Department of Molecular & Clinical Cancer Medicine

Institute of Translational Medicine

The Molecular & Clinical Implications of Human papillomavirus-16 Mediated Oropharyngeal Squamous Cell Carcinoma

Thesis submitted in accordance with the requirements
of the University of Liverpool for the degree of Doctor in Philosophy by
Andrew Graeme Schache

August 2013

Supervisors: Triantafillos Liloglou, Janet M Risk & Richard J Shaw

To my wonderful family

Disclaimer:

The entirety of the work presented in this thesis, unless otherwise stated, is that of the author.

ABSTRACT

The Molecular & Clinical Implications of Human papillomavirus-16 Mediated Oropharyngeal Squamous Cell Carcinoma

Andrew Graeme Schache

The last three decades have seen a fundamental change in the profile of oropharyngeal squamous cell carcinoma (OPSCC) within the developed world. The incidence of OPSCC attributable to tobacco and alcohol exposure has been gradually declining whilst Human papillomavirus (HPV)-related OPSCC has seen a rapid increase. Detection of High Risk HPV has profound prognostic significance as it correlates with both a disease-specific and an overall survival advantage. The stringency of testing, both in terms of diagnostic and prognostic capacity is therefore of increasing importance. This study sought to define the relative abilities of the diagnostic tests presently available in clinical practice and to explore the potential of a novel test in reaching the improved stringency called for by the clinical community.

Diagnostic biomarkers with prognostic capacity, such as those utilised in defining HPV status in this research have been well described, however, despite HPV positive OPSCC being biologically distinct from HPV negative malignancy, predictive biomarkers defining the transition from persistent to transforming infection are yet to be forthcoming. A lack of an apparent premalignant state, akin to that seen in

HPV-mediated cervical malignancy has restricted biomarker recognition. This research aimed to better define the epigenetic state and clarify the impact of viral integration for the virus and host in HPV positive OPSCC. Although detectable epigenetic alterations, within the genome of the virus and that of the host, were capable of providing an improved description of this burgeoning disease state, they fell short of providing clinically relevant biomarkers. It was however demonstrated that the previously held concept of preferential E2 cleavage during viral integration as a means to disrupt gene expression, is overstated and the model persists to the exclusion of other viral and host genome disruptions. A paradigm shift may be necessary in HPV positive OPSCC to an understanding of obligatory viral integration, the significance of which however, is yet to be fully elucidated.

Abstract iv

List of Tables xii

List of Figures xvi

List of Appendices xix

Acknowledgements xxi

1 Thesis Aims 1

2 Introduction: Human Papillomavirus Related Malignancy 3

2.1 Human papillomavirus 3

Classification 3

Structure & Function 5

Viral Lifecycle 11

Host Immune Response, Natural History of Infection & Clearance 12

2.2 Oral & Oropharyngeal Squamous Cell Carcinoma 13

Epidemiology 13

Aetiology 15

Regional Anatomy 16

2.3 Epidemiology of HPV-mediated Malignancy 18

Global Incidence & Subsites 18

HPV Carriage, Clearance and Consequences of Persistence. 20

Human papillomavirus-16 in Head & Neck Squamous Cell Carcinoma 23

Clinical Implications of HPV-positive OPSCC 25

2.4 HPV-mediated Tumourgenesis 28

Molecular Pathogenesis 28

Immune system response in HPV-positive OPSCC 33

2.5 HPV detection techniques in HNSCC 35

Viral Oncogene Expression (HPV mRNA qPCR) – The "Gold Standard" 36
Viral DNA Detection (DNA PCR, DNA in situ hybridisation) 37
p16 immunohistochemistry (p16 IHC) as a Surrogate Marker of HPV-mediated Malignancy. 39
Diagnostic Algorithms 40
2.6 DNA Methylation in HNSCC 42
DNA methylation in malignancy 43
Regulation of DNA methylation 46
DNA Methylation in HNSCC 47
2.7 HPV16 Integration State in Malignancy 49
Viral E2 gene Disruption 50
Altered viral oncogene expression and stabilization 51
Insertion mutagenesis 51
Integration in HPV positive HNSCC 53
3 Material & Methods 55
3.1 Human Tissue Procurement & Characterisation 55
Ethical Agreements for Clinical Sample & Data Collection 55
3.2 Tissue Microarray (TMA) Construction & Utilisation 57
TMA Construction 57
TMA Utilisation 58
3.3 p16 Immunohistochemistry 58
3.4 High Risk HPV DNA in situ Hybridisation 59
3.5 High Risk HPV RNA in situ Hybridisation (RNAscope) 59
3.6 High Risk HPV Test Interpretation Techniques 61
3.7 Cell Lines 61
Established Cell Lines 62
3.8 Nucleic Acid Extraction 63

Fresh Frozen Tissue Samples 63
Cell Line Derived Samples 64
Cell Line Verification Procedure 65
3.9 Nucleic Acid Quantification 67
3.10 Complimentary DNA Synthesis 67
3.11 Quantitative Real-Time PCR Analysis 68
HR HPV qPCR Assays Design and Optimisation 68
DNA qPCR Assays 73
Expression (mRNA qPCR) Assays 73
3.12 Bisulphite Treatment of DNA 76
3.13 Pyrosequencing Methylation Analysis 78
LINE-1 Assay 78
HPV16 Gene Methylation Assays 80
3.14 Methylation MicroArray Analysis 83
Bioinfomatic Analysis & Validation Pyrosequencing Methylation Assay Design 84
3.15 DNA Artificially Methylated Controls 88
3.16 HPV16 Integration Analysis 89
E2 Gene Integrity Analysis 89
Next Generation Sequencing (NGS) Analysis 91
4 Evaluation of Diagnostic Testing in Oropharyngeal Squamous Cell Carcinoma 96
4.1 Introduction 96
4.2 Diagnostic Test Analysis Aims & Methods 97
Hypotheses 97
Testing Methods & Statistical Analyses 98
4.3 Diagnostic & Prognostic Capacity Results 100
Cohort characteristics 100

Availability for testing, sample quality and consistency between repeats. 102
HPV status 105
Implications of Diagnostic Thresholds 110
HPV16 Status in Tumours & Adjacent Normal Marginal Tissues 111
HPV16 status vs. clinical characteristics 112
4.4 Discussion & Implications for Clinical Practice 117
5 Is the Reference "Gold Standard" Test Restricted to Fresh Frozen Samples? 125
5.1 Introduction 125
High Risk HPV RNA In situ Hybridisation Aims 126
Test Methods, Interpretation & Analysis 126
5.2 Results 127
Tissue sample quality and consistency 127
High Risk HPV Detection in Normal Tissues 130
Oropharyngeal SCC Test Analysis 130
False positive and false negative reporting 135
5.3 Discussion & Implications for Clinical Practice 136
6 Epigenetic Regulation through DNA Methylation in HPV Mediated Oropharyngeal Squamous Cell Carcinoma (OPSCC) 141
6.1 Introduction 141
Analysis of Epigenetic Regulation in HPV positive OPSCC Aims 141
6.2 HPV E2 & Long Control Region Methylation Status and implications for viral oncogene expression 143
Analysis of Epigenetic Regulation in HPV positive OPSCC Methods 143
Epigenetic Regulation in HPV positive OPSCC Results 144

6.3 Comparative Host Genome-wide methylation State in HPV Positive and HPV Negative OPSCC 149

Genome-wide Methylation Analysis Methods 149

Genome-wide Methylation Analysis Results 149

6.4 Detection of Differential Host Gene Promoter Methylation in HPV postive & HPV Negative OPSCC 152

Detection of Promoter Methylation State in OPSCC: Methods 152

Host Tumour Suppressor Gene promoter sequence hypermethylation in HPV Positive & HPV Negative OPSCC: Results 152

6.5 Expression of key regulators of Methylation state in HPV positive HNSCC 160

DNA Methyltransferase & UHRF1 Expression Results 161

6.6 Discussion 165

Viral Methylome 165

Genome-wide Host Methylation State 168

Host Gene Promoter Methylation in OPSCC 169

Genome-wide Assay Validity 172

Regulators of Methylation in HPV positive OPSCC 172

7 Viral Integration State in HPV positive OPSCC 175

7.1 Introduction 175

Integration Analysis Aims 175

7.2 Viral E2 Gene Integrity Analysis 176

E2 Integrity Assay - Methods 176

E2 Integrity Assay - Results 176

7.3 Specific Viral Capture & Next Generation Sequencing For Determination of Viral Cleavage & Host Insertion Positions 180

Viral sequence capture and sequencing – Results 180

7.4 Discussion 186

Detection of viral integration – E2 gene integrity 186

Next Generation Sequencing as an Analytical Tool 187

Viral Cleavage Position Detection 190

Host Integration Position 191

8 Discussion & Future Directions 194

HPV Diagnostics in HNSCC – Prognostic Biomarker & Disease Stratification 194

Biomarkers for HPV positive OPSCC 197

9 Publications Supporting This Thesis 201

Associated Publications 201

10 Appendices 202

11 References 218

LIST OF TABLES

Table 1: Viral Protein Structure & Function 9

Table 2: United Kingdom annual Incidence for OSCC & OPSCC 14

Table 3: Selective list of gene previously reported to be silenced by promoter hypermethylation in malignancy 45

Table 4: Thermal Cycling Conditions for Cell ID System 66

Table 5: Reported and Actual STR Profiles for Cell Lines Utilised..... 66

Table 6: Primer and probe sequences for HPV-16 & -18 gene specific assays 69

Table 7: HPV16 E2, E6 & E7 qPCR Thermal Conditions..... 72

Table 8: HPV18 E6 qPCR Thermal Conditions..... 72

Table 9: Genesig HPV33 qPCR Thermal Conditions..... 73

Table 10: Microsatellite Marker Analysis PCR Thermal conditions 74

Table 11 Identification and additional information for proprietary gene expression assays DNMT1, -3A, -3B & UHRF1..... 75

