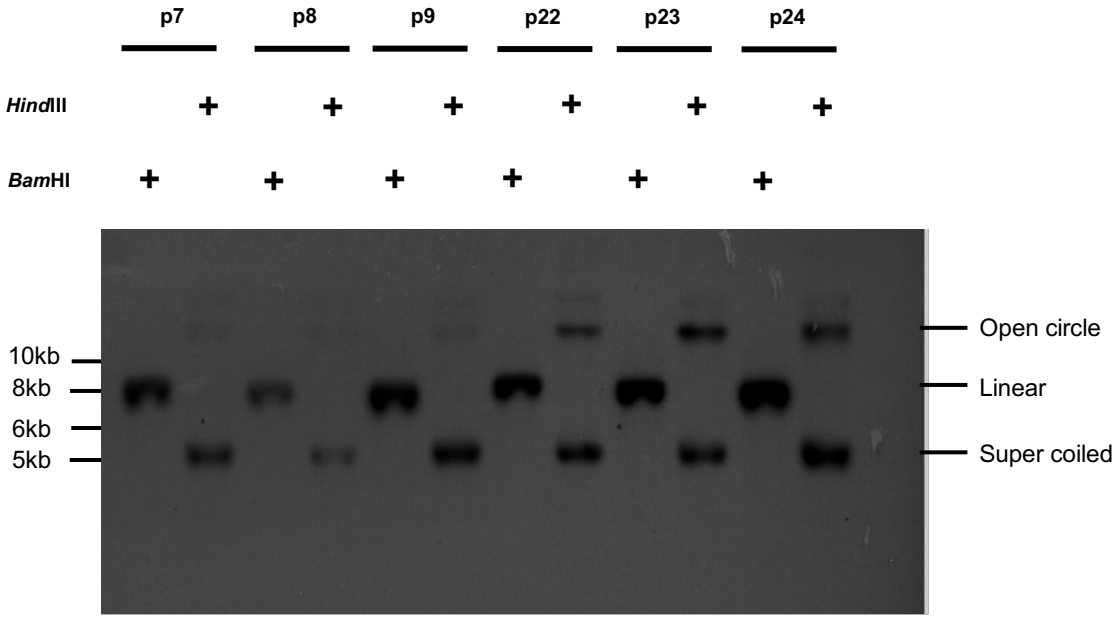


Figure 4.1 HPV16 genomes remain episomal in high-passage HTKs. HPV16 genome status was assessed in low- and high-passage 16-HTKs using Southern blotting. DNA was extracted using phenol: chloroform: isoamylalcohol (25:24:1) from three low-passage and three high-passage 16-HTKs. A total of 5 µg of DNA was restriction enzyme digested with either *Bam*HI or *Hind*III and subsequently separated on a 0.8% agarose gel. The DNA was transferred to a GeneScreen nitrocellulose membrane and UV crosslinked. Viral DNA was labelled using a HPV16 ^{32}P- labelled probe and visualised using autoradiography. Molecular weight (kbp) is indicated on the left. Data show genome status from three independent repeats. P, passage.

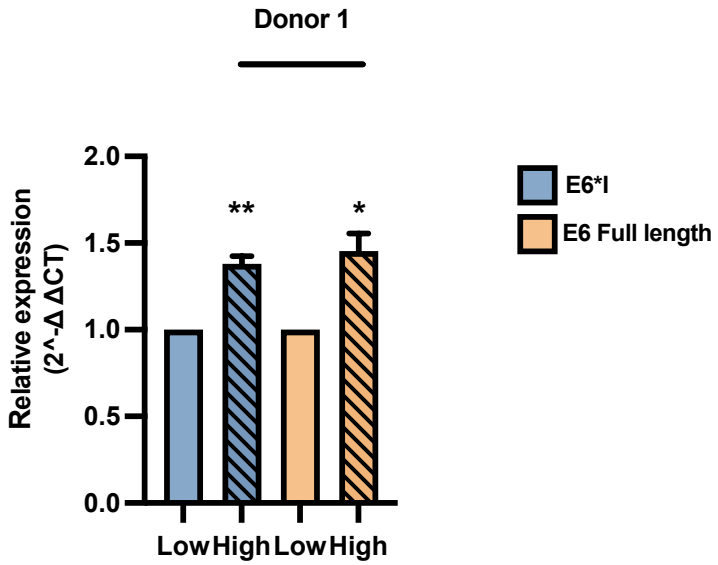


Figure 4.2 Increased E6/E7 viral oncogene transcription in high-passage 16-HTKs. Relative fold change of E6/E7 transcripts in high-passage 16-HTKs normalised to β-actin (2^-ΔΔCt). Data show the mean ± SEM of three independent repeats. * $p < 0.05$, ** $p < 0.01$ as determined by an unpaired *t* test.

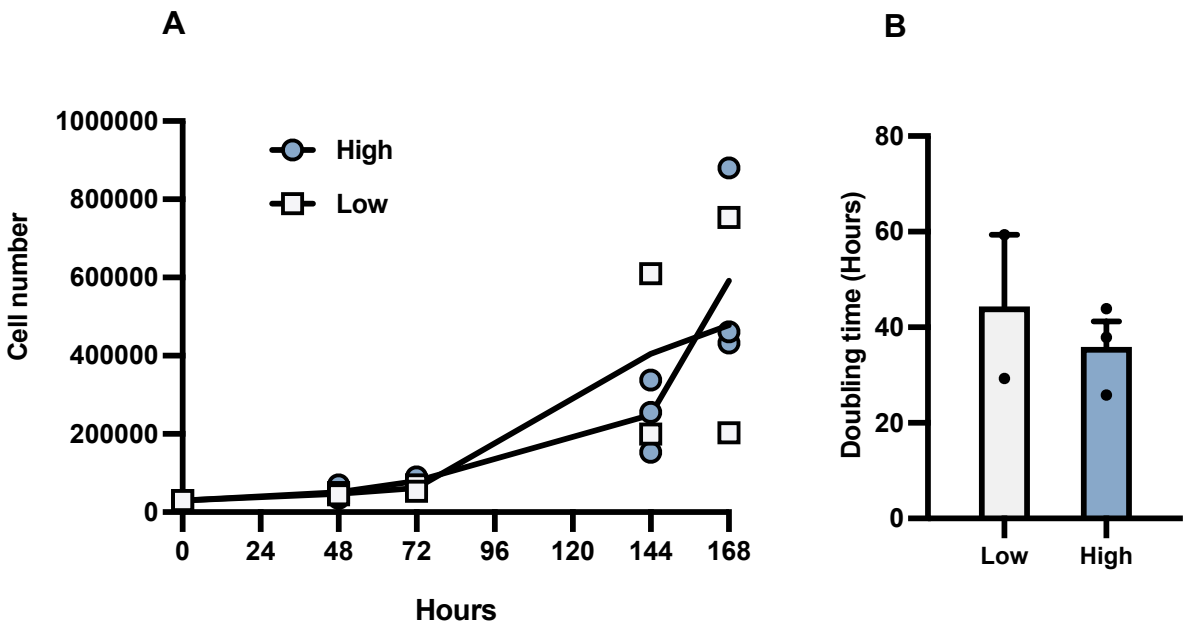


Figure 4.3 Late passage 16-HTKs show a non-significant increase in growth rate. (A) Cell growth assays were carried out using Trypan Blue staining as an indicator of viable cells. Cells were seeded at a density of 3 x 10^4 and viable cells were counted at 48, 72, 144 and 168 hours postseeding. (B) Doubling time of low- and high-passage 16-HTKs. Data shown are the mean ± SEM from ≥ two independent repeats.

4.3 Altered binding of CTCF and YY1 in a HPV16-driven disease progression model

Having assessed changes to E6/E7 transcription and growth rate following the long term culture of the HPV16 HTKs, we were next interested to understand whether there were any changes to the expression and enrichment of CTCF and YY1 that might be consistent with those seen in HPV18 (Chapter 3). We first aimed to assess whether there were any changes to transcript expression between the low- and high-passage 16-HTKs using qRT-PCR. Transcription of CTCF was quantified relative to β-actin using primers that anneal to exon 4, exon 5 and the exon-exon junction between 5 and 6 of CTCF. No change to CTCF transcript expression between the low-passage and high-passage 16-HTKs was detected **(Fig 4.4A).** The expression of YY1 transcripts was also assessed by qRT-PCR with primer sets that anneal to exon 1 and exon 2 of YY1. A non-significant increase in YY1 transcript expression was observed with all three primer sets used however, the magnitude of this increased expression was greater with primer set 1, although this did not reach significance **(Fig 4.4B).**

We next explored whether there were any changes to CTCF and YY1 at the protein level. The expression of CTCF and YY1 was assessed using western blotting in the untransfected HTKs, low-passage and high-passage 16-HTKs. As shown in **Figure 4.5** transfection of HTKs with HPV16 episomes results in increased protein expression of both CTCF and YY1 which is consistent with our previous findings in HPV18 containing HTKs. This confirms that establishment with HPV16 results in an

increase in both CTCF and YY1 protein abundance. Interestingly, as the 16-HTKs are serially passaged, there is a further 2-fold increase in both CTCF and YY1 protein expression when normalised to β-actin **(Fig 4.5).**

In order to understand whether the changes in protein expression might result in changes to binding to the HPV16 genome, we used ChIP-qPCR to quantify enrichment of CTCF in low-passage and high-passage 16-HTKs. Again, we were also interested to explore whether binding of CTCF might differ between HPV16 and HPV18. To assess this we used ChIP-qPCR to look at enrichment of CTCF within the low- and high-passage 16-HTKs. HPV16 containing HTKs were fixed in 2% formaldehyde, Dounce homogenised and sheared via sonication. Shearing efficiency was assessed as previously described and the resolved DNA gels can be seen in **Figure 4.6.** Efficient shearing should result in fragmented chromatin of approximately 200 bp, however, the resulting fragment sizes following shearing appear to be larger than this **(Fig 4.6).** This indicates that whilst there are some smaller sheared fragments there are also larger non-sheared fragments of chromatin. This is likely as a result of inefficient nuclei release, perhaps due to the smaller cell size of the HPV16 containing HTKs. However, because the shearing efficiency was consistent across all three independent repeats, and because there were some fragments of sheared chromatin, we subsequently performed ChIP-qPCR.

Sheared chromatin was incubated overnight at 4ºC with 8 µg of polyclonal CTCF antibody, 5 µg of anti-Flag M2 antibody and protein G magnetic beads. Enrichment (as % Input) was assessed using qPCR with primers that anneal at multiple sites

across the HPV16 genome. The forward and reverse primer sites used for ChIP-qPCR are shown in **Figure 4.7.** As shown in **Figure 4.8** the low-passage 16-HTKs show enrichment of CTCF across the genome. This is particularly prevalent within the E2 open reading frame at nucleotides 2893-2950. However, there also appears to be peaks representing CTCF binding within the L2 open reading frame, namely at nucleotides 5123-5309, where there is a known CTCF binding site **(Fig 4.8).** Consistent with the results shown in HPV18 (#donor 1) serially passaging the HPV16 containing HTKs results in a significant reduction in CTCF enrichment across the HPV16 genome. This is particularly prominent within the E2 open reading frame (p<0.0001) **(Fig 4.8).** Whilst we attempted to also assess enrichment of YY1 binding, we were unable to achieve reproducible results. Often, the anti-Flag M2 control showed significantly higher enrichment that YY1. It is possible that this is as a result of inefficient shearing**.** We were therefore unable to confirm whether enrichment of YY1 binding is also increased in a HPV16-HTK progression model.

Taken together, this data suggests that in a HPV16 progression model the expression of both CTCF and YY1 is altered. As with HPV18, this appears to be a post-translational upregulation as the transcription of both genes was relatively stable. Additionally, even though the protein levels are significantly increased, there is a significant reduction in the enrichment of CTCF to the HPV16 genome as the cells are serially passaged.

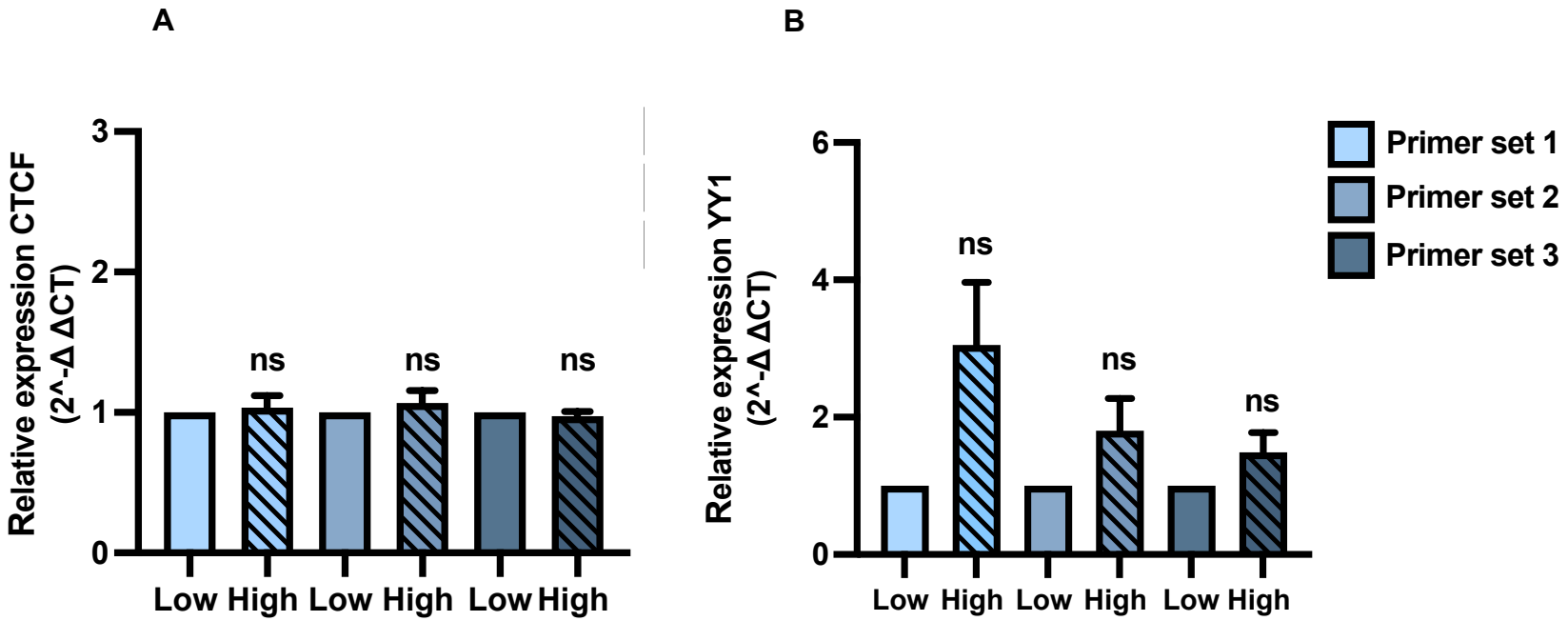


Figure 4.4 CTCF and YY1 transcription is not significantly altered in high-passage 16-HTKs. Relative fold change of (A) CTCF and (B) YY1 transcripts in high-passage HPV16 containing HTKs normalised to β-actin (2^-ΔΔCt). Data show the mean ± SEM of three independent repeats. *p* values were determined using an unpaired *t* test.

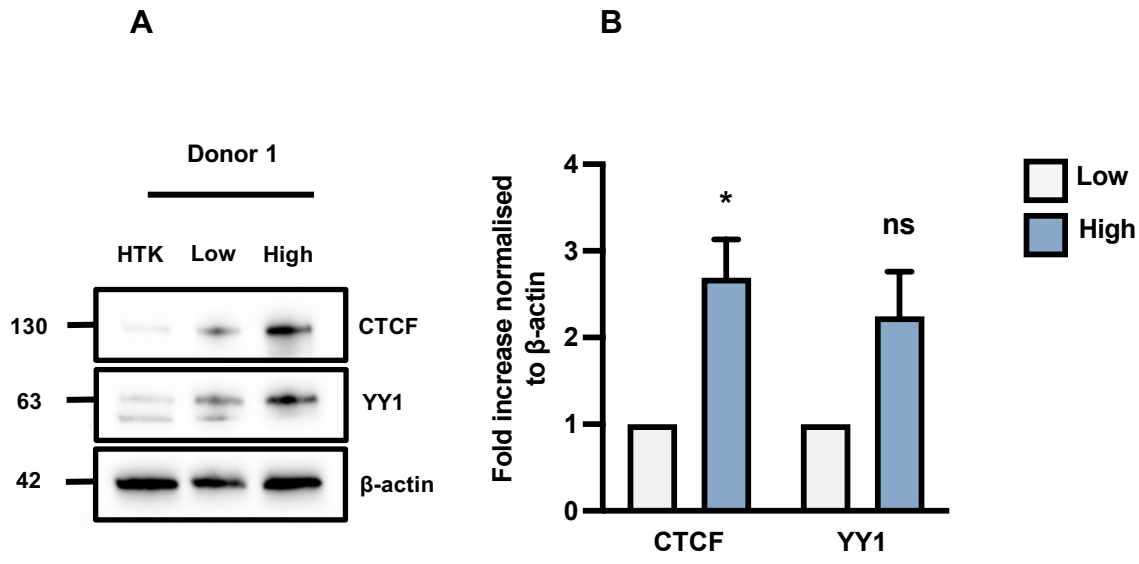


Figure 4.5 Increased CTCF protein expression in high-passage 16-HTKs. (A) CTCF and YY1 protein expression was assessed in untransfected, low-passage and high-passage 16-HTKs and normalised to β-actin. A total of 10 µg of protein was separated using a 4-12% polyacrylamide gel. Protein was transferred onto a PDVF membrane and imaged using a Fusion FX7 digital detection system. Data shown are one representative image from three independent repeats. (B) Protein densitometry was carried out using the Fusion FX7 digital detection system and relative fold increase normalised to β-actin was calculated. Data show the mean ± SEM from three independent repeats. * $p < 0.05$ as determined by an unpaired *t* test.

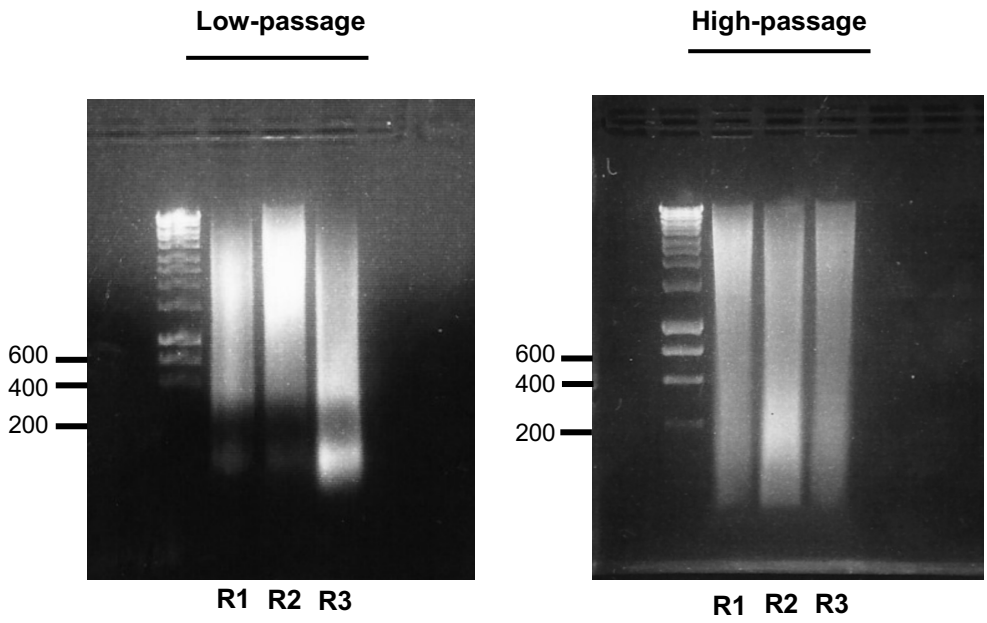


Figure 4.6 Assessment of shearing efficiency in low- and high-passage 16-HTKs. Cell pellets were fixed in 2% formaldehyde, lysed and Dounce homogenized for the preparation of intact nuclei. Chromatin was fragmented via sonication for 15 minutes at 25% amplitude (30 seconds on, 30 seconds off). DNA was extracted from sheared chromatin using phenol: chloroform: isoamylalcohol (25:24:1). A total of 10 µg of sheared DNA and a 1 kbp DNA ladder and was separated in a 1.5% agarose gel and visualised using a Gene Flash UV lightbox to assess sheared chromatin. Data show shearing efficiency from three independent repeats. R1, repeat 1. R2, repeat 2. R3, repeat 3.

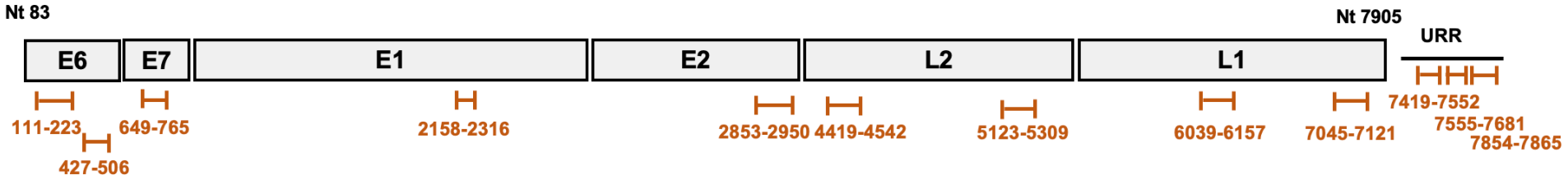


Figure 4.7 Linearised HPV16 genome with forward and reverse primer binding sites. HPV16 linearised genome with marked primer binding sites used for ChIP-qPCR displayed in orange. Primer details are the forward start site and reverse start site.

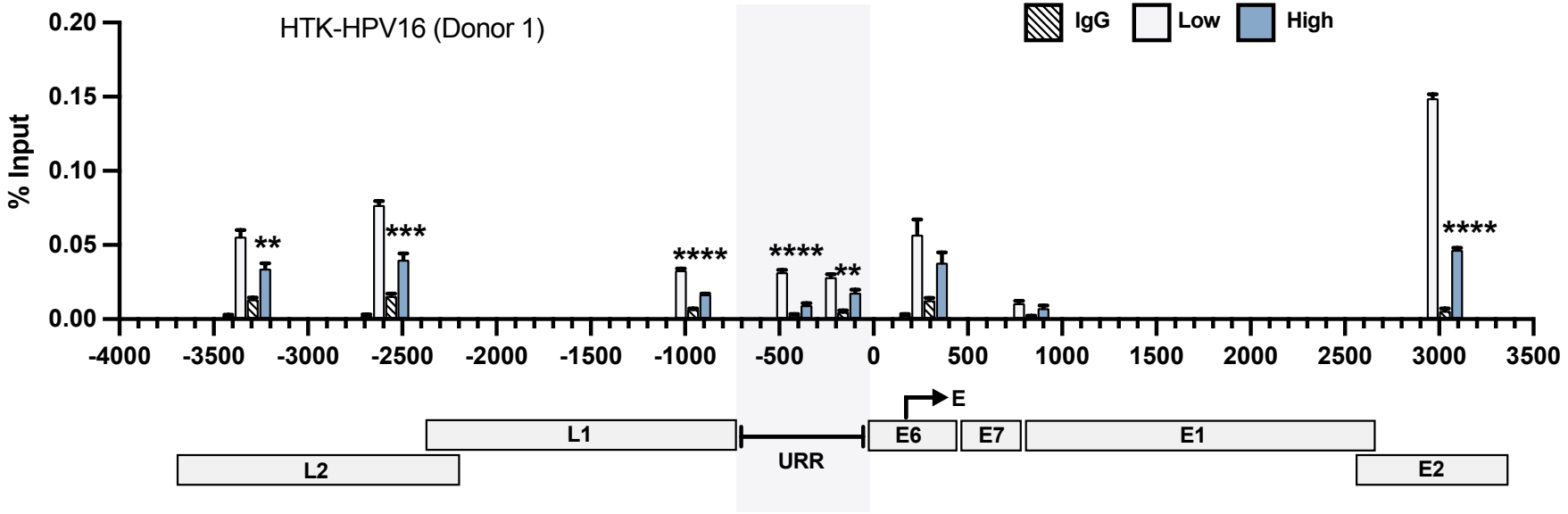


Figure 4.8 Reduced CTCF enrichment in the E2 ORF of high-passage 16-HTKs. CTCF association to the HPV16 genome was assessed by ChIP-qPCR in low- and high-passage 16-HTKs. Cell pellets were fixed in 2% formaldehyde, lysed and Dounce homogenized for the preparation of intact nuclei. Chromatin was fragmented via sonication and 25 µg was immunoprecipitated with 8 µg of CTCF polyclonal antibody or 10 µg of anti-Flag M2 antibody. Eluted DNA was amplified using qPCR with primers that anneal at multiple sites along the HPV18 genome. The linearized HPV18 genome is displayed on the x-axis based on the 1/7905 bp sequence (NC_001526.2). Peaks represent the midpoint of the PCR amplicon. Data shown are the mean ± SD (% Input) from three technical replicates and are representative of three independent repeats. ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$ as determined by an unpaired *t* test comparing low-passage and high-passage 16-HTKs.

4.4 Absence of CTCF-YY1-mediated regulation of the HPV16 early promoter in high-passage 16-HTKs

YY1 functions to regulate transcription from the HPV16 early promoter, p97, and this is primarily in the form of transcription repression (O'Connor et al., 1996, Lace et al., 2009b). It is unclear as to whether YY1 mediated repression of transcription is through an epigenetically repressive chromatin loop mediated by CTCF and through recruitment of PcG proteins. Given that both proteins are known to bind to the HPV16 genome and regulate oncogene expression, it can be postulated that this is the potential mechanism. In order to understand how changes to protein expression and enrichment of CTCF and YY1 might function to alter transcription we firstly assessed whether there were any changes to chromatin topology between the low-passage and high-passage HPV16 containing HTKs. Low-passage and high-passage 16-HTKs were fixed in 2% formaldehyde chromatin topology was assessed by FAIRE as previously described (see section 3.5). The chromatin accessibility was subsequently assessed using qPCR with primers that amplify multiple sites across the length of the HPV16 genome. As shown in **Figure 4.8**, there is a distinct peak at the viral enhancer and a less prominent peak at the 5'URR which represents open chromatin within the low-passage 16-HTKs. The remainder of the genome appears to have relatively closed chromatin **(Fig 4.9).** Surprisingly, in the high-passage 16-HTKs we show that the chromatin accessibility at the enhancer region is reduced – although this did not reach significance **(Fig 4.9).** The remainder of the genome shows very small peaks which represent closed chromatin conformation which is expected of non-regulatory regions of DNA **(Fig 4.9).**

Interestingly, the low-passage 16-HTKs appear to have a far more open chromatin configuration when compared to the low-passage 18-HTKs, perhaps suggesting that HPV16 containing HTKs are generally more transcriptionally active that HPV18-HTKs (see section 3.5). This data suggests that prolonged culture of HPV16 containing HTKs does not result in any significant changes to nucleosome occupancy at the viral URR, again, in contrast to the chromatin landscape detailed in the high-passage HPV18-HTKs. However, given that we were unable to show increased YY1 binding in our high-passage 16-HTKs this supports the hypothesis that YY1 functions to alter chromatin topology to favour active transcription. Because we have shown an increase in E6 spliced and unspliced transcription in the high-passage 16-HTKs, it is also possible that the chromatin landscape as shown by FAIRE may not represent the events happening at the single cell level. We are looking at the bulk population and it is possible that we have one episome which is open and highly transcribed giving rise to this increase in E6/E7 expression and yet we might not see a difference on the whole population by FAIRE.

Prompted by this, we wanted to establish whether CTCF or YY1 function to repress HPV16 oncogene transcription within the high-passage keratinocytes. To assess this, the high-passage 16-HTKs were either left untransfected (UT) or treated with scrambled (Scr), CTCF-specific (CTCF) or YY1-specific (YY1) siRNA at 10 nM for 48 hours. To assess the efficiency of the knockdown the expression of CTCF protein was assessed using western blotting. As seen in **Figure 4.10A** treatment with siCTCF at 10 nM for 48 hours results in a 3-fold reduction in CTCF protein expression when compared to siScr **(Fig 4.10A).** Next we assessed whether CTCF depletion alters the transcription of the HPV16 oncogenes within the high-passage

16-HTKs using qRT-PCR. As expected, depletion of CTCF resulted in no change to the expression of either full length or spliced E6 transcripts **(Fig 4.10A).** Following treatment with 10 nM siYY1 for 48 hours western blotting confirmed a 3-fold reduction in the expression of YY1 protein, confirming that the knockdown was effective **(Fig 4.10B).** Assessment of spliced and unspliced E6 transcripts by qRT-PCR demonstrated that the knockdown of YY1 results in a very slight (non-significant) reduction in oncogene transcription **(Fig 4.10B).** Taken together, this data demonstrates that CTCF and YY1 are likely no longer acting to repress HPV16 oncogene transcription in the high-passage HPV16 containing keratinocytes.

4.5 Conclusions

Infection with HPV16 is linked to the development of ~90% of HPV associated cancers. As such, it was important that we assessed whether CTCF and YY1 function to regulate HPV16 transcription from the early promoter in a disease progression model. In order to compare virus-specific differences we opted to use the same human tonsil keratinocyte donor as in Chapter 3 (#Donor 1) however, transfected with recircularised HPV16. This allowed us to compare and contrast the differences between HPV16 and HPV18 transcription regulation in a disease-progression model and to understand whether CTCF and YY1 are recruited to the viral genome in order to mediate repression of the viral oncogenes. Using this model, we found that prolonged culture of HPV16 containing HTKs does not result in integration of viral genomes into host cell chromatin. This was in contrast to HPV18 (#Donor 1). Importantly, this reflects a scenario that mimics the events seen in HPV-driven cancer of the oropharynx where the majority of viral genomes exist

as an episome. Despite the fact that we saw no integration in these cells, we did show a significant increase in E6 unspliced and E6 spliced transcripts which resulted in a non-significant increase in the growth rate of the high-passage HPV16 HTKs. This data supports the use of our HPV16 progression model as a means to recapitulate the events which occur during the natural history of a HPV16 infection.

Prolonged culture of the HPV16-HTKs results in significantly increased expression of both CTCF and YY1 protein. This was independent of any changes to transcription and again corroborates the findings shown in Chapter 3. Using ChIP-qPCR we show that CTCF binding to the HPV16 genome was significantly reduced as the cells were serially passaged. This was most apparent at nucleotides 2893-2950 which is where the major CTCF-binding site is located. However, we also found binding of CTCF to the late region of the viral genome which again was reduced following long-term culture. Unfortunately, we were unable to confirm whether there were any changes to binding of YY1 in our HPV16 progression model as our results were not reproducible. Using FAIRE we assessed how increased protein expression of CTCF and YY1 and reduced CTCF binding would function to alter chromatin topology. We showed that there were no significant changes to chromatin accessibility at the viral enhancer as the HPV16-HTKs were serially passaged. When we depleted both CTCF and YY1 using siRNA we found that neither proteins were functioned to repress transcription of E6 spliced or unspliced transcripts, suggesting that following prolonged culture of the HPV containing HTKs there is a loss of CTCF and YY1 mediated transcription regulation.

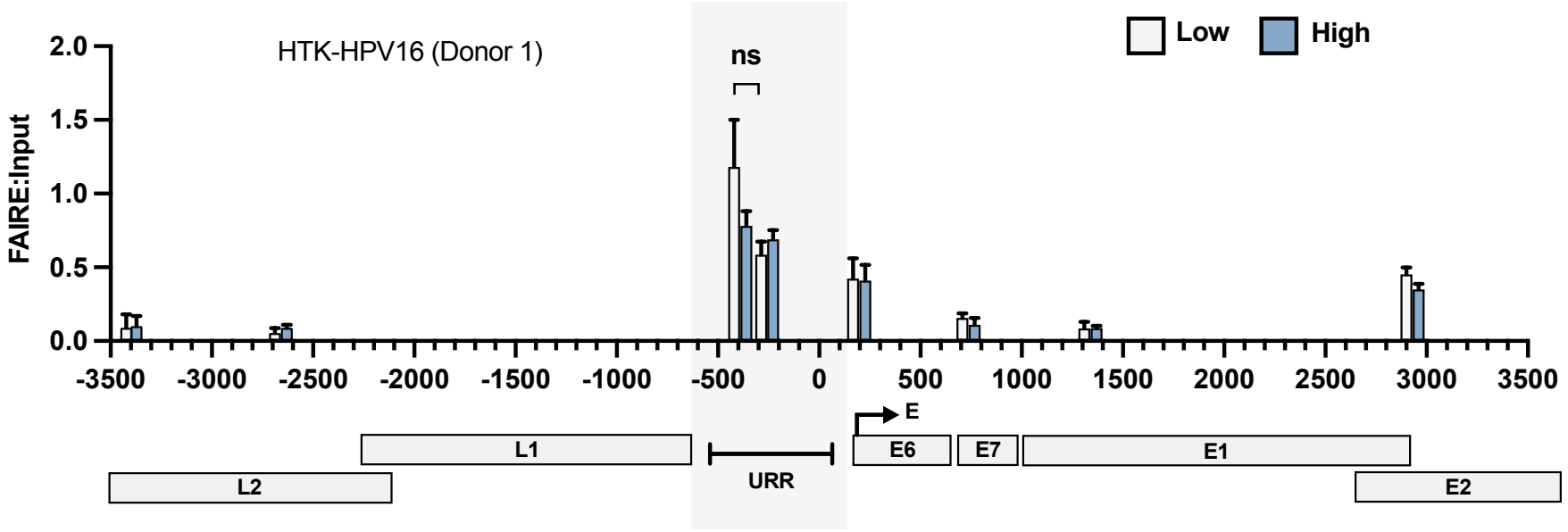


Figure 4.9 Chromatin accessibility is unchanged at the viral URR in high-passage 16-HTKs. Chromatin topology was assessed using FAIRE in low- and high-passage 16-HTKs. Cell pellets were fixed in 2% formaldehyde, lysed and Dounce homogenized for the preparation of intact nuclei. Chromatin was fragmented via sonication and samples were split into FAIRE or Input. Crosslinks were reversed in the Input sample and DNA was extracted from both FAIRE and Input using phenol: chloroform: isoamylalcohol (25:24:1). Isolated DNA was amplified using qPCR with primers that anneal at different points on the HPV16 genome. Peaks represent the midpoint of the PCR amplicon. The linearized HPV16 genome is displayed on the x-axis based on the 1/7905 bp sequence (NC_001526.2). Data show the mean ± SEM of FAIRE: Input (2^- ΔCt) from three independent repeats. *p* values were determined by an unpaired *t* test.

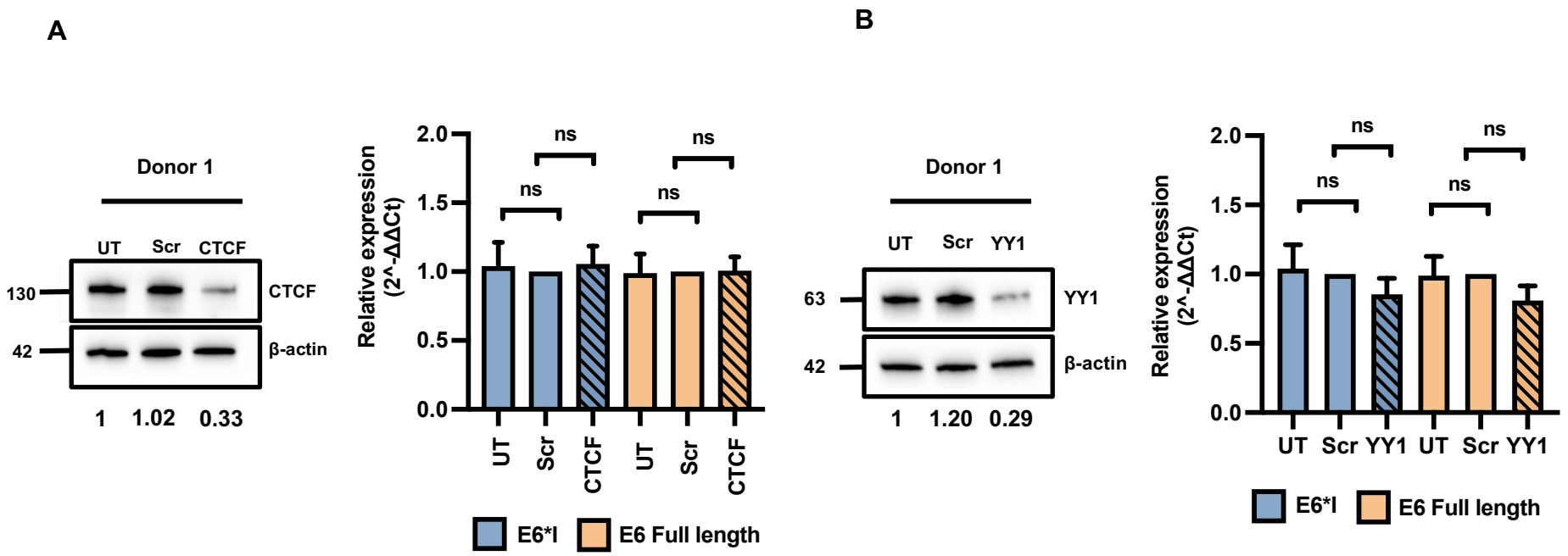


Figure 4.10 Depletion of CTCF and YY1 in high-passage 16-HTKs does not alter E6/E7 transcript expression. High-passage 16-HTKs were left untransfected (UT) or transfected with scrambled (Scr), CTCF-specific siRNA (CTCF) or YY1-specific siRNA at 10 nM for 48 hours. CTCF, YY1 and β-actin protein expression analysed by western blotting (A; siCTCF, B; siYY1). Relative protein expression normalised to β-actin was assessed using densitometry. E6 unspliced and spliced transcripts were assessed via qRT-PCR (A; siCTCF, B; siYY1). Data shown are the mean ± SEM fold change normalised to β-actin (2^-ΔΔCt) from three independent repeats. *p* values (n.s.) were determined using one-way ANOVA.

Chapter 5 YY1 functions as an activator of E6/E7 oncogene transcription to contribute to HPV-driven oropharyngeal squamous cell carcinoma

5.1 Introduction

Using our physiologically relevant HPV progression model we have begun to understand how the function of CTCF and YY1 in regulating E6/E7 oncogene transcription is altered from the HPV life cycle through to early stages of disease. However, one of the main aims of this study was to understand whether CTCF-YY1-mediated regulation of HPV transcription becomes deregulated during HPV-driven cancer. To begin to answer this question, we opted to use a series of HPV-positive oropharyngeal squamous cell carcinoma cell lines derived from patients with either primary or recurrent tumours of the oropharynx that contain naturally occurring HPV16. The majority of this work was carried out in SCC147 and SCC104 which are both derived from patients with floor of the mouth squamous cell carcinoma (Hermsen et al., 1996, Tang et al., 2012). Both cell lines strongly express the viral oncogene transcripts including E6*I, E6*II and E7 (Tang et al., 2012, Olthof et al., 2015, Walline et al., 2017). These cell lines also display altered expression of E1, E2 and E5. SCC147 shows significantly reduced expression of both E1, E2 and E5 transcripts in comparison to E6/E7. The expression of E2 transcripts in SCC104 is comparable to that of E6*I and E6*II, however, E1 and E5 transcript expression is significantly reduced (Walline et al., 2017). Previous groups have assessed the genome status in SCC147 and SCC104 using various methods including fluorescent in situ hybridisation (FISH), Amplification of Papillomavirus Oncogene Transcripts (APOT) and detection of integrated papillomavirus sequences by

ligation mediated PCR (DIPS-PCR). However, the result of this analysis was slightly varied and inconsistent. It was originally reported that SCC147 contains both episomal and integrated copies of HPV16 mapped to chromosome 17q12. In contrast, SCC104 was found to contain only episomal HPV16 DNA (Olthof et al., 2015). This was subsequently contradicted when it was reported that SCC147 contained integrated HPV16 DNA mapped to four distinct chromosomal loci and that SCC104 also contained integrated HPV16 DNA mapped to two distant chromosomal loci (Walline et al., 2017). Our initial aim was to therefore characterise the genome status in the OPSCCs.

5.2 Characterisation of the genome status in SCC147 and SCC104

To assess the genome status in SCC147 and SCC104 we opted to use Southern blotting analysis. DNA was phenol-chloroform extracted and restriction enzyme digested with either *Bam*HI*,* which linearises the HPV16 genome, or *Hind*III*,* which is a non-cutter of the HPV16 genome. HPV16 DNA was subsequently visualised using a HPV16 ^{32}P labelled probe. The result of this is shown in **Figure 5.1**. In **Figure 5.1A,** upon digestion of SCC147 DNA with *Bam*HI, a linear 8 kbp band can be seen, however, there is also a band visible which runs at a higher molecular weight, which might represent host-viral fusions. In addition, following digestion by *Hind*III there is one very high molecular weight band present, suggesting viral-host concatemers (**Figure 5.1A**). In **Figure 5.1B** digestion of SCC104 DNA with *Bam*HI does not generate an 8 kbp linearised band and again in the *Hind*III restricted lanes there is a large molecular weight band at around 10 kbp **(figure 5.1B).** Based on this, we conclude that SCC147 and SCC104 both contain integrated HPV16 DNA.

Whilst Southern blotting enables visualisation of the genome status, we were also interested in characterising the integration events at the molecular level. HPV-driven cancers can contain multiple integration events; however, it is often the case that only a minority of these integrants are transcriptionally active and as such are considered the drivers of oncogenesis (Van Tine et al., 2004, Xu et al., 2013, Warburton et al., 2021). Integration has been mapped previously in SCC147 – firstly, using DIPS-PCR which identified three individual genomic regions that contained integrated HPV16. These were mapped to the E1 open reading frame and chromosome 17q21, to the L2 open reading frame and chromosome 3p21 and to the E1 open reading frame and chromosome 5p15.31 (Walline et al., 2017). More recently, RNA-seq has been used to identify the integration sites in SCC147. This identified seven individual genomic regions that contained integrated HPV16. These were located on chromosome 15q13.1, 15q11.2, 16p11.2, 17q12, 17q21.32, 17q25.3 and 3p21.2 (Karimzadeh et al., 2022).

Because these techniques only detail transcriptionally active integrants, we opted to use a different approach in order to map all virus-host fusions, using a method that allows for the detection of both transcriptionally silent and transcriptionally active integration events. SCC147 DNA was barcoded and subsequently sequenced using the NextSeq 500 (Illumina). Single end reads were aligned to the HPV16 genome using HISAT. The sequencing reads were then soft-clipped which removes any portions of the reads that do not align to the HPV16 reference genome. All soft-clipped reads were recovered and subsequently aligned to the human genome reference sequence (Hg19) again using HISAT. Finally, the reads were filtered, merged and overlapped to generate a bed file. Visualisation of the

integration events was carried out using IGV. Details of the identified host-virus fusions are shown in **Figure 5.2.** We firstly wanted to validate previously identified integrants. Out of the seven events identified using RNA-seq and the three identified using DIPS-PCR, we were only able to validate one event which was located on chromosome 17q12 **(Fig 5.2).** We found that this integration event could be mapped to nucleotide 1066 of the HPV16 genome which is situated within the E1 open reading frame **(Fig 5.2A).** We also identified two novel integration events that had not previously been characterised within these cells; given that they were not identified previously using RNA-sequencing, it is possible that these integrants are transcriptionally inactive. We identified a fusion event between HPV16 at nucleotide 903 and chromosome 8q21 **(Fig 5.2B**) and HPV16 at nucleotide 891 and chromosome 18p11.22 **(Fig 5.2C).** Taken together, this data highlights that SCC104 and SCC147 both contain integrated HPV DNA. Importantly, we have identified multiple integration events in SCC147, most of which disrupt the E1/E2 open reading frame.

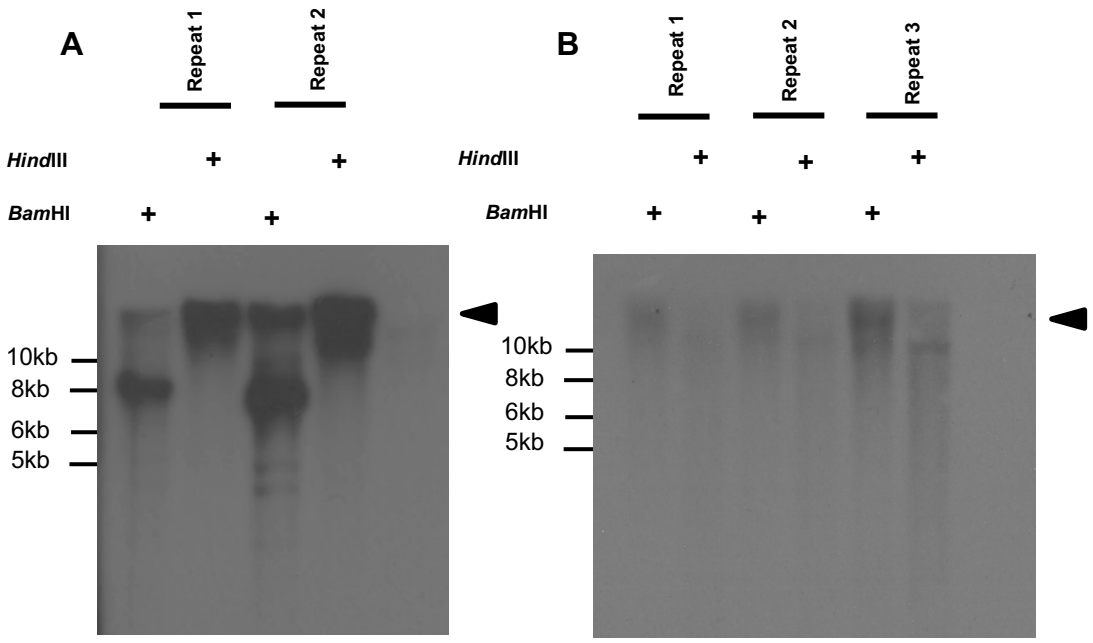


Figure 5.1 Loss of episomal HPV16 genomes in SCC147 and SCC104. HPV16 genome status was assessed in OPCSCC cell lines (SCC147; A, SCC104; B) using Southern blotting. DNA was phenol-chloroform extracted and 5 µg was restriction enzyme digested with either *Bam*HI or *Hind*III and subsequently separated in a 0.8% agarose gel. The DNA was transferred to a GeneScreen nitrocellulose membrane and UV crosslinked. Viral DNA was labelled using a HPV16 ^{32}P probe and visualised using autoradiography. Molecular weight (kbp) is indicated on the left. Data show genome status from two independent repeats (SCC147) and three independent repeats (SCC104).

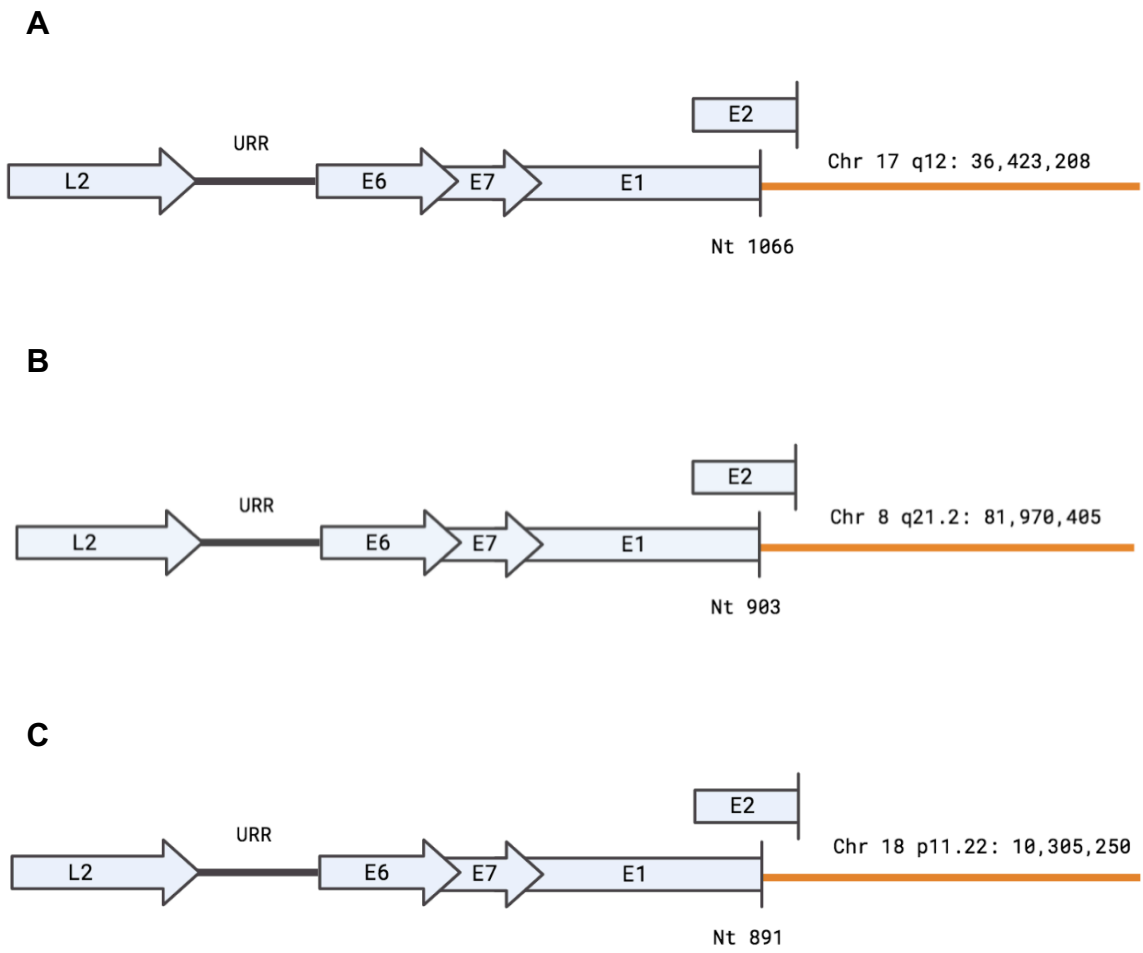


Figure 5.2 Identification of the virus-host fusion integration events in SCC147 cell lines. SCC147 DNA was barcoded and sequenced using the NextSeq 500 (Illumina). Single end reads were aligned to the HPV16 genome and were soft-clipped. The soft-clipped reads were recovered and aligned to the human genome. Reads were filtered, merged and overlapped to generate a bed file. IGV was used to visualise host-viral integration events. The HPV genome is indicated on the left (blue) and the host genome is shown in orange.

5.3 Enrichment of CTCF and YY1 binding to the HPV16 genome in OPSCCs

Cervical cancer and oropharyngeal cancer often have overexpression of and/or mutations to both YY1 and CTCF (Wang et al., 2018, Schnoell et al., 2021, Fu et al., 2022, Haft et al., 2019). We have previously assessed whether there is any apparent increase in the expression of CTCF and YY1 in HPV16 driven oropharyngeal cancers using immunohistochemistry on HPV-negative and HPV-positive tonsil tumour tissue sections. This demonstrated a robust increase in the expression of both CTCF and YY1 in HPV-positive oropharyngeal tumours – albeit in a small dataset (unpublished data). Having also shown that CTCF and YY1 protein is significantly increased following prolonged culture of both 16- and 18-HTKs, we hypothesised that upregulation of CTCF and YY1 during HPV-driven disease might function to drive the overexpression of E6/E7 oncogenes. This would represent a switch in function from a repressor of transcription during the early life cycle. We therefore wanted to analyse the expression and enrichment of CTCF and YY1 and to understand whether they function to regulate oncogene transcription, either positively or negatively, in HPV16 oropharyngeal squamous cell carcinoma cell lines.

Having previously shown an increase in protein expression of CTCF and YY1 in HPV positive tonsil tumour sections, we wanted to confirm whether this was also consistent in vitro. We therefore assessed this using western blotting in three HPV-negative (FaDu, CAL27, 040) and three HPV-positive (147, 154, 047) oropharyngeal squamous cell carcinoma lines. We showed that there was no significant change in either CTCF or YY1 protein expression between the HPV-negative and HPV-

positive samples **(Fig 5.3).** Whilst we were unable to corroborate our findings in vitro, it is important to note the HPV-negative cell lines are not normal tissue. As such, it is probable that these lines also display some aspect of aberrant CTCF and YY1 expression. The differences between the ex vivo and in vitro findings might also reflect differences in sample heterogeneity.

Given that both CTCF and YY1 regulate gene expression through direct binding to the viral genome we wanted to assess whether binding of either protein is enriched in the HPV16 genome in the OPSCC lines, despite no clear change in protein expression. Cells were fixed in 2% formaldehyde, Dounce homogenised and sheared using sonication. A subset of sheared chromatin was purified, and the shearing efficiency was assessed by resolving 10 µg of sheared DNA on a 1.5% agarose gel. As shown in **Figure 5.4 (A; SCC147, B; SCC104)** both representative samples show a DNA smear that runs from 200-600 bp, indicating efficient shearing. Enrichment of CTCF and YY1 to the HPV16 genome was then assessed using ChIP-qPCR. Chromatin was incubated overnight with 8 µg of CTCF polyclonal antibody, 10 µg of YY1 polyclonal antibody or 5 µg of anti-Flag M2 antibody and protein G magnetic beads. Precipitated DNA was eluted and enrichment on the HPV16 genome was assessed using qPCR. As there was not an internal control in the form of low-passage HTKs, the eluted DNA was assessed using a negative and positive control primer set to ensure that binding (or absence of binding) was not a result of experimental errors. These primers anneal within the human genome at sites known to be strongly enriched by CTCF and sites that are void of CTCF binding (Nakamoto et al., 2017). In both tested cell lines there was a significant increase in

enrichment of CTCF using the positive control when compared to the negative (SCC147 p <0.01 and SCC104 p <0.001) **(Fig 5.5).**

To assess enrichment of CTCF and YY1 to the viral genome the eluted DNA was analysed using qPCR with primers that amplify multiple sites across the HPV16 genome. As shown in **Figure 5.6** and **Figure 5.8** there is significant enrichment of CTCF compared to the anti-Flag control at the E2 open reading frame at nucleotides 2853-2950 in both tested cell lines (p <0.01). In addition to this, there is also a robust enrichment of CTCF at the late region of the genome, particularly across the L2 open reading frame in both tested cell lines (SCC147 p<0.01, p<0.001, SCC104 p <0.05). **(Fig 5.6, Fig 5.8).** In SCC147, enrichment of YY1 is observed at the URR, at the early promoter and at the L2 open reading frame (p<0.01) **(Fig 5.7)**. In comparison, in SCC104 there is robust enrichment of YY1 at the URR and the early promoter, however, it is not so apparent across the late region of the genome (p <0.05) **(Fig 5.9).** The subtle differences in enrichment across the late region of the HPV16 genome may be attributable to differences in shearing efficiency between the two cell lines. Taken together, this data indicates that both CTCF and YY1 show significantly enriched binding to the HPV16 genome in oropharyngeal squamous cell carcinoma cell lines. A summary comparing CTCF and YY1 enrichment, and how this correlates to E6/E7 expression, from the primary HTKs and OPSCC cell lines can be found in **Table 5.1**.

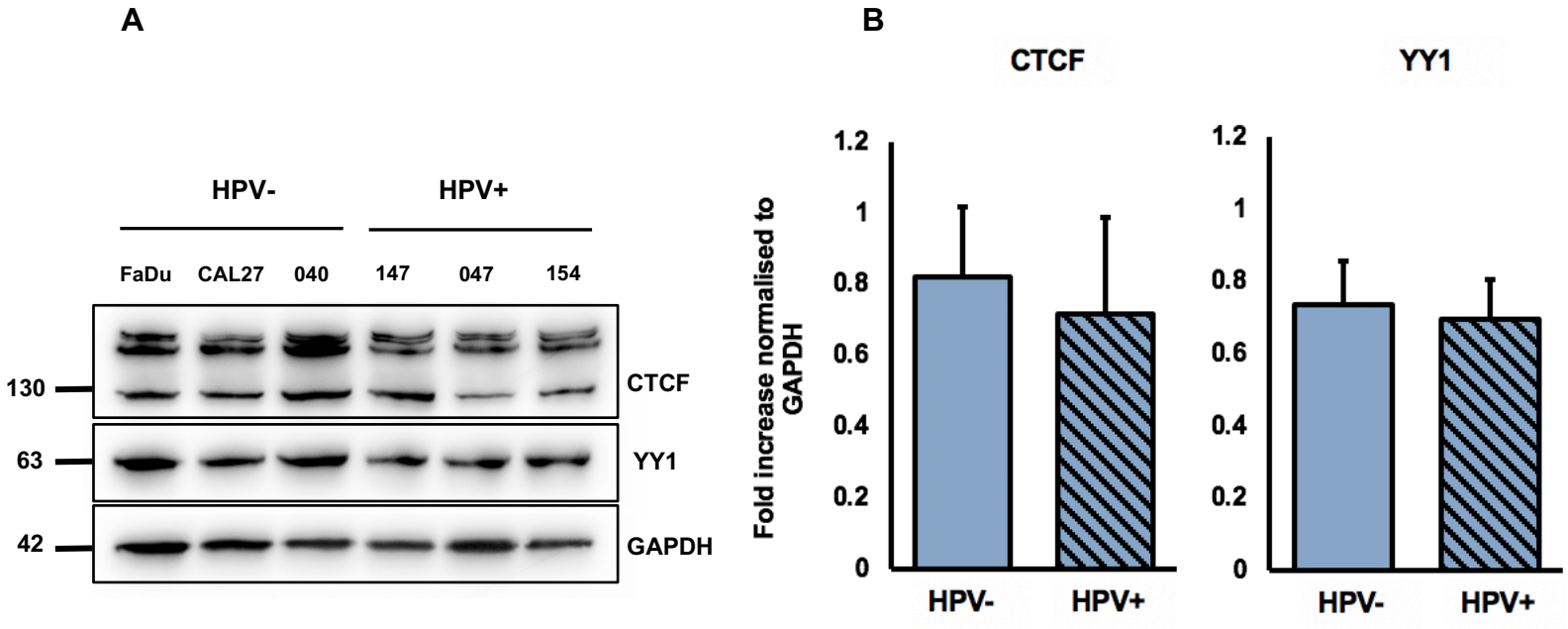


Figure 5.3 CTCF and YY1 protein expression is unchanged between HPV- and HPV+ OPSCC cell lines. (A) The expression of CTCF and YY1 was assessed in three HPV- (FaDu, CAL27, SCC040) and three HPV+ (SCC147, SCC047, SCC154) oropharyngeal squamous cell carcinoma cell lines using western blotting. A total of 20 µg of protein was separated using a 4-20% polyacrylamide gel. Protein was transferred onto a PDVF membrane and imaged using a Fusion FX7 digital detection system. (B) Protein densitometry was carried out using the Fusion FX7 digital detection system and relative fold increase normalised to β-actin was calculated. Data represent mean ±SD from one independent repeat.

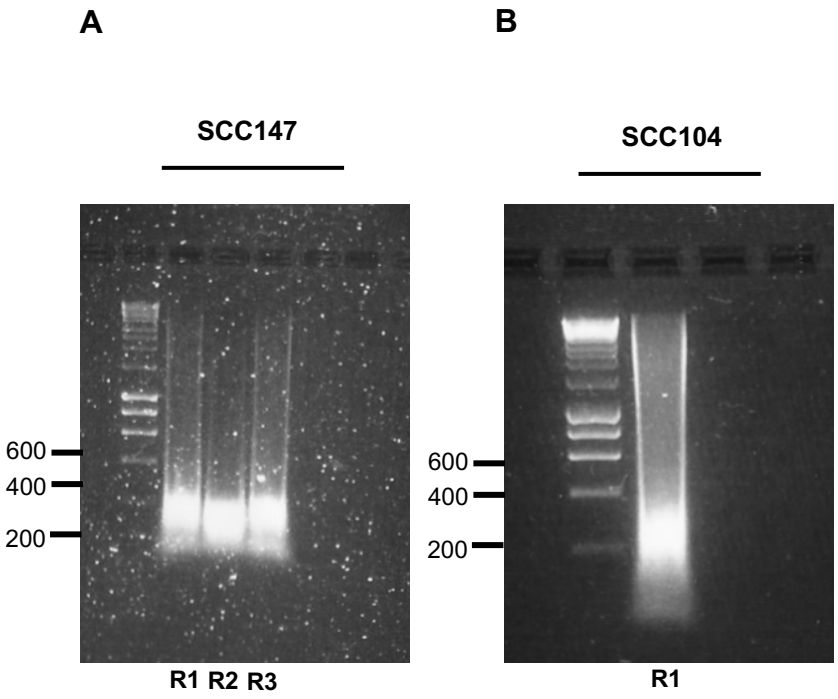


Figure 5.4 Assessment of chromatin shearing efficiency in SCC147 and SCC104. Cell pellets were fixed in 2% formaldehyde, lysed and Dounce homogenized for the preparation of intact nuclei. Chromatin was fragmented via sonication for 15 minutes at 25% amplitude (30 seconds on, 30 seconds off). DNA was extracted from sheared chromatin using phenol: chloroform: isoamylalcohol (25:24:1). A total of 10 µg of sheared DNA and a 1 kbp DNA ladder and was separated in a 1.5% agarose gel and visualised using a Gene Flash UV lightbox to assess sheared chromatin (SCC147; A, SCC104; B). Left: data show shearing efficiency from three independent repeats (SCC147). Right: data shown are shearing efficiency from one independent repeat and are representative of three independent repeats (SCC104). s R1, repeat 1. R2, repeat 2. R3, repeat 3.

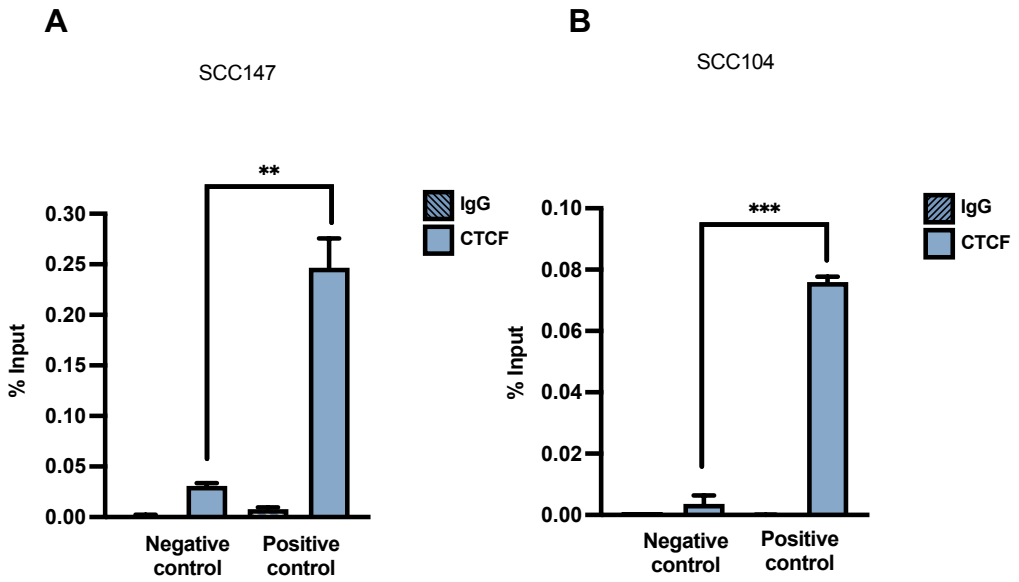


Figure 5.5 CTCF binding to the human genome in SCC147 and SCC104 cell lines. Assessment of CTCF association to the human genome in OPSCC cell lines (SCC147; A and SCC104; B). Cell pellets were fixed in 2% formaldehyde lysed and Dounce homogenized for the preparation of intact nuclei. Chromatin was fragmented via sonication and 25 µg was immunoprecipitated with 8 µg of CTCF polyclonal antibody or 5 µg anti-Flag M2 antibody. Eluted DNA was amplified using qPCR with primers that anneal within a site that CTCF does not bind (negative control) and within a site enriched for CTCF binding (positive control). Data shown are the mean ±SD (% Input) from three technical replicates and are representative of three independent repeats. ** $p < 0.01$ and *** $p < 0.001$ as determined by a paired *t* test.

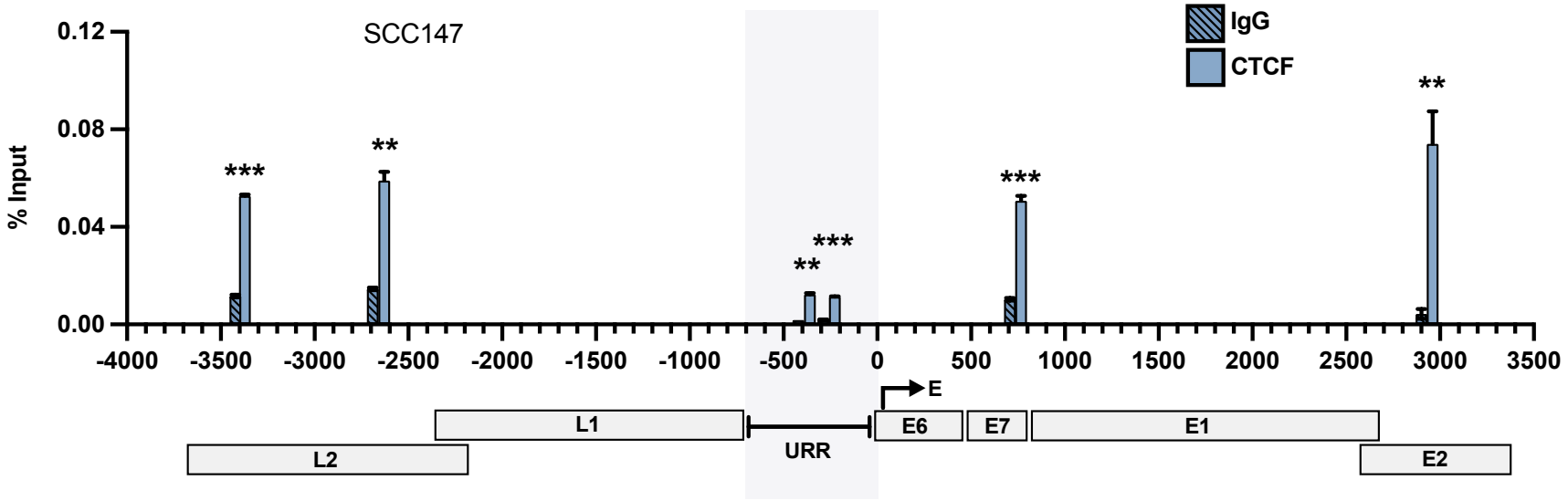


Figure 5.6 CTCF binding to the HPV16 genome of SCC147 cell lines. CTCF association to the HPV16 genome was assessed by ChIP-qPCR in SCC147 cell lines. Cell pellets were fixed in 2% formaldehyde lysed and Dounce homogenized for the preparation of intact nuclei. Chromatin was fragmented via sonication and 25 µg was immunoprecipitated with 8 µg of CTCF polyclonal antibody or 5 µg anti-flag M2 antibody. Eluted DNA was amplified using qPCR with primers that anneal at multiple sites along the HPV16 genome. The linearized HPV16 genome is displayed on the x-axis based on the 1/7905 bp sequence (NC_001526.2). Peaks represent the midpoint of the PCR amplicon. Data shown are the mean ±SD (% Input) from three technical replicates and are representative of three independent repeats. ** $p < 0.01$, *** $p < 0.001$ as determined by a paired *t* test.

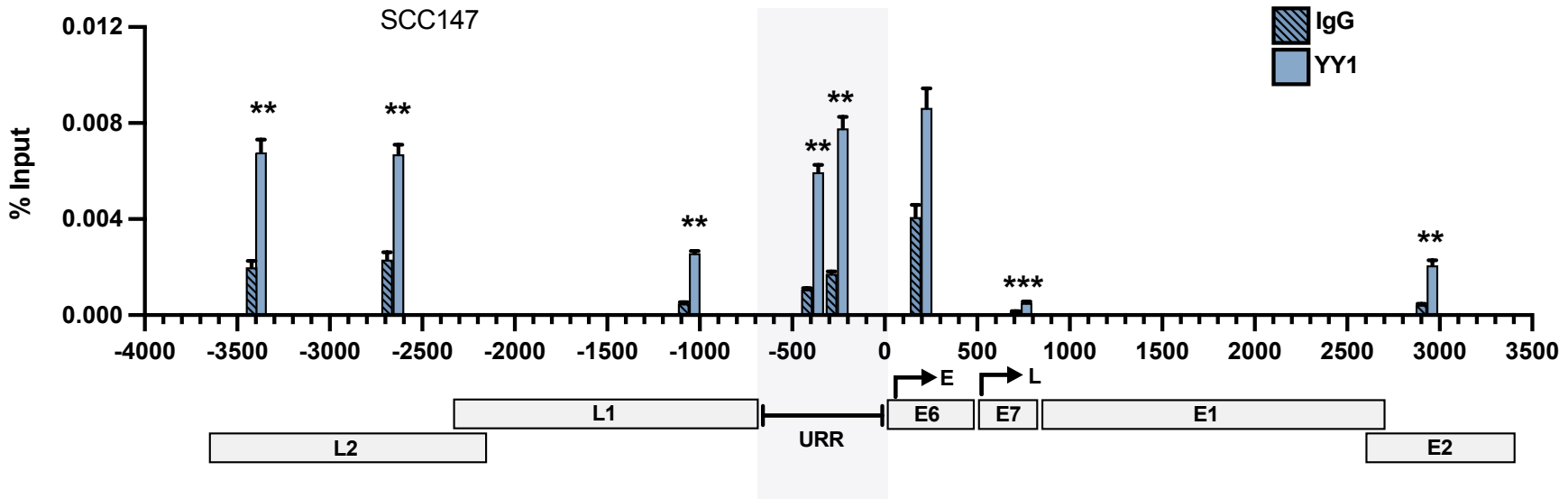


Figure 5.7 YY1 binding to the HPV16 genome of SCC147 cell lines. YY1 association to the HPV16 genome was assessed by ChIP-qPCR in SCC147 cell lines. Cell pellets were fixed in 2% formaldehyde lysed and Dounce homogenized for the preparation of intact nuclei. Chromatin was fragmented via sonication and 25 µg was immunoprecipitated with 10 µg of YY1 polyclonal antibody or 5 µg anti-Flag M2 antibody. Eluted DNA was amplified using qPCR with primers that anneal at multiple sites along the HPV16 genome. The linearized HPV16 genome is displayed on the x-axis based on the 1/7905 bp sequence (NC_001526.2). Peaks represent the midpoint of the PCR amplicon. Data shown are the mean ±SD (% Input) from three technical replicates and are representative of three independent repeats. ** $p < 0.01$, *** $p < 0.001$ as determined by a paired *t* test.