Table 12: DNMT & UHRF1 qPCR Thermal Conditions..... 76

Table 13: LINE-1 Primer Sequences..... 78

Table 14: PCR conditions for LINE-1 Assay 79

Table 15: HPV16 Pyrosequencing Methylation Assay Primer Sequences 80

Table 16: HPV16 E2 PMA Thermal Conditions 82

Table 17: HPV16 Long Control Region 1 PMA Conditions 82

Table 18: HPV16 Long Control Region 2 PMA Conditions 82

Table 19: Infinium Technical Validation PMA Primer Sequences..... 85

Table 20: Infinium Technical Validation PMA Conditions..... 85

Table 21: Infinium PMA Gene Specific Annealing Temperatures & CpG Inclusions..... 86

Table 22: Thermal Cycling Conditions for HPV16 E2 Whole Gene 90

Table 23: Thermal Cycling Conditions for HPV16 E2 Component Parts (P1 – P5) 90

Table 24: Cohort Characteristics 101

Table 25: Tumour HPV Status determination Over Time and HPV test Diagnostic Stringency 107

Table 26: Patient & Tumour Individual Case Diagnostic Analysis..... 108

Table 27: Patient & Tumour Individual Case Diagnostic Analysis..... 109

Table 28: HPV Status in Matched Tumour and Adjacent Normals..... 111

Table 29: Clinical Characteristics by HPV status 113

Table 30: Kaplan-Meier estimates of survival at 36 Months by HPV status 115

Table 31: HR HPV ISH Analysis Cohort Characteristics 129

Table 32: Kaplan-Meier survival estimates of Disease Specific Survival with associated Hazard Ratios 134

Table 33: Kaplan-Meier survival estimates of Overall Survival with associated Hazard Ratios 134

Table 34: Diagnostic capabilities of individual tests by comparison to HR HPV qPCR 135

Table 35: Control Sample Compiled Viral Expression and Methylation Analysis Results 144

Table 36: Proportion of HPV samples displaying positive methylation at viral targets 145

Table 37: Correlation between Viral methylation and Viral Gene Expression 147

Table 38: Correlation between Viral E2 Expression and Viral Oncogene Expression (E6 & E7) 148

Table 39: Evidence for role of Target Genes in Oncogenesis (HPV related or otherwise) 154

Table 40: PMA Internal validation analysis - Correlation Coefficients for Individual PMAs using methylated control DNA samples. 155

Table 41: Spearman's Correlation Coefficients for Methylation Status (%) between Infinium HumanMethylation450 BeadChip Array and Individual Gene Pyrosequencing Assay Analysis by target gene. 158

Table 42: Regulation of DNA Methylation Analysis Cohort 160

Table 43: Correlation in levels of expression of DNMTs/UHRF1 within OPSCC tumour samples. 162

Table 44: Correlation in levels of expression of DNMTs/UHRF1 within OPSCC tumour samples stratified by HPV status. 163

Table 45: Relative viral gene expression (RQ) and E2 gene Integrity Analysis for HPV16 DNA positive OPSCC. 179

Table 46: Compiled results for samples analysed using next generation sequencing (NGS) 185

LIST OF FIGURES

Figure 1: Phylogenic Tree of the *Papillomaviridae* family 4

Figure 2 Component parts of the circular HPV genome and gene functions..... 6

Figure 3 The Natural History of Cervical HPV infection and Carcinogenesis..... 20

Figure 4: Cellular influences of HPV oncoproteins in the malignant transformation of keratinocyte 32

Figure 5: Algorithm for detection of HPV in FFPE head and neck biopsies 41

Figure 6: Molecular Consequences of Viral integration 52

Figure 7: HPV16 Assay Detection Sensitivity 70

Figure 8: Establishment of the Linear Dynamic Range of the HPV16 Assay..... 71

Figure 9: Schematic representation of HPV genome with detailed LCR and Pyrosequencing Methylation Assay sites. 81

Figure 10: Schematic Diagram of HPV16 E2 Integrity Overlapping Primer Analysis..... 90

Figure 11: Overall sequencing sample preparation workflow..... 92

Figure 12: Sample Hybridisation Schematic 93

Figure 13: Optimal DNA shearing profile from Agilent 2100 Bioanalyzer electropherogram (12k chip) 94
Figure 14: DNA Quality Over Time 103
Figure 15: RNA Quality Over Time 103
Figure 16: Plot of Individuals case HPV16 ΔCT DNA qPCR results by HPV status 110
Figure 17: 36 Month Kaplan-Meir Plots of Disease Specific Survival and Overall Survival by HPV status 116
Figure 18: Photomicrographs of OPSCC cases stained using RNAscope 132
Figure 19: Kaplan-Meier Survival plots by HPV status as determined by HR HPV RNA ISH (RNAscope) 133
Figure 20: Correlation between methylation of viral regions and related viral gene expression 147
Figure 21: Scatter plot of correlation between HPV16 E2 and E6 gene expression (RQ) 148
Figure 22: Boxplot representation of Average LINE-1 Methylation values in OPSCC and Adjacent Normal Tissues 150
Figure 23: Boxplot representations of Average LINE-1 Methylation values in OPSCC stratified by HPV status. 151

Figure 24: Representation Internal Validation Plot for C12orf42 Pyrosequencing Methylation Assay. Detection of artificially methylated DNA control samples. 155

Figure 25: Comparative Methylation Analysis: Average Methylation in promoter sequence of eight differentially methylated genes. 158

Figure 26: Scatter plots depicting correlation between Infinium Array methylation results and Pyrosequencing Methylation Array average methylation values. 159

Figure 27: Comparative expression for UHRF1 between paired OPSCC tumour and adjacent normal tissues..... 161

Figure 28: Kaplan Meier Estimates of Overall Survival at 36 Months on the basis of DNMT3b expression. 164

LIST OF APPENDICES

Appendix I: Mapping Statistics for individual samples, including duplicate reads where apparent. 203

Appendix II: Mapping Statistics for individual samples, excluding duplicate reads. 205

Appendix III: Viral & Host Sequencing Read Representations for the HPV16 positive Cell Line CaSki..... 207

Appendix IV: Viral & Host Sequencing Read Representations for the HPV16 positive Cell Line SiHa 208

Appendix V: Viral & Host Sequencing Read Representations for the HPV16 positive OPSCC sample No. 13 (311T)..... 209

Appendix VI: Viral & Host Sequencing Read Representations for the HPV16 positive OPSCC sample No. 34 (427T)..... 210

Appendix VII: Viral & Host Sequencing Read Representations for the HPV16 positive OPSCC sample No. 106 (045-09T) 211

Appendix VIII: Viral & Host Sequencing Read Representations for the HPV16 positive OPSCC sample No. 95 (270-08T) 212

Appendix IX: Viral & Host Sequencing Read Representations for the HPV16 positive OPSCC samples No. 105 (043-09T)..... 213

Appendix X: Viral & Host Sequencing Read Representations for the HPV16 positive OPSCC sample No. 88 (077-08T) 214

Appendix XI: Viral & Host Sequencing Read Representations for the HPV16 positive OPSCC sample No. 108 (239-08T) 215

Appendix XII: Viral & Host Sequencing Read Representations for the HPV16 positive OPSCC sample No. 87 (075-08T) 216

Appendix XIII: Viral & Host Sequencing Read Representations for the HPV16 positive OPSCC sample No. 101 (035-09T) 217

ACKNOWLEDGEMENTS

I am grateful for the support of The Wellcome Trust and The Faculty of Dental Surgery of the Royal College of Surgeons of England. Their support, both financial and practical, has been invaluable in driving this research.

There are several people to whom I am indebted for the support, guidance and inspiration that they have shown during the process of this research and the ensuing thesis.

In particular, Lakis Liloglou and Janet Risk, who have offered calm and insightful guidance during my scientific development and have always provided understanding whilst at times I have generated more questions than answers.

Within their research group, I would specifically like to thank George Nikolaidis and Amelia Acha-Sagredo for their frequent assistance in the lab and tolerance of a surgeon in their midst. I have also been fortunate to collaborate with colleagues, now friends, like Max Robinson & Benjamin Lallemant.

Particular mention is warranted for Richard Shaw who has become a mentor both in research and clinical practice. I consider myself fortunate to have benefited from his friendship, guidance and inspiration from the inception of this project to the final printed word.

To my family in both hemispheres I owe everything. More specifically, to my beautiful children, Elizabeth, Oscar & Sophia; you always give me perspective, and to Rosie; who never ceases to be there for me. I dedicate this thesis to them.

1 THESIS AIMS

Oropharyngeal Squamous Cell Carcinoma (OPSCC) incidence has shown a dramatic increase over the last two decades in both United Kingdom clinical practice[1,2] and beyond[3]. Evidence for the role of Human papillomavirus-16 (HPV16) in the rapidly evolving increase varies between geographical regions yet it is clear that those tumours demonstrated to be HPV positive are biologically distinct from HPV negative malignancies[4].

Detection of HPV, and the tests with which detection is made, in OPSCC serves to provide capacity for disease stratification and also effective prognostication, however the optimal test has not been defined.

Although these tests provide utility once malignant transformation has occurred, predictive biomarkers have not been forthcoming for application in the variable latent period between viral infection and the development of cancer. Following establishment of cellular viral infection and a variable latent period, both epigenetic alterations and detectable changes in the viral state (integration) have been demonstrated during the progression to invasive disease[5-7].

It is hypothesised that in United Kingdom Head & Neck Cancer clinical practice, HPV16 plays an aetiological role in a significant proportion of Oropharyngeal Squamous Cell Carcinomas and that the currently available clinical tests demonstrate wide variation in both diagnostic and prognostic capacity when detecting HPV positive malignancy.

Further, it is proposed that developments in diagnostic tests are capable of levels of diagnostic stringency on clinical samples comparable to current research-based gold standard testing.