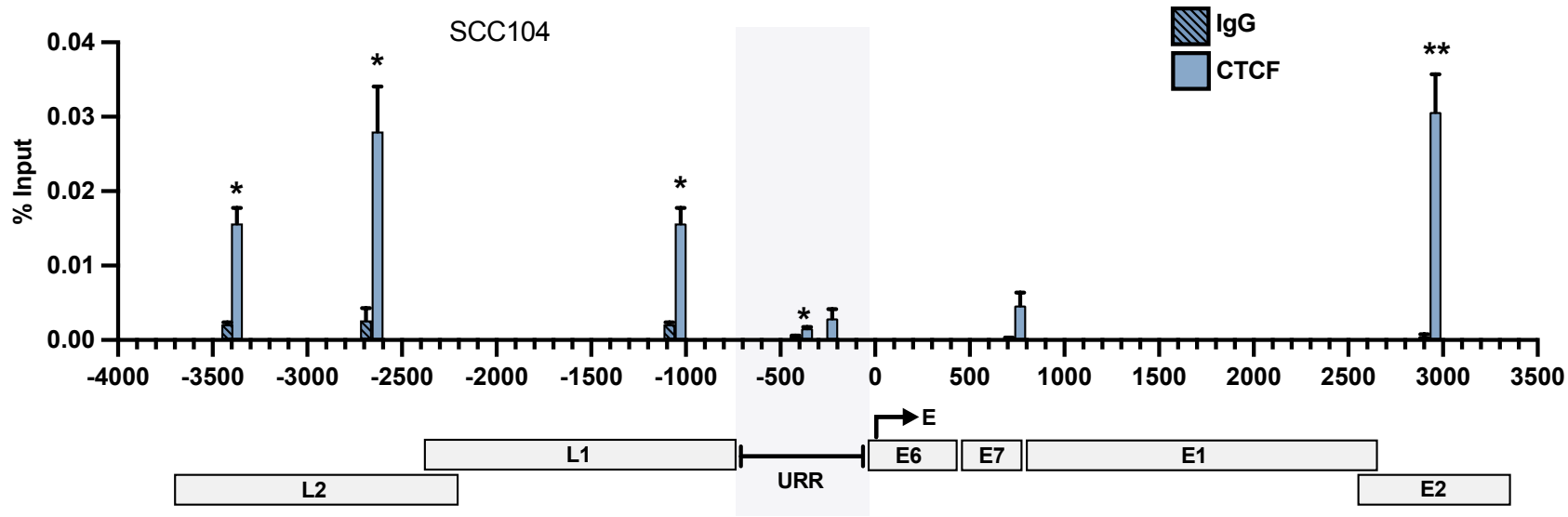


Figure 5.8 CTCF binding to the HPV16 genome of SCC104 cell lines. CTCF association to the HPV16 genome was assessed by ChIP-qPCR in SCC104 cell lines. Cell pellets were fixed in 2% formaldehyde lysed and Dounce homogenized for the preparation of intact nuclei. Chromatin was fragmented via sonication and 25µg was immunoprecipitated with 8µg of CTCF polyclonal antibody or 5 µg anti-Flag M2 antibody. Eluted DNA was amplified using qPCR with primers that anneal at multiple sites along the HPV16 genome. The linearized HPV16 genome is displayed on the x-axis based on the 1/7905 bp sequence (NC_001526.2). Peaks represent the midpoint of the PCR amplicon. Data shown are the mean ±SD (% Input) from three technical replicates and are representative of three independent repeats. * $p < 0.05$, ** $p < 0.01$ as determined by a paired *t* test.

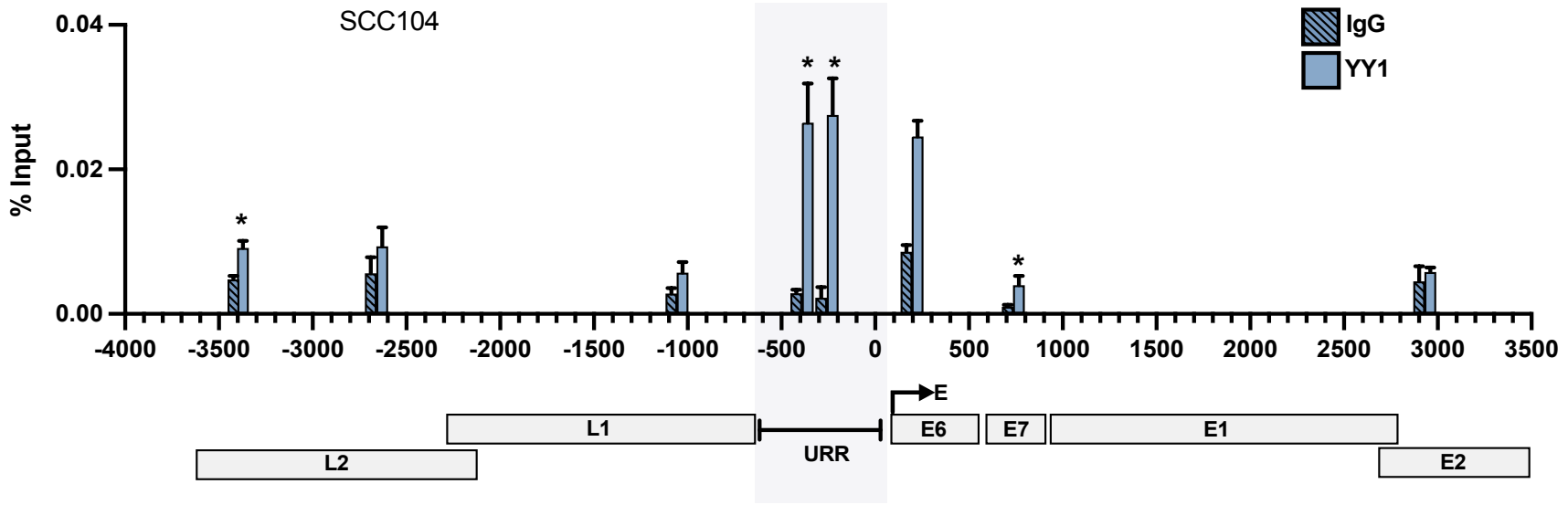


Figure 5.9 YY1 binding to the HPV16 genome of SCC104 cell lines. YY1 association to the HPV16 genome was assessed by ChIP-qPCR in SCC104 cell lines. Cell pellets were fixed in 2% formaldehyde lysed and Dounce homogenized for the preparation of intact nuclei. Chromatin was fragmented via sonication and 25 µg was immunoprecipitated with 10 µg of YY1 polyclonal antibody or 5 µg anti-Flag M2 antibody. Eluted DNA was amplified using qPCR with primers that anneal at multiple sites along the HPV16 genome. The linearized HPV16 genome is displayed on the x-axis based on the 1/7905 bp sequence (NC_001526.2). Peaks represent the midpoint of the PCR amplicon. Data shown are the mean ±SD (% Input) from three technical replicates and are representative of four independent repeats. * $p < 0.05$ as determined by a paired *t* test.

	HPV16 high-passage HTKs	HPV18 high-passage HTKs	HPV16 positive OPSCC cell lines
Genome status	Integrated	Episomal	Integrated
CTCF binding	Yes	No	Yes
YY1 binding	Unknown	Yes	Yes
Increased E6/E7 expression	Yes	Yes	Yes
Increased CTCF/YY1 expression	Yes	Yes	No

Table 5.1 Summary and comparison of the genome status, CTCF enrichment, YY1 enrichment, E6/E7 expression and CTCF/YY1 protein expression in HPV16 high-passage HTKs, HPV18 high-passage HTKs and HPV16 positive OPSCC cell lines.

5.4 Enrichment of YY1 and CTCF binding to the HPV16 genome in OPSCCs is consistent with a transcriptionally active epigenetic landscape

We have previously demonstrated that in undifferentiated HPV18-HFKs, enrichment of CTCF and YY1 binding results in a high abundance of trimethylation of lysine 27 (H2K27Me3) and mono-ubiquitination of lysine 119 (H2AK119Ub) across the viral genome but predominantly at the URR and early promoter region. As a result of this chromatin is compacted across the URR and at the early promoter, and as such, there is reduced promoter activity and low expression of E6/E7 transcripts (Pentland et al., 2018). Having established that there is robust enrichment of CTCF and YY1 to the HPV16 genome in SCC147 and SCC104, we next wanted to determine whether this resulted in epigenetic changes that would regulate transcription from the early promoter.

Given the high expression of the oncogene transcription in these cell lines we firstly assessed enrichment of active histone marks including H3K4 trimethylation (H3K4Me3) and H3K27 acetylation (H3K27Ac). ChIP-qPCR revealed one significant peak of enrichment of H3K27Ac at the HPV16 early promoter in both tested cell lines (SCC147 $p<0.01$, SCC104 $p<0.05$) **(Fig 5.10, Fig 5.12)**. This was also true for H3K4Me3 where again there was robust enrichment at the early promoter (SCC147 n.s., SCC104 $p<0.05$) **(Fig 5.11, Fig 5.13)**. Consistent with this, ChIP-qPCR also demonstrated that there was a robust peak of enrichment of H3K27Me3 at the HPV16 late promoter in SCC147 when compared to the

remainder of the viral genome ($p < 0.05$) (**Fig 5.14)** indicative of high early promoter activity.

Next, we examined whether the epigenetic status of these resulted in recruitment of RNA PolII to the early promoter, again assessed by ChIP-qPCR. We found this to be the case and enrichment of active histone marks strongly correlated with recruitment of RNA PolII in both tested cell lines (SCC147 $p<0.001$, SCC104 $p<0.05$) **(Fig 5.15, Fig 5.16)**. Taken together, this data confirms that unlike in the HPV life cycle where binding of CTCF and YY1 is coincident with an epigenetically repressed chromatin landscape, in HPV-driven OPSCCs, enrichment of both YY1 and CTCF is consistent with active histone modifications at the early promoter and recruitment of RNA PolII.

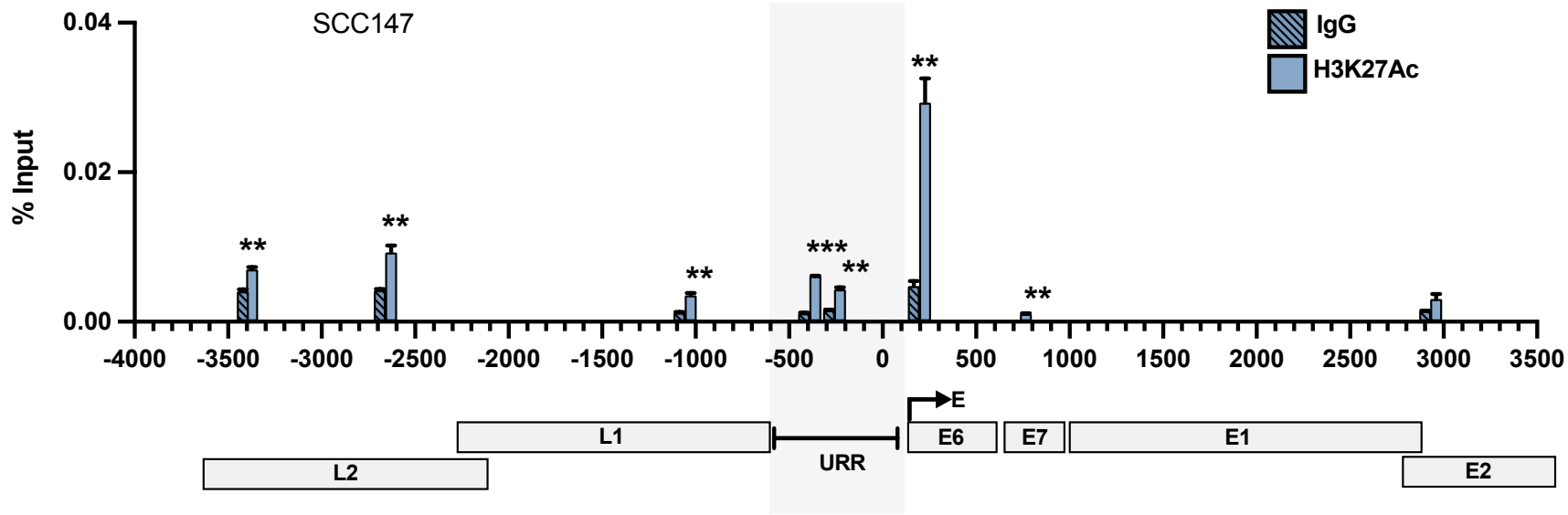


Figure 5.10 Enrichment of H3K27Ac at the HPV16 early promoter in SCC147. H3K27Ac association to the HPV16 genome was assessed by ChIP-qPCR in SCC147 cell lines. Cell pellets were fixed in 2% formaldehyde, lysed and Dounce homogenized for the preparation of intact nuclei. Chromatin was fragmented via sonication and 25 µg was immunoprecipitated with 5 µg of H3K27Ac polyclonal antibody or 5 µg of anti-Flag M2 antibody. Eluted DNA was amplified using qPCR with primers that anneal at multiple sites along the HPV16 genome. The linearized HPV16 genome is displayed on the x-axis based on the 1/7905 bp sequence (NC_001526.2). Peaks represent the midpoint of the PCR amplicon. Data shown are the mean ±SD (% Input) from three technical replicates and are representative of three independent repeats. ** $p < 0.01$, *** $p < 0.001$ as determined by a paired *t* test.

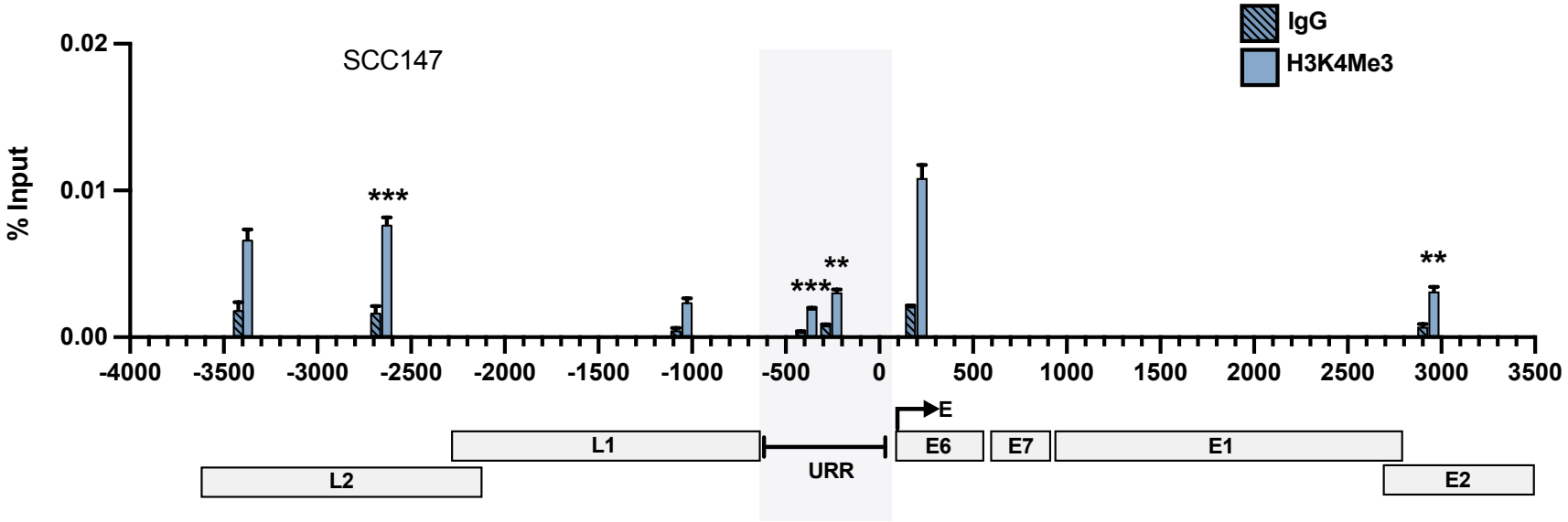


Figure 5.11 Enrichment of H3K4Me3 at the HPV16 early promoter in SCC147. H3K4Me3 association to the HPV16 genome was assessed by ChIP-qPCR in SCC147 cell lines. Cell pellets were fixed in 2% formaldehyde, lysed and Dounce homogenized for the preparation of intact nuclei. Chromatin was fragmented via sonication and 25 µg was immunoprecipitated with 5 µg of H4K4Me3 polyclonal antibody or 5 µg of anti-Flag M2 antibody. Eluted DNA was amplified using qPCR with primers that anneal at multiple sites along the HPV16 genome. The linearized HPV16 genome is displayed on the x-axis based on the 1/7905 bp sequence (NC_001526.2). Peaks represent the midpoint of the PCR amplicon. Data shown are the mean ±SD (% Input) from three technical replicates and are representative of three independent repeats. ** $p < 0.01$, *** $p < 0.001$ as determined by a paired *t* test.

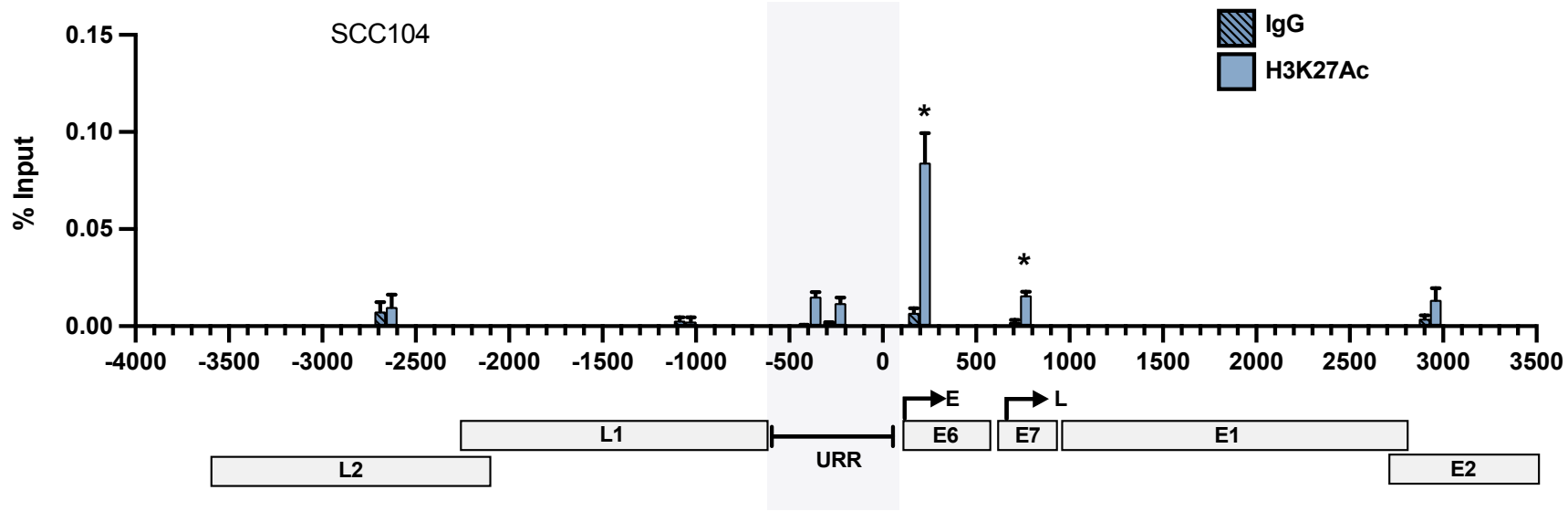


Figure 5.12 Enrichment of H3K27Ac at the HPV16 early promoter in SCC104. H3K27Ac association to the HPV16 genome was assessed by ChIP-qPCR in SCC104 cell lines. Cell pellets were fixed in 2% formaldehyde, lysed and Dounce homogenized for the preparation of intact nuclei. Chromatin was fragmented via sonication and 25 µg was immunoprecipitated with 5 µg of H3K27Ac polyclonal antibody or 5 µg of anti-Flag M2 antibody. Eluted DNA was amplified using qPCR with primers that anneal at multiple sites along the HPV16 genome. The linearized HPV16 genome is displayed on the x-axis based on the 1/7905 bp sequence (NC_001526.2). Peaks represent the midpoint of the PCR amplicon. Data shown are the mean ±SD (% Input) from three technical replicates and are representative of three independent repeats. * $p < 0.05$ as determined by a paired *t* test.

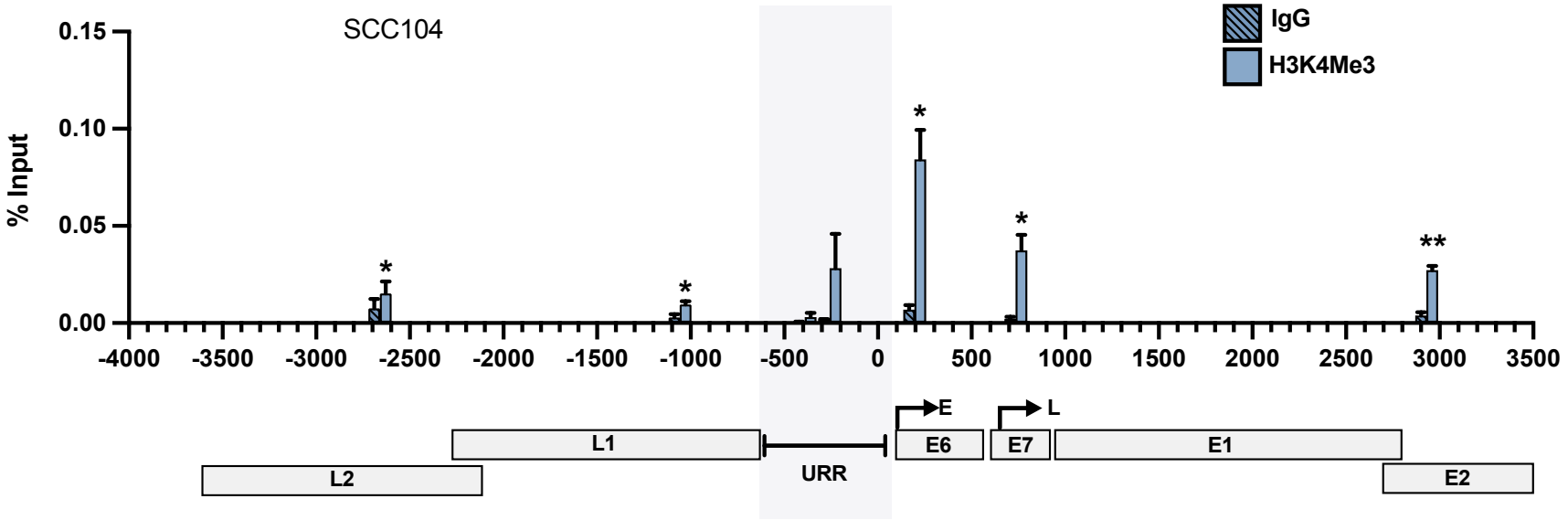


Figure 5.13 Enrichment of H3K4Me3 at the HPV16 early promoter in SCC104. H3K4Me3 association to the HPV16 genome was assessed by ChIP-qPCR in SCC104 cell lines. Cell pellets were fixed in 2% formaldehyde, lysed and Dounce homogenized for the preparation of intact nuclei. Chromatin was fragmented via sonication and 25 µg was immunoprecipitated with 5 µg of H4K4Me3 polyclonal antibody or 5 µg of anti-Flag M2 antibody. Eluted DNA was amplified using qPCR with primers that anneal at multiple sites along the HPV16 genome. The linearized HPV16 genome is displayed on the x-axis based on the 1/7905 bp sequence (NC_001526.2). Peaks represent the midpoint of the PCR amplicon. Data shown are the mean ±SD (% Input) from three technical replicates and are representative of three independent repeats. * $p < 0.05$, ** $p < 0.01$ as determined by a paired *t* test.

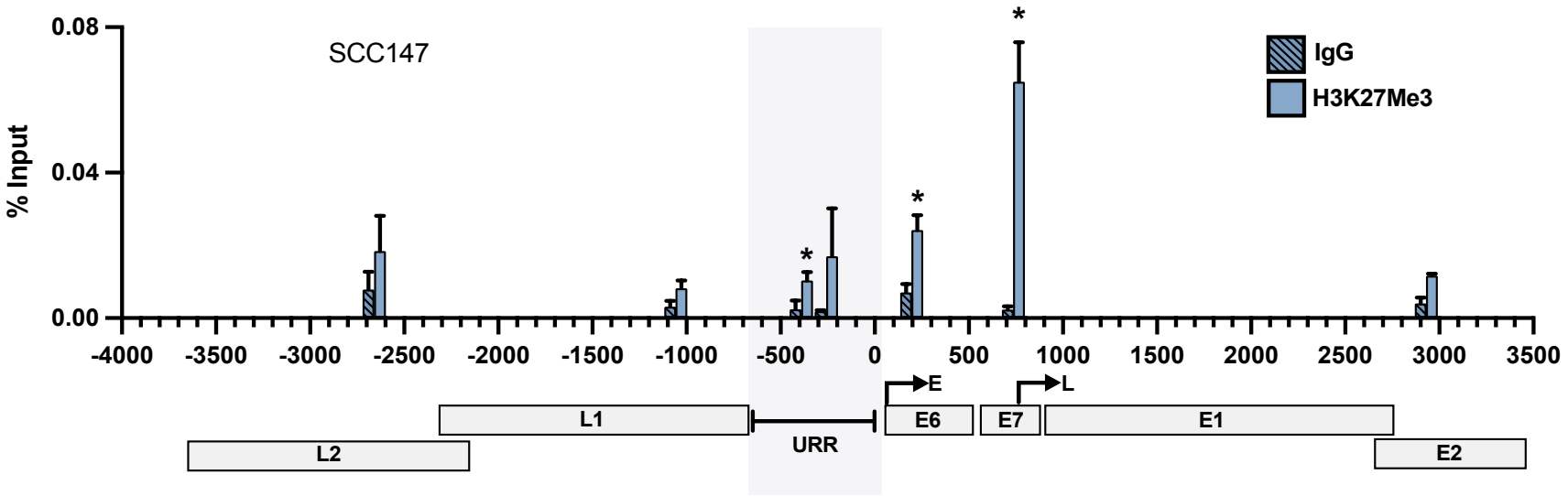


Figure 5.14 Enrichment of H3K27Me3 at the HPV16 late promoter in SCC147. H3K27Me3 association to the HPV16 genome was assessed by ChIP-qPCR in SCC147 cell lines. Cell pellets were fixed in 2% formaldehyde, lysed and Dounce homogenized for the preparation of intact nuclei. Chromatin was fragmented via sonication and 25 µg was immunoprecipitated with 5 µg of H3K27Me3 polyclonal antibody or 5 µg of anti-Flag M2 antibody. Eluted DNA was amplified using qPCR with primers that anneal at multiple sites along the HPV16 genome. The linearized HPV16 genome is displayed on the x-axis based on the 1/7905 bp sequence (NC_001526.2). Peaks represent the midpoint of the PCR amplicon. Data shown are the mean ±SD (% Input) from three technical replicates and are representative of three independent repeats. * $p < 0.05$ as determined by a paired *t* test.

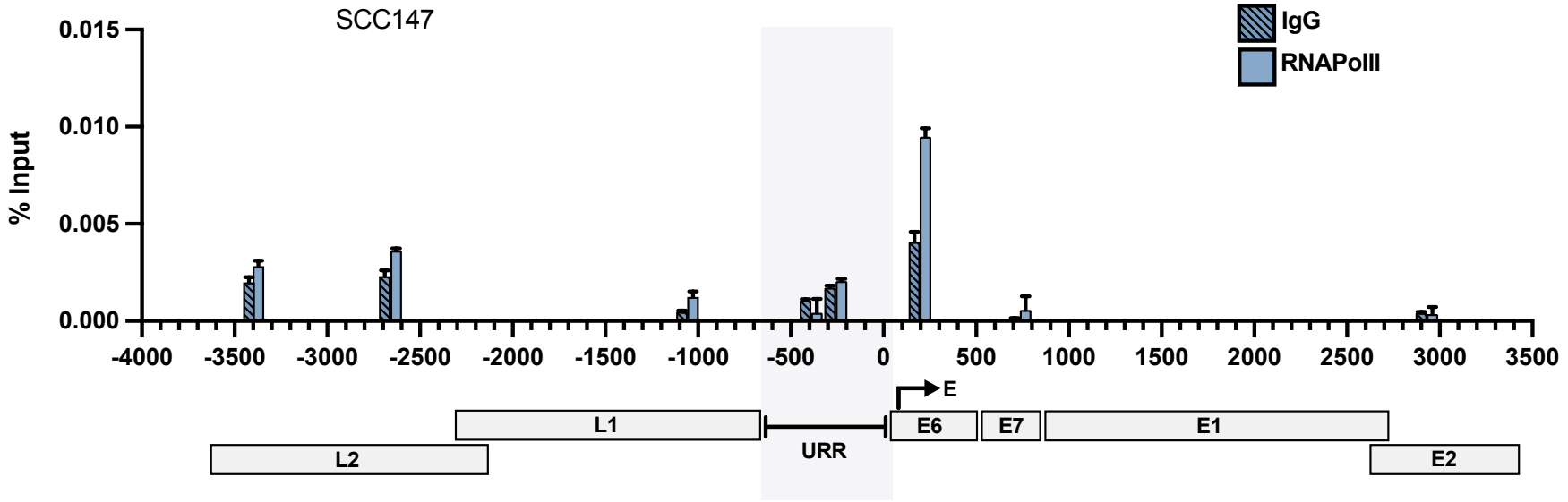


Figure 5.15 Enrichment of RNAPolII at the HPV16 early promoter in SCC147. RNAPolII association to the HPV16 genome was assessed by ChIP-qPCR in SCC147 cell lines. Cell pellets were fixed in 2% formaldehyde, lysed and Dounce homogenized for the preparation of intact nuclei. Chromatin was fragmented via sonication and 25 µg was immunoprecipitated with 5 µg of RNAPolII polyclonal antibody or 5 µg of anti-Flag M2 antibody. Eluted DNA was amplified using qPCR with primers that anneal at multiple sites along the HPV16 genome. The linearized HPV16 genome is displayed on the x-axis based on the 1/7905 bp sequence (NC_001526.2). Peaks represent the midpoint of the PCR amplicon. Data shown are the mean ±SD (% Input) from three technical replicates and are representative of two independent repeats.

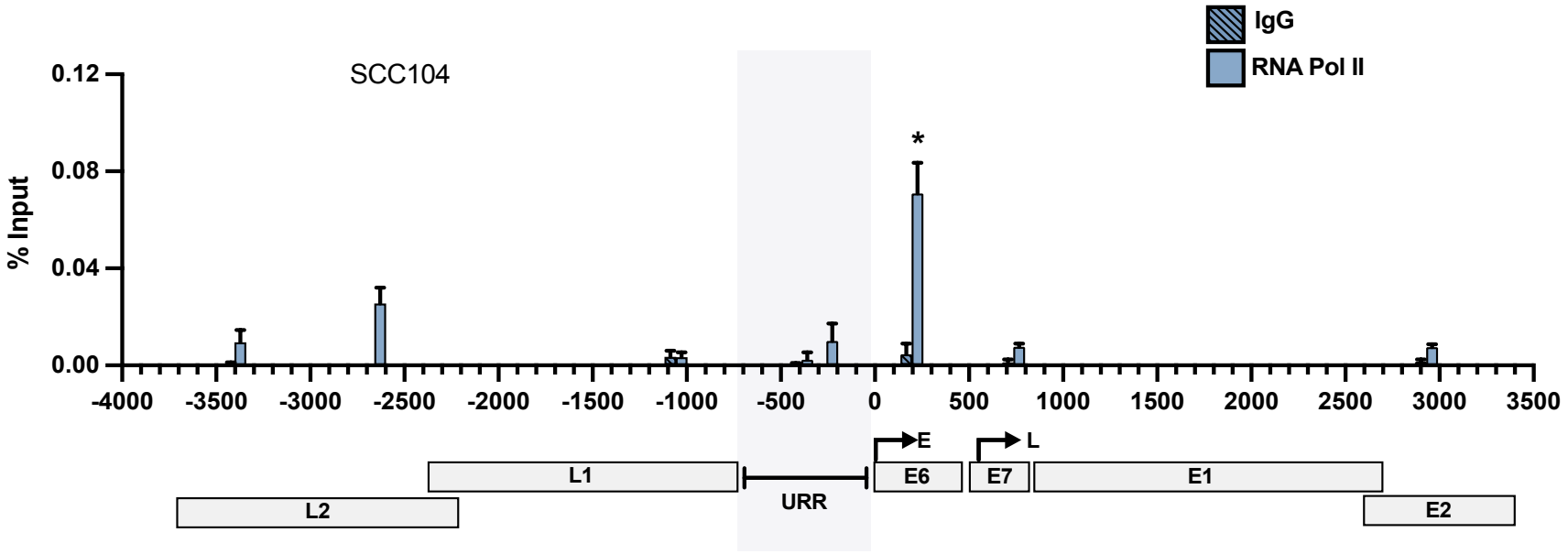


Figure 5.16 Enrichment of RNAPolII at the HPV16 early promoter in SCC104. RNAPolII association to the HPV16 genome was assessed by ChIP-qPCR in SCC104 cell lines. Cell pellets were fixed in 2% formaldehyde, lysed and Dounce homogenized for the preparation of intact nuclei. Chromatin was fragmented via sonication and 25 µg was immunoprecipitated with 5 µg of RNAPolII polyclonal antibody or 5 µg of anti-Flag M2 antibody. Eluted DNA was amplified using qPCR with primers that anneal at multiple sites along the HPV16 genome. The linearized HPV16 genome is displayed on the x-axis based on the 1/7905 bp sequence (NC_001526.2). Peaks represent the midpoint of the PCR amplicon. Data shown are the mean ±SD (% Input) from three technical replicates and are representative of three independent repeats. * $p < 0.05$ as determined by a paired *t* test.

5.5 Recruitment of PRC1 component RING1B to the viral genome in HPV16 OPSCCs is independent of transcriptional repression

When acting as a repressor of transcription YY1 can mediate the recruitment of PcG proteins to DNA (Basu et al., 2014). The PcG proteins can be categorised into two main groups: PRC1 and PRC2. We have shown previously that both PRC1 component RING1B and PRC2 component EED are recruited to the HPV18 genome by YY1 in undifferentiated HPV18-HFKs to facilitate the formation of heterochromatin and mediate transcriptional repression of viral oncogenes (Pentland et al., 2018). Having shown that binding of YY1 was coincident with an open chromatin landscape, we next wanted to assess whether this was as a result of loss of PcG recruitment to the HPV16 URR. To analyse this we assessed enrichment of PRC1 component RING1B to the HPV16 genome using ChIP-qPCR. RING1B catalyses the deposition of monoubiquitination on lysine 119 on histone 2A – this is a mark indicative of repressed gene transcription. RING1B is recruited directly by YY1 via its REPO domain which facilitates an interaction with the RYBP/RING1B complex. We therefore wanted to assess whether the failure of YY1 to facilitate an epigenetically repressed chromatin landscape was due to loss of RYBP/RING1B recruitment to the HPV16 genome. Intriguingly this was not the case – as seen in **Figure 5.17** and **Figure 5.19** there are peaks of enrichment of RING1B to multiple sites across the HPV16 genome in both SCC147 and SCC104. This is most apparent at the early promoter (SCC147 $p<0.01$, SCC104 $p<0.01$) and interestingly also at the L2 open reading frame (SCC147 $p<0.01$, SCC104 $p<0.01$). This data confirms that RING1B is still being recruited to the HPV16 genome, either

through YY1 or through an independent mechanism. Because RING1B functions to deposit H2AK119Ub we also used ChIP-qPCR to assess enrichment of this mark across the genome. Consistent with strong enrichment of RING1B we showed that the deposition of H2AK119Ub is also present within the OSPCCs. There is robust enrichment of H2AK119Ub at the early promoter (SCC147 $p<0.001$, SCC104 $p<0.01$) and at the L2 open reading frame (SCC147 $p<0.001$, SCC104 $p<0.01$) **(Fig 5.18, Fig 5.20).** Taken together, this suggests that RING1B is still recruited to the HPV16 genome despite high transcriptional activity from the early promoter. This recruitment of RING1B is consistent with enrichment of the histone modification H2AK119Ub.

Alternative mechanisms exist for RING1B mediated transcriptional repression which are independent of its histone ubiquitination activity; specifically, through mediating chromatin compaction (Eskeland et al., 2010). Recently, it has been shown that the canonical function of RING1B can be lost during breast cancer and instead, recruitment of RING1B to oncogenic enhancers acts to facilitate gene activation (Chan et al., 2018). Human breast cancer cells express the estrogen receptor (ERα) – this receptor can be stimulated by 17β-estrodiol (E_2). Following E_2 mediated stimulation, ERα relocates to the nucleus and is recruited to chromatin where it regulates genes that are required for cellular proliferation and cell cycle regulation. Stimulation of breast cancer cells by E_2 is mostly involved in the upregulation of genes which is partially dependent on chromatin remodelling (Zhang et al., 2020). RING1B has been shown to be important for E_2 stimulated gene upregulation– this is mediated mostly through RING1B association with nucleosomes and less dependent on the enzymatic activity of RING1B. As a result of E_2 stimulation,

RING1B and ERα are recruited to promoters and enhancers of genes that involved in the oncogenesis of breast cancer, including BCL2L1 (Zhang et al., 2020). Depletion of RING1B was also found to reduce global levels of the active histone modification H3K27Ac (Zhang et al., 2020). RING1B regulates enhancer-promoter contacts of E_2 responsive genes, initiates the formation of R-loops and regulates RNA polymerase II elongation, all of which are important to promote transcription of E_2 target genes (Zhang et al., 2021). Importantly, this alternative function of RING1B is not just true in breast cancer cells – RING1B has been shown to bind active promoters within quiescent lymphocytes and was found to recruit RNA polymerase II to these sites. Again, depletion of RING1B results in reduced transcription highlighting a role for RING1B in activating gene expression within multiple cell types (Frangini et al., 2013).