In addition to prognostic discrimination and tumour stratification on the basis of HPV status, this thesis aims to test the hypothesis that both the distinct epigenetic profile of HPV+ OPSCC and the occurrence of viral DNA integration play a critical role in oncogenesis through their influence on viral oncogene expression. Further, it is hypothesised that epigenetic changes offer clinically relevant predictive biomarkers in HPV mediated oncogenesis.

2 INTRODUCTION: HUMAN PAPILLOMAVIRUS RELATED MALIGNANCY

2.1 HUMAN PAPILLOMAVIRUS

Classification

Papillomaviruses are a heterogeneous group of viruses traditionally classified as part of the *Papoviviridae* family, however under recently adopted general criteria established by the International Committee on the Taxonomy of Viruses in 2004, they are now exclusively recognised as part of the *Papillomaviridae* family [8]. Papillomaviruses are classified by genus and type, of which in excess of 120 have been fully sequenced[9]. The human papillomaviruses (HPV) are ubiquitous infectious agents characterised by strict species specificity. They are obligate epitheliotropic human pathogens capable of causing both benign and malignant disease in mucosal or cutaneous tissues[10]. Figure 1 (reproduced from de Villiers et al) demonstrates a phylogenetic tree of the papillomaviruses based on the nucleotide sequence of the major capsid protein, L1, which classifies viruses into genera. By definition, each

type shares less than 90% DNA sequence homology within the L1 gene[11].

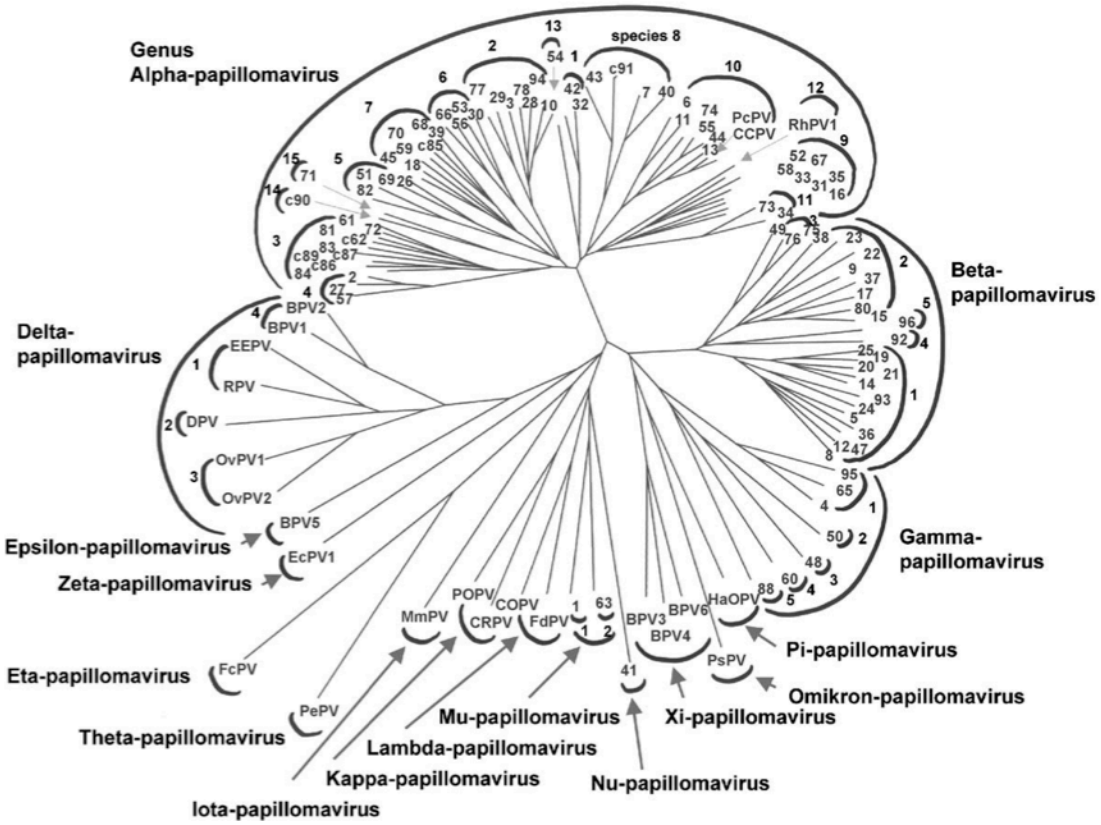


Figure 1: Phylogenic Tree of the *Papillomaviridae* family

The numbers adjacent to each branch denote the HPV type whilst the outermost numbers refer to HPV species, each of which contains more than HPV type. All other abbreviations refer to animal papillomavirus types (modified from de Villiers et al 2005[8]).

The genus with greatest clinical relevance is referred to as the alpha-papillomaviruses[12]. It contains all HPV types associated with mucosal and genital lesions, however for reasons of clinical utility the more common subtypes are subdivided into high and low risk groups to reflect their relative risk of inducing malignancy within the infected tissues. There are 12 low-risk types (6, 11, 40, 42, 43, 44, 54, 61, 70, 72, 81 and CP6108) and 15 known high-risk types (16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, 68, 73 and 82)[11], the latter being those types which are found preferentially in precancerous and cancerous lesions[9].

The beta genus comprises viruses that preferentially infect cutaneous sites, many of

which have been implicated in non-melanoma skin cancers[13]. HPV types within the remaining genera are typically associated with cutaneous papillomata.

Structure & Function

Viral Structure

HPV is a small (50-55 nm in diameter), double stranded, circular DNA virus with an icosahedral capsid coat derived from major and minor elements, L1 and L2 respectively. The genome contains approximately 7900 base pairs with 8 or 9 open reading frames (ORFs) within the individual DNA strand[14]. The viral genome can be considered as having three distinct regions according to the location of coding regions for early genes, late genes and a further non-coding region located between the L1 and E6 open reading frames, termed the long control region (LCR) (Figure 2).

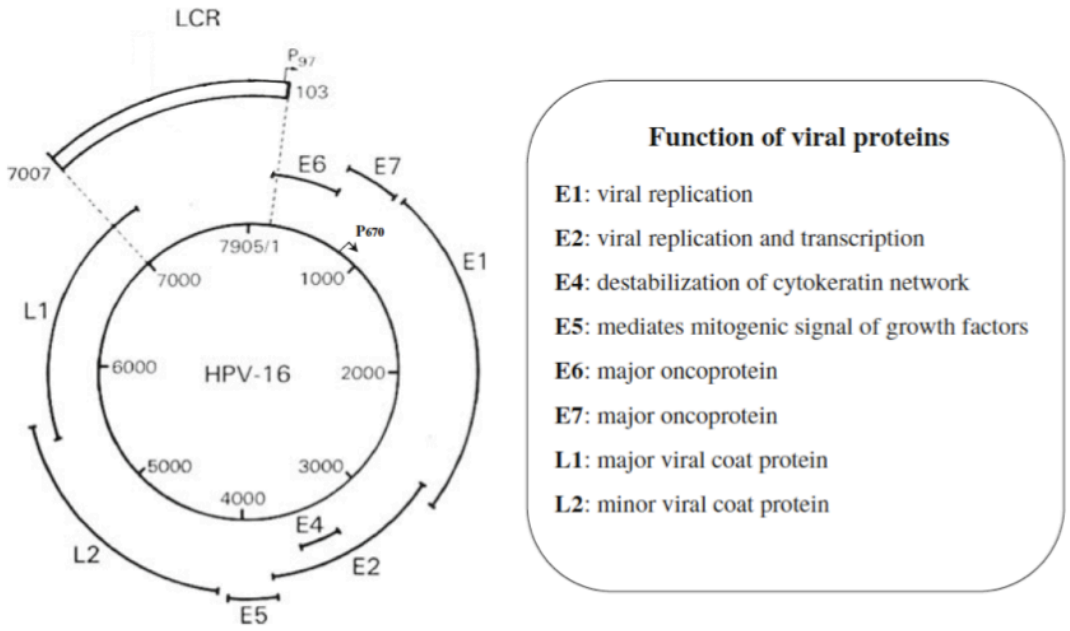


Figure 2 Component parts of the circular HPV genome and gene functions

The three distinct regions are comprised of the early genes (E1, E2 & E4-7), the late genes (L1 & L2) and the long control region (LCR). The internal numbering reflects the nucleotide numbers. The early and late promoters are also identified; p97 & p670 respectively. (Modified from Ghittoni et al 2010[14]).

The LCR is approximately 800 bp in length and contains both the origin of replication (ori) for the virus and several transcriptional binding sites thus extending control over the expression of viral genes[15]. Transcriptional modulation of early gene expression is a central regulatory event initially activated by host cell transcription factors. Subsequently, self-regulation ensues through the effects of the viral E2 gene[15] on four conserved binding sites specific to the E2 viral gene product (E2-BS) each within the LCR of high-risk HPV genomes. Each site confers replication and transcriptional effects during the viral life cycle[15,16], the variability of which has been suggested to be influenced by both binding site methylation and order of E2-BS occupancy[15,16]. The E2BSs are located immediately upstream to the early promoter (P_{97}) that regulates early viral gene expression, including that of the

viral oncogenes E6 and E7[15]. In addition to the early promoter located in the LCR, HPV contains a late promoter (P_{670}) that is located within the E7 ORF. Transcripts originating at the P_{97} promoter are polycistronic, with the potential to encode both E6 and E7 proteins as well as the replication and transcription control proteins E1 and E2 respectively[9].

Six genes are expressed in the early stages of infection (E1, E2, E4, E5, E6 & E7). Variation in the presence of particular early genes and in the length of the LCR has been described, according to different viral subtypes[14]. Whilst E6 and E7 genes are highly conserved for almost all subtypes studied thus far, E5 presence shows greater variability and is generally not found in the majority of beta types.

The E1 and E2 genes encode for regulatory proteins that have fundamental roles in both viral replication and in transcription of the remaining genes. E6 and E7, and to a lesser extent E5, are involved in host cell transformation, and are termed major oncoproteins to reflect their role in tumourgenesis[4,17]. All viral gene products are detailed in Table 1 including their relative quantities and intracellular locations, a more detailed discussion of the functions of the key early genes involved in viral oncogenesis (E6 and E7) is made below (Molecular Pathogenesis).

Viral E2 gene plays a pivotal role in the balance between optimal or controlled early gene transcription necessary for regulation of the viral life and infection cycles and uncontrolled permissive replication of oncogenes associated with malignant progression[15]. As already mentioned, E2 is a natural transcriptional repressor of other early genes, E6 & E7 through its effects on the LCR.

Viral capsid formation requires the protein products of the late genes, L1 and L2. The L1 major capsid protein self-assembles to complete the 72 pentamers of the

capsid whilst the minor protein, L2 is principally located internally within the virion[18].

Protein	Molecular Mass	Cellular Location	Cellular quantity	Function
Early				
E1	68-75 kDa	Nuclear	+	Initiator of viral replication. Activation of helicase and maintainer of episomal viral DNA
E2	50 kDa	Nuclear	+	Viral transcription (inc E6 & E7 repression) and DNA replication. Segregation of viral genomes.
E4	17 kDa Mostly fusion protein with E1	Cytoplasm	+++	Facilitator of viral genome packing and maturation of the viral particles. Destruction of cytokeratin filaments. Interaction with RNA helicase
E5	8-10 kDa	Cytoplasm	+	Interaction with EGF-receptor, activates PDGF-receptor. Oncoprotein, allows continuous cellular proliferation, delays differentiation
E6	16-18 kDa	Nuclear	+	Major Oncoprotein. Deregulation of cell division/cell cycle control. Degrades p53 in presence of E6-AP. Multiple host protein interactions
E7	11 kDa	Nuclear	++	Major Oncoprotein. Deregulation of cell division/cell cycle control. Degrades pRb. Multiple host protein interactions
Late				
L1	55-60 kDa	Nuclear	++++	Major Capsid Protein
L2	70 kDa	Nuclear	++	Minor Capsid Protein. Aids in viral internalisation and localisation to nucleus.

Table 1: Viral Protein Structure & Function

(Modified from Ruatava et al[9]).

Viral Cellular Internalisation

For non-enveloped viruses, such as HPVs, the capsid coat provides essential protection for the viral nucleic acid whilst also being instrumental in the initial

phase of cell infection. Viral binding and internalization into a target keratinocyte is reliant upon both L1 and L2 components of the viral capsid[19]. However, before attachment is possible there is a pre-requisite for basement membrane exposure beneath either cutaneous or mucosal keratinocytes, either by chemical or mechanical microtrauma[20]. Subsequently, it is believed that, the viral particles bind to exposed heparin sulphate proteoglycans (HSPG) on the basement membrane inducing a conformational change in the capsid. This alteration allows exposure and cleavage of the L2 protein from the viral particle permitting a previously unexposed region of the L1 capsid protein to bind to receptors on adjacent keratinocytes[19]. Laterally migrating epithelial cells, such as those involved in mucosal wound repair, seem to be particularly susceptible to HPV binding as a result of high levels of expression of the putative HPV receptor α6β4-integrin[21,22].

The understanding of the mechanisms of viral cell entry remains incomplete and continues to be a source of scientific interest and debate, however that which is presently understood was reviewed with clarity by Hovarth et al[23]. Following an extended, and as yet undefined time period following viral-receptor interaction, internalisation of the virus occurs most probably via an endocytotic mechanism[23,24]. Subsequently L2 facilitates egress from the endosome and allowing viral DNA to remain within the cytoplasm for a protracted period before a small proportion of viral DNA transits to the cell nucleus[9].

Viral Lifecycle

Following infection, the dsDNA viral genome remains in its circular form within the infected keratinocytes cytoplasm and only limited viral replication occurs, independent of normal cell cycle, to produce a low viral copy number of approximately 50-100 copies/cell[25]. HPV encodes only a single DNA replication enzyme (E1), a DNA helicase[11] and as such must harness the cellular transcriptional and translational machinery of the host cell to permit increased DNA amplification and ultimately production of encapsidated progeny virions for subsequent release[26].

To this end, HPV exploits the natural differentiation of keratinocytes, from basal stem cells through to the terminally differentiated keratinocytes, present in the most-superficial layers of the stratified squamous epithelium, to achieve genome amplification and expression of capsid proteins[27]. HPV DNA replicates during the S phase of the cell cycle, in concert with host cell chromosomal replication. And the initiation of replication is the same as that for all eukaryotic chromosomes. E2 is the initiating factor at the HPV origin of replication (ori) within the LCR and facilitates E1 recruitment, which in turn utilises cellular molecular of the replication machinery (polymerases and replication proteins)[28].

Ultimately, the infected keratinocyte will have a viral copy number exceeding 1000 viral copies/cell with associated abundant expression of the viral oncogenes E6 and E7 and the late genes L1 and L2. It is the viral oncogenes that play the most significant role at this stage, as the cell will no longer be naturally mitotically active and hence it is reactivation of the cell cycle by E6 and E7 that creates an

environment permissive of viral DNA replication. It is this feature too that, albeit rarely, allows disregulation of growth control and creates the potential for malignant transformation[26], although in this situation it is both spatial and quantitative deregulation of tight oncogene transcriptional control that occurs, leading to production of E6 and E7 throughout the epithelium including the basal layer[6]. The latter, "transforming" infection is discussed in greater detail later (2.4). The duration of the process, from infection to viral sheading, follows the normal squamous epithelial turnover of approximately 3 weeks, however this is dependent up on the specific site and as such may take considerably longer.

Host Immune Response, Natural History of Infection & Clearance

HPV infection can be characterised as acute, chronic or latent, however it is the persistent infection, whether chronic or latent, that exposes the cell to an environment suitable for cellular transformation.

In establishing a persistent infection, evasion of the host immune surveillance mechanisms is essential[29]. The essential role of the immune system in controlling HPV infections has been deduced from studies of immunocompromised women[30]. In an intact immune system it is the three oncoproteins, E5, E6 and E7, that orchestrate immune escape both individually and in concert with one and other[29]. In HPV positive tumour cells, immunogenic peptides from both E6 ad E7 proteins are not processed or transported to the cell surface effectively, nor presented successfully. HPV E6 and E7 target type I interferon (IFN) that is produced by cells in response to viral infection. E7 inhibits induction of IFN- α whilst E6 binds to the IFN-

associated transcription complex, ISG factor-3, preventing transcription of IFN-β. E5 protein reduces antigen presentation through selective down regulation of components of the MHC/HLC Class I such that cytotoxic T lymphocytes antigen presentation is diminished whilst natural killer cell inhibitory ligands are still apparent[31].

Although there is a survival advantage to the virus being able to evade the host immune response, this advantage is not preserved once development of malignancy has occurred and completion of the viral life cycle fails, due to a stalled progression to late phase viral proteins.

2.2 ORAL & OROPHARYNGEAL SQUAMOUS CELL CARCINOMA

Epidemiology

Head & Neck Squamous Cell Carcinoma (HNSCC) is the sixth commonest cancer worldwide with approximately 600,000 new cases diagnosed annually, accounting for 5% of all tumours[32]. Tumours from oral and oropharyngeal subsites account for 400,000 of these malignancies and, by comparison to other head and neck subsites, there has been a disproportionate increase in incidence over recent years [1]. A similar picture has become apparent in the United Kingdom; indeed a considerable proportion of the recent increase in HNSCC incidence is due directly to the influence of changes in OPSCC incidence. OPSCC incidence has risen from a direct standardised population rate of 1:100,000 to 2.5:100,000 in the period from 1990 – 2006, whilst in the same period Oral Squamous Cell Carcinoma (OSCC) incidence remained relatively stable between 2.5 and 3:100,000[2]. Data released from the US

National Cancer Institute's Surveillance, Epidemiology, and End Results (SEER) indicate a similar steady rise in OPSCC incidence since 1973 during which time evidence of exposure to tobacco smoke has declined[33].

Wide variation in rates of OSCC and OPSCC are apparent across the globe, with areas of particularly high incidence centered in South and Southeast Asia, Latin America and small regions within the Eastern Europe[34,35]. When considered in a global context, the United Kingdom has comparatively low rates of oral and oropharyngeal SCC with the combined diagnoses accounting for approximately 3% of all malignancies. Table 1 demonstrates the incidence in absolute terms for the United Kingdom from 2009.

Number of UK new cases of Oral & Oropharyngeal SCC (2009)				
Site	**Males**	**Females**	**Persons**	**M:F ratio**
Lip (ICD10 C00)	217	124	341	1.8:1
Tongue (ICD10 C01-02)	1239	675	1914	1.8:1
Mouth (ICD10 C03-06)	1074	762	1836	1.4:1
Oropharynx (ICD10 C09-10)	989	357	1346	2.8:1
Total OSCC & OPSCC	**3,519**	**1918**	**5,437**	**2:1**

Table 2: United Kingdom annual Incidence for OSCC & OPSCC
(Modified from Cancer Research UK Cancer Stats, 2009).

Until relatively recently survival from HNSCC has displayed only modest improvements[36] despite refinement to surgical techniques, introduction of variations in radiation therapy delivery and the advent of new generation chemotherapeutics[37]. As will be discussed in greater depth below (aetiology) the

contribution of factors influencing causation other than tobacco and alcohol usage have become more apparent recently. The evolving role of human papillomavirus in head and neck cancer has the potential to drastically change incidence and survival particularly in subsites most likely to harbor HPV-mediated malignancy[3,38].