Given these additional and non-canonical functions of RING1B it is therefore plausible that recruitment of a YY1/RING1B complex to the HPV16 URR in OPSCCs might represent a mechanism as to which these PcG proteins cooperate to activate transcription from the early promoter and drive high oncogene expression through chromatin remodelling. Based on this, we set out to assess whether recruitment of RING1B can function to alter chromatin topology in a non-canonical manner as such to create a transcriptionally active chromatin landscape in OSPCCs. SCC147s were either left untransfected (UT) or transfected with scrambled (Scr) or RNF2-specific (RNF2) siRNA at a final concentration of 10 nM for 48 hours. The resulting chromatin changes were then assessed using FAIRE. After 48 hours, a small proportion of cells were harvested for protein whilst the remainder were fixed with 2% formaldehyde. Fixed samples were then Dounce homogenised and sheared using

sonication. Chromatin was split into two conditions FAIRE and Input. The input sample underwent an additional step in which crosslinks were reversed and DNA was extracted from both FAIRE and Input using phenol-chloroform. Chromatin accessibility was quantified using qPCR with primers that amplify multiple sites across the HPV16 genome. As shown in **Figure 5.21A** treatment with siRNF2 for 48 hours resulted in a 1.7-fold reduction in the expression of RING1B protein as determined by western blotting. Intriguingly, assessment by FAIRE showed that there was a significant difference in chromatin accessibility at the URR in the siRNF2 treated cells compared to the siScr treated cells ($p<0.05$) (**Fig 5.21B).** To try and corroborate these findings and understand the biological relevance, we also assessed whether depletion of RING1B would alter E6/E7 transcription from the early promoter. However, we found that depletion of RING1B with siRNF2 at a final concentration of 10nM for 48 hours results in no significant change to E6 spliced or E6 unspliced transcript expression as confirmed by qRT-PCR **(Fig 5.22).** We therefore concluded that RING1B recruitment to the viral genome was not acting to positively regulate transcription through chromatin remodelling. Taken together, these results show that RING1B recruitment to the HPV16 genome in OPSCCs results in the formation of open chromatin. However, RING1B does not appear to directly function to repress E6/E7 transcription from the early promoter.

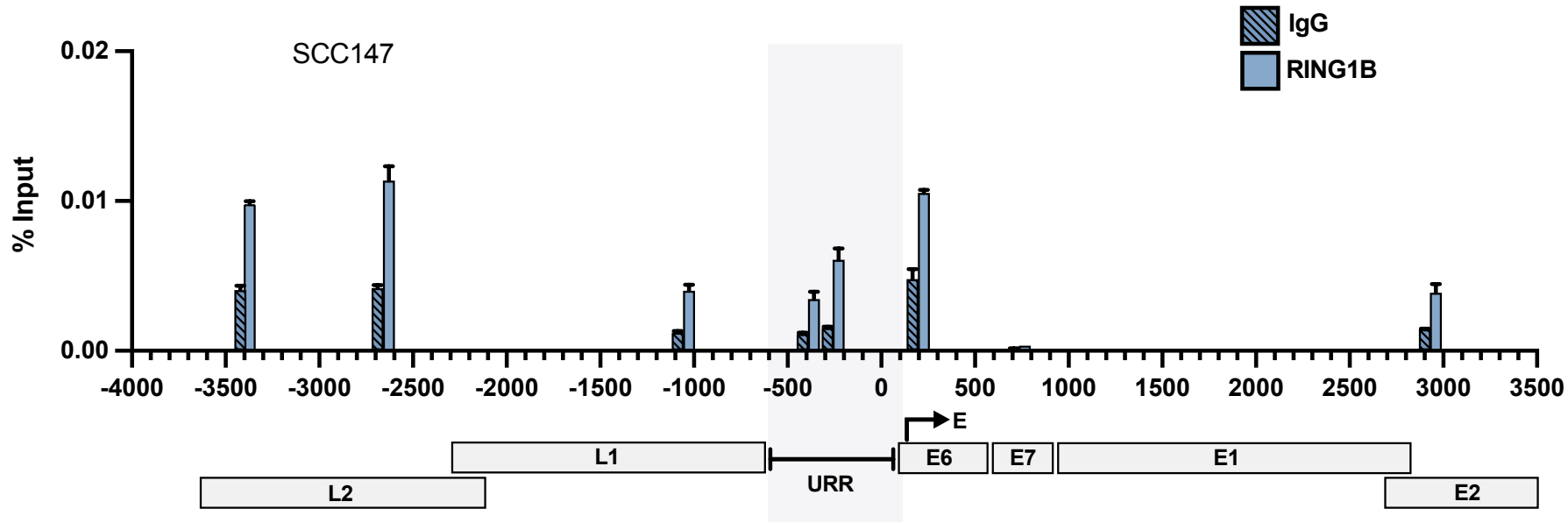


Figure 5.17 Enrichment of RING1B on the HPV16 genome in SCC147 cell lines. RING1B association to the HPV16 genome was assessed by ChIP-qPCR in SCC147 cell lines. Cell pellets were fixed in 2% formaldehyde, lysed and Dounce homogenized for the preparation of intact nuclei. Chromatin was fragmented via sonication and 25 µg was immunoprecipitated with 5 µg of RING1B monoclonal antibody or 5 µg of anti-Flag M2 antibody. Eluted DNA was amplified using qPCR with primers that anneal at multiple sites along the HPV16 genome. The linearized HPV16 genome is displayed on the x-axis based on the 1/7905 bp sequence (NC_001526.2). Peaks represent the midpoint of the PCR amplicon. Data shown are the mean ±SD (% Input) from three technical replicates and are representative of two independent repeats.

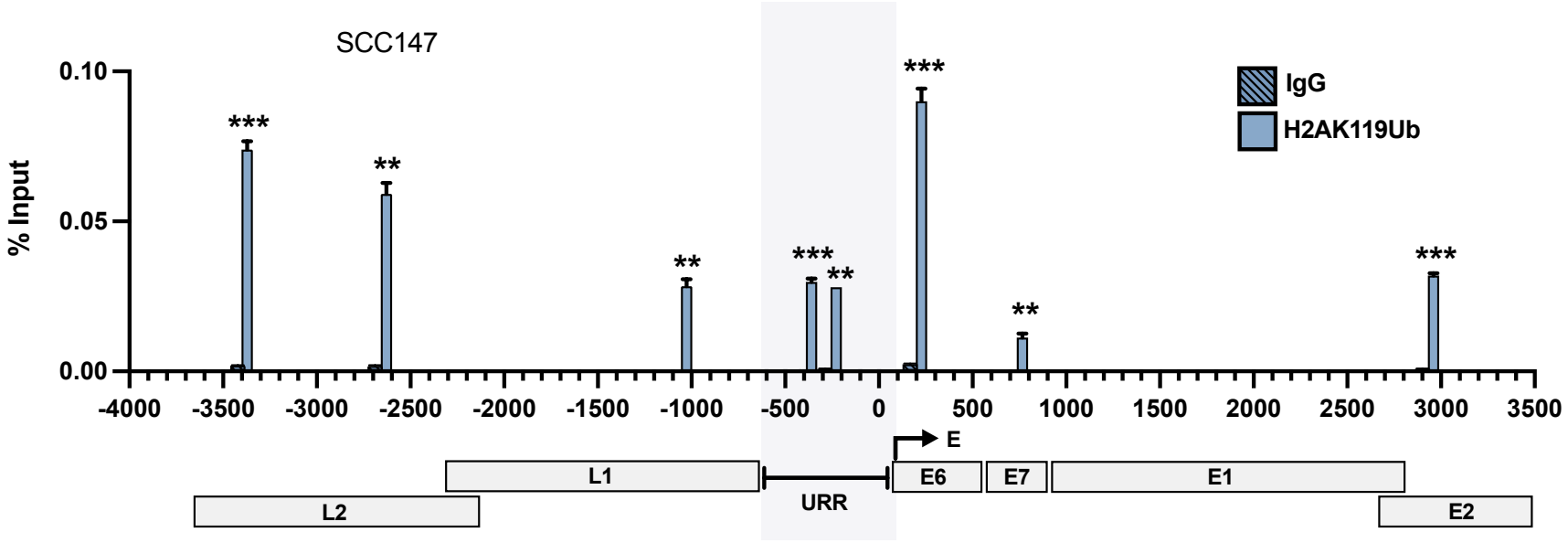


Figure 5.18 Enrichment of H2AK119Ub on the HPV16 genome in SCC147 cell lines. H2AK119Ub association to the HPV16 genome was assessed by ChIP-qPCR in SCC147 cell lines. Cell pellets were fixed in 2% formaldehyde, lysed and Dounce homogenised for the preparation of intact nuclei. Chromatin was fragmented via sonication and 25 µg was immunoprecipitated with 5 µg of monoclonal H2AK119Ub antibody or 5 µg of anti-Flag M2 antibody. Eluted DNA was amplified using qPCR with primers that anneal at multiple sites along the HPV16 genome. The linearized HPV16 genome is displayed on the x-axis based on the 1/7905 bp sequence (NC_001526.2). Peaks represent the midpoint of the PCR amplicon. Data shown are the mean ±SD (% Input) from three technical replicates and are representative of three independent repeats. ** $p < 0.01$, *** $p < 0.001$ as determined by a paired *t* test.

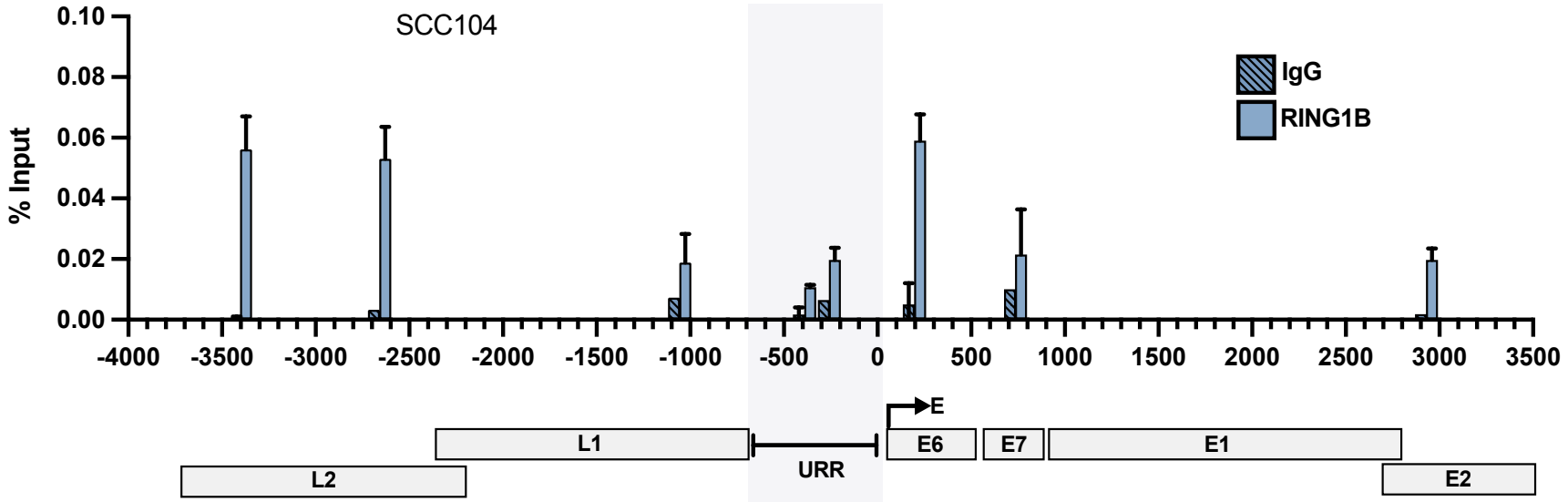


Figure 5.19 Enrichment of RING1B on the HPV16 genome in SCC104 cell lines. RING1B association to the HPV16 genome was assessed by ChIP-qPCR in SCC147 cell lines. Cell pellets were fixed in 2% formaldehyde, lysed and Dounce homogenized for the preparation of intact nuclei. Chromatin was fragmented via sonication and 25 µg was immunoprecipitated with 5 µg of RING1B monoclonal antibody or 5 µg of anti-Flag M2 antibody. Eluted DNA was amplified using qPCR with primers that anneal at multiple sites along the HPV16 genome. The linearized HPV16 genome is displayed on the x-axis based on the 1/7905 bp sequence (NC_001526.2). Peaks represent the midpoint of the PCR amplicon. Data shown are the mean ±SD (% Input) from three technical replicates and are representative of three independent repeats. *p* values (n.s.) were determined using a paired *t* test.

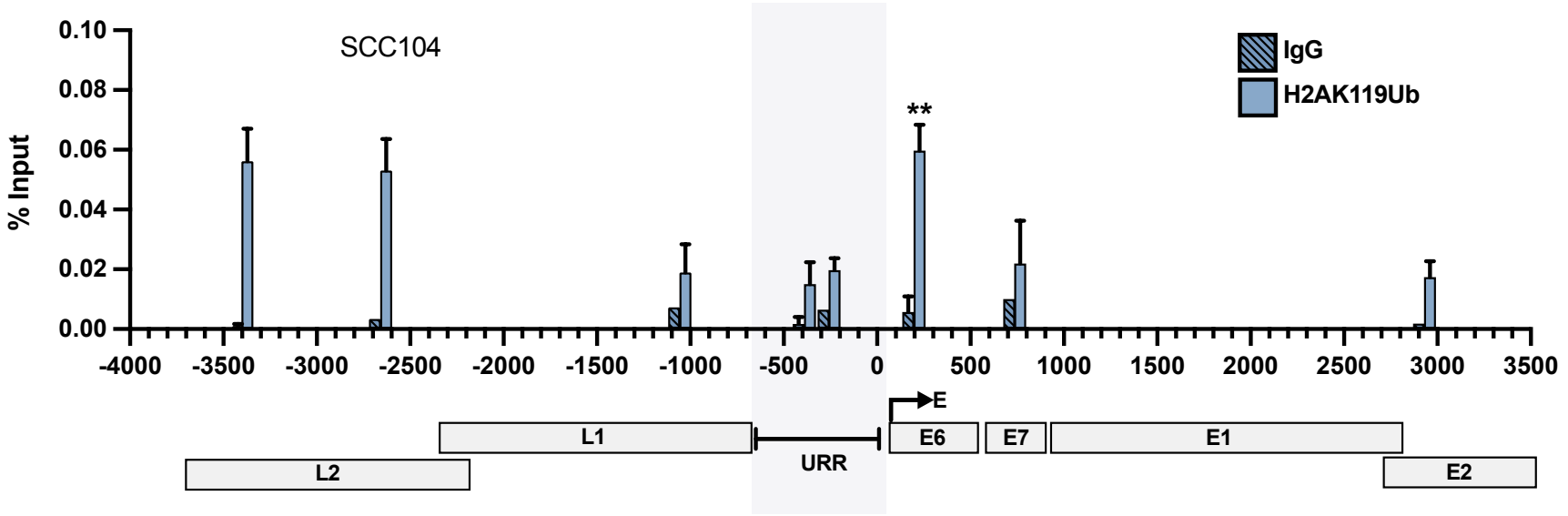


Figure 5.20 Enrichment of H2AK119Ub on the HPV16 genome in SCC104 cell lines. H2AK119Ub association to the HPV16 genome was assessed by ChIP-qPCR in SCC104 cell lines. Cell pellets were fixed in 2% formaldehyde, lysed and Dounce homogenised for the preparation of intact nuclei. Chromatin was fragmented via sonication and 25 µg was immunoprecipitated with 5 µg of monoclonal H2AK119Ub antibody or 5 µg of anti-Flag M2 antibody. Eluted DNA was amplified using qPCR with primers that anneal at multiple sites along the HPV16 genome. The linearized HPV16 genome is displayed on the x-axis based on the 1/7905 bp sequence (NC_001526.2). Peaks represent the midpoint of the PCR amplicon. Data shown are the mean ±SD (% Input) from three technical replicates and are representative of three independent repeats. ** $p < 0.01$ as determined by a paired *t* test.

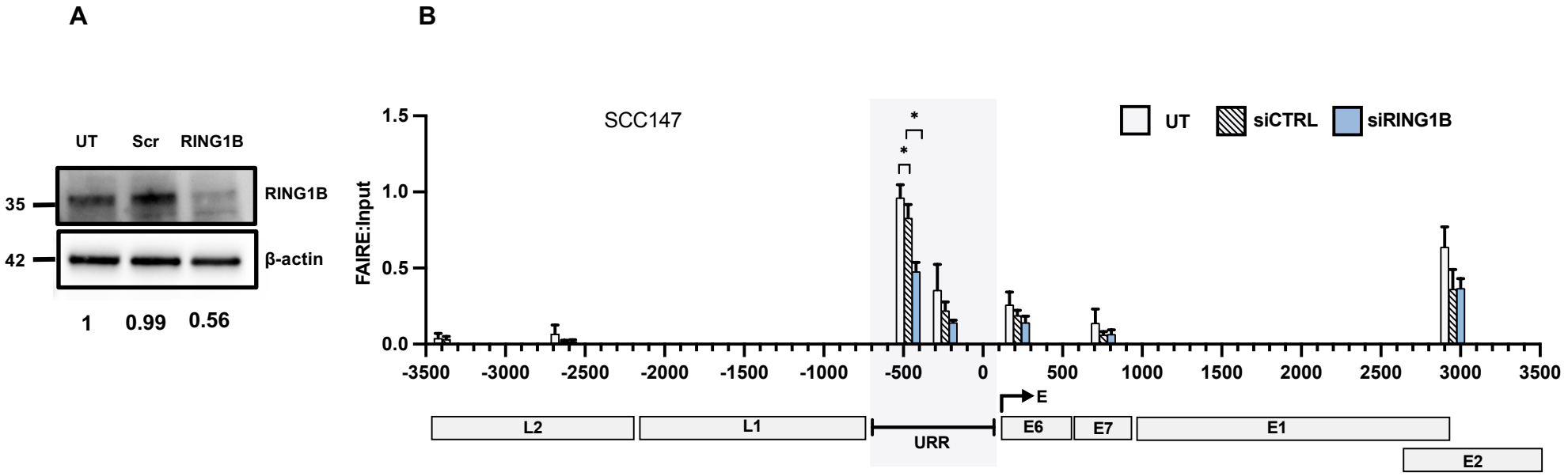


Figure 5.21 Depletion of RING1B in SCC147 results in reduced chromatin accessibility at the URR. (A) HPV16 containing SCC147 cell lines were left untransfected (UT) or transfected with scrambled (Scr) or RING1B-specific siRNA (YY1) at 10 nM for 48 hours and RING1B and β-actin protein expression analysed by western blotting. Relative protein expression normalised to β-actin was assessed using densitometry (n-1). (B) Chromatin topology was then assessed using FAIRE. Cell pellets were fixed in 2% formaldehyde, lysed and Dounce homogenized for the preparation of intact nuclei. Chromatin was fragmented via sonication and samples were split into FAIRE or Input. Crosslinks were reversed in the Input sample and DNA was extracted from both FAIRE and Input using phenol: chloroform: isoamylalcohol (25:24:1). Isolated DNA was amplified using qPCR with primers that anneal at different points on the HPV16 genome. Peaks represent the midpoint of the PCR amplicon. The linearized HPV16 genome is displayed on the x-axis based on the 1/7905 bp sequence (NC_001526.2). Data show the mean ± SEM of FAIRE: Input (2^- ΔCt) from three independent repeats. * $p < 0.05$ as determined by a one-way ANOVA.

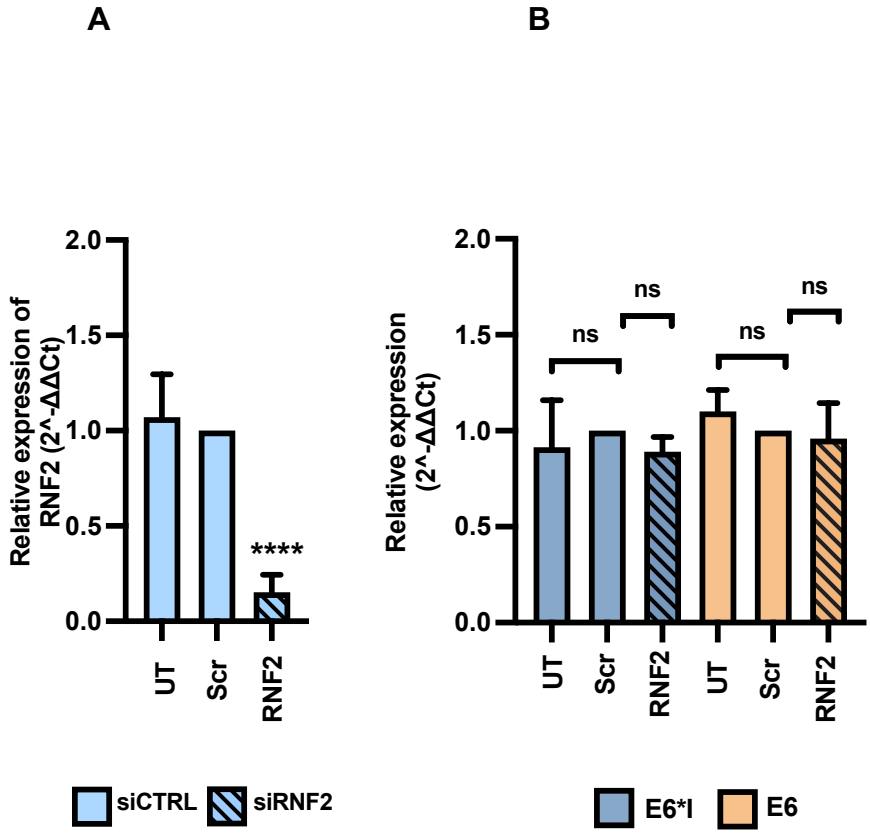


Figure 5.22 Depletion of RNF2 in SCC147 does not alter E6/E7 transcript expression. HPV16 containing SCC147 cell lines were left untransfected (UT) or transfected with scrambled (Scr) or RNF2-specific siRNA (RNF2) at 10 nM for 48 hours and (A) RNF2, (B) E6 unspliced and E6 spliced transcripts were assessed via qRT-PCR. Data shown are the mean ± SEM fold change normalised to β-actin (2^-ΔΔCt) from three independent repeats. **** $p < 0.0001$ as determined by a one-way ANOVA.

5.6 YY1 functions as a transactivator of the HPV16 early promoter to drive high expression of E6/E7 transcripts in OPSCCs

Having demonstrated that enrichment of CTCF and YY1 on the HPV16 genome is consistent with an active transcriptional landscape in HPV16 OPSCC cell lines, we were interested to understand whether these cellular proteins were functioning to directly alter oncogene transcription. To assess this, both OPSCC lines were either left untransfected (UT) or treated with SMARTpools of scrambled (Scr), CTCF-specific or YY1-specific siRNA at a final concentration of 10 nM for 48 hours. Cells were harvested at 48 hours and the knockdown efficiency was assessed using western blotting and qRT-PCR. As shown in **Figure 5.23A** and **Figure 5.24A** 10 nM of siCTCF in SCC147 and SCC104 resulted in a 4-fold and 3-fold reduction in CTCF protein expression, respectively. This was also consistent at the transcript level where a significant reduction in CTCF transcript expression was observed between siScr and siCTCF using three different primer sets to detect CTCF mRNA **(Fig 5.23B, Fig 5.24B).** To explore the effects that CTCF depletion has on full length E6 and E6*I we assessed transcript expression using qRT-PCR. Depletion of CTCF in SCC147 resulted in no significant alteration to E6 or E6*I transcript expression **(Fig 5.23C).** Whilst there did appear to be an increase in E6 and E6*I transcription upon CTCF depletion in SCC104, this did not reach significance **(Fig 5.24C)** suggesting that the ability of CTCF to function as a repressor of HPV oncogene transcription may become lost during HPV-driven cancer. As shown in **Figure 5.25A** and **Figure 5.26A** treatment with siYY1 at 10 nM for 48 hours in SCC147 and SCC104 results in a 30-fold and 5-fold reduction in YY1 protein expression, respectively. Transcript

expression of YY1 was then assessed using qRT-PCR which again showed a significant reduction in YY1 transcript expression between siScr and siYY1 treated cells using three different primer sets to detect YY1 mRNA **(Fig 5.25B, Fig 5.26B)**. Intriguingly, in both tested OPSCC lines, depletion of YY1 resulted in a significant reduction in both full length E6 and E6*I expression (SCC147 $p<0.01$, SCC104 $p<0.01$) **(Fig 5.25C, Fig 5.26C).** When compared to our previous work which found that depletion of YY1 using shRNA in HPV18-HFKs resulted in a significant increase in viral oncogene expression (Pentland et al., 2018) and results listed in earlier chapters highlighting that depletion of YY1 in high-passage 16-HTKs and 18-HTKs results in no significant alteration to E6/E7 transcript expression, it is plausible that there may be a switch from the canonical function of YY1 acting as a transcriptional repressor to an activator in the context of HPV infection and HPV driven cancers. This is also consistent with enrichment of YY1 at the HPV16 early promoter and enrichment of active histone modifications.

Prompted by this, we next used siRNA and FAIRE to understand how YY1 might function to activate HPV transcription within the OPSCCs. To explore this, SCC147 cell lines were either left untransfected (UT) or treated with scrambled (Scr) or YY1-specific siRNA at 10 nM for 48 hours. Any resulting topological changes to the chromatin was subsequently assessed using FAIRE. After 48 hours of siRNA treatment, a small proportion of cells were harvested for protein whilst the remainder were fixed with 2% formaldehyde. They were then Dounce homogenised and sheared using sonication. Chromatin was split into two conditions FAIRE and Input. The input sample underwent an additional step in which crosslinks were reversed and DNA was extracted from both FAIRE and Input using phenol-chloroform.

Chromatin accessibility was quantified using qPCR. As shown in **Figure 5.27A**, treatment with 10 nM of siYY1 results in an 11-fold reduction in YY1 protein expression as determined by western blotting. Analysis by FAIRE showed a distinct peak in the UT and siScr treated samples which is indicative of open chromatin upstream of the URR, within the URR and also within the E2 open reading frame **(Fig 5.27B).** Following depletion of YY1 we show a non-significant reduction in open chromatin between the UT and control treated cells, and a significant reduction in open chromatin between the UT and siYY1 treated cells at the viral enhancer (p<0.05) **(Fig 5.27B)**. Just upstream from the viral enhancer, within the L1 open reading frame, we also show a significant reduction in open chromatin between the UT and siYY1 treated cells and also between the siScr and siYY1 treated cells (p<0.05, p<0.01) **(Fig 5.27B)**. Taken together, this data highlights that YY1 functions to create a transcriptionally active chromatin landscape in OPSCCs.

5.7 Conclusions

During the HPV life cycle early gene transcription is tightly regulated through a number of mechanisms. This includes the formation of an epigenetically repressed, long-range chromatin loop between CTCF bound at the E2 open reading frame and YY1 bound at the viral enhancer. This mechanism contributes to the temporal regulation of HPV oncogenes throughout the viral life cycle to prevent constitutively high expression of E6/E7 and therefore reduce the likelihood of host immune clearance. With constitutively high expression of HPV oncogenes being the driver of HPV-driven cancer we hypothesised that CTCF-YY1-mediated transcription regulation would become deregulated during HPV-driven cancers. Using HPV16

containing OPSCC cell lines we explored whether CTCF and/or YY1 are deregulated during HPV-driven cancers to uncouple oncogene transcription regulation from the HPV life cycle. Using ChIP-qPCR we show a robust peak of enrichment of CTCF at the conserved binding site within the E2 open reading frame and a robust peak of enrichment of YY1 at the viral enhancer, 5'URR and early promoter. Importantly, binding of both host cell proteins was consistent with a transcriptionally active landscape. Again using ChIP-qPCR we show significant enrichment of H3K4Me3, H3K27Ac and recruitment of RNA PolII to the HPV16 early promoter. This was despite enrichment of YY1 which ordinarily functions to recruit PcG proteins that are writers of transcriptionally repressive histone modifications.

We hypothesised that loss of PcG recruitment was likely the mechanism involved in creating a transcriptionally active landscape, despite high YY1 binding. However, we show that within the OPSCC cell lines there is robust enrichment of both RING1B and H2AK119Ub – the repressive chromatin modification it lays down. The reason for strong enrichment of RING1B is still unclear. However, we continued to explore the function of both CTCF and YY1 in driving HPV-driven cancers by depleting both YY1 and CTCF using siRNA. Importantly, we found that whilst depletion of CTCF had no impact on E6/E7 transcription in these cells, the depletion of YY1 resulted in significant reduction of E6/E7 oncogene transcription. Using FAIRE, we show that this is a result of YY1-mediated chromatin alteration to favour a transcriptionally active landscape. Taken together, we show that in HPV16-driven cancers of the oropharynx, the canonical function of YY1 is altered. Instead, YY1 functions as a transactivator of the HPV16 early promoter to drive high E6/E7 oncogene transcription.

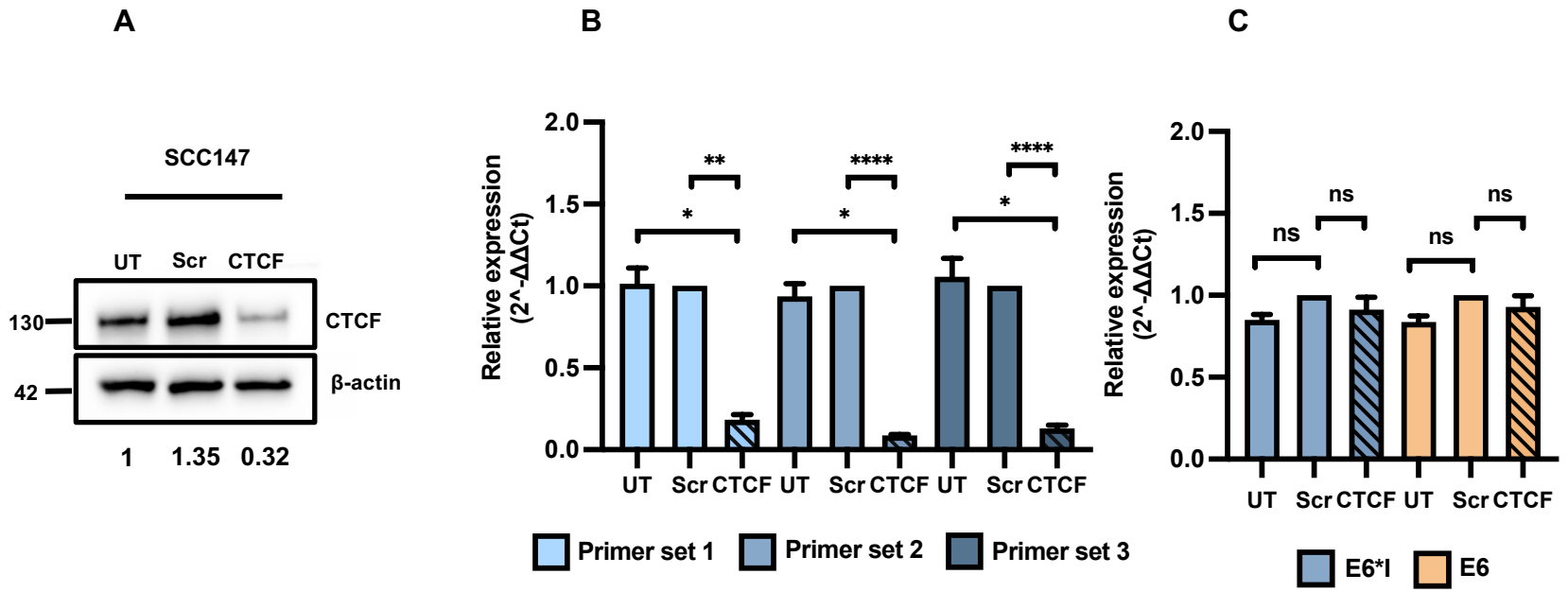


Figure 5.23 Depletion of CTCF in SCC147 results in unchanged E6/E7 transcript expression. (A) HPV16 containing SCC147 cell lines were left untransfected (UT) or transfected with scrambled (Scr) or CTCF-specific siRNA (CTCF) at 10 nM for 48 hours and CTCF and β-actin protein expression analysed by western blotting. Relative protein expression normalised to β-actin was assessed using densitometry. (B) CTCF, (C) E6 unspliced and E6 spliced transcripts were assessed via qRT-PCR. Data shown are the mean ± SEM fold change normalised to β-actin (2^-ΔΔCt) from three independent repeats. * $p < 0.05$, ** $p < 0.01$, **** $p < 0.0001$ as determined by a one-way ANOVA.

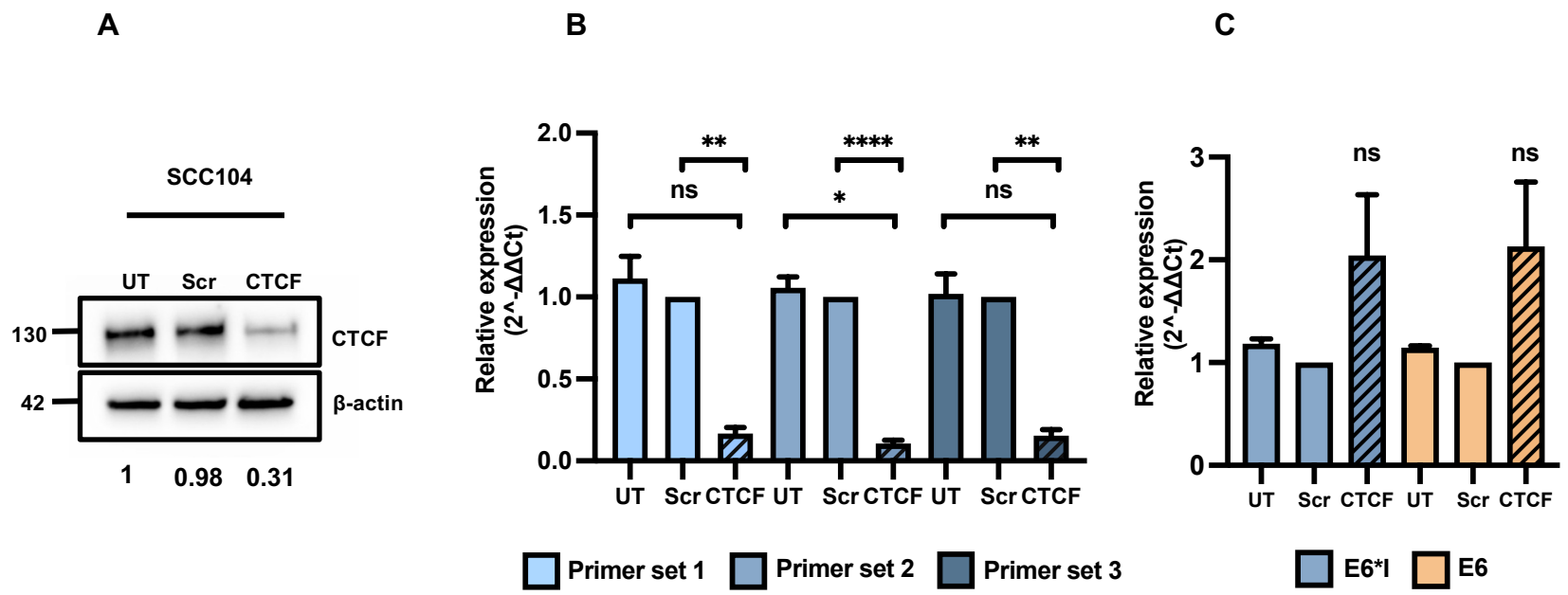


Figure 5.24 Depletion of CTCF in SCC104 results in a non-significant increase in E6/E7 transcript expression. (A) HPV16 containing SCC104 cell lines were left untransfected (UT) or transfected with scrambled (Scr) or CTCF-specific siRNA (CTCF) at 10 nM for 48 hours and CTCF and β-actin protein expression analysed by western blotting. Relative protein expression normalised to β-actin was assessed using densitometry. (B) CTCF, (C) E6 unspliced and E6 spliced transcripts were assessed via qRT-PCR. Data shown are the mean ± SEM fold change normalised to β-actin (2^-ΔΔCt) from three independent repeats. * $p < 0.05$, ** $p < 0.01$, **** $p < 0.0001$ as determined by a one-way ANOVA.