Aetiology

The aetiology of squamous malignancies of the head and neck is multifactorial however the majority of these malignancies are related to either consumption or usage of tobacco in its various forms, areca nut/betel quid chewing and or alcohol[35]. There is also mounting evidence to suggest that deficiency in particular dietary micronutrients has an increasing influence on aetiology, a trend that may persist particularly in developing countries[39]. Factors specific to particular sites within the head and neck have also been reported, in particular UV light inducing malignant change particularly in sun exposed areas (lip)[35], and, within the main oral cavity, poor oral hygiene with its associated bacterial infections with or without chronic trauma from so called dental factors have been cited[40]. In addition to environmental factors, an inherited genetic predisposition is involved in a small proportion of upper aero-digestive cancers[35].

Awareness of histological similarities between anogenital and upper aerodigestive mucosal surfaces combined with an increasing understanding of the aetiological role of HPV in anogenital malignancy, particularly in cervical cancer, lead investigators to explore the role of HPV in head and neck mucosal malignancy and premalignancy[41-43]. In 2009, the International Agency for Research on Cancer (IARC)

published an important monograph[44] summarising those infectious agents for which causation has been established beyond doubt. Consistent with a mounting body of evidence, the IARC found Human papillomavirus type 16 (HPV-16) to be causal in squamous cell cancer (SCC) of the oral cavity and oropharynx.

At a molecular level, head and neck carcinogenesis is a complex multistep, multifactorial process containing a myriad of genetic and epigenetic abnormalities in DNA repair, cell signaling, cellular differentiation, angiogenesis, apoptosis and cell cycle regulation[45]. Whilst recognising the considerable role that genetic alterations play, this research sought to explore the key epigenetic alterations in HPV mediated HNSCC which to date have received little attention.

Regional Anatomy

Oral and Oropharyngeal Squamous Cell Carcinoma (OSCC and OPSCC) describes the squamous malignancies arising from mucosal surfaces of the lip, tongue and mouth (oral cavity [ICD-10: C00, C02-06]) and oropharynx [ICD-10: C01, C09-10][46,47]. This subgroup of head and neck cancer excludes salivary malignancies and tumours arising within other subsites of the pharynx.

Despite the close physical proximity and similar embryological development[48,49], the distinction between oral cavity and oropharynx is important when considering the role of virally mediated malignancy. A considerable proportion, although by no means all, of the evidence for HPV positive malignancy in the head and neck refers to the oropharynx[50-52]. The oral cavity is the mucosal lined region extending from the lips anteriorly to the junction of the anterior 2/3 and posterior 1/3 of the

tongue posteriorly, the palatoglossal arch posterolaterally and the soft palate superiorly[49]. The oropharynx is bounded by the soft palate superiorly, the base of the tongue (posterior 1/3 of the tongue) to the level of the epiglottis inferiorly and the palatoglossal and palatopharyngeal arches laterally. Posteriorly, it is enclosed by the posterior pharyngeal wall[49]. A particularly important constituent part of the oropharynx, which contrasts markedly with the oral cavity, is the inclusion of abundant lymphoid tissue arranged in an incomplete ring or arch superiorly – *Waldeyer's ring*[49]. It has been suggested that regional variation in HPV related malignancy within the head and neck may, at least in part, be related to the lymphoid tissue of the oropharynx[53]. The palatine tonsils in particular seem to be disproportionally effected by HPV positive tumours as will be discussed in greater depth below (2.3 Epidemiology of HPV-mediated Malignancy). These collections of lymphoid tissue, on each side of the oropharynx in the interval or intertonsillar cleft between the palatoglossal and palatopharyngeal arches, are covered by stratified squamous epithelium and incorporate multiple invaginations or crypts. The reticular crypts greatly increase the tonsils surface area and, importantly, are composed of specialised epithelia with both immune and secretory features[54] that may facilitate access of oral pathogens, including HPV, to the basement membrane[55].

Functionally, both the oral cavity and the oropharynx play critical roles in speech, mastication and swallowing. The detrimental influence on these functions caused by both the tumour and subsequent cancer treatment varies, with more anterior lesions altering speech and mastication to a larger extent and more posterior lesions causing greater swallowing impairment[56,57].

2.3 EPIDEMIOLOGY OF HPV-MEDIATED MALIGNANCY

Global Incidence & Subsites

Global cancer incidence in 2008 was estimated to exceed 12.7 million new cases accounting for 7.6 million cancer related deaths[58]. Although there is considerable global inter-regional variation, an estimated 16% of all human malignancies result from chronic infection, one third of which are attributable directly to HPV infection. In addition to the human costs of HPV related malignancy, there is also a substantial and mounting financial cost to tax payers for HPV related medical care. In 2000 alone, the direct costs associated with HPV infection approached US$3 billion, the majority of which was spent on monitoring and initial management of HPV-related cervical premalignant and malignant disease[59].

As described above, HPV induces malignancy in both cutaneous and mucosal surfaces.

At present, the greatest burden of HPV related malignancy is cervical cancer, which is the second most common cancer amongst women with approximately 500,000 new cases and 274,000 deaths annually[60]. Cancers of the vagina and the external genitalia, namely the penis and vulva, are likewise frequently virally mediated however by comparison to the cervix, where virtually all cases are HPV DNA positive, vaginal cancer is 64-91% HPV positive and only 40-60% of all penile and vulval cancers HPV positive[60]. Cancer of the mucosal lining of the anal canal have a similarly strong association with HPV DNA infection, with 88-94% of these cancers proving positive for HPV16[61,62]. HPV positive malignancy in the head and neck region is greatest in the oropharynx (OPSCC) where approximately 40% of all

tumours globally are HPV positive albeit with considerable regional variation[3]. Although presently cervical cancer is the largest group of HPV positive malignancy by number, extrapolations of current trends in HPV positive OPSCC incidence suggest that this group may exceed cervical cancer by 2020[63]. In head and neck subsites outwith the oropharynx the incidence and role of HPV in cancers remains contentious and controversial[52,64] however it was deemed sufficient by the International Agency for Research on Cancer (IARC) to conclude that HPV was causative in both oral and oropharyngeal cancer[65].

Evidence for HPV as an aetiological factor in cutaneous malignancy has been available for more than 30 years, particularly within those individuals with pre-existing immunosuppression, however a direct causal relationship has yet to be formally established[66]. It may be that the influence of HPV in skin malignancy comes as a co-carcinogen with UV radiation or immunosuppression[62].

Of the 15 high risk subtypes of HPV, HPV16 & HPV18 are the two most common subtypes involved in mucosal malignancy. It is generally accepted that all cervical cancers are HPV positive and HPV16 & HPV18 are found in 70-75% of cases[62]. Within HPV positive anal, oropharyngeal, vulval, vaginal and penile malignancies the dominant subtype is invariably HPV16 although there is some evidence for the role of HPV18 in vulval, anal and penile cancer and to lesser extent OPSCC[62]. Remaining high-risk subtypes have variable influence by site and, in the head and neck in particular, invariably are isolated as co-infections with HPV16 or HPV18[52].

HPV Carriage, Clearance and Consequences of Persistence.

The prevalence of genital HPV infection in women within the US population has been estimated at greater than 40% for all subtypes and 4.7% specifically for HPV16[67]. Although it is without question that HPV infection is a fundamental necessity for the development of HPV-related malignancy, the relationship between HPV carriage, HPV clearance or persistence and malignancy is yet to be elucidated fully in all sites.

The greatest clarity exists in cervical cancer where it is considered that in virtually all cases, cancer arises in a sequential or step-wise fashion; acute viral infection is followed by detectable viral persistence that over decades leads to cervical precancer and invasion[68].

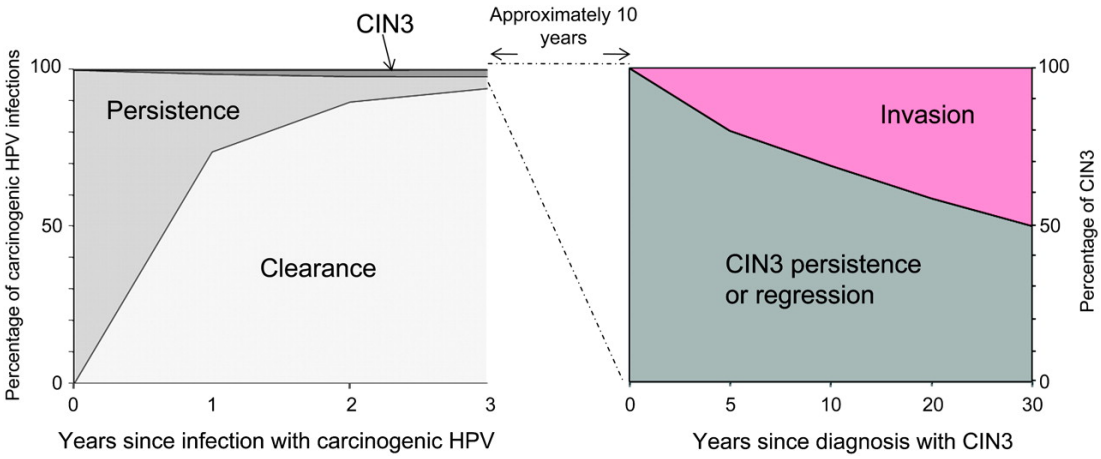


Figure 3 The Natural History of Cervical HPV infection and Carcinogenesis

Representation of the risk associated with viral persistence and progression. Left graph demonstrates proportional population based viral infection outcomes from infection to either clearance or cervical dysplasia. Right graph depicts the implications of virally induced dysplasia (CIN3) untreated over 30 years following diagnosis (modified from Schiffman et al 2011[68]).

However, viral clearance by half of newly infected individuals interrupts this process within the first 6 months, and 90% of women will be clear of HPV infection by 12 months[68]. As Figure 3 indicates, the rate of HPV clearance from the cervix of infected women does subsequently diminish over time and, by contrast to infections that clear, cancer risk increases substantially for the 5% of women whose infection persists.