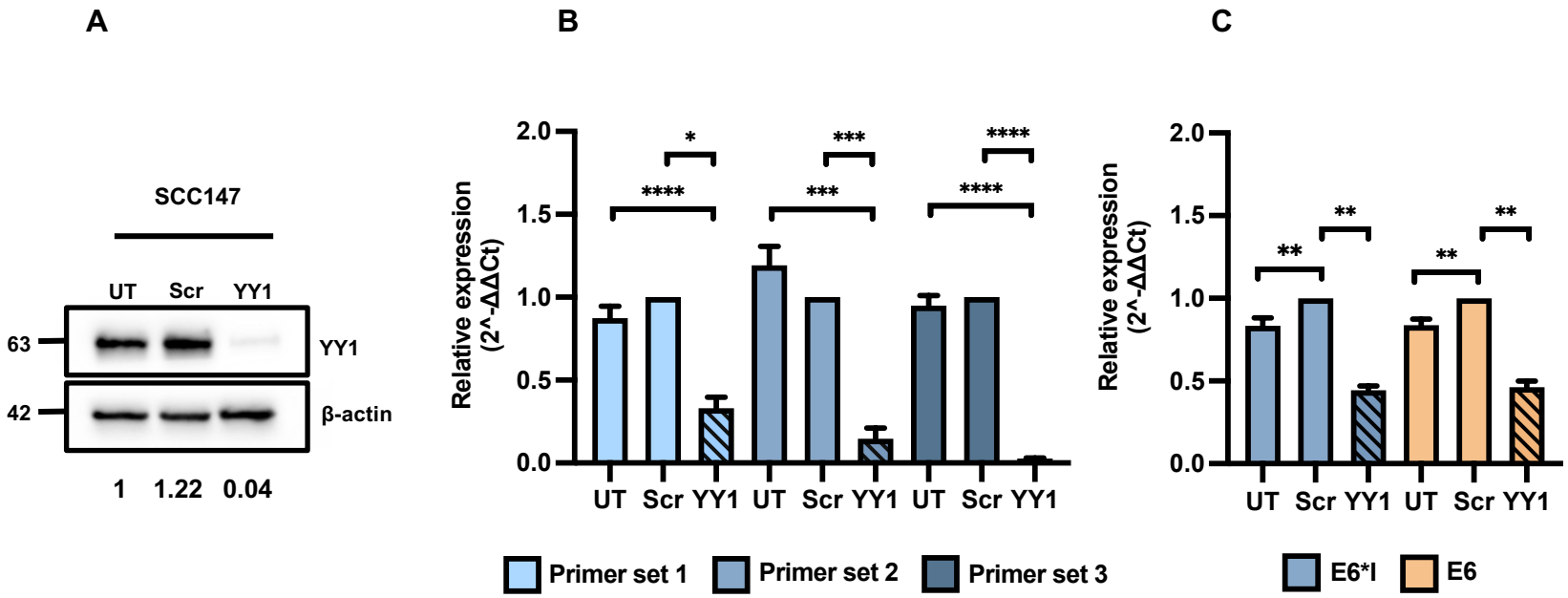


Figure 5.25 Depletion of YY1 in SCC147 results in significantly reduced E6/E7 transcript expression. (A) HPV16 containing SCC147 cell lines were left untransfected (UT) or transfected with scrambled (Scr) or YY1-specific siRNA (YY1) at 10 nM for 48 hours and YY1 and β-actin protein expression analysed by western blotting. Relative protein expression normalised to β-actin was assessed using densitometry. (B) YY1, (C) E6 unspliced and E6 spliced transcripts were assessed via qRT-PCR. Data shown are the mean ± SEM fold change normalised to β-actin (2^-ΔΔCt) from three independent repeats. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$ as determined by a one-way ANOVA.

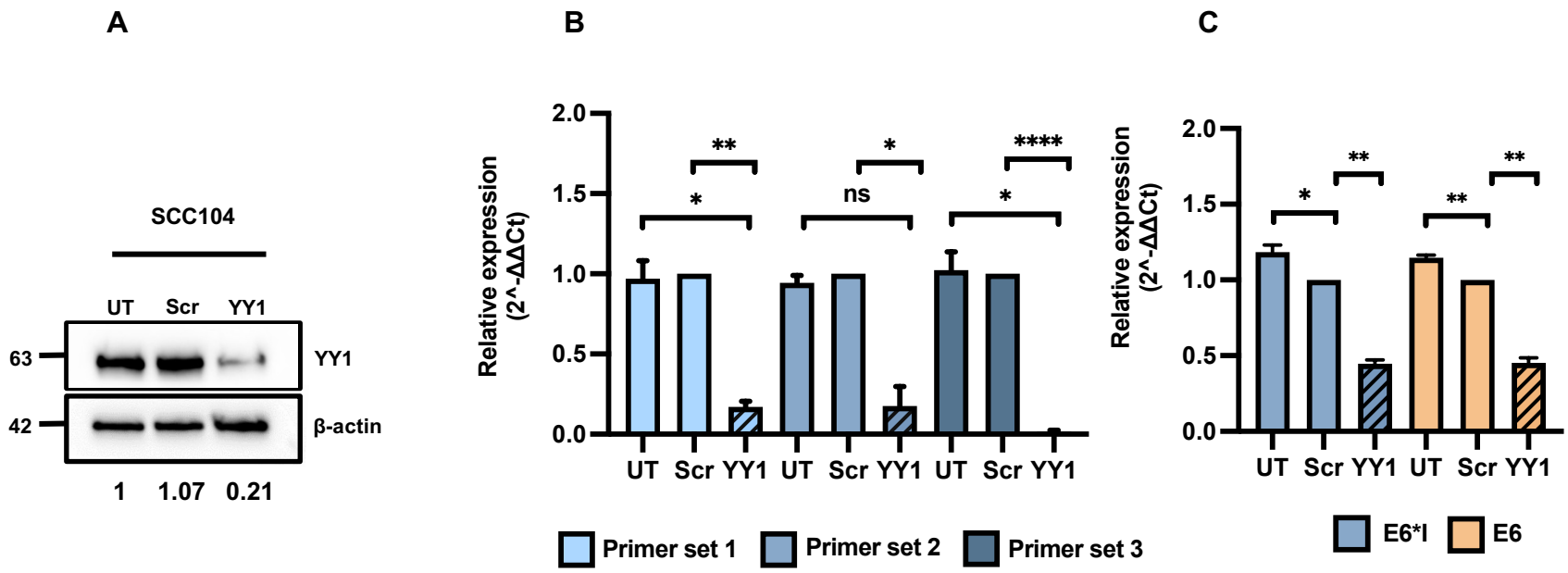


Figure 5.26 Depletion of YY1 in SCC104 results in significantly reduced E6/E7 transcript expression. (A) HPV16 containing SCC104 cell lines were left untransfected (UT) or transfected with scrambled (Scr) or YY1-specific siRNA (YY1) at 10 nM for 48 hours and YY1 and β-actin protein expression analysed by western blotting. Relative protein expression normalised to β-actin was assessed using densitometry. (B) YY1, (C) E6 unspliced and E6 spliced transcripts were assessed via qRT-PCR. Data shown are the mean ± SEM fold change normalised to β-actin (2^-ΔΔCt) from three independent repeats. * $p < 0.05$, ** $p < 0.01$, **** $p < 0.0001$ as determined by a one-way ANOVA.

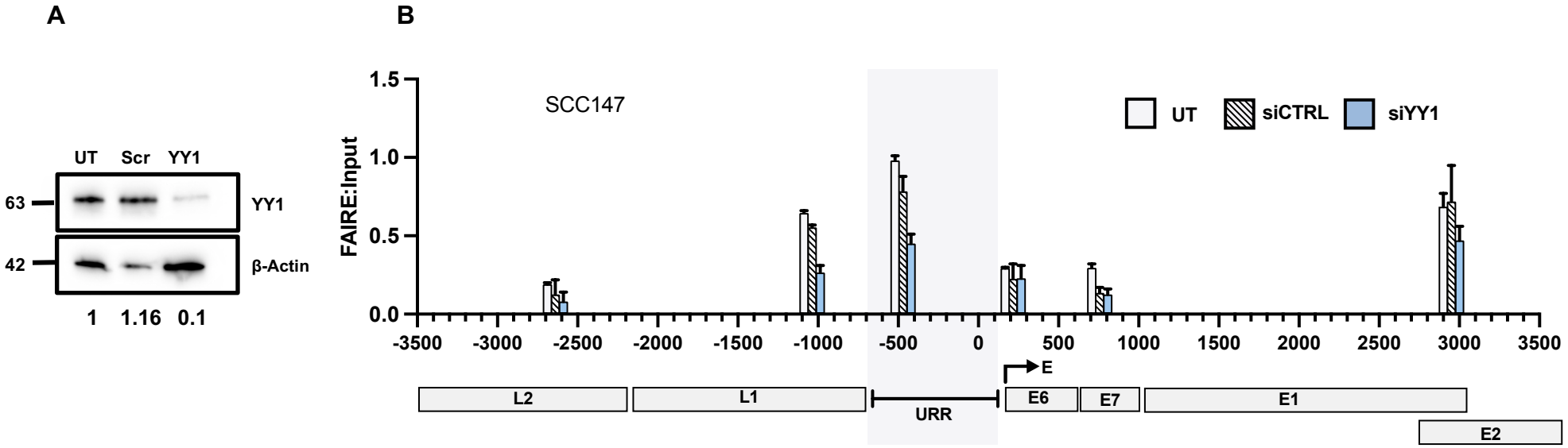


Figure 5.27 Depletion of YY1 in SCC147 cell lines results in reduced chromatin accessibility at the URR. (A) HPV16 containing SCC147 cell lines were left untransfected (UT) or transfected with scrambled (Scr) or YY1-specific siRNA (YY1) at 10 nM for 48 hours and YY1 and β-actin protein expression analysed by western blotting. Relative protein expression normalised to β-actin was assessed using densitometry. Data shown are a representative image from two independent repeats. (B) Chromatin topology was then assessed using FAIRE. Cell pellets were fixed in 2% formaldehyde, lysed and Dounce homogenized for the preparation of intact nuclei. Chromatin was fragmented via sonication and samples were split into FAIRE or Input. Crosslinks were reversed in the Input sample and DNA was extracted from both FAIRE and Input using phenol: chloroform: isoamylalcohol (25:24:1). Isolated DNA was amplified using qPCR with primers that anneal at different points on the HPV16 genome. Peaks represent the midpoint of the PCR amplicon. The linearized HPV16 genome is displayed on the x-axis based on the 1/7905 bp sequence (NC_001526.2). Data show the mean ± SEM of FAIRE: Input (2^- ΔCt) from two independent repeats.

Chapter 6 Discussion

6.1 Modelling HPV-driven disease progression

It is well established that infection with oncogenic HPV is a risk factor for development of both cervical and oropharyngeal cancer. This is as a result of expression of high-risk oncoproteins E6 and E7 which target multiple host proteins and disrupt the normal process of cell proliferation. The process of transformation is as a result of a combination of factors, namely genetic instability, overexpression of oncogenes and aberrant cell growth. The development of cervical cancer is a process which takes many years and there is a well-established staging within the cervix from infection with HPV, progressing through to CIN and eventually cervical cancer. This process is less well characterised within the oropharynx and identification of dysplasia at this site is rare (Roberts et al., 2019). This highlights the importance of establishing a model that would characterise the changes that occur following a persistent infection with oncogenic HPV.

There are many available cell lines for studying HPV-driven cancer in the head and neck. These are established from patients with primary or recurrent oropharyngeal squamous cell carcinoma containing naturally occurring HPV16 and rarely HPV18. Establishment of OPSCCs is essential for understanding the changes that occur as a result of persistent HPV infection. Whilst this is useful, there are currently no available cell lines that have been established from premalignant lesions isolated from the oropharynx. Given that cell lines established from premalignant cervical lesions, such as W12 or CIN612, have provided insight into the process of cervical

cancer progression, we were interested in developing a model which would enable us to pinpoint the exact events which take place during the very initial stages of disease in the oropharynx and throughout the natural history of an HPV infection. Given that the use of our primary keratinocytes is a physiologically relevant system to model the HPV life cycle, we opted to use this system to also model HPV-disease progression in the tonsil.

In low-passage primary tonsil keratinocytes, the expression of E6 and E7 oncoproteins extends the lifespan of the cell by targeting pRb and p53. Because this is also essential in HPV-transformed cancer cell lines, we hypothesised that continued growth of the HTKs would recapitulate the cellular events which occur in vivo. In order to validate this model we assessed some of virus induced changes that we would expect to find in HPV-driven cancer. We wanted to ensure that we were able to model the selection pressures that exist in vivo including integration and constitutively expressed E6/E7. Because HPV16 is so prevalent in the oropharynx, we were also interested to understand whether there were any differences in how each virus behaves that might contribute to higher pathogenesis. During a persistent infection with oncogenic HPV increased transcription of E6 unspliced and spliced transcripts results in high expression of E6 and E7 oncoproteins. We show this to be consistent in our high-passage 16- and 18-HTKs. Long term culture results in increased transcription and protein expression of both oncogenes. We found that this was coincident with reduced doubling time which again we would expect to see in a premalignant lesion.

As an HPV infection persists, the population of keratinocytes that harbour episomal HPV genomes are reduced as integration of viral DNA into host cell chromatin increases. This inhibits completion of the HPV life cycle due to disruption of the circular genome and often results in constitutively high expression of E6 and E7. As such, during the progression of cervical cancer, as dysplasia becomes more severe, this is often correlated with high levels of viral integration. However, in the oropharynx, integration is not always as prevalent. We were therefore interested to assess the genome status of the high-passage 16- and 18-HTKs. In doing so, we found that prolonged culture of HPV18 containing HTKs results in viral integration, whilst prolonged culture of HPV16 containing HTKs does not. The reason for the difference in integration status between virus type (within the same tonsil donor) is unclear. Different HPV types often display different frequencies of viral integration. Integration of viral genomes within cervical carcinoma lesions has previously been assessed across the five most frequent HPV types (HPV16, HPV18, HPV31, HPV33 and HPV45) using the APOT assay. This demonstrated clear differences in integration frequencies with high levels of integration found in HPV16, HPV18 and HPV45 and significantly lower frequencies of integrated viral genomes found in lesions induced by HPV31 and HPV33 (Vinokurova et al., 2008). The pattern of integration that we show is also consistent with the integration events that exist in vivo, in which HPV18 integration is more common compared to HPV16 (Pirami et al., 1997, Lagström et al., 2021, Roberts et al., 2019). Despite this, the exact reason for the differences in integration frequencies between different virus types still remains unknown and has not previously been explored. It has been shown that increased DNA damage can increase the likelihood of viral integration – an

accumulation of double strand breaks, induced by the loss of the helicase and mediator of non-homologous end joining protein, ku70, has been shown to be associated with the loss of HPV16 episomes and initiation of viral integration into cervical keratinocytes (Winder et al., 2007). Additionally, proton-induced DNA damage has been shown to promote integration of HPV DNA into host cell genomes (Katerji et al., 2022). Infection with high-risk HPV often results in genomic instability as a result of E6/E7 expression. There is evidence to suggest that this process precedes viral integration and as such, integration occurs as a consequence of chromosomal instability (Melsheimer et al., 2004, Vinokurova et al., 2008, Labarge et al., 2022). Because we can transfect the primary HTKs with episomal HPV DNA and assess the genome status over time this provides a useful tool to compare differences in the events which occur pre-integration to try and understand why one type integrates more frequently. It would be interesting to assess the role of DNA damage on HPV integration and compare whether HPV16 or HPV18 might display different integration frequencies. We can compare the differences in integration frequencies by co-transfecting HTKs with both HPV16 and HPV18 plasmids. To assess the role of DNA damage we can induce breaks in the DNA through methods such as ionizing-radiation or the use of DNA damaging agents such as Etoposide and then by using Southern blotting we can then compare at different passages whether the induction of DNA damage results in integration and to compare the frequencies of each virus. Given that HPV16 did not integrate into the host genome in the primary HTK mode it is plausible that this virus displays differences in the levels of DNA damage or chromosomal instability, at least within the early stages of infection, when compared to HPV18. Using our primary HTKs we could explore this

possibility by transfecting the cells with either HPV16 or HPV18 and tracking over passages any changes that occur before viral genomes have integrated. Comparing the levels of DNA damage markers such as yH2AX for example or measuring chromosomal instability using techniques such as single cell whole genome sequencing would enable us to pinpoint any changes that might precede integration. It has previously been shown that keratinocytes which harboured integrated HPV16 DNA had a growth advantage over the cells containing non-integrated forms of the viral genome. However, despite this, these cells were still able to differentiate and could grow in agar, suggesting that they were not transformed (Jeon et al., 1995). It would therefore be interesting to elucidate whether the high-passage HTKs would be able to successfully differentiate when grown as three-dimensional organotypic raft cultures.

6.2 Upregulation of CTCF and YY1 protein expression in HPV-driven disease

We have previously shown that establishment of HPV18 containing human foreskin keratinocytes results in the significant upregulation of CTCF protein expression (Pentland et al., 2018). Using human tonsil keratinocytes we show that this is also true in the context of the oropharynx – transfection of primary HTKs with HPV16 and HPV18 results in increased CTCF protein expression. We have now shown that establishment of HTKs with HPV16/18 also results in upregulation of YY1 protein. Given that in the normal viral life cycle, CTCF and YY1 function to repress E6/E7 transcription, we hypothesised that prolonged culture of the HPV16/18 HTKs would result in loss of expression of both proteins so to deregulate repression of oncogene

transcription. However, we found that following the prolonged culture of HPV16 and HPV18 containing HTKs there is a further 2-fold increase in both CTCF and YY1 protein expression.

Previously, we have analysed the abundance of both CTCF and YY1 within HPV-negative and HPV-positive tonsil tumour tissues using immunohistochemistry. We identified that there is a significant increase in both CTCF and YY1 in the HPV-driven samples compared to the HPV-negative samples (data not published). During the current study, we were unable to corroborate these findings in vitro. Using HPV-negative and HPV-positive OPSCC cell lines we show that there is no significant increase in either CTCF or YY1 protein expression in HPV+ OPSCCs. However, given that both cell lines are derived from oropharyngeal carcinomas, it is also possible that the expression of CTCF and YY1 has become deregulated within the HPV- cells in a mechanism independent of HPV establishment. It would be interesting to study this further and explore whether there are any changes in CTCF and/or YY1 expression using normal tonsil tissue and tonsil tumour sections as a comparison. This could be done using immunohistochemistry and would indicate whether either proteins likely have a role during HPV-driven cancers.

Although upregulation of CTCF and YY1 was not consistent with our initial hypothesis, this phenomenon is supported by the literature. In multiple HPV-driven cancers there is evidence of overexpression and/or mutation to both CTCF and YY1. Analysis of oropharyngeal squamous cell carcinoma samples revealed the mutation profile of HPV-driven OPSCCs. The authors identified that 33% of these mutations were within genes that are associated with chromatin regulation. They used

mutational analysis of HPV-driven OPSCCs available from the cancer genome cancer genome atlas (TCGA) and a separate analysis using whole exome sequencing of HPV-positive OPSCCs that were matched to normal tissue controls. Chromatin modifiers that were altered included the CREB-binding protein and E1A-associated protein. One of the identified chromatin regulators that was found to be mutated in this analysis was CTCF and the authors showed that CTCF displayed a missense mutation (Haft et al., 2019). This mutation resulted in a glutamic acid to lysine mutation (E432K) – whilst the protein function of CTCF was not directly tested, this mutation falls within the seventh zinc-finger of CTCF raising the possibility that it could alter protein-DNA interactions although this is speculative.

A systematic pan-cancer analysis combined both the cancer genome atlas (TCGA) and gene expression omnibus (GEO) databases to explore aberrations to YY1 expression. This was combined with analysis of both YY1 RNA and protein expression in multiple cancer types. This revealed that YY1 is often highly expressed in multiple cancers and this is closely associated with poor patient prognosis (Fu et al., 2022). Expression of YY1 in head and neck carcinoma and cervical epithelial squamous carcinoma was significantly higher than in corresponding control tissues. Analysis of human tissues revealed high expression of YY1 in head and neck tumours (Fu et al., 2022). Within cervical intraepithelial neoplasia tissues the expression of YY1, E6 and E-cadherin was assessed using immunohistochemistry. YY1 was found to be significantly upregulated within cervical squamous cell carcinoma tissues when compared to both normal cervical tissue and CIN1 (Wang et al., 2018). Correlational analysis revealed that YY1 was negatively associated with the expression of E-cadherin (Wang et al., 2018).

Interestingly, this also revealed an association between the expression of YY1 and HPV16 E6, supporting the data presented in the current study whereby YY1 can act to promote the expression of E6 during HPV-driven cancer (Wang et al., 2018). Other studies have used the cancer genome atlas to assess the effects of YY1 mRNA expression as a prognostic marker for head and neck carcinoma. Increased expression of YY1 mRNA in combination with transcription factor CP2 (TFCP2) was a prognostic marker for worse overall survival (Schnoell et al., 2021).

Intriguingly, using the primary tonsil keratinocytes, we have found that the upregulation of both YY1 and CTCF appears to be independent of any increase in transcript expression, highlighting a possible post-transcriptional modification. This also corroborates our previous work in which CTCF-upregulation was independent of changes to CTCF-encoding transcripts. It could be that HPV alters the proteosome in such a way to result in increased stabilisation of either proteins. YY1 can be regulated by the calcium activated protease Calpain and conversely can be stabilised by the use of Calpain inhibitors or proteosome inhibitors such as MG132 (Walowitz et al., 1998). It is also possible that the levels of YY1 protein are being altered by microRNAs. Ordinarily, microRNAs target the untranslated region of YY1, resulting in degradation of YY1 mRNA. In HPV infected cells, HPV E7 targets microRNAs for degradation, as such, YY1 expression is increased (Liu et al., 2021).

6.3 Deregulation of CTCF-YY1-mediated HPV oncogene transcription regulation in a primary human tonsil keratinocyte disease progression model

In our previous work we mutated three amino acids within the HPV18 E2 open reading frame resulting in abrogated binding of CTCF at its conserved binding site. This led us to discover the role that CTCF plays in regulating E6/E7 transcription. Abrogation of CTCF binding resulted in significantly increased activity from the HPV18 early promoter, p102. This was as a result of enhanced recruitment of RNA PolII and led to significantly increased E6/E7 oncogene transcription (Pentland et al., 2018). This prompted us to explore the idea that CTCF-mediated repression of oncogene transcription might become deregulated during HPV-driven cancer to drive high oncogene expression.

We analysed changes to CTCF binding in our disease progression model using ChIP-qPCR. To assess whether there were any virus specific differences in CTCF binding we used both HPV16 and HPV18 containing human tonsil keratinocytes. Although we did not assess CTCF-mediated transcription regulation of the HPV16 promoter in our previous work, it has been shown that the HPV16 genome contains a conserved binding site for CTCF (Paris et al., 2015). When bound to the HPV16 episome, CTCF can function to regulate E6 transcript expression (Groves et al., 2021). Following the long-term culture of the HTKs, we found that CTCF enrichment becomes significantly abrogated. Interestingly, despite the difference in virus type, the pattern of CTCF enrichment was consistent in both HPV16 and HPV18.

In our HPV18 model, loss of CTCF binding is likely linked to viral genome integration. Often, HPV integration results in disruption to the both the E1 and E2

open reading frame. In high-passage 18-HTKs a loss of episomal DNA indicates that the viral genome has integrated into host cell chromatin (Roberts et al., 2019). It was therefore important to know whether this had disrupted the CTCF binding site within the E2 open reading frame. To assess this we extracted genomic DNA and used qPCR to analyse the abundance of the E2 open reading frame within these cells – specifically, the region that contains the conserved CTCF binding site. We found that in the high-passage HPV18 containing keratinocytes the abundance of the E2 open reading frame was significantly reduced, likely explaining the loss of CTCF binding in these cells. This significant abrogation to CTCF binding was consistent with high E6 and E7 oncogene expression within the high-passage HPV18 HTKs. Because integration of viral genomes is common in HPV-driven cancers, this might highlight an alternative mechanism for loss of oncogene regulation. Ordinarily, in integrant driven cancers, disruption to the E1 and E2 open reading frame results in loss of E2-mediated repression of viral transcription. However, it is also plausible that loss of CTCF-mediated regulation of HPV oncogenes, as a result of integration, might also be an important driver of carcinogenesis.

Within our HPV16 progression model, the viral genomes remain extrachromosomal. Therefore, these cells presumably have no disruption to the E2 open reading frame – although this is something that would be important to confirm using qPCR and western blotting. This considered, it is likely that there are also other factors involved in regulating CTCF binding to the HPV episome during disease progression. Aberrant DNA methylation is common in HPV-driven cancers. Importantly, increasing DNA methylation during cancer can disrupt CTCF binding and directly

modulate CTCF mediated gene silencing (Lai et al., 2010, Wang et al., 2012). Based on this, it would be interesting to assess whether the low-passage and high-passage HTKs display differential patterns of DNA methylation and importantly whether DNA demethylation would reverse the phenotype of abrogated CTCF binding. One way to assess this would be through bisulphate genomic sequencing – a technique that enables identification of 5-methylcytosine (5mC) at a base pair resolution (Li and Tollefsbol, 2011). Sequencing of the low-passage and high-passage HPV16 and HPV18 HTKs would enable us to identify whether DNA methylation is significantly increased during HPV-driven disease progression, specifically at sites of or near CTCF binding. The use of a demethylating agent such as 5-aza-2′-deoxycytidine (DAC) would also indicate whether methylation is involved in regulating CTCF binding. Interestingly, the use of DAC treatment has been found to result in the significant downregulation of E6/E7 oncogene expression, resulting in reduced proliferation and colony formation capabilities, in HPV-driven cervical and oropharyngeal squamous cell carcinoma lines (Stich et al., 2017). In combination with bisulphate sequencing, it would also be interesting to treat the low-passage and high-passage HTKs with DAC to assess whether we can reverse the loss of CTCF binding within these cells.

CTCF-mediated repression of HPV oncogene transcription is dependent on a chromatin loop formed with transcription factor YY1 bound at the viral enhancer. YY1 functions to recruit PcG proteins that mediate the deposition of repressive histone modifications. Following the finding that CTCF binding is significantly reduced with prolonged culture of HPV16 and HPV18 HTKs, we also assessed whether the pattern of YY1 binding was consistent. Using ChIP-qPCR we show that

in high-passage HPV18 containing HTKs there is strong enrichment of YY1 binding at the viral URR. The HPV18 genome contains seven individual binding sites for YY1 – several of these span the enhancer, the 5'URR and proximal promoter. The ability of YY1 to regulate HPV transcription is very intricately linked to binding to specific sites in the genome. We and several others have shown that YY1 functions to repress transcription of E6 and E7 transcription from both the HPV16 early promoter (p97) and the HPV18 early promoter (p102) (Bauknecht et al., 1992, O'Connor et al., 1996, May et al., 1994, Lace et al., 2009b, Pentland et al., 2018). It was therefore somewhat surprising to identify such significant enrichment of YY1 binding given the high expression of E6 and E7.

Given that YY1 functions to recruit chromatin modifying proteins to the viral genome, we assessed whether enrichment of YY1 at the HPV18 URR would result in the formation of condensed chromatin to repress transcription. In HPV18, using FAIRE, we found that enrichment of YY1 is consistent with an open chromatin landscape that would favour transcription. This suggests that the canonical function of YY1 as a repressor of HPV transcription has been altered following long-term growth of HPV18-HTKs. Whilst YY1 has been shown to predominantly act as a repressor of transcription from both the HPV16 and HPV18 early promoters, there is evidence to show a switch in function from repressor to transactivator. The mechanism behind the switch in function of YY1 from repressor to activator is unclear. One key function of YY1 is to initiate long-distance DNA interactions – in addition to interacting with other architectural proteins such as CTCF, YY1 is also able to form long-range interactions with itself between enhancers and promoters (Weintraub et al., 2017). Having shown that CTCF binding is significantly abrogated, at least in our high-

passage HPV18 model, it is plausible that this results in a rearrangement in chromatin architecture in such a way that YY1 instead forms a loop with itself. This could either be a viral DNA to viral DNA interaction or host DNA to viral DNA interaction and might represent a scenario where YY1 can initiate expression of certain genes by facilitating their enhancer-promoter contacts, which would not have been possible whilst bound to CTCF. YY1 is able to form both dimers and oligomers. Interestingly, dimerization of YY1 has been shown to be important in enhancing the process of DNA loop formation (Li et al., 2021, Weintraub et al., 2017). It is therefore plausible that increased enrichment of YY1 to the HPV18 genome might be a result of structural changes to the protein that enhance binding affinity and alter chromatin looping capabilities. In the HPV18 URR there is a region termed the switch region which is located upstream of the proximal promoter. This region functions to facilitate YY1-mediated activation of the HPV18 promoter within HeLa cells – if this site is abrogated then YY1 instead acts as a repressor of the HPV18 URR (Bauknecht et al., 1995). Protein-protein interactions between YY1 and co-factors are important for its function. This includes interactions with co-repressor proteins, mediated through the transcriptional repressor domain of YY1, or through recruitment of polycomb proteins, mediated through the REPO domain of YY1 (Wilkinson et al., 2006). The function of YY1 can also be heavily altered by different PTMs including phosphorylation, acetylation, and methylation. Phosphorylation of YY1 on serine 184 by Aurora kinase inhibits binding of YY1 to DNA (Kassardjian et al., 2012). Acetylation by p300 is required for the repressive activity of YY1 (Yao et al., 2001). It is therefore possible that post-translational modifications hinder YY1's function as a transcriptional repressor and initiate the switch to transcriptional

activator. In an attempt to understand if and how YY1 is modified we aim to use Mass Spectrometry comparing the covalent modifications present in the low-passage and high-passage 18-HTKs.

In HPV16 we were unable to confirm whether patterns of YY1 binding were altered following long-term culture of these cells. This is something that will be important to assess. Intriguingly, we also found that there did not appear to be any increase in chromatin accessibility at the viral URR. This was inconsistent with the high E6/E7 expression and increased growth rate in the high-passage 16-HTKs. This might represent a situation in which only a subset of clones are driving high oncogene expression. We would therefore would not detect any change in chromatin accessibility unless looking at the single cell level. Whilst we have shown how CTCF-YY1-mediated HPV transcription repression becomes deregulated during HPV-driven disease, it is also important to understand whether these proteins function to regulate host gene expression. Given the lack of established cell lines derived from premalignant oropharyngeal lesions, deciphering host-cell gene expression changes that occur as a result of persistent HPV infection in the tonsil will also be of benefit. It would be interesting to use ATAC-seq on the low-passage and high-passage HPV16 and HPV18 HTKs. This would enable us to look at changes in chromatin accessibility across the human genome, in particular at enhancer regions, in order to assess whether long-term culture of HTKs results in changes to host cell gene regulation.

6.4 Loss of CTCF-mediated repression of HPV E6 and E7 transcription in HPV-driven OPSCC cell lines

We opted to explore the function of CTCF during HPV16-driven cancer using patient derived oropharyngeal squamous cell carcinoma lines. We used ChIP-qPCR to assess CTCF binding to the HPV16 genome within SCC147 and SCC104. Interestingly, this showed slightly varying results to what we had shown in our primary disease progression model. At the E2 open reading frame in the OPSCCs there were significant peaks of enrichment of CTCF binding. Interestingly, another group has recently corroborated this finding. They found that in a number of HPV-driven cancer cell lines, including SCC147 and SiHa, that CTCF remains bound to the viral genome following integration (Karimzadeh et al., 2022). It is not clear why we see loss of enrichment of CTCF binding in our primary progression model and yet significant enrichment of binding in the OPSCCs. Although it is important to note that the high-passage HTKs are not transformed and therefore only give an insight into the early changes that occur following persistent HPV infection. It is possible that within HPV-driven oropharyngeal carcinomas, CTCF no longer functions to repress E6/E7 transcription. This idea is supported by the siRNA treatment of CTCF in the OPSCC cell lines in which depletion of CTCF results in no significant change to E6/E7 expression.

The mechanism behind the loss of CTCF-mediated transcription repression is unclear. As previously described, chromatin modifiers such as CTCF are mutated during HPV-transformed OPSCCs (Haft et al., 2019). It is also possible that whilst CTCF does not function to regulate viral transcription in the OPSCCs, it might

instead function to alter host transcription to favour carcinogenesis. In SCC147, recruitment of CTCF to integration sites was found to correlate with increased expression of viral-host chimera and chromatin accessibility (Karimzadeh et al., 2022). It would be interesting to perform ATAC-seq on these cells to see if CTCF binding functions to regulate transcription of host genes that might function to drive oncogenesis. It is also important to note the caveat that the OPSCC cell lines both contain integrated HPV DNA – we know that SCC147 contains multiple different integration events which are most likely a mix of transcriptionally active and transcriptionally inactive. This is something that should be considered when analysing ChIP data. It might be possible to assess this using ChIP followed by long read sequencing to look at enrichment of specific virus-host junctions. Alternatively, by using a combination of immunofluorescence and fluorescent in situ hybridisation it may be possible to elucidate whether CTCF and YY1 are localised at sites of integration.

6.5 YY1 functions as a transactivator of the HPV16 p97 promoter in HPV-driven OPSCC cell lines

Assessment of YY1 binding in OPSCC cell lines revealed robust enrichment of YY1 binding at the enhancer, 5'URR and the early promoter. Because YY1 ordinarily functions to recruit chromatin modifiers to sites of DNA, the epigenetic landscape of SCC147 and SCC104 was subsequently analysed using ChIP-qPCR. We initially assessed enrichment of active histone modifications given the high expression of E6 and E6*I transcripts expressed in these cells. ChIP-qPCR revealed two strong peaks of enrichment of the active histone marks H3K27Ac and H3K4Me3 at the

HPV16 early promoter. Interestingly, H3K27Ac is considered a marker of super-enhancers and has previously been shown to be enriched at HPV integration loci. Using a clone derived from W12 cells (20861), which predominantly contains integrated HPV16, H3K27Ac was shown to be enriched at HPV16 integration locus along with BRD4 and MED1 to drive high expression of E6/E7 transcripts from the integrated HPV16 promoter (Dooley et al., 2016).

YY1 has been shown to be enriched at sites of super-enhancers and often co-occupies these sites with BRD4, MED1 and active RNA PolII – YY1 can also form a direct complex with BRD4 (Weintraub et al., 2017, Wang et al., 2022, Xu et al., 2021). Having shown that YY1 is bound at the HPV16 early promoter p97, which is also strongly enriched with H3K27Ac, it is possible that other super-enhancer markers might also be enriched. This might reflect a mechanism where YY1 recruits and/or directly interacts with these marks at integrated HPV16 promoters in order to drive high expression of E6 and E7 within OSPCCs. Importantly, we also show that there is strong enrichment of RNA PolII at the HPV16 early promoter in the OPSCC cell lines, again supporting the idea that YY1, in combination with these marks, functions as a transactivator of the early promoter.YY1 can interact with RNA PolII to facilitate transcription initiation (Austen et al., 1997). Interestingly, preliminary ChIP-qPCR data suggests that depletion of YY1 by siRNA results in a significant reduction in recruitment of RNA PolII to the HPV16 early promoter within OPSCCs (data not shown). Given that during the HPV life cycle RNA PolII is only recruited to the early promoter following depletion of YY1 binding, this supports the hypothesis that there is a switch from the canonical function of YY1 during disease. It is possible that during HPV-driven cancer of the head and neck, YY1 functions to recruit RNA

PolII to transcription start sites in order to drive high expression of the viral oncogenes. Given that this data is from one independent repeat, it will be important to ensure that this is consistent. It would also be interesting to see if YY1 and RNA PolII directly interact using co-immunoprecipitation, to try and elucidate a mechanism for YY1-mediated transactivation of the HPV16 early promoter.

Given that YY1-mediated repression of HPV oncogenes in the early virus life cycle is dependent on long-range chromatin looping with CTCF it is important to consider whether YY1 and/or CTCF might alter chromatin architecture in the high-passage HTKs in order to favour transcription of either viral or host genes. One possibility is the involvement of the human genome in the formation of long-range viral-host chromatin loops. Excluding integration of viral DNA into host chromatin, long-range interactions between viral episomes and host-DNA can also be established. This is true for several DNA viruses including EBV, KSHV and HBV – all of which form long-range interactions with host chromatin, often at active chromatin regions (Yang et al., 2020, Campbell et al., 2018, Wang et al., 2020). HBV DNA forms long-range loop structures with the host genome which are preferential for regions including enhancers containing active histone modifications (H3K4Me3, H3K9Ac, H3K27Ac), transcription start sites, promoters and CpG islands (Yang et al., 2020). EBV episomes were found to associate preferentially with active gene regions including super enhancers which were also enriched with transcription factors – EBV was found to utilise the host transcription factors to initiate transcription of viral genes (Wang et al., 2020). It is therefore possible that YY1 or CTCF might facilitate HPV DNA to host DNA contacts which could favour transcription of either HPV oncogenes or host genes that are upregulated within HPV-driven OPSCCs. To test

this possibility we could use Chromatin Interaction Analysis by Paired-End Tag Sequencing (ChIA-PET).