Infection with high-risk HPV is almost ubiquitous for the sexually active population[69,70] as HPV infections are easily transmitted by a variety of sexual contacts, either to or from the anogenital tract to the oral cavity. Gillison et al[71] reported the first population-based study to concurrently examine the epidemiology of oral HPV infection among both men and women following analysis of data from the National Health and Nutrition Examination Survey (NHANES). The authors found infection with any HPV type amongst men and women aged 14 – 69 was 6.9% and for HPV16 specifically, only 1%. Of particular interest was a 5-fold increase in HPV16 prevalence in the oral cavity of men and a bimodal distribution of incidence by age, with peaks early in the forth decade of life and again in the early seventh decade[71]. The oral infection rate is considerably lower than that seen in the genital tract and it remains to be seen whether these static data collections are stable or more dynamic, as such reflecting a fundamental change in global infection and carriage providing an insight into the mechanisms behind rising HPV positive OPSCC rates. It is similarly unclear why males have a significantly elevated oral HPV carriage rate, however one hypothesis centers on the elevated burden of viral copy number within an infected cervix by comparison to the penis, and follows therefore that viral transmission via oral sex is more likely from a woman to a man (cervix to

mouth) than visa versa (penis to mouth). Evidence of increased HPV transmission amongst heterosexuals from cervix to penis, rather than penis to cervix would seem to support this hypothesis[72].

An increasing body of evidence suggests that oral HPV infection is unlikely to be associated with casual non-sexual contact. Infection seems consistently associated with sexual behavior, exemplified by an 8-fold increase in infection in sexually experienced individuals and a significant increase in infection rates as number of sexual partners increases[71].

The natural history of oral HPV infection is similar to that of anogential disease with most prevalent infections having been cleared within the first 12 months[73,74]. Due to the lack of a clinically apparent dysplastic lesion in HPV positive OPSCC, analogous to that seen in the cervix (cervical intraepithelial neoplasia, CIN), it is difficult to describe a temporal relationship between persistent infection and neoplastic change. As is the case for the natural history of HPV infection, much of the understanding of viral persistence and epithelial transformation therefore is gleaned from cervical cancer data.

Analysis of pooled data from eight large observation studies comprising 5642 cases of HNSCC and 6069 controls demonstrated that the risk of oropharynx cancer in particular was associated with six or more lifetime sexual partners (odds ratio 1.25; 95% CI 1.01 – 1.54) and of greater significance, four or more oral sex partners of their lifetime (odds ratio 3.36; 95% CI 1.32 – 8.53)[75].

Human papillomavirus-16 in Head & Neck Squamous Cell Carcinoma

Several case-series conducted in the late 1990s and early 2000s evaluated the point prevalence of HPV infection within oropharyngeal cancers and culminated in a systematic review by Kreimer et al.[76] in 2005 that concluded that over one third (35.6%) of OPSCC contained HPV DNA (87% of which were HPV16). In addition to viral DNA presence, several studies have added to the burden of proof necessary to conclude that HPV has a causal relationship in these tumours. These studies demonstrated localization of HPV DNA to the cell nucleus[77], evidence of viral DNA integration[77,78], elevated viral copy number in malignancy[79] and, perhaps most importantly, evidence of viral oncogene expression[80].

Outwith the oropharynx, HPV DNA presence in HNSCC cases was reported to be lower yet still a significant minority of cases. In the systematic review mentioned previously, 23.5% of Oral SCC and 24% of laryngeal SCC were similarly HPV DNA positive[76]. To date, the same supporting molecular evidence as is available for the oropharynx has been lacking. In addition to sample testing inadequacies, there has been a recognised failure to adhere to strict site classification[64]. The literature has a variety of HNSCC site terms that are applied differently by different authors leading to potential confusion; oral SCC (OSCC) can be used to imply the oral cavity or the wider oral and oropharyngeal region and similarly OSCC has been used in specific reference to the oropharynx alone. There is therefore, an understandable potential for misclassification of some HPV positive oropharynx cancers as “oral” either clinically or literally in publication[64].

As such, additional molecular and epidemiologic studies with strict site classification have been called for to further evaluate the association of HPV infection with oral cavity cancer and larynx cancer[38]. This carries particular relevance when consideration is given to the substantial contribution that these subsites make to total global HNSCC cases. Even a modest HPV-attributable proportion to either oral or larynx cancer would translate to a substantial tumour burden.

Whilst the proportion of HNSCC cases that are HPV positive varies by head and neck subsite, the same is also true more generally between different populations and global geographical regions[38]. A case control study conducted in Latin America and Central Europe between 1998 and 2003 reported HPV prevalence in OPSCC of 4.4%[81] whilst Scandinavian data from 2006-7 indicated that OPSCC, specifically from the tonsil, demonstrated HPV positive malignancy in 93% of cases[82]. When considered collectively, HPV prevalence in OPSCC from North American studies was 47%, whilst in Asia this was 46% and 28% for Western Europe. This geographical heterogeneity in HPV positive OPSCC rates is likely to be influenced both by sexual practices in the differing ethnic and cultural groups and also the extent to which tobacco and other traditional HNSCC risk factors play in populations[38].

Despite variations in the HPV positive fraction of OPSCC in various regions, there is consistency in the trend of increasing overall OPSCC incidence over time irrespective of geography[63,83-90]. The combination of falling tobacco consumption in combination with evidence of increasing HPV proportions coinciding with the incidence rise has lead to use of the term “virus-related epidemic”[82]. Certainly retrospective analyses of OPSCC cohorts in the USA[63], Australia[84] and Sweden[82]

would support this suggestion, each demonstrating substantial, sustained elevations in HPV positive OPSCC.

Clinical Implications of HPV-positive OPSCC

Clinical Features

The clinical features of HPV positive OPSCC are distinct from HPV negative counterparts. HPV positive tumours tend to present at a more advanced stage, typically with a small primary tumour with advanced nodal disease in the neck[3,91,92], a significant proportion of which display cystic degeneration within cervical metastatic deposits[93]. Individuals with HPV positive OPSCC display differing sociodemographic and behavioral characteristics. HPV positive patients tend to be 5-10 years younger than HPV negative individuals and are less likely to drink alcohol or smoke tobacco[94], whilst having had a higher number of sexual partners (both generally and for oral sex)[95]. In developed nations, individuals are more likely to be of white ethnicity than any other[95].

Prognostic Significance of HPV positive OPSCC

Tumour HPV status is an important and independent predictor of both disease free survival and overall survival in OPSCC. The first indications of such a survival advantage became apparent in a single-institution case series of tonsillar SCC[96]. At 3 years, survival for HPV positive individuals was 65.3% compared with 31.5% in the HPV negative group (odds ration 4.18).

This level III evidence was subsequently supported by a meta-analysis of several individual case series[97] and importantly by analysis of outcomes of individuals stratified by HPV status in the Eastern Cooperative Oncology Group (ECOG) 2399 phase II trial[98]. The latter indicating that following a median survival of 39 months, HPV positive individuals demonstrated a 73% (HR, 0.27; 95% CI, 0.1 – 0.75) reduction in risk of progression and 64% (HR, 0.36; 95% CI, 0.15 – 0.85) reduction in risk of mortality by comparison to HPV negative individuals following adjustment for age, tumour stage and performance status.

The potential for differential survival benefit beyond the confines of an aggressive, multimodality therapeutic trial has been demonstrated for surgically treated patients with or without postoperative radiotherapy[99], conventional radiotherapy alone[100] and chemoradiation (in a large phase III trial with variation of radiotherapy delivery; fractionated and accelerated fractionation)[91]. Interestingly the evidence from Ang et al[91], in reporting the results of the RTOG0129 trial, suggests that further stratification of disease outcomes can be made when consideration of both HPV status and tobacco exposure is made. They demonstrated that tobacco smoking was independently associated with overall survival and progression-free survival in both HPV subgroups and the magnitude of tobacco effect was similar in both groups with a resultant survival classification (high, intermediate and low risk categories) based on HPV status, tobacco consumption and tumour stage. The authors concluded that this evidence suggested HPV status and tobacco smoking are major independent risk factors in OPSCC and go on to infer that the observed survival differences are a consequence of differing molecular profiles and the resultant difference in biological behavior of tumours.

In the broader context of all HPV positive OPSCC, a biological rationale for the improved survival outcomes has yet to be defined. It is conceivable that the presence of wild type p53 and Rb tumour suppressor genes in HPV-transformed cells makes them more susceptible to treatments which induce additional cellular stress, sufficient to tip the sensitive balance between p53/Rb production and their sequestration through the effects of E6 or E7. This would in turn reactivate apoptotic and cell cycle regulatory pathways in a way not generally possible in tobacco- or alcohol carcinogen-transformed cells with mutated p53 and or Rb genes[55].

The absence of field cancerisation in HPV positive cohorts may have reduced the incidence of second primary tumours with resultant improvement in outcomes[99]. Furthermore, HPV positive tumours have been shown to demonstrate reduced expression of prognostic biomarkers of poor outcome such as epidermal growth factor receptor (EGFR), although this may be a function of reduced tobacco-usage associated tissue hypoxia rather than a direct HPV-associated effect[101].

Finally, activation of the host immune response by unexplained treatment associated factors may also contribute to improved outcomes[102].

2.4 HPV-MEDIATED TUMOURGENESIS

Molecular Pathogenesis

HPV is reliant on the host cells replication machinery to copy its own DNA however the cellular proteins necessary for this replication are only apparent in actively dividing cells. As a consequence HPV encodes proteins of its own to maintain the host cell in a dividing state but in doing so the potential for cell cycle disregulation and inappropriate cell division becomes a possibility.

The viral proteins responsible for this alteration of cellular homeostasis are the transforming oncoproteins E6 and E7[4] and to a lesser extent E5[103]. As previously discussed (above) they are also of importance in immune evasion, being involved in both innate and adaptive immune response.

The oncogenic role of the HPV E5 protein occurs early in the course of infection, as it promotes cellular proliferation through binding to the epidermal growth factor receptor (EGFR), platelet derived growth factor β receptor and colony stimulating factor 1 receptor[103]. The coding sequence for this protein is frequently deleted or disrupted in the process of viral DNA integration in established or later infection, leading to the presumption that its persistent expression is not fundamental requirement or necessity for ongoing oncogenesis[4]. The maintenance of the malignant phenotype is predominantly a function therefore of the major oncogenes E6 and E7.

HPV E6 protein facilitates proteosomal degredation of p53 leading to loss of cell cycle arrest and apoptosis in response to DNA damage[104]. The formation of a complex between E6 and the ubiquitin ligase, E6-associated protein (E6AP), initially facilitates ubiquitination of p53 causing its rapid degredation and removes the cell cycle checkpoint control (G1/S and G2/M) usually afforded by p53. As a result, DNA damage and other cellular stresses have an increased opportunity to be translated into persistent genomic instability through loss of p53-mediated DNA repair or, where replication of damaged DNA has already occurred, the loss of p53-induced apoptosis[105] (Figure 4D). Interestingly, E6 from low risk HPV types also binds E6-AP and p53 however the low risk type have the capacity to produce only a single variant of E6, whilst high risk types such as HPV16 and HPV18 produce both the full E6 transcript and two splice variants (E6^I and E6^II). It is through these E6 variants that high risk HPV mediates its fundamentally different, transformative impact on cells[9,14].

p53-independent anti-apoptotic influence occurs through a variety of mechanisms including E6's downregulation of BAK-induced apoptosis. BAK, a member of the Bcl-2 family normally highly expressed in fully differentiated cells, will induce a caspase apoptotic cascade, however E6 mediates its degredation via ubiquitination[14]. This would clearly be of benefit to the virus in completing its life cycle in differentiated cells but similarly reduces opportunities for exclusion of damaged DNA from further replication.

E6 also contributes to induction of telomerase activity through activation of hTERT and in doing so promotes cellular immortalization though telomere length maintenance and indefinite proliferation[14]. hTERT, in normal conditions, is under

the transcriptional repressive control of NFX-91 however the E6-E6AP promotes NFX-91 degredation leading to hTERT transcription.

Interactions of E6 with several other cellular factors have been documented[106,107]. These interactions affect the keratinocytes transcription and differentiation, induce telomerase activation and/or lengthen cell life span thus maximizing HPV amplification but also contributing to conditions favoring malignant change/progression.

The HPV E7 protein binds to the cullin 2 ubiquitin ligase complex and causes ubiquitination of the retinoblastoma protein (pRb) tumour suppressor through proteosomal degredation[108]. The consequence is once more seen in disregulation of cell cycle control, on this occasion resulting in unrestricted progression through the G1/S cell cycle checkpoint[4]. Without pRb, the transcription factors from the E2F family of proteins induce transcription of S phase promoting genes that leads to further cellular proliferation[4] (Figure 4).

E7 from low risk HPV type will still also bind to pRB although generally with significantly reduced affinity. This reduced affinity to low risk types is not exclusive however, as HPV1 E7 displays binding comparable to that of HPV16, indicating that mechanisms other than pRB binding are of importance in HPV-mediated cellular transformation[9].

E7 interacts with pRb-related pocket proteins, p107 and p130, and the cyclin-dependent kinase (CDK) inhibitors having a further inhibitory effect on cell cycle arrest[4] (Figure 4B & C). Under normal cellular conditions, cyclin dependent kinase

activity maintains control over phosphorylation of pRb and reducing resultant release of pRB from the pRB/E2F complex. E7 binds to the pocket proteins and in a similar fashion to CDKs, results in active E2F release that in turn promotes transcription of cell cycle proteins such as cyclin E & A. The process of positive feedback ensues driving the cell through the G1/S restriction point[14].

E7 targeting of the CDK inhibitors p21 and p27 causes neutralization of their normal inhibitory effects on cell cycle, once again promoting cell cycle progression.

In addition to their individual autonomous impact on the cell, the viral proteins also work in concert to compliment one and other. Whilst the impact of E7 on the cell may ordinarily increase the likelihood of apoptosis, E6 counteracts this through p53-dependent and –independent anti-apoptotic influence. Similarly E7 rescues E6 from CDKN2A (p16) inhibition that follows Rb knockdown, by direct activation of cyclins A and E and functional inactivation of p16, bypassing its cell cycle regulation[4]. It is this marked, albeit ineffective, upregulation of p16 expression[109], occurring as a cellular feedback mechanism intended to restrict cell cycle progression in the face of pRb function loss, that has been used as a surrogate for HPV positive status in diagnostics tests (2.5 HPV Detection Techniques in HNSCC).

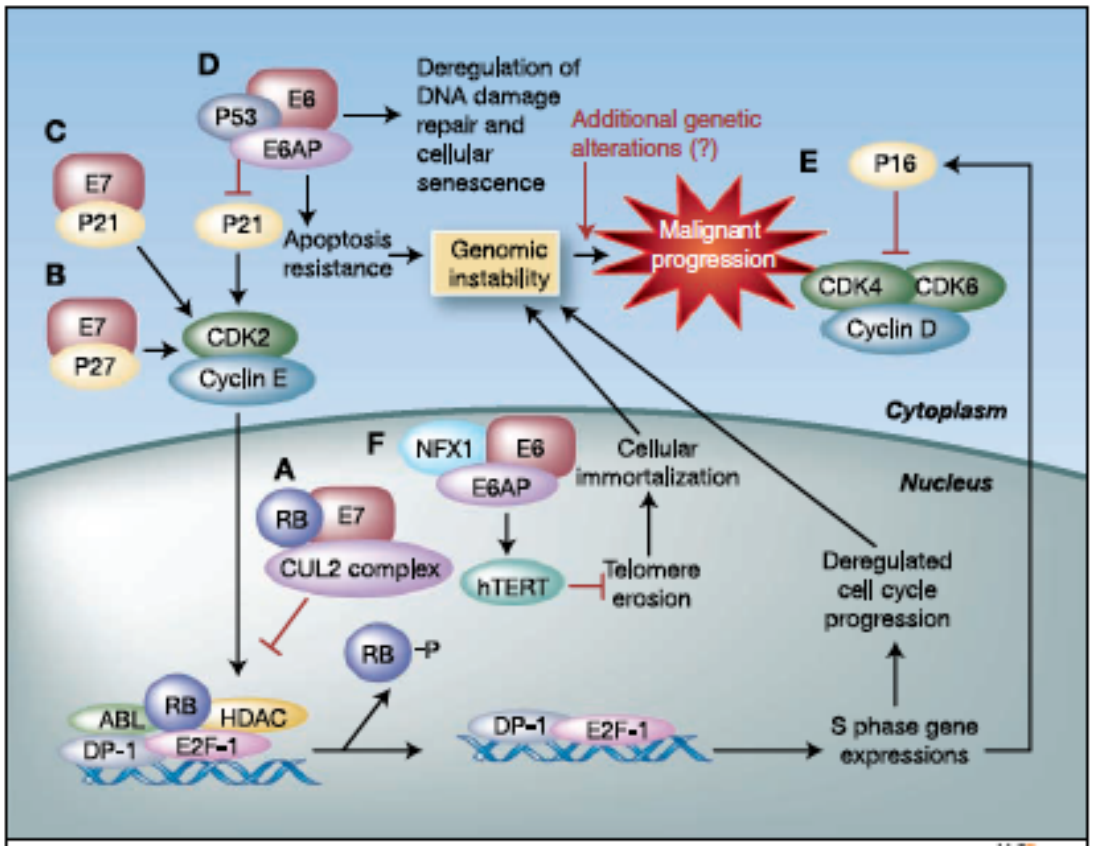


Figure 4: Cellular influences of HPV oncoproteins in the malignant transformation of keratinocyte

From clockwise inferiorly, A – ubiquitination of pRb thought the actions E7 and the cullin 2 ubiquitin ligase complex (CUL2); B and C – interaction between E7 and p27 & p21 (respectively) with resultant inhibition of cell cycle arrest contributing to carcinogenesis; D – ubiquitination of p53 by E6 and the ubiquitin ligase E6AP leading to p53 degredation; E – overexpression of p16 in response to downregulation of pRb and F – degredation of the hTERT transcriptional repressor, NFX1, following the association of E6 and E6AP. Consequently hTERT is activated leading in turn to cellular immortalization.
(Modified from Chung, CH and Gillison, ML[4])

E6 and E7 expression is primarily under tight self control through the regulatory influence of the early viral gene transcriptional repressor E2[15].

The function of the E2 gene is frequently disrupted during the process of carcinogenesis as the circular HPV genome linearises and inserts into the host genome[110]. Although the point of insertion into the host genome appears to be relatively random, with a predilection for chromosomal fragile sites, within the

circular viral genome the E1/2 region of HPV appears to be preferentially interrupted[110]. The consequence of which is release of the viral oncogenes, E6 and E7, from transcriptional repression by E2 leading to E6/7-mediated alteration of key tumour suppressor pathways, as already discussed. In vitro evidence would suggest that, rather than merely an inconsequential event, integration appears to be a fundamental step in cervical oncogenesis. Restoration of E2 expression, in integrant cell lines, results in repression of E6 and E7 expression and a detrimental effect on cell proliferation via induction of arrest in the cell cycle in G1[111,112]. Additionally, E2 produces an apoptotic effect independent of the E6-p53 interaction well as both cellular senescence and apoptosis[113]. As will be discussed in greater detail (7.1), there remains conflict within the literature as to the true rate of viral integration and relevance of integration where oncogene expressing episomal transcripts exists in malignant HNSCC[97].

Despite the potential to induce substantial perturbations in both DNA repair and cell cycle regulation, HPV E6 and E7 proteins are not sufficient to induce malignancy alone, however additional genetic events are, as yet, unclear[4].