In the current study we have only assessed how YY1 might function to regulate viral oncogenes, however, given that YY1 has been implicated in the regulation of multiple cellular genes, it is plausible that YY1 might also positively or negatively regulate certain host cell genes to favour cancer progression. A recent study identified a role for YY1 in HPV-driven cancers to upregulate the expression of *lnc-FANCI-2* to drive HPV driven carcinogenesis. The expression of YY1 is regulated by miR-29a which targets the 3'UTR of YY1 mRNA. The expression of HPV E7 downregulates miR-29a resulting in increased expression of YY1. Interestingly YY1 is able to directly interact with E7 which stabilises the binding of YY1 to DNA. The YY1-E7 complex binds to the *FANCI-2* promoter resulting in its transactivation (Liu et al., 2021). It is therefore also possible that in the OPSCCs, YY1 might be interacting with HPV E7 to create a positive feedback loop that upregulates transcription from the early promoter. This was briefly assessed in the current study using co-immunoprecipitation, but the results were inconclusive. It would therefore be interesting to determine whether YY1 and E7 are directly interacting to elucidate a potential mechanism for YY1-induced upregulation of E6/E7 transcription. It would also be interesting to assess whether there are any changes to enrichment of YY1 binding to the human genome in HPV-driven disease using ChIP-qPCR or ChIP-seq.

HPV16 contains ten binding sites for YY1 yet only five are required to regulate transcription from the HPV16 early promoter, p97 (O'Connor et al., 1996). Analysis

of primary and metastasised tumours derived from cervical cancer biopsies has identified that YY1 binding sites in the HPV genome are often mutated in cervical cancer. Interestingly, these mutations to the YY1 binding sites within the URR are found in episomal HPV16 DNA – suggesting that regardless of integration status, YY1 binding can become altered. Interestingly, point mutations to the YY1 binding site #4 (which regulates YY1 binding to DNA) had no impact on YY1-mediated regulation of oncogene transcription. However, point mutations to YY1 binding site #2, which have been identified in both primary tumour and metastasis, results in a 6.5-fold increase in the activity of the HPV16 early promoter (Dong et al., 1994b). Something that should therefore be considered is assessing on a single base pair resolution where YY1 is bound within the viral URR and early promoter. It might be that YY1 binding and/or loss of binding to one site within the URR is responsible for upregulation of E6/E7 transcription. Because ChIP-qPCR works on ~100-200 bp fragments we are unable to pinpoint exactly where YY1 is binding. Though we show three distinct peaks of enrichment at the 5'URR, enhancer and early promoter, it could be the case that YY1 is bound to only one of these sites. Using a technique such as ChIP-exo would show near base-pair resolution to enable us to identify precisely where YY1 is bound.

Surprisingly, we saw enrichment of PRC1 component RING1B within the HPV16 genome in the OPSCCs – this was consistent with enrichment of the histone modification H2A monoubiquitination which is a repressive mark. We initially presumed that RING1B was functioning in a non-canonical manner to activate transcription through chromatin re-arrangements. In breast cancer, RNF2, the gene encoding RING1B, is highly overexpressed. In the bulk population, RING1B exerts

its canonical repressive function, but is also recruited to oncogenic enhancers where it functions in a non-canonical manner to transactivate the enhancer (Chan et al., 2018). We hypothesised that RING1B might be functioning in a similar way in the HPV16 positive OSPCCs whereby, through binding to the HPV16 enhancer, RING1B could alter chromatin conformation to create a transcriptionally active landscape. We opted to deplete RNF2 using siRNA and subsequently use FAIRE to assess the resulting chromatin topology. Intriguingly, we found that depletion of RING1B resulted in a significant reduction of open chromatin at the HPV16 URR, suggesting that RING1B functions to alter the chromatin landscape to favour transcription rather than repress. However, when we assessed the expression of E6/E7 transcripts following depletion of RING1B, it was apparent that RING1B was unable to directly regulate transcription from the early promoter.

Exactly why we see such robust enrichment of both RING1B and H2AK119Ub in our HPV16-driven OPSCCs is still not yet clear especially given that the OPSCCs express high levels of E6/E7 and that H2AK119Ub is considered a repressive modification. One possibility is lack of antibody specificity – although the antibody has been raised to recognise endogenous levels of ubiquitinated H2A protein at lysine 119, the site-specificity is not detailed. The antibody needs to recognise the attachment of ubiquitin to lysine 119 – if the epitope is not site-specific, i.e. between the lysine and the ubiquitin protein, then there is a possibility that the robust enrichment of H2AK119Ub that we see could be due to non-site-specific enrichment. It would therefore be important to test the specificity of the antibody by depleting RING1B using siRNA and then assessing the levels of H2AK119Ub. It is however important to note that previous groups have shown that depletion of

RING1B has no effect on global levels of H2AK119Ub (Zhang et al., 2020). Instead, RING1A was found to be the ligase that predominantly ubiquitinated lysine 119 (Chan et al., 2018). The clonality of the OPSCC cell lines has not been tested and is currently unclear – it is possible that the population is relatively heterogenous. If the OPSCCs are polyclonal then it is difficult to distinguish between the effects of different integration events within a population. Some integration events might be transcriptionally active whilst others might be transcriptionally repressed. It would therefore make sense that a transcriptionally inactive integrant be strongly enriched by the repressive modification H2AK119Ub. If only a small proportion of cells within the polyclonal population are transcriptionally active, then it is likely not possible to distinguish those few cells that are devoid of H2AK119Ub using ChIP-qPCR as this represents the bulk population. This is not just relevant for the RING1B and H2AK119Ub ChIP-qPCR data but is also a caveat to all of the ChIP-qPCR data displayed. It would therefore be important to elucidate within a population of cells which genomes are repressed, and which are transcriptionally active and to then assess the epigenetic landscape and enrichment of CTCF and YY1 of individual integrants. Having shown that we can use whole genome sequencing to identify the different integration events in a population it would be useful to pair this with RNA-sequencing which would enable us to identify transcriptionally active virus-host fusions. It might then be possible to design ChIP primers that bridge the host-viral fusion in order to determine whether certain integrants are enriched for active or repressive histone marks, and if possible YY1 and CTCF.

Given the alteration to chromatin topology, it is possible that RING1B functions to alter chromatin accessibility to facilitate binding of transcription factors such as YY1.

Given that we identified multiple integration events in SCC147 it is also possible that this is not a reflection of the events taking place at a single integrant level. It is therefore possible that RING1B is bound to only transcriptionally inactive integrants. It would be interesting to assess this on a single-integrant level using long-read ChIP-sequencing to elucidate whether transcriptionally active integrants which are driving high expression of E6/E7 are also devoid of RING1B/H2AK119Ub binding. Given that PRC2 components are also essential in repressing transcription it would also be important to assess whether there is recruitment of other PcG proteins to the HPV16 enhancer in the OPSCCs.

6.6 Clinical implications

In the current study we have provided multiple lines of evidence suggesting that YY1 functions to positively regulate E6/E7 oncogene expression – we have shown that YY1 protein expression is significantly upregulated in HPV-driven disease, that there is robust enrichment of YY1 to the HPV genome and that depletion of YY1 results in significantly reduced expression of E6/E7 transcripts**.** In order to validate this work techniques such as immunohistochemistry or ChIP-Immunohistochemistry could be used to assess the expression of YY1 and/or the enrichment of YY1 binding in different grades of HPV-positive oropharyngeal tumour sections. This could be performed as a retrospective study or in real-time. This may provide insight into whether high expression of YY1 is linked to more aggressive forms of cancer, increased likelihood of recurrence or reduced life expectancy. It would also enable treatment options, including the dose of Cisplatin, to be adjusted based on

expression levels of YY1. This would ultimately result in patients having a better quality of life and possibly a better response to treatment.

Chemotherapy in the form of Cisplatin is the first line of treatment for patients with head and neck squamous cell carcinoma. Not only is this very aggressive resulting in unfavourable side-effects, but chemoresistance is also an increasing issue with Cisplatin chemotherapy. The recurrence rate of head and neck cancer is high and often patients have a low five-year survival rate. Additionally, cancers at this site are often associated with resistance to chemotherapy. Interestingly, YY1 has been implicated in enhancing resistance to Cisplatin. Using squamous cell carcinoma lines (HPV-negative) it has been found that knockdown of YY1 can significantly enhance Cisplatin induced apoptosis therefore resulting in reduced cell proliferation (Zhao et al., 2018). This enhancement of the anti-cancer effects of Cisplatin was also reproducible within xenograft models where depletion of YY1 resulted in significant inhibition of tumour growth (Zhao et al., 2018). This is also true for other cancer types – using U87MG cell lines which are isolated from gliomas, it has been shown that silencing YY1 sensitised U87MG cells to Cisplatin. Interestingly, the expression of YY1 was found to be regulated by miR-186 and by restoring miR-186 and therefore inhibiting YY1, the cells were able to respond to Cisplatin treatment (Li et al., 2019). Targeting of YY1 in HPV-driven head and neck cancers therefore presents as a potential therapeutic strategy to reduce chemoresistance to Cisplatin and therefore increase survival rate. Alternatively, high expression of YY1 in patients with HPV-driven cancers of the head and neck could be used as a biomarker for risk stratification.

The use of siRNA as a potential therapeutic modality is gaining interest. The major limitation with this so far has been effectively delivering siRNAs into tumours. One such method for this is the use of liposomal delivery – a preclinical safety study found that a lipid construct known as 1,2-Dioleoyl-sn-Glycero-3-Phosphatidylcholine (DOPC) can be used to facilitate nucleic acid delivery into a range of cancers with no adverse reactions (Landen et al., 2005, Wagner et al., 2017). This was then progressed into an ongoing phase I clinical trial with the objective to determine the safety, tolerability and best dose for the use of an EPH receptor A2-targeting DOPC-encapsulated siRNA to target tumours. Because we have shown that targeting YY1 using siRNA results in significantly reduced E6/E7 transcription and reduced chromatin accessibility, this might represent a possibility for a therapeutic intervention for HPV-driven oropharyngeal carcinomas.

6.7 Limitations

Whilst we and others have shown that YY1 and CTCF function to repress the activity of the HPV18 early promoter during the HPV life cycle, in the current study we were only able to confirm this once using siRNA. Using low-passage HTKs transfected with both HPV16 and HPV18 we attempted to deplete both YY1 and CTCF with SMARTpool siRNAs using several different conditions. However, doing so appeared to elicit an anti-viral response and treatment with scrambled siRNA resulted in a significant reduction in E6/E7 expression. Although we attempted to optimise this as best we could, we were only successful in repeating this experiment once in the 18-HTKs. It is therefore important that this experiment is repeated to ensure it is reproducible. Interestingly, this response was not found in the high-passage HTKs.

Using the same control, we found little to no effect on viral oncogene expression following transfection with the scrambled siRNA. This might indicate that the high-passage HTKs have less of an anti-viral response. There are other methods that could be used to deplete both proteins – previously we have used a short hairpin RNA (shRNA) system and found that depletion of both CTCF and YY1 in low-passage HFKs results in significant upregulation of E6/E7 oncogenes. It might therefore be worthwhile using this system in the low-passage HTKs in order to support our hypothesis that YY1-mediated oncogene repression is lost during cancer.

Our work supports the function of YY1 as a transactivator of the HPV16 early promoter during HPV-driven disease so to drive high E6/E7 transcription. However, the ability to distinguish between individual integration events is limited. HPV-infected keratinocytes contain multiple integration events, and it is often the case that only one or a few of these are transcriptionally active and drive high oncogene expression. Throughout this study we have used Southern blotting to assess the genome status of all of tested HTKs and OPSCCs that have been used. In the SCC147 we also used DNA sequencing to characterise where these events occur in both the viral and host genome. Despite this, we have not yet determined which integration events are transcriptionally active, and importantly, whether we see enrichment of CTCF and/or YY1 at a single integration event. Going forward it will be important to combine whole genome sequencing with RNA-sequencing in order to confirm which integrants are transcriptionally active. We aim to design ChIP-qPCR primers for individual integration events to strengthen the results shown in the current study.

6.8 Concluding remarks

The aim of this thesis was to assess whether CTCF-YY1-mediated repression of HPV oncogene transcription is deregulated during HPV-driven cancers to drive high expression of E6/E7. We have used primary tonsil keratinocytes to model HPV-driven disease progression and patient derived HPV16 oropharyngeal squamous cell carcinoma lines to show that CTCF-mediated repression of HPV oncogene transcription is absent during HPV-driven cancers and that the canonical function of YY1 as a transcriptional repressor is switched during HPV-driven cancer to drive transcription from the early promoter. Binding of CTCF to the HPV genome is conserved in high-risk HPV types. Importantly, CTCF regulates oncogene transcription during the normal viral life cycle in order to repress constitutively high expression of E6/E7. Therefore, loss of CTCF binding during HPV-driven disease progression and loss of CTCF-mediated repression of oncogene transcription is likely a critical aspect of HPV-driven disease. YY1 can function to both activate and repress transcription depending on the context. There is strong evidence to suggest that YY1 can function to repress the activity of the HPV early promoter and this is consistent across multiple different HPV types. The proposed model for YY1-mediated oncogene transcription activation in HPV-driven OPSCCs is shown in **Figure 6.1.** In HPV-driven cancers of the oropharynx, the canonical function of YY1 is altered. Instead, YY1 functions as a transactivator of the HPV16 early promoter to drive the overexpression of E6 and E7 transcription. Enrichment of YY1 to the HPV16 URR is coincident with enrichment of active histone marks and RNA PolII. Importantly, depletion of YY1 results in a significant reduction of both E6 and E7 transcripts. This could highlight the use of YY1 as a potential biomarker for risk

stratification for patients with HPV-driven oropharyngeal cancer. Future studies to assess whether our data is also consistent in tonsil tumour sections would prove valuable for identifying potential therapeutic targets for HPV-driven disease.

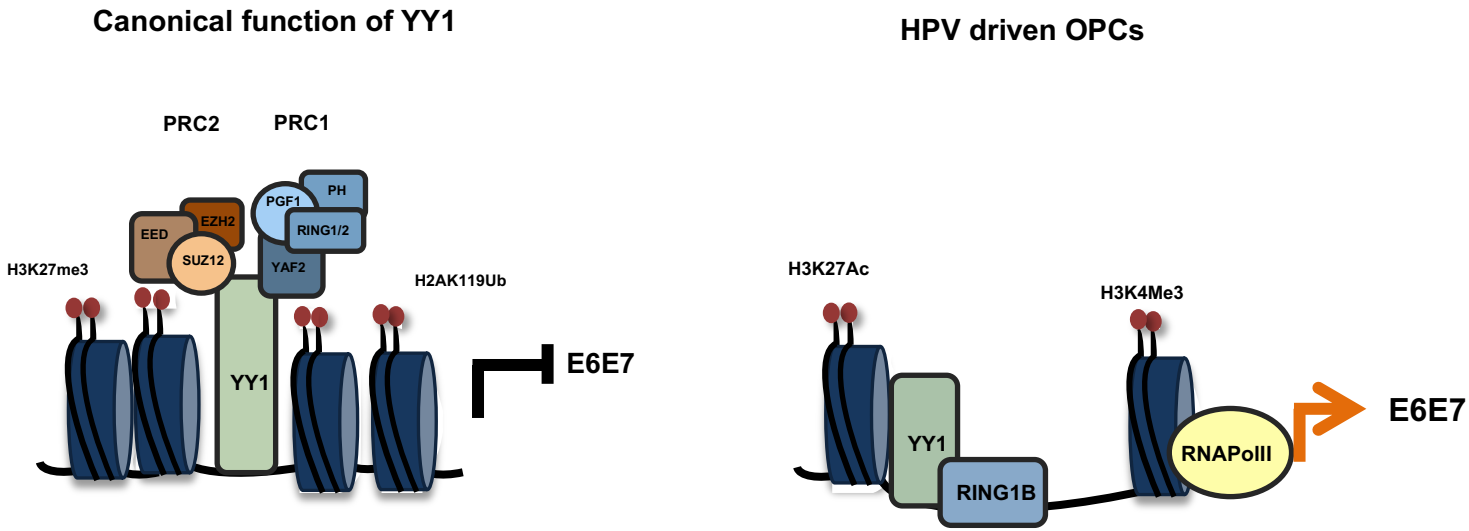


Figure 6.1 Proposed model for switch in function of YY1 from a transcriptional repressor to transactivator of the HPV16 early promoter in OPSCCs. Left: YY1 binds the HPV URR in undifferentiated keratinocytes during the normal viral life cycle. YY1 functions to recruit histone modifying complexes known as PRC1 and PRC2 that deposit H3K27Me3 and H2AK119Ub. This results in condensed chromatin and repression of transcription from the early promoter. Right: YY1 enrichment to the HPV16 genome in OPSCCs is consistent with enrichment of active histone marks including H3K27Ac and H3K4Me3. Although RING1B is also bound, this does not appear to repress E6/E7 transcription. YY1 binding is consistent with recruitment of RNA PolII. Loss of YY1 binding results in reduced oncogene transcription and closed chromatin conformation.

References

AHN, J., PENG, S., HUNG, C. F., RODEN, R. B. S. & BEST, S. R. 2018. Prophylactic immunization with human papillomavirus vaccines induces oral immunity in mice. *Laryngoscope,* 128**,** E16-e20.

ALAZAWI, W., PETT, M., STRAUSS, S., MOSELEY, R., GRAY, J., STANLEY, M. & COLEMAN, N. 2004. Genomic imbalances in 70 snap-frozen cervical squamous intraepithelial lesions: associations with lesion grade, state of the HPV16 E2 gene and clinical outcome. *Br J Cancer,* 91**,** 2063-70.

ALEXANDER, K. E. & RIZKALLAH, R. 2017. Aurora A Phosphorylation of YY1 during Mitosis Inactivates its DNA Binding Activity. *Sci Rep,* 7**,** 10084.

ANDL, T., KAHN, T., PFUHL, A., NICOLA, T., ERBER, R., CONRADT, C., KLEIN, W., HELBIG, M., DIETZ, A., WEIDAUER, H. & BOSCH, F. X. 1998. Etiological involvement of oncogenic human papillomavirus in tonsillar squamous cell carcinomas lacking retinoblastoma cell cycle control. *Cancer Res,* 58**,** 5-13.

ANDRAE, B., ANDERSSON, T. M., LAMBERT, P. C., KEMETLI, L., SILFVERDAL, L., STRANDER, B., RYD, W., DILLNER, J., TÖRNBERG, S. & SPARÉN, P. 2012. Screening and cervical cancer cure: population based cohort study. *Bmj,* 344**,** e900.

ANDROPHY, E. J., LOWY, D. R. & SCHILLER, J. T. 1987. Bovine papillomavirus E2 trans-activating gene product binds to specific sites in papillomavirus DNA. *Nature,* 325**,** 70-3.

ANG, K. K., HARRIS, J., WHEELER, R., WEBER, R., ROSENTHAL, D. I., NGUYEN-TÂN, P. F., WESTRA, W. H., CHUNG, C. H., JORDAN, R. C., LU, C., KIM, H., AXELROD, R., SILVERMAN, C. C., REDMOND, K. P. & GILLISON, M. L. 2010. Human papillomavirus and survival of patients with oropharyngeal cancer. *N Engl J Med,* 363**,** 24-35.

ARBYN, M., WEIDERPASS, E., BRUNI, L., DE SANJOSÉ, S., SARAIYA, M., FERLAY, J. & BRAY, F. 2020. Estimates of incidence and mortality of cervical cancer in 2018: a worldwide analysis. *Lancet Glob Health,* 8**,** e191-e203.

ASHRAFI, G. H., HAGHSHENAS, M., MARCHETTI, B. & CAMPO, M. S. 2006. E5 protein of human papillomavirus 16 downregulates HLA class I and interacts with the heavy chain via its first hydrophobic domain. *Int J Cancer,* 119**,** 2105-12.

ASHRAFI, G. H., HAGHSHENAS, M. R., MARCHETTI, B., O'BRIEN, P. M. & CAMPO, M. S. 2005. E5 protein of human papillomavirus type 16 selectively downregulates surface HLA class I. *Int J Cancer,* 113**,** 276-83.

AUSTEN, M., LÜSCHER, B. & LÜSCHER-FIRZLAFF, J. M. 1997. Characterization of the transcriptional regulator YY1. The bipartite transactivation domain is independent of interaction with the TATA box-binding protein, transcription factor IIB, TAFII55, or cAMP-responsive element-binding protein (CPB)-binding protein. *J Biol Chem,* 272**,** 1709-17.

AZENE, G. K. 2021. Visual inspection with acetic-acid (VIA) service utilization and associated factors among women in Hawassa city, southern Ethiopia: a community based cross-sectional study. *Womens Midlife Health,* 7**,** 6.

BAGCHI, S., WEINMANN, R. & RAYCHAUDHURI, P. 1991. The retinoblastoma protein copurifies with E2F-I, an E1A-regulated inhibitor of the transcription factor E2F. *Cell,* 65**,** 1063-72.

BAKER, C. C. C., C. 1997. Maps of Papillomavirus mRNA Transcripts. *Los Alamos National Labratories; Los Alamos, NM USA*, pp. III-3-III10.
BAKER, T. S., NEWCOMB, W. W., OLSON, N. H., COWSERT, L. M., OLSON, C. & BROWN, J. C. 1991. Structures of bovine and human papillomaviruses. Analysis by cryoelectron microscopy and three-dimensional image reconstruction. *Biophys J,* 60**,** 1445-56.
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.
BASU, A., WILKINSON, F. H., COLAVITA, K., FENNELLY, C. & ATCHISON, M. L. 2014. YY1 DNA binding and interaction with YAF2 is essential for Polycomb recruitment. *Nucleic Acids Res,* 42**,** 2208-23.
BASUKALA, O. & BANKS, L. 2021. The Not-So-Good, the Bad and the Ugly: HPV E5, E6 and E7 Oncoproteins in the Orchestration of Carcinogenesis. *Viruses,* 13.
BAUKNECHT, T., ANGEL, P., ROYER, H. D. & ZUR HAUSEN, H. 1992. Identification of a negative regulatory domain in the human papillomavirus type 18 promoter: interaction with the transcriptional repressor YY1. *Embo j,* 11**,** 4607-17.
BAUKNECHT, T., JUNDT, F., HERR, I., OEHLER, T., DELIUS, H., SHI, Y., ANGEL, P. & ZUR HAUSEN, H. 1995. A switch region determines the cell type-specific positive or negative action of YY1 on the activity of the human papillomavirus type 18 promoter. *J Virol,* 69**,** 1-12.
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,* 27**,** 1139-1152.
BEDELL, M. A., HUDSON, J. B., GOLUB, T. R., TURYK, M. E., HOSKEN, M., WILBANKS, G. D. & LAIMINS, L. A. 1991. Amplification of human papillomavirus genomes in vitro is dependent on epithelial differentiation. *J Virol,* 65**,** 2254-60.
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.
BEREZUTSKAYA, E., YU, B., MOROZOV, A., RAYCHAUDHURI, P. & BAGCHI, S. 1997. Differential regulation of the pocket domains of the retinoblastoma family proteins by the HPV16 E7 oncoprotein. *Cell Growth Differ,* 8**,** 1277-86.
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. *J Virol,* 63**,** 4317-24.
BIENKOWSKA-HABA, M., PATEL, H. D. & SAPP, M. 2009. Target cell cyclophilins facilitate human papillomavirus type 16 infection. *PLoS Pathog,* 5**,** e1000524.
BODELON, C., UNTEREINER, M. E., MACHIELA, M. J., VINOKUROVA, S. & WENTZENSEN, N. 2016. Genomic characterization of viral integration sites in HPV-related cancers. *Int J Cancer,* 139**,** 2001-11.
BOSHART, M., GISSMANN, L., IKENBERG, H., KLEINHEINZ, A., SCHEURLEN, W. & ZUR HAUSEN, H. 1984. A new type of papillomavirus DNA, its presence in genital cancer biopsies and in cell lines derived from cervical cancer. *Embo j,* 3**,** 1151-7.

BOUVARD, V., STOREY, A., PIM, D. & BANKS, L. 1994. Characterization of the human papillomavirus E2 protein: evidence of trans-activation and trans-repression in cervical keratinocytes. *Embo j,* 13**,** 5451-9.
BOYER, S. N., WAZER, D. E. & BAND, V. 1996. E7 protein of human papilloma virus-16 induces degradation of retinoblastoma protein through the ubiquitin-proteasome pathway. *Cancer Res,* 56**,** 4620-4.
BUBB, V., MCCANCE, D. J. & SCHLEGEL, R. 1988. DNA sequence of the HPV-16 E5 ORF and the structural conservation of its encoded protein. *Virology,* 163**,** 243-6.
BURLEY, M., ROBERTS, S. & PARISH, J. L. 2020. Epigenetic regulation of human papillomavirus transcription in the productive virus life cycle. *Semin Immunopathol,* 42**,** 159-171.
BUSHMEYER, S., PARK, K. & ATCHISON, M. L. 1995. Characterization of functional domains within the multifunctional transcription factor, YY1. *J Biol Chem,* 270**,** 30213-20.
CAMPBELL, M., WATANABE, T., NAKANO, K., DAVIS, R. R., LYU, Y., TEPPER, C. G., DURBIN-JOHNSON, B., FUJIMURO, M. & IZUMIYA, Y. 2018. KSHV episomes reveal dynamic chromatin loop formation with domain-specific gene regulation. *Nat Commun,* 9**,** 49.
CAMPO, M. S., GRAHAM, S. V., CORTESE, M. S., ASHRAFI, G. H., ARAIBI, E. H., DORNAN, E. S., MINERS, K., NUNES, C. & MAN, S. 2010. HPV-16 E5 down-regulates expression of surface HLA class I and reduces recognition by CD8 T cells. *Virology,* 407**,** 137-42.
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.
CAO, R. & ZHANG, Y. 2004. SUZ12 is required for both the histone methyltransferase activity and the silencing function of the EED-EZH2 complex. *Mol Cell,* 15**,** 57-67.
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.
CASTELLSAGUÉ, X., ALEMANY, L., QUER, M., HALEC, G., QUIRÓS, B., TOUS, S., CLAVERO, O., ALÒS, L., BIEGNER, T., SZAFAROWSKI, T., ALEJO, M., HOLZINGER, D., CADENA, E., CLAROS, E., HALL, G., LACO, J., POLJAK, M., BENEVOLO, M., KASAMATSU, E., MEHANNA, H., NDIAYE, C., GUIMERÀ, N., LLOVERAS, B., LEÓN, X., RUIZ-CABEZAS, J. C., ALVARADO-CABRERO, I., KANG, C. S., OH, J. K., GARCIA-ROJO, M., ILJAZOVIC, E., AJAYI, O. F., DUARTE, F., NESSA, A., TINOCO, L., DURAN-PADILLA, M. A., PIROG, E. C., VIARHEICHYK, H., MORALES, H., COSTES, V., FÉLIX, A., GERMAR, M. J., MENA, M., RUACAN, A., JAIN, A., MEHROTRA, R., GOODMAN, M. T., LOMBARDI, L. E., FERRERA, A., MALAMI, S., ALBANESI, E. I., DABED, P., MOLINA, C., LÓPEZ-REVILLA, R., MANDYS, V., GONZÁLEZ, M. E., VELASCO, J., BRAVO, I. G., QUINT, W., PAWLITA, M., MUÑOZ, N., DE SANJOSÉ, S. & XAVIER BOSCH, F. 2016. HPV Involvement in Head and Neck Cancers: Comprehensive Assessment of Biomarkers in 3680 Patients. *J Natl Cancer Inst,* 108**,** djv403.
CHAN, H. L., BECKEDORFF, F., ZHANG, Y., GARCIA-HUIDOBRO, J., JIANG, H., COLAPRICO, A., BILBAO, D., FIGUEROA, M. E., LACAVA, J.,

SHIEKHATTAR, R. & MOREY, L. 2018. Polycomb complexes associate with enhancers and promote oncogenic transcriptional programs in cancer through multiple mechanisms. *Nat Commun,* 9**,** 3377.
CHAN, W.-K., CHONG, T., BERNARD, H.-U. & KIOCK, 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.
CHELLAPPAN, S. P., HIEBERT, S., MUDRYJ, M., HOROWITZ, J. M. & NEVINS, J. R. 1991. The E2F transcription factor is a cellular target for the RB protein. *Cell,* 65**,** 1053-61.
CHEN, H. S., WIKRAMASINGHE, P., SHOWE, L. & LIEBERMAN, P. M. 2012. Cohesins repress Kaposi's sarcoma-associated herpesvirus immediate early gene transcription during latency. *J Virol,* 86**,** 9454-64.
CHEN, J., XUE, Y., POIDINGER, M., LIM, T., CHEW, S. H., PANG, C. L., ABASTADO, J. P. & THIERRY, F. 2014. Mapping of HPV transcripts in four human cervical lesions using RNAseq suggests quantitative rearrangements during carcinogenic progression. *Virology,* 462-463**,** 14-24.
CHEN, S. L. & MOUNTS, P. 1990. Transforming activity of E5a protein of human papillomavirus type 6 in NIH 3T3 and C127 cells. *J Virol,* 64**,** 3226-33.
CHIANG, C. M., DONG, G., BROKER, T. R. & CHOW, L. T. 1992. Control of human papillomavirus type 11 origin of replication by the E2 family of transcription regulatory proteins. *J Virol,* 66**,** 5224-31.
CHIN, M. T., HIROCHIKA, R., HIROCHIKA, H., BROKER, T. R. & CHOW, L. T. 1988. Regulation of human papillomavirus type 11 enhancer and E6 promoter by activating and repressing proteins from the E2 open reading frame: functional and biochemical studies. *J Virol,* 62**,** 2994-3002.
CHOE, J., VAILLANCOURT, P., STENLUND, A. & BOTCHAN, M. 1989. Bovine papillomavirus type 1 encodes two forms of a transcriptional repressor: structural and functional analysis of new viral cDNAs. *J Virol,* 63**,** 1743-55.
CHONG, T., CHAN, W. K. & BERNARD, H. U. 1990. Transcriptional activation of human papillomavirus 16 by nuclear factor I, AP1, steroid receptors and a possibly novel transcription factor, PVF: a model for the composition of genital papillomavirus enhancers. *Nucleic Acids Res,* 18**,** 465-70.
COBRINIK, D., WHYTE, P., PEEPER, D. S., JACKS, T. & WEINBERG, R. A. 1993. Cell cycle-specific association of E2F with the p130 E1A-binding protein. *Genes Dev,* 7**,** 2392-404.
CONGER, K. L., LIU, J. S., KUO, S. R., CHOW, L. T. & WANG, T. S. 1999. Human papillomavirus DNA replication. Interactions between the viral E1 protein and two subunits of human dna polymerase alpha/primase. *J Biol Chem,* 274**,** 2696-705.
CRUZ, L. & MEYERS, C. 2013. Differential dependence on host cell glycosaminoglycans for infection of epithelial cells by high-risk HPV types. *PLoS One,* 8**,** e68379.
CULP, T. D., BUDGEON, L. R., MARINKOVICH, M. P., MENEGUZZI, G. & CHRISTENSEN, N. D. 2006. Keratinocyte-secreted laminin 5 can function as a transient receptor for human papillomaviruses by binding virions and transferring them to adjacent cells. *J Virol,* 80**,** 8940-50.
CUZICK, J. 2015. Gardasil 9 joins the fight against cervix cancer. *Expert Rev Vaccines,* 14**,** 1047-9.

D'ARIENZO, V., FERGUSON, J., GIRAUD, G., CHAPUS, F., HARRIS, J. M., WING, P. A. C., CLAYDON, A., BEGUM, S., ZHUANG, X., BALFE, P., TESTONI, B., MCKEATING, J. A. & PARISH, J. L. 2021. The CCCTC-binding factor CTCF represses hepatitis B virus enhancer I and regulates viral transcription. *Cell Microbiol,* 23**,** e13274.

DAY, P. M., BAKER, C. C., LOWY, D. R. & SCHILLER, J. T. 2004. Establishment of papillomavirus infection is enhanced by promyelocytic leukemia protein (PML) expression. *Proc Natl Acad Sci U S A,* 101**,** 14252-7.

DAY, P. M., LOWY, D. R. & SCHILLER, J. T. 2003. Papillomaviruses infect cells via a clathrin-dependent pathway. *Virology,* 307**,** 1-11.

DAY, P. M., LOWY, D. R. & SCHILLER, J. T. 2008. Heparan sulfate-independent cell binding and infection with furin-precleaved papillomavirus capsids. *J Virol,* 82**,** 12565-8.

DE GEEST, K., TURYK, M. E., HOSKEN, M. I., HUDSON, J. B., LAIMINS, L. A. & WILBANKS, G. D. 1993. Growth and differentiation of human papillomavirus type 31b positive human cervical cell lines. *Gynecol Oncol,* 49**,** 303-10.

DE MARTEL, C., GEORGES, D., BRAY, F., FERLAY, J. & CLIFFORD, G. M. 2020. Global burden of cancer attributable to infections in 2018: a worldwide incidence analysis. *Lancet Glob Health,* 8**,** e180-e190.

DE MARTEL, C., PLUMMER, M., VIGNAT, J. & FRANCESCHI, S. 2017. Worldwide burden of cancer attributable to HPV by site, country and HPV type. *Int J Cancer,* 141**,** 664-670.

DE SANJOSÉ, S., DIAZ, M., CASTELLSAGUÉ, X., CLIFFORD, G., BRUNI, L., MUÑOZ, N. & BOSCH, F. X. 2007. Worldwide prevalence and genotype distribution of cervical human papillomavirus DNA in women with normal cytology: a meta-analysis. *Lancet Infect Dis,* 7**,** 453-9.

DE SANJOSE, S., QUINT, W. G., ALEMANY, L., GERAETS, D. T., KLAUSTERMEIER, J. E., LLOVERAS, B., TOUS, S., FELIX, A., BRAVO, L. E., SHIN, H. R., VALLEJOS, C. S., DE RUIZ, P. A., LIMA, M. A., GUIMERA, N., CLAVERO, O., ALEJO, M., LLOMBART-BOSCH, A., CHENG-YANG, C., TATTI, S. A., KASAMATSU, E., ILJAZOVIC, E., ODIDA, M., PRADO, R., SEOUD, M., GRCE, M., USUBUTUN, A., JAIN, A., SUAREZ, G. A., LOMBARDI, L. E., BANJO, A., MENÉNDEZ, C., DOMINGO, E. J., VELASCO, J., NESSA, A., CHICHAREON, S. C., QIAO, Y. L., LERMA, E., GARLAND, S. M., SASAGAWA, T., FERRERA, A., HAMMOUDA, D., MARIANI, L., PELAYO, A., STEINER, I., OLIVA, E., MEIJER, C. J., AL-JASSAR, W. F., CRUZ, E., WRIGHT, T. C., PURAS, A., LLAVE, C. L., TZARDI, M., AGORASTOS, T., GARCIA-BARRIOLA, V., CLAVEL, C., ORDI, J., ANDÚJAR, M., CASTELLSAGUÉ, X., SÁNCHEZ, G. I., NOWAKOWSKI, A. M., BORNSTEIN, J., MUÑOZ, N. & BOSCH, F. X. 2010. Human papillomavirus genotype attribution in invasive cervical cancer: a retrospective cross-sectional worldwide study. *Lancet Oncol,* 11**,** 1048-56.

DE VILLIERS, E. M., FAUQUET, C., BROKER, T. R., BERNARD, H. U. & ZUR HAUSEN, H. 2004. Classification of papillomaviruses. *Virology,* 324**,** 17-27.

DEL MAR PEÑA, L. M. & LAIMINS, L. A. 2001. Differentiation-dependent chromatin rearrangement coincides with activation of human papillomavirus type 31 late gene expression. *J Virol,* 75**,** 10005-13.

DELURY, C. P., MARSH, E. K., JAMES, C. D., BOON, S. S., BANKS, L., KNIGHT, G. L. & ROBERTS, S. 2013. The role of protein kinase A regulation of the E6

PDZ-binding domain during the differentiation-dependent life cycle of human papillomavirus type 18. *J Virol,* 87**,** 9463-72.
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, Z., HASEGAWA, M., KIYUNA, A., MATAYOSHI, S., UEHARA, T., AGENA, S., YAMASHITA, Y., OGAWA, K., MAEDA, H. & SUZUKI, M. 2013. Viral load, physical status, and E6/E7 mRNA expression of human papillomavirus in head and neck squamous cell carcinoma. *Head Neck,* 35**,** 800-8.
DILLNER, J., KJAER, S. K., WHEELER, C. M., SIGURDSSON, K., IVERSEN, O. E., HERNANDEZ-AVILA, M., PEREZ, G., BROWN, D. R., KOUTSKY, L. A., TAY, E. H., GARCÍA, P., AULT, K. A., GARLAND, S. M., LEODOLTER, S., OLSSON, S. E., TANG, G. W., FERRIS, D. G., PAAVONEN, J., LEHTINEN, M., STEBEN, M., BOSCH, F. X., JOURA, E. A., MAJEWSKI, S., MUÑOZ, N., MYERS, E. R., VILLA, L. L., TADDEO, F. J., ROBERTS, C., TADESSE, A., BRYAN, J. T., MAANSSON, R., LU, S., VUOCOLO, S., HESLEY, T. M., BARR, E. & HAUPT, R. 2010. Four year efficacy of prophylactic human papillomavirus quadrivalent vaccine against low grade cervical, vulvar, and vaginal intraepithelial neoplasia and anogenital warts: randomised controlled trial. *Bmj,* 341**,** c3493.
DONG, G., BROKER, T. R. & CHOW, L. T. 1994a. Human papillomavirus type 11 E2 proteins repress the homologous E6 promoter by interfering with the binding of host transcription factors to adjacent elements. *J Virol,* 68**,** 1115-27.
DONG, X. P., STUBENRAUCH, F., BEYER-FINKLER, E. & PFISTER, H. 1994b. Prevalence of deletions of YY1-binding sites in episomal HPV 16 DNA from cervical cancers. *Int J Cancer,* 58**,** 803-8.
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.
DOOLEY, K. E., WARBURTON, A. & MCBRIDE, A. A. 2016. Tandemly Integrated HPV16 Can Form a Brd4-Dependent Super-Enhancer-Like Element That Drives Transcription of Viral Oncogenes. *mBio,* 7.
DOORBAR, J., EGAWA, N., GRIFFIN, H., KRANJEC, C. & MURAKAMI, I. 2015. Human papillomavirus molecular biology and disease association. *Rev Med Virol,* 25 Suppl 1**,** 2-23.
DROLET, M., BÉNARD, É., PÉREZ, N. & BRISSON, M. 2019. Population-level impact and herd effects following the introduction of human papillomavirus vaccination programmes: updated systematic review and meta-analysis. *Lancet,* 394**,** 497-509.
DÜRST, M., GISSMANN, L., IKENBERG, H. & ZUR HAUSEN, H. 1983. A papillomavirus DNA from a cervical carcinoma and its prevalence in cancer biopsy samples from different geographic regions. *Proc Natl Acad Sci U S A,* 80**,** 3812-5.
DYSON, N. 1998. The regulation of E2F by pRB-family proteins. *Genes Dev,* 12**,** 2245-62.
DYSON, N., GUIDA, P., MÜNGER, K. & HARLOW, E. 1992. Homologous sequences in adenovirus E1A and human papillomavirus E7 proteins mediate interaction with the same set of cellular proteins. *J Virol,* 66**,** 6893-902.

DYSON, N., HOWLEY, P. M., MÜNGER, 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., EGAWA, K., GRIFFIN, H. & DOORBAR, J. 2015. Human Papillomaviruses; Epithelial Tropisms, and the Development of Neoplasia. *Viruses,* 7**,** 3863-90.
ESKELAND, R., LEEB, M., GRIMES, G. R., KRESS, C., BOYLE, S., SPROUL, D., GILBERT, N., FAN, Y., SKOULTCHI, A. I., WUTZ, A. & BICKMORE, W. A. 2010. Ring1B compacts chromatin structure and represses gene expression independent of histone ubiquitination. *Mol Cell,* 38**,** 452-64.
FAVRE, M., BREITBURD, F., CROISSANT, O. & ORTH, G. 1977. Chromatin-like structures obtained after alkaline disruption of bovine and human papillomaviruses. *J Virol,* 21**,** 1205-9.
FEHRMANN, F. & LAIMINS, L. A. 2003. Human papillomaviruses: targeting differentiating epithelial cells for malignant transformation. *Oncogene,* 22**,** 5201-7.
FERGUSON, J., CAMPOS-LEÓN, K., PENTLAND, I., STOCKTON, J. D., GÜNTHER, T., BEGGS, A. D., GRUNDHOFF, A., ROBERTS, S., NOYVERT, B. & PARISH, J. L. 2021. The chromatin insulator CTCF regulates HPV18 transcript splicing and differentiation-dependent late gene expression. *PLoS Pathog,* 17**,** e1010032.
FILIPPOVA, G. N., FAGERLIE, S., KLENOVA, E. M., MYERS, C., DEHNER, Y., GOODWIN, G., NEIMAN, P. E., COLLINS, S. J. & LOBANENKOV, V. V. 1996. An exceptionally conserved transcriptional repressor, CTCF, employs different combinations of zinc fingers to bind diverged promoter sequences of avian and mammalian c-myc oncogenes. *Mol Cell Biol,* 16**,** 2802-13.
FLORES, E. R. & LAMBERT, P. F. 1997. Evidence for a switch in the mode of human papillomavirus type 16 DNA replication during the viral life cycle. *J Virol,* 71**,** 7167-79.
FORCIER, M. & MUSACCHIO, N. 2010. An overview of human papillomavirus infection for the dermatologist: disease, diagnosis, management, and prevention. *Dermatol Ther,* 23**,** 458-76.
FRANGINI, A., SJÖBERG, M., ROMAN-TRUFERO, M., DHARMALINGAM, G., HABERLE, V., BARTKE, T., LENHARD, B., MALUMBRES, M., VIDAL, M. & DILLON, N. 2013. The aurora B kinase and the polycomb protein ring1B combine to regulate active promoters in quiescent lymphocytes. *Mol Cell,* 51**,** 647-61.
FU, X., JI, F., HE, Q. & QIU, X. 2022. A Systematic Pan-Cancer Analysis of YY1 Aberrations and their Relationship with Clinical Outcome, Tumor Microenvironment, and Therapeutic Targets. *J Immunol Res,* 2022**,** 5826741.
FUTURE, G. S. I. 2007. Quadrivalent vaccine against human papillomavirus to prevent high-grade cervical lesions. *N Engl J Med,* 356**,** 1915-27.
GAO, Q., WANG, S. & ZHANG, Z. 2021. E3 ubiquitin ligase SMURF2 prevents colorectal cancer by reducing the stability of the YY1 protein and inhibiting the SENP1/c-myc axis. *Gene Ther*.
GAO, Z., ZHANG, J., BONASIO, R., STRINO, F., SAWAI, A., PARISI, F., KLUGER, Y. & REINBERG, D. 2012. PCGF homologs, CBX proteins, and RYBP define functionally distinct PRC1 family complexes. *Mol Cell,* 45**,** 344-56.

GARCÍA, E., MARCOS-GUTIÉRREZ, C., DEL MAR LORENTE, M., MORENO, J. C. & VIDAL, M. 1999. RYBP, a new repressor protein that interacts with components of the mammalian Polycomb complex, and with the transcription factor YY1. *Embo j,* 18**,** 3404-18.

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. 2007. Quadrivalent vaccine against human papillomavirus to prevent anogenital diseases. *N Engl J Med,* 356**,** 1928-43.

GIACINTI, C. & GIORDANO, A. 2006. RB and cell cycle progression. *Oncogene,* 25**,** 5220-7.

GILLISON, M. L., TROTTI, A. M., HARRIS, J., EISBRUCH, A., HARARI, P. M., ADELSTEIN, D. J., JORDAN, R. C. K., ZHAO, W., STURGIS, E. M., BURTNESS, B., RIDGE, J. A., RINGASH, J., GALVIN, J., YAO, M., KOYFMAN, S. A., BLAKAJ, D. M., RAZAQ, M. A., COLEVAS, A. D., BEITLER, J. J., JONES, C. U., DUNLAP, N. E., SEAWARD, S. A., SPENCER, S., GALLOWAY, T. J., PHAN, J., DIGNAM, J. J. & LE, Q. T. 2019. Radiotherapy plus cetuximab or cisplatin in human papillomavirus-positive oropharyngeal cancer (NRG Oncology RTOG 1016): a randomised, multicentre, non-inferiority trial. *Lancet,* 393**,** 40-50.

GIRI, I. & YANIV, M. 1988. Structural and mutational analysis of E2 trans-activating proteins of papillomaviruses reveals three distinct functional domains. *Embo j,* 7**,** 2823-9.

GIROGLOU, T., FLORIN, L., SCHÄFER, 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., SCHNÜRCH, 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.

GISSMANN, L. & ZUR HAUSEN, H. 1980. Partial characterization of viral DNA from human genital warts (Condylomata acuminata). *Int J Cancer,* 25**,** 605-9.

GIULIANO, A. R., JOURA, E. A., GARLAND, S. M., HUH, W. K., IVERSEN, O. E., KJAER, S. K., FERENCZY, A., KURMAN, R. J., RONNETT, B. M., STOLER, M. H., BAUTISTA, O. M., MOELLER, E., RITTER, M., SHIELDS, C. & LUXEMBOURG, A. 2019. Nine-valent HPV vaccine efficacy against related diseases and definitive therapy: comparison with historic placebo population. *Gynecol Oncol,* 154**,** 110-117.

GIULIANO, A. R., WILKIN, T., BAUTISTA, O. M., CHEON, K., CONNOR, L., DUBEY, S., LUXEMBOURG, A., RAWAT, S., SHAW, A., VELICER, C., VENDETTI, N. & TU, Y. 2022. Design of a phase III efficacy, immunogenicity, and safety study of 9-valent human papillomavirus vaccine in prevention of oral persistent infection in men. *Contemp Clin Trials,* 115**,** 106592.

GLAUNSINGER, B. A., LEE, S. S., THOMAS, M., BANKS, L. & JAVIER, R. 2000. Interactions of the PDZ-protein MAGI-1 with adenovirus E4-ORF1 and high-risk papillomavirus E6 oncoproteins. *Oncogene,* 19**,** 5270-80.

GLOSS, B. & BERNARD, H. U. 1990. The E6/E7 promoter of human papillomavirus type 16 is activated in the absence of E2 proteins by a sequence-aberrant Sp1 distal element. *J Virol,* 64**,** 5577-84.
GLOSS, B., CHONG, T. & BERNARD, H. U. 1989a. Numerous nuclear proteins bind the long control region of human papillomavirus type 16: a subset of 6 of 23 DNase I-protected segments coincides with the location of the cell-type-specific enhancer. *J Virol,* 63**,** 1142-52.
GLOSS, B., YEO-GLOSS, M., MEISTERENST, M., ROGGE, L., WINNACKER, E. L. & BERNARD, H. U. 1989b. Clusters of nuclear factor I binding sites identify enhancers of several papillomaviruses but alone are not sufficient for enhancer function. *Nucleic Acids Res,* 17**,** 3519-33.
GRAHAM, S. V. 2017. The human papillomavirus replication cycle, and its links to cancer progression: a comprehensive review. *Clin Sci (Lond),* 131**,** 2201-2221.
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.
GROVES, I. J., KNIGHT, E. L., ANG, Q. Y., SCARPINI, C. G. & COLEMAN, N. 2016. HPV16 oncogene expression levels during early cervical carcinogenesis are determined by the balance of epigenetic chromatin modifications at the integrated virus genome. *Oncogene,* 35**,** 4773-86.
GROVES, I. J., TANG, G., PENTLAND, I., PARISH, J. L. & COLEMAN, N. 2021. CTCF association with episomal HPV16 genomes regulates viral oncogene transcription and splicing. *bioRxiv***,** 2021.02.18.431881.
HAEGGBLOM, L., RAMQVIST, T., TOMMASINO, M., DALIANIS, T. & NÄSMAN, A. 2017. Time to change perspectives on HPV in oropharyngeal cancer. A systematic review of HPV prevalence per oropharyngeal sub-site the last 3 years. *Papillomavirus Res,* 4**,** 1-11.
HAFKAMP, H. C., SPEEL, E. J., HAESEVOETS, A., BOT, F. J., DINJENS, W. N., RAMAEKERS, F. C., HOPMAN, A. H. & MANNI, J. J. 2003. A subset of head and neck squamous cell carcinomas exhibits integration of HPV 16/18 DNA and overexpression of p16INK4A and p53 in the absence of mutations in p53 exons 5-8. *Int J Cancer,* 107**,** 394-400.
HAFT, S., REN, S., XU, G., MARK, A., FISCH, K., GUO, T. W., KHAN, Z., PANG, J., ANDO, M., LIU, C., SAKAI, A., FUKUSUMI, T. & CALIFANO, J. A. 2019. Mutation of chromatin regulators and focal hotspot alterations characterize human papillomavirus-positive oropharyngeal squamous cell carcinoma. *Cancer,* 125**,** 2423-2434.
HAGENSEE, M. E., YAEGASHI, N. & GALLOWAY, D. A. 1993. Self-assembly of human papillomavirus type 1 capsids by expression of the L1 protein alone or by coexpression of the L1 and L2 capsid proteins. *J Virol,* 67**,** 315-22.
HANSEN, K. H., BRACKEN, A. P., PASINI, D., DIETRICH, N., GEHANI, S. S., MONRAD, A., RAPPSILBER, J., LERDRUP, M. & HELIN, K. 2008. A model for transmission of the H3K27me3 epigenetic mark. *Nat Cell Biol,* 10**,** 1291-300.

HEGDE, R. S., GROSSMAN, S. R., LAIMINS, L. A. & SIGLER, P. B. 1992. Crystal structure at 1.7 A of the bovine papillomavirus-1 E2 DNA-binding domain bound to its DNA target. *Nature,* 359**,** 505-12.
HERFS, M., YAMAMOTO, Y., LAURY, A., WANG, X., NUCCI, M. R., MCLAUGHLIN-DRUBIN, M. E., MÜNGER, 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. *Proc Natl Acad Sci U S A,* 109**,** 10516-21.
HERMSEN, M. A., JOENJE, H., ARWERT, F., WELTERS, M. J., BRAAKHUIS, B. J., BAGNAY, M., WESTERVELD, A. & SLATER, R. 1996. Centromeric breakage as a major cause of cytogenetic abnormalities in oral squamous cell carcinoma. *Genes Chromosomes Cancer,* 15**,** 1-9.
HIROCHIKA, H., HIROCHIKA, R., BROKER, T. R. & CHOW, L. T. 1988. Functional mapping of the human papillomavirus type 11 transcriptional enhancer and its interaction with the trans-acting E2 proteins. *Genes Dev,* 2**,** 54-67.
HOPPE-SEYLER, F. & BUTZ, K. 1992. Activation of human papillomavirus type 18 E6-E7 oncogene expression by transcription factor Sp1. *Nucleic Acids Res,* 20**,** 6701-6.
HOPPE-SEYLER, F., BUTZ, K. & ZUR HAUSEN, H. 1991. Repression of the human papillomavirus type 18 enhancer by the cellular transcription factor Oct-1. *J Virol,* 65**,** 5613-8.
HOU, S. Y., WU, S. Y., ZHOU, T., THOMAS, M. C. & CHIANG, C. M. 2000. Alleviation of human papillomavirus E2-mediated transcriptional repression via formation of a TATA binding protein (or TFIID)-TFIIB-RNA polymerase II-TFIIF preinitiation complex. *Mol Cell Biol,* 20**,** 113-25.
HOUBAVIY, H. B., USHEVA, A., SHENK, T. & BURLEY, S. K. 1996. Cocrystal structure of YY1 bound to the adeno-associated virus P5 initiator. *Proc Natl Acad Sci U S A,* 93**,** 13577-82.
HUANG, P. S., PATRICK, D. R., EDWARDS, G., GOODHART, P. J., HUBER, H. E., MILES, L., GARSKY, V. M., OLIFF, A. & HEIMBROOK, D. C. 1993. Protein domains governing interactions between E2F, the retinoblastoma gene product, and human papillomavirus type 16 E7 protein. *Mol Cell Biol,* 13**,** 953-60.
HUH, K., ZHOU, X., HAYAKAWA, H., CHO, J. Y., LIBERMANN, T. A., JIN, J., HARPER, J. W. & MUNGER, K. 2007. Human papillomavirus type 16 E7 oncoprotein associates with the cullin 2 ubiquitin ligase complex, which contributes to degradation of the retinoblastoma tumor suppressor. *J Virol,* 81**,** 9737-47.
HUIBREGTSE, J. M., SCHEFFNER, M. & HOWLEY, P. M. 1991. A cellular protein mediates association of p53 with the E6 oncoprotein of human papillomavirus types 16 or 18. *Embo j,* 10**,** 4129-35.
HUIBREGTSE, J. M., SCHEFFNER, M. & HOWLEY, P. M. 1993. Localization of the E6-AP regions that direct human papillomavirus E6 binding, association with p53, and ubiquitination of associated proteins. *Mol Cell Biol,* 13**,** 4918-27.
HYDE-DERUYSCHER, R. P., JENNINGS, E. & SHENK, T. 1995. DNA binding sites for the transcriptional activator/repressor YY1. *Nucleic Acids Res,* 23**,** 4457-65.

IARC, G. W. 2007. IARC Working Group on the Evaluation of Carcinogenic Risks to Humans. Human Papillomaviruses. *IARC Monographs on the Evaluation of Carcinogenic Risks to Humans,* 90**,** 1-636.
IARC, G. W. 2020. IARC Monographs On The Identification Of Carcinogenic Hazards To Humans. 1-127.
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. 1992. 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.
JAMES, C. D., SAINI, S., SESAY, F., KO, K., FELTHOUSEN-RUSBASAN, J., INESS, A. N., NULTON, T., WINDLE, B., DOZMOROV, M. G., MORGAN, I. M. & LITOVCHICK, L. 2021. Restoring the DREAM Complex Inhibits the Proliferation of High-Risk HPV Positive Human Cells. *Cancers (Basel),* 13.
JEON, S., ALLEN-HOFFMANN, B. L. & LAMBERT, P. F. 1995. Integration of human papillomavirus type 16 into the human genome correlates with a selective growth advantage of cells. *J Virol,* 69**,** 2989-97.
JEON, S. & LAMBERT, P. F. 1995. Integration of human papillomavirus type 16 DNA into the human genome leads to increased stability of E6 and E7 mRNAs: implications for cervical carcinogenesis. *Proc Natl Acad Sci U S A,* 92**,** 1654-8.
JEONG, H. M., LEE, S. H., YUM, J., YEO, C. Y. & LEE, K. Y. 2014. Smurf2 regulates the degradation of YY1. *Biochim Biophys Acta,* 1843**,** 2005-11.
JERONIMO, J., CASTLE, P. E., TEMIN, S., DENNY, L., GUPTA, V., KIM, J. J., LUCIANI, S., MUROKORA, D., NGOMA, T., QIAO, Y., QUINN, M., SANKARANARAYANAN, R., SASIENI, P., SCHMELER, K. M. & SHASTRI, S. S. 2017. Secondary Prevention of Cervical Cancer: ASCO Resource-Stratified Clinical Practice Guideline. *J Glob Oncol,* 3**,** 635-657.
JOHANSSON, C., SOMBERG, M., LI, X., BACKSTRÖM 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.
JOHNSON, D. E., BURTNESS, B., LEEMANS, C. R., LUI, V. W. Y., BAUMAN, J. E. & GRANDIS, J. R. 2020. Head and neck squamous cell carcinoma. *Nat Rev Dis Primers,* 6**,** 92.
JOURA, E. A., GIULIANO, A. R., IVERSEN, O. E., BOUCHARD, C., MAO, C., MEHLSEN, J., MOREIRA, E. D., JR., NGAN, Y., PETERSEN, L. K., LAZCANO-PONCE, E., PITISUTTITHUM, P., RESTREPO, J. A., STUART, G., WOELBER, L., YANG, Y. C., CUZICK, J., GARLAND, S. M., HUH, W., KJAER, S. K., BAUTISTA, O. M., CHAN, I. S., CHEN, J., GESSER, R., MOELLER, E., RITTER, M., VUOCOLO, S. & LUXEMBOURG, A. 2015. A 9-valent HPV vaccine against infection and intraepithelial neoplasia in women. *N Engl J Med,* 372**,** 711-23.
JOYCE, J. G., TUNG, J. S., PRZYSIECKI, C. T., COOK, J. C., LEHMAN, E. D., SANDS, J. A., JANSEN, 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. *J Biol Chem,* 274**,** 5810-22.

KALENIK, J. L., CHEN, D., BRADLEY, M. E., CHEN, S. J. & LEE, T. C. 1997. Yeast two-hybrid cloning of a novel zinc finger protein that interacts with the multifunctional transcription factor YY1. *Nucleic Acids Res,* 25**,** 843-9.
KARIMZADEH, M., ARLIDGE, C., ROSTAMI, A., LUPIEN, M., BRATMAN, S. V. & HOFFMAN, M. M. 2022. Human papillomavirus integration transforms chromatin to drive oncogenesis. *bioRxiv***,** 2020.02.12.942755.
KARLEN, S. & BEARD, P. 1993. Identification and characterization of novel promoters in the genome of human papillomavirus type 18. *J Virol,* 67**,** 4296-306.
KASSARDJIAN, A., RIZKALLAH, R., RIMAN, S., RENFRO, S. H., ALEXANDER, K. E. & HURT, M. M. 2012. The transcription factor YY1 is a novel substrate for Aurora B kinase at G2/M transition of the cell cycle. *PLoS One,* 7**,** e50645.
KATERJI, M., BERTUCCI, A., FILIPPOV, V., VAZQUEZ, M., CHEN, X. & DUERKSEN-HUGHES, P. J. 2022. Proton-induced DNA damage promotes integration of foreign plasmid DNA into human genome. *Front Oncol,* 12**,** 928545.
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. *Proc Natl Acad Sci U S A,* 94**,** 11612-6.
KJAER, S. K., SIGURDSSON, K., IVERSEN, O. E., HERNANDEZ-AVILA, M., WHEELER, C. M., PEREZ, G., BROWN, D. R., KOUTSKY, L. A., TAY, E. H., GARCÍA, P., AULT, K. A., GARLAND, S. M., LEODOLTER, S., OLSSON, S. E., TANG, G. W., FERRIS, D. G., PAAVONEN, J., LEHTINEN, M., STEBEN, M., BOSCH, F. X., DILLNER, J., JOURA, E. A., MAJEWSKI, S., MUÑOZ, N., MYERS, E. R., VILLA, L. L., TADDEO, F. J., ROBERTS, C., TADESSE, A., BRYAN, J., MAANSSON, R., LU, S., VUOCOLO, S., HESLEY, T. M., SAAH, A., BARR, E. & HAUPT, R. M. 2009. A pooled analysis of continued prophylactic efficacy of quadrivalent human papillomavirus (Types 6/11/16/18) vaccine against high-grade cervical and external genital lesions. *Cancer Prev Res (Phila),* 2**,** 868-78.
KLAES, R., WOERNER, S. M., RIDDER, R., WENTZENSEN, N., DUERST, M., SCHNEIDER, A., LOTZ, B., MELSHEIMER, P. & VON KNEBEL DOEBERITZ, M. 1999. Detection of high-risk cervical intraepithelial neoplasia and cervical cancer by amplification of transcripts derived from integrated papillomavirus oncogenes. *Cancer Res,* 59**,** 6132-6.
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.
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 Dev,* 16**,** 2893-905.
LABARGE, B., HENNESSY, M., ZHANG, L., GOLDRICH, D., CHARTRAND, S., PURNELL, C., WRIGHT, S., GOLDENBERG, D. & BROACH, J. R. 2022. Human Papillomavirus Integration Strictly Correlates with Global Genome Instability in Head and Neck Cancer. *Mol Cancer Res,* 20**,** 1420-1428.
LACE, M. J., ANSON, J. R., THOMAS, G. S., TUREK, L. P. & HAUGEN, T. H. 2008. The E8--E2 gene product of human papillomavirus type 16 represses early

transcription and replication but is dispensable for viral plasmid persistence in keratinocytes. *J Virol,* 82**,** 10841-53.
LACE, M. J., ISACSON, C., ANSON, J. R., LÖRINCZ, A. T., WILCZYNSKI, S. P., HAUGEN, T. H. & TUREK, L. P. 2009a. Upstream regulatory region alterations found in human papillomavirus type 16 (HPV-16) isolates from cervical carcinomas increase transcription, ori function, and HPV immortalization capacity in culture. *J Virol,* 83**,** 7457-66.
LACE, M. J., YAMAKAWA, Y., USHIKAI, M., ANSON, J. R., HAUGEN, T. H. & TUREK, L. P. 2009b. 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**,** 2402-2412.
LAGSTRÖM, S., LØVESTAD, A. H., UMU, S. U., AMBUR, O. H., NYGÅRD, M., ROUNGE, T. B. & CHRISTIANSEN, I. K. 2021. HPV16 and HPV18 type-specific APOBEC3 and integration profiles in different diagnostic categories of cervical samples. *Tumour Virus Res,* 12**,** 200221.
LAI, A. Y., FATEMI, M., DHASARATHY, A., MALONE, C., SOBOL, S. E., GEIGERMAN, C., JAYE, D. L., MAV, D., SHAH, R., LI, L. & WADE, P. A. 2010. DNA methylation prevents CTCF-mediated silencing of the oncogene BCL6 in B cell lymphomas. *J Exp Med,* 207**,** 1939-50.
LANDEN, C. N., JR., CHAVEZ-REYES, A., BUCANA, C., SCHMANDT, R., DEAVERS, M. T., LOPEZ-BERESTEIN, G. & SOOD, A. K. 2005. Therapeutic EphA2 gene targeting in vivo using neutral liposomal small interfering RNA delivery. *Cancer Res,* 65**,** 6910-8.
LANDY, R., PESOLA, F., CASTAÑÓN, A. & SASIENI, P. 2016. Impact of cervical screening on cervical cancer mortality: estimation using stage-specific results from a nested case-control study. *Br J Cancer,* 115**,** 1140-1146.
LEE, J. S., GALVIN, K. M., SEE, R. H., ECKNER, R., LIVINGSTON, D., MORAN, E. & SHI, Y. 1995a. Relief of YY1 transcriptional repression by adenovirus E1A is mediated by E1A-associated protein p300. *Genes Dev,* 9**,** 1188-98.
LEE, J. S., SEE, R. H., GALVIN, K. M., WANG, J. & SHI, Y. 1995b. Functional interactions between YY1 and adenovirus E1A. *Nucleic Acids Res,* 23**,** 925-31.
LEE, S. S., GLAUNSINGER, B., MANTOVANI, F., BANKS, L. & JAVIER, R. T. 2000. Multi-PDZ domain protein MUPP1 is a cellular target for both adenovirus E4-ORF1 and high-risk papillomavirus type 18 E6 oncoproteins. *J Virol,* 74**,** 9680-93.
LEE, T. C., ZHANG, Y. & SCHWARTZ, R. J. 1994. Bifunctional transcriptional properties of YY1 in regulating muscle actin and c-myc gene expression during myogenesis. *Oncogene,* 9**,** 1047-52.
LEECHANACHAI, P., BANKS, L., MOREAU, F. & MATLASHEWSKI, G. 1992. The E5 gene from human papillomavirus type 16 is an oncogene which enhances growth factor-mediated signal transduction to the nucleus. *Oncogene,* 7**,** 19-25.
LEI, J., PLONER, A., ELFSTRÖM, K. M., WANG, J., ROTH, A., FANG, F., SUNDSTRÖM, K., DILLNER, J. & SPARÉN, P. 2020. HPV Vaccination and the Risk of Invasive Cervical Cancer. *N Engl J Med,* 383**,** 1340-1348.
LEIENDECKER, L., NEUMANN, T., JUNG, P. S., CRONIN, S. M., STEINACKER, T. L., SCHLEIFFER, A., SCHUTZBIER, M., MECHTLER, K., KERVARREC, T.,

LAURENT, E., BACHIRI, K., COYAUD, E., MURALI, R., BUSAM, K. J., ITZINGER-MONSHI, B., KIRNBAUER, R., CERRONI, L., CALONJE, E., RÜTTEN, A., STUBENRAUCH, F., GRIEWANK, K. G., WIESNER, T. & OBENAUF, A. C. 2023. Human Papillomavirus 42 Drives Digital Papillary Adenocarcinoma and Elicits a Germ Cell-like Program Conserved in HPV-Positive Cancers. *Cancer Discov,* 13**,** 70-84.

LI, H., TORABI, S. J., YARBROUGH, W. G., MEHRA, S., OSBORN, H. A. & JUDSON, B. 2018. Association of Human Papillomavirus Status at Head and Neck Carcinoma Subsites With Overall Survival. *JAMA Otolaryngol Head Neck Surg,* 144**,** 519-525.

LI, J., SONG, J. & GUO, F. 2019. miR-186 reverses cisplatin resistance and inhibits the formation of the glioblastoma-initiating cell phenotype by degrading Yin Yang 1 in glioblastoma. *Int J Mol Med,* 43**,** 517-524.

LI, L., WILLIAMS, P., REN, W., WANG, M. Y., GAO, Z., MIAO, W., HUANG, M., SONG, J. & WANG, Y. 2021. YY1 interacts with guanine quadruplexes to regulate DNA looping and gene expression. *Nat Chem Biol,* 17**,** 161-168.

LI, N., FRANCESCHI, S., HOWELL-JONES, R., SNIJDERS, P. J. & CLIFFORD, G. M. 2011a. Human papillomavirus type distribution in 30,848 invasive cervical cancers worldwide: Variation by geographical region, histological type and year of publication. *Int J Cancer,* 128**,** 927-35.

LI, Y. & TOLLEFSBOL, T. O. 2011. DNA methylation detection: bisulfite genomic sequencing analysis. *Methods Mol Biol,* 791**,** 11-21.

LI, Y., WANG, F., XU, J., YE, F., SHEN, Y., ZHOU, J., LU, W., WAN, X., MA, D. & XIE, X. 2011b. Progressive miRNA expression profiles in cervical carcinogenesis and identification of HPV-related target genes for miR-29. *J Pathol,* 224**,** 484-95.

LIN, C., FRANCESCHI, S. & CLIFFORD, G. M. 2018. Human papillomavirus types from infection to cancer in the anus, according to sex and HIV status: a systematic review and meta-analysis. *Lancet Infect Dis,* 18**,** 198-206.

LIPOVSKY, A., POPA, A., PIMIENTA, G., WYLER, M., BHAN, A., KURUVILLA, L., GUIE, M. A., POFFENBERGER, A. C., NELSON, C. D., ATWOOD, W. J. & DIMAIO, D. 2013. Genome-wide siRNA screen identifies the retromer as a cellular entry factor for human papillomavirus. *Proc Natl Acad Sci U S A,* 110**,** 7452-7.

LITOVCHICK, L., SADASIVAM, S., FLORENS, L., ZHU, X., SWANSON, S. K., VELMURUGAN, S., CHEN, R., WASHBURN, M. P., LIU, X. S. & DECAPRIO, J. A. 2007. Evolutionarily conserved multisubunit RBL2/p130 and E2F4 protein complex represses human cell cycle-dependent genes in quiescence. *Mol Cell,* 26**,** 539-51.

LIU, H., XU, J., YANG, Y., WANG, X., WU, E., MAJERCIAK, V., ZHANG, T., STEENBERGEN, R. D. M., WANG, H. K., BANERJEE, N. S., LI, Y., LU, W., MEYERS, C., ZHU, J., XIE, X., CHOW, L. T. & ZHENG, Z. M. 2021. Oncogenic HPV promotes the expression of the long noncoding RNA lnc-FANCI-2 through E7 and YY1. *Proc Natl Acad Sci U S A,* 118.

LIU, J. S., KUO, S. R., BROKER, T. R. & CHOW, L. T. 1995. The functions of human papillomavirus type 11 E1, E2, and E2C proteins in cell-free DNA replication. *J Biol Chem,* 270**,** 27283-91.

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.
LUGER, K., MÄDER, A. W., RICHMOND, R. K., SARGENT, D. F. & RICHMOND, T. J. 1997. Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature,* 389**,** 251-60.
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., SCHLEGEL, R. & HOWLEY, P. M. 1988. The carboxy-terminal domain shared by the bovine papillomavirus E2 transactivator and repressor proteins contains a specific DNA binding activity. *Embo j,* 7**,** 533-9.
MEHANNA, H., ROBINSON, M., HARTLEY, A., KONG, A., FORAN, B., FULTON-LIEUW, T., DALBY, M., MISTRY, P., SEN, M., O'TOOLE, L., AL BOOZ, H., DYKER, K., MOLERON, R., WHITAKER, S., BRENNAN, S., COOK, A., GRIFFIN, M., AYNSLEY, E., ROLLES, M., DE WINTON, E., CHAN, A., SRINIVASAN, D., NIXON, I., GRUMETT, J., LEEMANS, C. R., BUTER, J., HENDERSON, J., HARRINGTON, K., MCCONKEY, C., GRAY, A. & DUNN, J. 2019. Radiotherapy plus cisplatin or cetuximab in low-risk human papillomavirus-positive oropharyngeal cancer (De-ESCALaTE HPV): an open-label randomised controlled phase 3 trial. *Lancet,* 393**,** 51-60.
MEHTA, K. & LAIMINS, L. 2018. Human Papillomaviruses Preferentially Recruit DNA Repair Factors to Viral Genomes for Rapid Repair and Amplification. *mBio,* 9.
MELENDY, T., SEDMAN, J. & STENLUND, A. 1995. Cellular factors required for papillomavirus DNA replication. *J Virol,* 69**,** 7857-67.
MELSHEIMER, P., VINOKUROVA, S., WENTZENSEN, N., BASTERT, G. & VON KNEBEL DOEBERITZ, M. 2004. DNA aneuploidy and integration of human papillomavirus type 16 e6/e7 oncogenes in intraepithelial neoplasia and invasive squamous cell carcinoma of the cervix uteri. *Clin Cancer Res,* 10**,** 3059-63.
MILLIGAN, S. G., VEERAPRADITSIN, T., AHAMET, B., MOLE, S. & GRAHAM, S. V. 2007. Analysis of novel human papillomavirus type 16 late mRNAs in differentiated W12 cervical epithelial cells. *Virology,* 360**,** 172-81.
MIRKOVIC, J., HOWITT, B. E., RONCARATI, P., DEMOULIN, S., SUAREZ-CARMONA, M., HUBERT, P., MCKEON, F. D., XIAN, W., LI, A., DELVENNE, P., CRUM, C. P. & HERFS, M. 2015. Carcinogenic HPV infection in the cervical squamo-columnar junction. *J Pathol,* 236**,** 265-71.
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**,** 1694-9.
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. 2009. Human papillomaviruses activate the ATM DNA damage pathway for viral genome amplification upon differentiation. *PLoS Pathog,* 5**,** e1000605.

MORRISON, J., BALDWIN, P., BUCKLEY, L., COGSWELL, L., EDEY, K., FARUQI, A., GANESAN, R., HALL, M., HILLABY, K., REED, N., ROLLAND, P. & FOTOPOULOU, C. 2020. British Gynaecological Cancer Society (BGCS) vulval cancer guidelines: Recommendations for practice. *Eur J Obstet Gynecol Reprod Biol,* 252**,** 502-525.

MÜNGER, K., PHELPS, W. C., BUBB, V., HOWLEY, P. M. & SCHLEGEL, R. 1989a. 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.

MÜNGER, K., SCHEFFNER, M., HUIBREGTSE, J. M. & HOWLEY, P. M. 1992. Interactions of HPV E6 and E7 oncoproteins with tumour suppressor gene products. *Cancer Surv,* 12**,** 197-217.

MÜNGER, K., WERNESS, B. A., DYSON, N., PHELPS, W. C., HARLOW, E. & HOWLEY, P. M. 1989b. Complex formation of human papillomavirus E7 proteins with the retinoblastoma tumor suppressor gene product. *Embo j,* 8**,** 4099-105.

MUZAFFAR, J., BARI, S., KIRTANE, K. & CHUNG, C. H. 2021. Recent Advances and Future Directions in Clinical Management of Head and Neck Squamous Cell Carcinoma. *Cancers (Basel),* 13.

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. *Mol Cell Biol,* 20**,** 8244-53.

NAKAMOTO, M., ISHIHARA, K., WATANABE, T., HIROSUE, A., HINO, S., SHINOHARA, M., NAKAYAMA, H. & NAKAO, M. 2017. The Glucocorticoid Receptor Regulates the ANGPTL4 Gene in a CTCF-Mediated Chromatin Context in Human Hepatic Cells. *PLoS One,* 12**,** e0169225.

NDIAYE, C., MENA, M., ALEMANY, L., ARBYN, M., CASTELLSAGUÉ, X., LAPORTE, L., BOSCH, F. X., DE SANJOSÉ, S. & TROTTIER, H. 2014. HPV DNA, E6/E7 mRNA, and p16INK4a detection in head and neck cancers: a systematic review and meta-analysis. *Lancet Oncol,* 15**,** 1319-31.

NIEDOBITEK, G., PITTEROFF, S., HERBST, H., SHEPHERD, P., FINN, T., ANAGNOSTOPOULOS, I. & STEIN, H. 1990. Detection of human papillomavirus type 16 DNA in carcinomas of the palatine tonsil. *J Clin Pathol,* 43**,** 918-21.

NILSSON, K., WU, C., KAJITANI, N., YU, H., TSIMTSIRAKIS, E., GONG, L., WINQUIST, E. B., GLAHDER, J., EKBLAD, L., WENNERBERG, J. & SCHWARTZ, S. 2018. The DNA damage response activates HPV16 late gene expression at the level of RNA processing. *Nucleic Acids Res,* 46**,** 5029-5049.

O'BRIEN, J., HAYDER, H., ZAYED, Y. & PENG, C. 2018. Overview of MicroRNA Biogenesis, Mechanisms of Actions, and Circulation. *Front Endocrinol (Lausanne),* 9**,** 402.

O'CONNOR, M. & BERNARD, H. U. 1995. Oct-1 activates the epithelial-specific enhancer of human papillomavirus type 16 via a synergistic interaction with NFI at a conserved composite regulatory element. *Virology,* 207**,** 77-88.

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.

OLTHOF, N. C., HUEBBERS, C. U., KOLLIGS, J., HENFLING, M., RAMAEKERS, F. C., CORNET, I., VAN LENT-ALBRECHTS, J. A., STEGMANN, A. P.,

SILLING, S., WIELAND, U., CAREY, T. E., WALLINE, H. M., GOLLIN, S. M., HOFFMANN, T. K., DE WINTER, J., KREMER, B., KLUSSMANN, J. P. & SPEEL, E. J. 2015. Viral load, gene expression and mapping of viral integration sites in HPV16-associated HNSCC cell lines. *Int J Cancer,* 136**,** E207-18.

ORTH, G. 1986. Epidermodysplasia verruciformis: a model for understanding the oncogenicity of human papillomaviruses. *Ciba Found Symp,* 120**,** 157-74.

OZBUN, M. A. 2002. Human papillomavirus type 31b infection of human keratinocytes and the onset of early transcription. *J Virol,* 76**,** 11291-300.

PAAVONEN, J., JENKINS, D., BOSCH, F. X., NAUD, P., SALMERÓN, J., WHEELER, C. M., CHOW, S. N., APTER, D. L., KITCHENER, H. C., CASTELLSAGUE, X., DE CARVALHO, N. S., SKINNER, S. R., HARPER, D. M., HEDRICK, J. A., JAISAMRARN, U., LIMSON, G. A., DIONNE, M., QUINT, W., SPIESSENS, B., PEETERS, P., STRUYF, F., WIETING, S. L., LEHTINEN, M. O. & DUBIN, G. 2007. Efficacy of a prophylactic adjuvanted bivalent L1 virus-like-particle vaccine against infection with human papillomavirus types 16 and 18 in young women: an interim analysis of a phase III double-blind, randomised controlled trial. *Lancet,* 369**,** 2161-2170.

PARFENOV, M., PEDAMALLU, C. S., GEHLENBORG, N., FREEMAN, S. S., DANILOVA, L., BRISTOW, C. A., LEE, S., HADJIPANAYIS, A. G., IVANOVA, E. V., WILKERSON, M. D., PROTOPOPOV, A., YANG, L., SETH, S., SONG, X., TANG, J., REN, X., ZHANG, J., PANTAZI, A., SANTOSO, N., XU, A. W., MAHADESHWAR, H., WHEELER, D. A., HADDAD, R. I., JUNG, J., OJESINA, A. I., ISSAEVA, N., YARBROUGH, W. G., HAYES, D. N., GRANDIS, J. R., EL-NAGGAR, A. K., MEYERSON, M., PARK, P. J., CHIN, L., SEIDMAN, J. G., HAMMERMAN, P. S. & KUCHERLAPATI, R. 2014. Characterization of HPV and host genome interactions in primary head and neck cancers. *Proc Natl Acad Sci U S A,* 111**,** 15544-9.

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. *J Virol,* 89**,** 4770-85.

PARISH, J. L., BEAN, A. M., PARK, R. B. & ANDROPHY, E. J. 2006. ChlR1 is required for loading papillomavirus E2 onto mitotic chromosomes and viral genome maintenance. *Mol Cell,* 24**,** 867-76.

PARK, K. & ATCHISON, M. L. 1991. Isolation of a candidate repressor/activator, NF-E1 (YY-1, delta), that binds to the immunoglobulin kappa 3' enhancer and the immunoglobulin heavy-chain mu E1 site. *Proc Natl Acad Sci U S A,* 88**,** 9804-8.

PARK, N. H., MIN, B. M., LI, S. L., HUANG, M. Z., CHERICK, H. M. & DONIGER, J. 1991. Immortalization of normal human oral keratinocytes with type 16 human papillomavirus. *Carcinogenesis,* 12**,** 1627-31.

PEDERSEN, C., PETAJA, T., STRAUSS, G., RUMKE, H. C., PODER, A., RICHARDUS, J. H., SPIESSENS, B., DESCAMPS, D., HARDT, K., LEHTINEN, M. & DUBIN, G. 2007. Immunization of early adolescent females with human papillomavirus type 16 and 18 L1 virus-like particle vaccine containing AS04 adjuvant. *J Adolesc Health,* 40**,** 564-71.

PENTLAND, I., CAMPOS-LEÓN, K., COTIC, M., DAVIES, K. J., WOOD, C. D., GROVES, I. J., BURLEY, M., COLEMAN, N., STOCKTON, J. D., NOYVERT,

B., BEGGS, A. D., WEST, M. J., ROBERTS, S. & PARISH, J. L. 2018. Disruption of CTCF-YY1-dependent looping of the human papillomavirus genome activates differentiation-induced viral oncogene transcription. *PLoS Biol,* 16**,** e2005752.
PENTLAND, I. & PARISH, J. L. 2015. Targeting CTCF to Control Virus Gene Expression: A Common Theme amongst Diverse DNA Viruses. *Viruses,* 7**,** 3574-85.
PERKINS, R. B., GUIDO, R. S., CASTLE, P. E., CHELMOW, D., EINSTEIN, M. H., GARCIA, F., HUH, W. K., KIM, J. J., MOSCICKI, A. B., NAYAR, R., SARAIYA, M., SAWAYA, G. F., WENTZENSEN, N. & SCHIFFMAN, M. 2020. 2019 ASCCP Risk-Based Management Consensus Guidelines for Abnormal Cervical Cancer Screening Tests and Cancer Precursors. *J Low Genit Tract Dis,* 24**,** 102-131.
PFISTER, D. G., SPENCER, S., ADELSTEIN, D., ADKINS, D., ANZAI, Y., BRIZEL, D. M., BRUCE, J. Y., BUSSE, P. M., CAUDELL, J. J., CMELAK, A. J., COLEVAS, A. D., EISELE, D. W., FENTON, M., FOOTE, R. L., GALLOWAY, T., GILLISON, M. L., HADDAD, R. I., HICKS, W. L., HITCHCOCK, Y. J., JIMENO, A., LEIZMAN, D., MAGHAMI, E., MELL, L. K., MITTAL, B. B., PINTO, H. A., RIDGE, J. A., ROCCO, J. W., RODRIGUEZ, C. P., SHAH, J. P., WEBER, R. S., WEINSTEIN, G., WITEK, M., WORDEN, F., YOM, S. S., ZHEN, W., BURNS, J. L. & DARLOW, S. D. 2020. Head and Neck Cancers, Version 2.2020, NCCN Clinical Practice Guidelines in Oncology. *J Natl Compr Canc Netw,* 18**,** 873-898.
PIM, D., COLLINS, M. & BANKS, L. 1992. Human papillomavirus type 16 E5 gene stimulates the transforming activity of the epidermal growth factor receptor. *Oncogene,* 7**,** 27-32.
PINTO, L. A., DILLNER, J., BEDDOWS, S. & UNGER, E. R. 2018. Immunogenicity of HPV prophylactic vaccines: Serology assays and their use in HPV vaccine evaluation and development. *Vaccine,* 36**,** 4792-4799.
PIRAMI, L., GIACHÈ, V. & BECCIOLINI, A. 1997. Analysis of HPV16, 18, 31, and 35 DNA in pre-invasive and invasive lesions of the uterine cervix. *J Clin Pathol,* 50**,** 600-4.
PIUNTI, A. & SHILATIFARD, A. 2022. Author Correction: The roles of Polycomb repressive complexes in mammalian development and cancer. *Nat Rev Mol Cell Biol,* 23**,** 444.
RAJARAM, S., MAHESHWARI, A. & SRIVASTAVA, A. 2015. Staging for vaginal cancer. *Best Pract Res Clin Obstet Gynaecol,* 29**,** 822-32.
RAMAZI, S. & ZAHIRI, J. 2021. Posttranslational modifications in proteins: resources, tools and prediction methods. *Database (Oxford),* 2021.
RAMKUMAR, C., CUI, H., KONG, Y., JONES, S. N., GERSTEIN, R. M. & ZHANG, H. 2013. Smurf2 suppresses B-cell proliferation and lymphomagenesis by mediating ubiquitination and degradation of YY1. *Nat Commun,* 4**,** 2598.
REBOLJ, M., PESOLA, F., MATHEWS, C., MESHER, D., SOLDAN, K. & KITCHENER, H. 2022. The impact of catch-up bivalent human papillomavirus vaccination on cervical screening outcomes: an observational study from the English HPV primary screening pilot. *Br J Cancer,* 127**,** 278-287.
REN, C., ZHANG, G., HAN, F., FU, S., CAO, Y., ZHANG, F., ZHANG, Q., MESLAMANI, J., XU, Y., JI, D., CAO, L., ZHOU, Q., CHEUNG, K. L., SHARMA, R., BABAULT, N., YI, Z., ZHANG, W., WALSH, M. J., ZENG, L. &

ZHOU, M. M. 2018. Spatially constrained tandem bromodomain inhibition bolsters sustained repression of BRD4 transcriptional activity for TNBC cell growth. *Proc Natl Acad Sci U S A,* 115**,** 7949-7954.
REZAI-ZADEH, N., ZHANG, X., NAMOUR, F., FEJER, G., WEN, Y. D., YAO, Y. L., GYORY, I., WRIGHT, K. & SETO, E. 2003. Targeted recruitment of a histone H4-specific methyltransferase by the transcription factor YY1. *Genes Dev,* 17**,** 1019-29.
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.
RIZKALLAH, R., ALEXANDER, K. E., KASSARDJIAN, A., LÜSCHER, B. & HURT, M. M. 2011. The transcription factor YY1 is a substrate for Polo-like kinase 1 at the G2/M transition of the cell cycle. *PLoS One,* 6**,** e15928.
RIZKALLAH, R. & HURT, M. M. 2009. Regulation of the transcription factor YY1 in mitosis through phosphorylation of its DNA-binding domain. *Mol Biol Cell,* 20**,** 4766-76.
ROBERTS, S., EVANS, D., MEHANNA, H. & PARISH, J. L. 2019. Modelling human papillomavirus biology in oropharyngeal keratinocytes. *Philos Trans R Soc Lond B Biol Sci,* 374**,** 20180289.
ROMMEL, O., DILLNER, J., FLIGGE, C., BERGSDORF, C., WANG, X., SELINKA, H. C. & SAPP, M. 2005. Heparan sulfate proteoglycans interact exclusively with conformationally intact HPV L1 assemblies: basis for a virus-like particle ELISA. *J Med Virol,* 75**,** 114-21.
RUESCH, M. N., STUBENRAUCH, F. & LAIMINS, L. A. 1998. Activation of papillomavirus late gene transcription and genome amplification upon differentiation in semisolid medium is coincident with expression of involucrin and transglutaminase but not keratin-10. *J Virol,* 72**,** 5016-24.
SADASIVAM, S. & DECAPRIO, J. A. 2013. The DREAM complex: master coordinator of cell cycle-dependent gene expression. *Nat Rev Cancer,* 13**,** 585-95.
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.
SATIJN, D. P., HAMER, K. M., DEN BLAAUWEN, J. & OTTE, A. P. 2001. The polycomb group protein EED interacts with YY1, and both proteins induce neural tissue in Xenopus embryos. *Mol Cell Biol,* 21**,** 1360-9.
SCHEFFER, K. D., GAWLITZA, A., SPODEN, G. A., ZHANG, X. A., LAMBERT, C., BERDITCHEVSKI, F. & FLORIN, L. 2013. Tetraspanin CD151 mediates papillomavirus type 16 endocytosis. *J Virol,* 87**,** 3435-46.
SCHEFFNER, M., HUIBREGTSE, J. M., VIERSTRA, R. D. & HOWLEY, P. M. 1993. The HPV-16 E6 and E6-AP complex functions as a ubiquitin-protein ligase in the ubiquitination of p53. *Cell,* 75**,** 495-505.
SCHELHAAS, M., SHAH, B., HOLZER, M., BLATTMANN, P., KÜHLING, L., DAY, P. M., SCHILLER, J. T. & HELENIUS, A. 2012. Entry of human papillomavirus type 16 by actin-dependent, clathrin- and lipid raft-independent endocytosis. *PLoS Pathog,* 8**,** e1002657.
SCHILLER, J. T., CASTELLSAGUÉ, X. & GARLAND, S. M. 2012. A review of clinical trials of human papillomavirus prophylactic vaccines. *Vaccine,* 30 Suppl 5**,** F123-38.

SCHNOELL, J., JANK, B. J., KADLETZ-WANKE, L., STOIBER, S., SPIELVOGEL, C. P., GURNHOFER, E., KENNER, L. & HEIDUSCHKA, G. 2021. Transcription factors CP2 and YY1 as prognostic markers in head and neck squamous cell carcinoma: analysis of The Cancer Genome Atlas and a second independent cohort. *J Cancer Res Clin Oncol,* 147**,** 755-765.
SCHWARZ, E., FREESE, U. K., GISSMANN, L., MAYER, W., ROGGENBUCK, B., STREMLAU, A. & ZUR HAUSEN, H. 1985. Structure and transcription of human papillomavirus sequences in cervical carcinoma cells. *Nature,* 314**,** 111-4.
SELINKA, H. C., FLORIN, L., PATEL, H. D., FREITAG, K., SCHMIDTKE, M., MAKAROV, V. A. & SAPP, M. 2007. Inhibition of transfer to secondary receptors by heparan sulfate-binding drug or antibody induces noninfectious uptake of human papillomavirus. *J Virol,* 81**,** 10970-80.
SEO, Y. S., MÜLLER, F., LUSKY, M. & HURWITZ, J. 1993. Bovine papilloma virus (BPV)-encoded E1 protein contains multiple activities required for BPV DNA replication. *Proc Natl Acad Sci U S A,* 90**,** 702-6.
SETO, E., LEWIS, B. & SHENK, T. 1993. Interaction between transcription factors Sp1 and YY1. *Nature,* 365**,** 462-4.
SEWALT, R. G., VAN DER VLAG, J., GUNSTER, M. J., HAMER, K. M., DEN BLAAUWEN, J. L., SATIJN, D. P., HENDRIX, T., VAN DRIEL, R. & OTTE, A. P. 1998. Characterization of interactions between the mammalian polycomb-group proteins Enx1/EZH2 and EED suggests the existence of different mammalian polycomb-group protein complexes. *Mol Cell Biol,* 18**,** 3586-95.
SHI, Y., SETO, E., CHANG, L. S. & SHENK, T. 1991. Transcriptional repression by YY1, a human GLI-Krüppel-related protein, and relief of repression by adenovirus E1A protein. *Cell,* 67**,** 377-88.
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.
SIEGEL, R. L., MILLER, K. D., FUCHS, H. E. & JEMAL, A. 2022. Cancer statistics, 2022. *CA Cancer J Clin,* 72**,** 7-33.
SIMON, J. M., GIRESI, P. G., DAVIS, I. J. & LIEB, J. D. 2012. Using formaldehyde-assisted isolation of regulatory elements (FAIRE) to isolate active regulatory DNA. *Nat Protoc,* 7**,** 256-67.
SINGG, N., ROUS, B AND GANESAN, R. 2018. 2018 FIGO staging system for cervical cancer: summary and comparison with 2009 FIGO staging system. *British Association of Gynaecological Pathologists*
SMITH, J. L., CAMPOS, S. K. & OZBUN, M. A. 2007. Human papillomavirus type 31 uses a caveolin 1- and dynamin 2-mediated entry pathway for infection of human keratinocytes. *J Virol,* 81**,** 9922-31.
SMOTKIN, D., PROKOPH, H. & WETTSTEIN, F. O. 1989. Oncogenic and nononcogenic human genital papillomaviruses generate the E7 mRNA by different mechanisms. *J Virol,* 63**,** 1441-7.
SPODEN, G., FREITAG, K., HUSMANN, M., BOLLER, K., SAPP, M., LAMBERT, C. & FLORIN, L. 2008. Clathrin- and caveolin-independent entry of human papillomavirus type 16--involvement of tetraspanin-enriched microdomains (TEMs). *PLoS One,* 3**,** e3313.

STANLEY, M. A., BROWNE, H. M., APPLEBY, M. & MINSON, A. C. 1989. Properties of a non-tumorigenic human cervical keratinocyte cell line. *Int J Cancer,* 43**,** 672-6.
STICH, M., GANSS, L., PUSCHHOF, J., PRIGGE, E. S., REUSCHENBACH, M., GUITERREZ, A., VINOKUROVA, S. & VON KNEBEL DOEBERITZ, M. 2017. 5-aza-2'-deoxycytidine (DAC) treatment downregulates the HPV E6 and E7 oncogene expression and blocks neoplastic growth of HPV-associated cancer cells. *Oncotarget,* 8**,** 52104-52117.
STRAIGHT, S. W., HINKLE, P. M., JEWERS, R. J. & MCCANCE, D. J. 1993. The E5 oncoprotein of human papillomavirus type 16 transforms fibroblasts and effects the downregulation of the epidermal growth factor receptor in keratinocytes. *J Virol,* 67**,** 4521-32.
STRAUB, E., FERTEY, J., DREER, M., IFTNER, T. & STUBENRAUCH, F. 2015. Characterization of the Human Papillomavirus 16 E8 Promoter. *J Virol,* 89**,** 7304-13.
STÜNKEL, 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.
SUNG, H., FERLAY, J., SIEGEL, R. L., LAVERSANNE, M., SOERJOMATARAM, I., JEMAL, A. & BRAY, F. 2021. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin,* 71**,** 209-249.
SYRJÄNEN, K., SYRJÄNEN, S., LAMBERG, M., PYRHÖNEN, S. & NUUTINEN, J. 1983. Morphological and immunohistochemical evidence suggesting human papillomavirus (HPV) involvement in oral squamous cell carcinogenesis. *Int J Oral Surg,* 12**,** 418-24.
SYRJÄNEN, S., SYRJÄNEN, K., MÄNTYJÄRVI, R., COLLAN, Y. & KÄRJÄ, J. 1987. Human papillomavirus DNA in squamous cell carcinomas of the larynx demonstrated by in situ DNA hybridization. *ORL J Otorhinolaryngol Relat Spec,* 49**,** 175-86.
TAKAGAKI, Y., RYNER, L. C. & MANLEY, J. L. 1989. Four factors are required for 3'-end cleavage of pre-mRNAs. *Genes Dev,* 3**,** 1711-24.
TAN, S. H., LEONG, L. E., WALKER, P. A. & BERNARD, H. U. 1994. The human papillomavirus type 16 E2 transcription factor binds with low cooperativity to two flanking sites and represses the E6 promoter through displacement of Sp1 and TFIID. *J Virol,* 68**,** 6411-20.
TANG, A. L., HAUFF, S. J., OWEN, J. H., GRAHAM, M. P., CZERWINSKI, M. J., PARK, J. J., WALLINE, H., PAPAGERAKIS, S., STOERKER, J., MCHUGH, J. B., CHEPEHA, D. B., BRADFORD, C. R., CAREY, T. E. & PRINCE, M. E. 2012. UM-SCC-104: a new human papillomavirus-16-positive cancer stem cell-containing head and neck squamous cell carcinoma cell line. *Head Neck,* 34**,** 1480-91.
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. *Cell,* 148**,** 664-78.
TEMPERA, I., KLICHINSKY, M. & LIEBERMAN, P. M. 2011. EBV latency types adopt alternative chromatin conformations. *PLoS Pathog,* 7**,** e1002180.

THIERRY, F., HEARD, J. M., DARTMANN, K. & YANIV, M. 1987. Characterization of a transcriptional promoter of human papillomavirus 18 and modulation of its expression by simian virus 40 and adenovirus early antigens. *J Virol,* 61**,** 134-42.
THIERRY, F. & YANIV, M. 1987. The BPV1-E2 trans-acting protein can be either an activator or a repressor of the HPV18 regulatory region. *Embo j,* 6**,** 3391-7.
THOMAS, M., LAURA, R., HEPNER, K., GUCCIONE, E., SAWYERS, C., LASKY, L. & BANKS, L. 2002. Oncogenic human papillomavirus E6 proteins target the MAGI-2 and MAGI-3 proteins for degradation. *Oncogene,* 21**,** 5088-96.
THOMAS, M., MYERS, M. P., MASSIMI, P., GUARNACCIA, C. & BANKS, L. 2016. Analysis of Multiple HPV E6 PDZ Interactions Defines Type-Specific PDZ Fingerprints That Predict Oncogenic Potential. *PLoS Pathog,* 12**,** e1005766.
THORVALDSDÓTTIR, H., ROBINSON, J. T. & MESIROV, J. P. 2013. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform,* 14**,** 178-92.
TIAN, B. & MANLEY, J. L. 2017. Alternative polyadenylation of mRNA precursors. *Nat Rev Mol Cell Biol,* 18**,** 18-30.
TOOTS, M., MÄNNIK, A., KIVI, G., USTAV, M., JR., USTAV, E. & USTAV, M. 2014. The transcription map of human papillomavirus type 18 during genome replication in U2OS cells. *PLoS One,* 9**,** e116151.
TRUS, B. L., RODEN, R. B., GREENSTONE, H. L., VRHEL, M., SCHILLER, J. T. & BOOY, F. P. 1997. Novel structural features of bovine papillomavirus capsid revealed by a three-dimensional reconstruction to 9 A resolution. *Nat Struct Biol,* 4**,** 413-20.
USHEVA, A. & SHENK, T. 1994. TATA-binding protein-independent initiation: YY1, TFIIB, and RNA polymerase II direct basal transcription on supercoiled template DNA. *Cell,* 76**,** 1115-21.
USHEVA, A. & SHENK, T. 1996. YY1 transcriptional initiator: protein interactions and association with a DNA site containing unpaired strands. *Proc Natl Acad Sci U S A,* 93**,** 13571-6.
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.
VAN DOORSLAER, K., CHEN, Z., BERNARD, H. U., CHAN, P. K. S., DESALLE, R., DILLNER, J., FORSLUND, O., HAGA, T., MCBRIDE, A. A., VILLA, L. L., BURK, R. D. & ICTV REPORT, C. 2018. ICTV Virus Taxonomy Profile: Papillomaviridae. *J Gen Virol,* 99**,** 989-990.
VAN TINE, B. A., KAPPES, J. C., BANERJEE, N. S., KNOPS, J., LAI, L., STEENBERGEN, R. D., MEIJER, C. L., SNIJDERS, P. J., CHATIS, P., BROKER, T. R., MOEN, P. T., JR. & CHOW, L. T. 2004. Clonal selection for transcriptionally active viral oncogenes during progression to cancer. *J Virol,* 78**,** 11172-86.
VERHEUL, T. C. J., VAN HIJFTE, L., PERENTHALER, E. & BARAKAT, T. S. 2020. The Why of YY1: Mechanisms of Transcriptional Regulation by Yin Yang 1. *Front Cell Dev Biol,* 8**,** 592164.
VILLA, L. L., COSTA, R. L., PETTA, C. A., ANDRADE, R. P., AULT, K. A., GIULIANO, A. R., WHEELER, C. M., KOUTSKY, L. A., MALM, C., LEHTINEN, M., SKJELDESTAD, F. E., OLSSON, S. E., STEINWALL, M., BROWN, D. R., KURMAN, R. J., RONNETT, B. M., STOLER, M. H., FERENCZY, A.,

HARPER, D. M., TAMMS, G. M., YU, J., LUPINACCI, L., RAILKAR, R., TADDEO, F. J., JANSEN, K. U., ESSER, M. T., SINGS, H. L., SAAH, A. J. & BARR, E. 2005. Prophylactic quadrivalent human papillomavirus (types 6, 11, 16, and 18) L1 virus-like particle vaccine in young women: a randomised double-blind placebo-controlled multicentre phase II efficacy trial. *Lancet Oncol,* 6**,** 271-8.

VINOKUROVA, S., WENTZENSEN, N., KRAUS, I., KLAES, R., DRIESCH, C., MELSHEIMER, P., KISSELJOV, F., DÜRST, M., SCHNEIDER, A. & VON KNEBEL DOEBERITZ, M. 2008. Type-dependent integration frequency of human papillomavirus genomes in cervical lesions. *Cancer Res,* 68**,** 307-13.

WAGNER, M. J., MITRA, R., MCARTHUR, M. J., BAZE, W., BARNHART, K., WU, S. Y., RODRIGUEZ-AGUAYO, C., ZHANG, X., COLEMAN, R. L., LOPEZ-BERESTEIN, G. & SOOD, A. K. 2017. Preclinical Mammalian Safety Studies of EPHARNA (DOPC Nanoliposomal EphA2-Targeted siRNA). *Mol Cancer Ther,* 16**,** 1114-1123.

WALBOOMERS, J. M., JACOBS, M. V., MANOS, M. M., BOSCH, F. X., KUMMER, J. A., SHAH, K. V., SNIJDERS, P. J., PETO, J., MEIJER, C. J. & MUÑOZ, N. 1999. Human papillomavirus is a necessary cause of invasive cervical cancer worldwide. *J Pathol,* 189**,** 12-9.

WALLINE, H. M., GOUDSMIT, C. M., MCHUGH, J. B., TANG, A. L., OWEN, J. H., TEH, B. T., MCKEAN, E., GLOVER, T. W., GRAHAM, M. P., PRINCE, M. E., CHEPEHA, D. B., CHINN, S. B., FERRIS, R. L., GOLLIN, S. M., HOFFMANN, T. K., BIER, H., BRAKENHOFF, R., BRADFORD, C. R. & CAREY, T. E. 2017. Integration of high-risk human papillomavirus into cellular cancer-related genes in head and neck cancer cell lines. *Head Neck,* 39**,** 840-852.

WALOWITZ, J. L., BRADLEY, M. E., CHEN, S. & LEE, T. 1998. Proteolytic regulation of the zinc finger transcription factor YY1, a repressor of muscle-restricted gene expression. *J Biol Chem,* 273**,** 6656-61.

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 Res,* 22**,** 1680-8.

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

WANG, L., LAING, J., YAN, B., ZHOU, H., KE, L., WANG, C., NARITA, Y., ZHANG, Z., OLSON, M. R., AFZALI, B., ZHAO, B. & KAZEMIAN, M. 2020. Epstein-Barr Virus Episome Physically Interacts with Active Regions of the Host Genome in Lymphoblastoid Cells. *J Virol,* 94.

WANG, W., QIAO, S., LI, G., CHENG, J., YANG, C., ZHONG, C., STOVALL, D. B., SHI, J., TENG, C., LI, D. & SUI, G. 2022. A histidine cluster determines YY1-compartmentalized coactivators and chromatin elements in phase-separated enhancer clusters. *Nucleic Acids Res,* 50**,** 4917-4937.

WANG, W., YUE, Z., TIAN, Z., XIE, Y., ZHANG, J., SHE, Y., YANG, B., YE, Y. & YANG, Y. 2018. Expression of Yin Yang 1 in cervical cancer and its correlation with E-cadherin expression and HPV16 E6. *PLoS One,* 13**,** e0193340.

WANG, X., 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. *J Virol,* 85**,** 8080-92.
WARBURTON, A., MARKOWITZ, T. E., KATZ, J. P., PIPAS, J. M. & MCBRIDE, A. A. 2021. Recurrent integration of human papillomavirus genomes at transcriptional regulatory hubs. *NPJ Genom Med,* 6**,** 101.
WARBURTON, A., REDMOND, C. J., DOOLEY, K. E., FU, H., GILLISON, M. L., AKAGI, K., SYMER, D. E., ALADJEM, M. I. & MCBRIDE, A. A. 2018. HPV integration hijacks and multimerizes a cellular enhancer to generate a viral-cellular super-enhancer that drives high viral oncogene expression. *PLoS Genet,* 14**,** e1007179.
WEINTRAUB, A. S., LI, C. H., ZAMUDIO, A. V., SIGOVA, A. A., HANNETT, N. M., DAY, D. S., ABRAHAM, B. J., COHEN, M. A., NABET, B., BUCKLEY, D. L., GUO, Y. E., HNISZ, D., JAENISCH, R., BRADNER, J. E., GRAY, N. S. & YOUNG, R. A. 2017. YY1 Is a Structural Regulator of Enhancer-Promoter Loops. *Cell,* 171**,** 1573-1588.e28.
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.
WIECZOREK, M., ABUALROUS, E. T., STICHT, J., ÁLVARO-BENITO, M., STOLZENBERG, S., NOÉ, F. & FREUND, C. 2017. Major Histocompatibility Complex (MHC) Class I and MHC Class II Proteins: Conformational Plasticity in Antigen Presentation. *Front Immunol,* 8**,** 292.
WILKINSON, F., PRATT, H. & ATCHISON, M. L. 2010. PcG recruitment by the YY1 REPO domain can be mediated by Yaf2. *J Cell Biochem,* 109**,** 478-86.
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.
WILSON, R., FEHRMANN, F. & LAIMINS, L. A. 2005. Role of the E1--E4 protein in the differentiation-dependent life cycle of human papillomavirus type 31. *J Virol,* 79**,** 6732-40.
WINDER, D. M., PETT, M. R., FOSTER, N., SHIVJI, M. K., HERDMAN, M. T., STANLEY, M. A., VENKITARAMAN, A. R. & COLEMAN, N. 2007. An increase in DNA double-strand breaks, induced by Ku70 depletion, is associated with human papillomavirus 16 episome loss and de novo viral integration events. *J Pathol,* 213**,** 27-34.
WOODHALL, S. C., JIT, M., SOLDAN, K., KINGHORN, G., GILSON, R., NATHAN, M., ROSS, J. D. & LACEY, C. J. 2011. The impact of genital warts: loss of quality of life and cost of treatment in eight sexual health clinics in the UK. *Sex Transm Infect,* 87**,** 458-63.
WOODMAN, C. B., COLLINS, S. I. & YOUNG, L. S. 2007. The natural history of cervical HPV infection: unresolved issues. *Nat Rev 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**,** 371-80.
WU, S. Y., LEE, A. Y., LAI, H. T., ZHANG, H. & CHIANG, C. M. 2013. Phospho switch triggers Brd4 chromatin binding and activator recruitment for gene-specific targeting. *Mol Cell,* 49**,** 843-57.

XU, B., CHOTEWUTMONTRI, S., WOLF, S., KLOS, U., SCHMITZ, M., DÜRST, M. & SCHWARZ, E. 2013. Multiplex Identification of Human Papillomavirus 16 DNA Integration Sites in Cervical Carcinomas. *PLoS One,* 8**,** e66693.

XU, C., TSAI, Y. H., GALBO, P. M., GONG, W., STOREY, A. J., XU, Y., BYRUM, S. D., XU, L., WHANG, Y. E., PARKER, J. S., MACKINTOSH, S. G., EDMONDSON, R. D., TACKETT, A. J., HUANG, J., ZHENG, D., EARP, H. S., WANG, G. G. & CAI, L. 2021. Cistrome analysis of YY1 uncovers a regulatory axis of YY1:BRD2/4-PFKP during tumorigenesis of advanced prostate cancer. *Nucleic Acids Res,* 49**,** 4971-4988.

YAMAMOTO, K., SONODA, M., INOKUCHI, J., SHIRASAWA, S. & SASAZUKI, T. 2004. Polycomb group suppressor of zeste 12 links heterochromatin protein 1alpha and enhancer of zeste 2. *J Biol Chem,* 279**,** 401-6.

YANG, B., LI, B., JIA, L., JIANG, Y., WANG, X., JIANG, S., DU, S., JI, X. & YANG, P. 2020. 3D landscape of Hepatitis B virus interactions with human chromatins. *Cell Discov,* 6**,** 95.

YAO, Y. L., YANG, W. M. & SETO, E. 2001. Regulation of transcription factor YY1 by acetylation and deacetylation. *Mol Cell Biol,* 21**,** 5979-91.

YOUSEFI, Z., ARIA, H., GHAEDRAHMATI, F., BAKHTIARI, T., AZIZI, M., BASTAN, R., HOSSEINI, R. & ESKANDARI, N. 2021. An Update on Human Papilloma Virus Vaccines: History, Types, Protection, and Efficacy. *Front Immunol,* 12**,** 805695.

ZHANG, W. J., WU, X. N., SHI, T. T., XU, H. T., YI, J., SHEN, H. F., HUANG, M. F., SHU, X. Y., WANG, F. F., PENG, B. L., XIAO, R. Q., GAO, W. W., DING, J. C. & LIU, W. 2016. Regulation of Transcription Factor Yin Yang 1 by SET7/9-mediated Lysine Methylation. *Sci Rep,* 6**,** 21718.

ZHANG, Y., CHAN, H. L., GARCIA-MARTINEZ, L., KARL, D. L., WEICH, N., SLINGERLAND, J. M., VERDUN, R. E. & MOREY, L. 2020. Estrogen induces dynamic ERα and RING1B recruitment to control gene and enhancer activities in luminal breast cancer. *Sci Adv,* 6**,** eaaz7249.

ZHANG, Y., LIU, T., YUAN, F., GARCIA-MARTINEZ, L., LEE, K. D., STRANSKY, S., SIDOLI, S., VERDUN, R. E., ZHANG, Y., WANG, Z. & MOREY, L. 2021. The Polycomb protein RING1B enables estrogen-mediated gene expression by promoting enhancer-promoter interaction and R-loop formation. *Nucleic Acids Res,* 49**,** 9768-9782.

ZHAO, J., HYMAN, L. & MOORE, C. 1999. Formation of mRNA 3' ends in eukaryotes: mechanism, regulation, and interrelationships with other steps in mRNA synthesis. *Microbiol Mol Biol Rev,* 63**,** 405-45.

ZHAO, L., LI, R. & GAN, Y. H. 2018. Knockdown of Yin Yang 1 enhances anticancer effects of cisplatin through protein phosphatase 2A-mediated T308 dephosphorylation of AKT. *Cell Death Dis,* 9**,** 747.

ZHOU, P., YU, Y. F., LIAN, C. L., WANG, J., ZHUO, R. G. & WU, S. G. 2021. Survival Outcomes and Treatment Decision by Human Papillomavirus Status Among Patients With Stage IVC Head and Neck Squamous Cell Carcinoma. *Front Oncol,* 11**,** 668066.

ZOU, N., LIN, B. Y., DUAN, F., LEE, K. Y., JIN, G., GUAN, R., YAO, G., LEFKOWITZ, E. J., BROKER, T. R. & CHOW, L. T. 2000. The hinge of the human papillomavirus type 11 E2 protein contains major determinants for nuclear localization and nuclear matrix association. *J Virol,* 74**,** 3761-70.