Immune system response in HPV-positive OPSCC

It is clear from previous studies of HNSCC as a whole, that evidence of immune response, in particular cytotoxic T cell infiltration at the tumour-stromal interface correlates with improved outcome measures such as risk of recurrence and death[114]. Recently, using a HNSCC cohort of limited size, Jung et al. produced

evidence of greater infiltration by cytotoxic CD8+ T cells within HPV positive tumours than was seen in HPV negative tumours[115]. Other studies have shown circulating markers of heightened immune response (raised CD8+ T cells and a lower CD4+/CD8+ ratio) in pretreatment HPV positive OPSCC, features seen in common with improved tumour response following induction chemotherapy[116]. It has therefore been suggested that such responses might play a part in the improved outcomes seen in HPV positive cohorts. Further there is evidence that within HPV positive HNSCC levels of expression of immune response genes are statistically higher by comparison to HPV negative cohorts, although the functional significance of this finding is yet to be fully elucidated[115].

2.5 HPV DETECTION TECHNIQUES IN HNSCC

As the recognition and understanding of the involvement of HPV in OPSCC has evolved, its importance as a prognostic biomarker has been clearly demonstrated[98,100].

Outwith the Oropharynx, the role that HPV plays has been a source of controversy both in malignant and premalignant mucosal disease[52,64,117]. This has been in part due to a lack of stringent site classification of included samples for analysis[52] but also as a consequence of considerable variability in the testing regimes applied[118,119]. Understandably calls have been made to standardise the definitions and clarify the best test or combination of tests for accurate diagnosis[118,120].

Recently, evidence-based clinical management guideline documents have been published detailing a recommendation that HPV testing be undertaken for head and neck squamous cell carcinomas; specifically for those arising in the oropharynx and where metastatic squamous cell carcinoma of unknown origin is evident (National Comprehensive Cancer Network, USA; College of American Pathologists; ENT UK; Royal College of Pathologists, UK). Of note however, is the fact that the guidelines are not prescriptive when detailing the laboratory tests required to establish HPV status since an 'international standard' for HPV testing in head and neck cancer is yet to be defined[118]. Currently, a variety of detection methods are available[119], each with specific benefits and detractions.

Detection techniques vary according to the tissue source being analysed. In clinical practice, diagnostic specimens are typically placed into fixative agents to preserve

the cellular morphology and tissue architecture primarily, although not exclusively, to allow histopathological analysis[119]. The fixation and embedding process induces undesirable degradation of nucleic acid, particularly RNA, resulting in considerably lower quality nucleic acid by comparison with fresh frozen tissue[121-123]. Additionally, duration of fixation, processing temperatures and post fixation/embedding storage my also contribute to reduced nucleic acid stability[122,124].

An alterative to FFPE tissue is fresh frozen tissue, either collected into protective media[123] or frozen directly to temperatures below -75 °C to protect nucleic integrity thus allowing extended storage times and utilisation at a later date, however an inability to conduct conventional histopathological assessment is a significant drawback[123].

It is unsurprising, given the reliance of routine histopathology services on FFPE samples, that the majority of HPV diagnostic test developments have focused on these samples rather than fresh frozen tissue resources. This reliance on FFPE has, to date, had the effect of precluding the use of the gold standard or reference test for HPV diagnostics; evidence of expression of viral oncogenes witnessed directly within tumour tissue, which is believed to require fresh tissue derived samples[125]. This introduction details the present understanding of HPV diagnostics in HNSCC and explores the clinical applicability of those tests, both diagnostic and prognostic.

Viral Oncogene Expression (HPV mRNA qPCR) – The "Gold Standard"

Sustained and persistent expression of high-risk HPV E6/E7 viral oncogene is a fundamental requirement for both the initiation and the maintenance of an HPV-

driven malignant phenotype[3]. It is known that the oncogenic effects of high-risk HPV E6/E7 driving OPSCC correlate with cellular genotoxic damage and gene expression changes which are the hallmarks of cancer[126]. As a consequence, demonstration of transcriptionally active oncogenic viral infection on samples derived from fresh tissue has been considered to be the reference or gold standard test. The practical application of this test is usually by means of quantitative reverse transcriptase polymerase chain reaction (qRT-PCR) amplifying high-risk HPV E6/E7 mRNA transcripts[127,128]. Such an approach is capable of providing precise quantitative assessment of both the viral oncogene transcript abundance within a single sample, by comparison to a constitutively expressed endogenous 'housekeeping' gene, and also relative expression levels between references (cell line samples) or other clinical samples.

Although its application to routine clinical diagnostic samples is limited, the use of HPV qRT-PCR in HNSCC samples was fundamental in framing the causal relationship of HPV16 in HNSCC[127,128].

Viral DNA Detection (DNA PCR, DNA in situ hybridisation)

Polymerase Chain Reaction (PCR) for HPV DNA target amplification can rely on either non-quantitative post amplification recognition of the target sequence or simultaneous quantitative detection utilising in-built reporting systems. Such a process can utilise type-specific primers or degenerate primers capable of recognising multiple subtypes in combination with a further step of type specific

PCR or hybridisation to custom array chips containing probes for a specific HPV types[129].

Given the high analytical sensitivity of PCR techniques there is a significant prospect of assays detecting HPV DNA presence that is merely a transient opportunistic contaminant rather than a driver of malignancy, this may particularly be the case for non-quantitative techniques[130]. Quantitative PCR (qPCR) has the specific advantage of reporting sequence presence in relation to concurrently amplified endogenous genes, for example β-actin. This allows reporting of viral detection in terms of copy number, from which a biologically relevant threshold can be applied[119]. This however, does not necessarily reflect the biological relevance of the viral DNA that is present, as has been highlighted in a combined analysis of both DNA detection and viral gene expression in clinical samples which found 50% (12/24) of cases with HPV DNA present lacked evidence of viral expression [128].

A further limitation of PCR-based techniques, when applied to HPV diagnostics, is the arbitrary nature of the diagnostic threshold for a positive test. Although these thresholds can be accommodated into a logical biological rationale (for example 1 viral copy per host genome) PCR is not capable of differentiating between detected copies originating from cells with multiple copies of virus (eg. productive HPV infection) or, the intended, single copy per malignant cell[130].

The application of DNA in situ hybridisation (DNA ISH) technology to FFPE tissues is a commonly utilised and clinically validated technique for HPV diagnostics[131]. Using nucleic acid probes which are specific for HPV sequences of interest (either for specific viral subtypes or groupings such as high-risk subgroups), viral DNA can be

detected at a cellular and subcellular level (localization to cytoplasm or nucleus)[123]. Introduction of various signal amplification steps have further increased the sensitivity of this technique, such that visualization of a single target sequence per cell is feasible[131]. Primarily the sensitivity, but also specificity of this test has been questioned given that probes may in certain conditions bind to similar sequences that are not a prefect match and it is difficult to exclude potential probe cross-reactivity[123].

p16 immunohistochemistry (p16 IHC) as a Surrogate Marker of HPV-mediated Malignancy.

The use of p16 immunohistochemistry to infer HPV status in tonsillar SCC was first described in 2003 by Klussmann et al.[132]. At a molecular level, p16 protein accumulates in HPV positive cells as a consequence of the effect of the viral E7 protein on Rb (Figure 4). In response to sequestered Rb, p16 is released from transcriptional repression with a consequent elevation of protein levels within HPV positive cells. Using standard immunohistochemistry techniques, the ensuing elevated protein levels are detectable.

The main advantage of p16 IHC is the applicability to FFPE specimens using techniques common to most clinical pathology laboratories. However, its analytical performance has been highlighted as suboptimal by some authors[125,133] and off target effects, such as staining in histologically normal tissues, have called into question its clinical utility as a stand alone HPV diagnostic test.

Despite concerns such as those detailed above, p16 IHC remains a frequently applied test both in routine clinical practice and it is the sole HPV diagnostic test used to determine case inclusion for several major clinical trails designed to analyse therapeutic regimes in HPV positive OPSCC[*].

Recent attempts have been made to refine the diagnostic stringency of p16 IHC through the application of scoring or grading systems based on staining features such as intracellular stain localisation[134] or novel staining scores such as the p16 *H* score validated by Jordan et al[133] which uses the cross product of the proportion of cells stained and the intensity with which they stain. It is, as yet, unclear whether such an application of p16 can correct for concerns surrounding test specificity.

Diagnostic Algorithms

Alternative techniques to overcome variation in sensitivity and specificity of single tests, combined tests and diagnostic algorithms, have been proposed[125,135,136]. Smeets et al.[125] classified HPV status in 48 HNSCC cases by detectable viral oncogene RNA (RTqPCR for HPV16 E6 and E7) derived from fresh frozen tissue samples and gauged performance of fixed tissue based tests on corresponding FFPE samples for each case. Using the fixed tissue results, they were able to generate an

[*] **RTOG 1016**: Radiation Therapy Oncology Group Phase III Trial of Radiotherapy Plus Cetuximab Versus Chemoradiotherapy in HPV-Associated Oropharynx Cancer;
De-ESCALaTE HPV: Determination of Epidermal growth factor receptor inhibitor (cetuximab) versus Standard Chemotherapy (cisplatin) early And Late Toxicity Events in Human Papillomavirus positive oropharyngeal squamous cell carcinoma, NCRN Portfolio Study ID: 11723;
ECOG 3311: Eastern Cooperative Oncology Group Low Risk OPSCC: Personalized adjuvant therapy based on pathologic staging of surgically excised HPV positive Oropharyngeal cancer

algorithm capable of achieving reliable HPV status detection, 100% sensitivity and specificity for each case using a combination of p16 IHC and GP5+/6+ HPV DNA PCR. Subsequent validation of this algorithm was undertaken on an independent series (n=86) from within the same institution using matched FFPE and frozen tumour confirming sensitivity of 96% and specificity of 98%. [90]

Based on broad experience of epidemiological studies in HSNCC and clinical trial design, the John Hopkins Institution published their diagnostic hierarchy[136], which was simplified to diagrammatical form subsequently[119] (Figure 5).

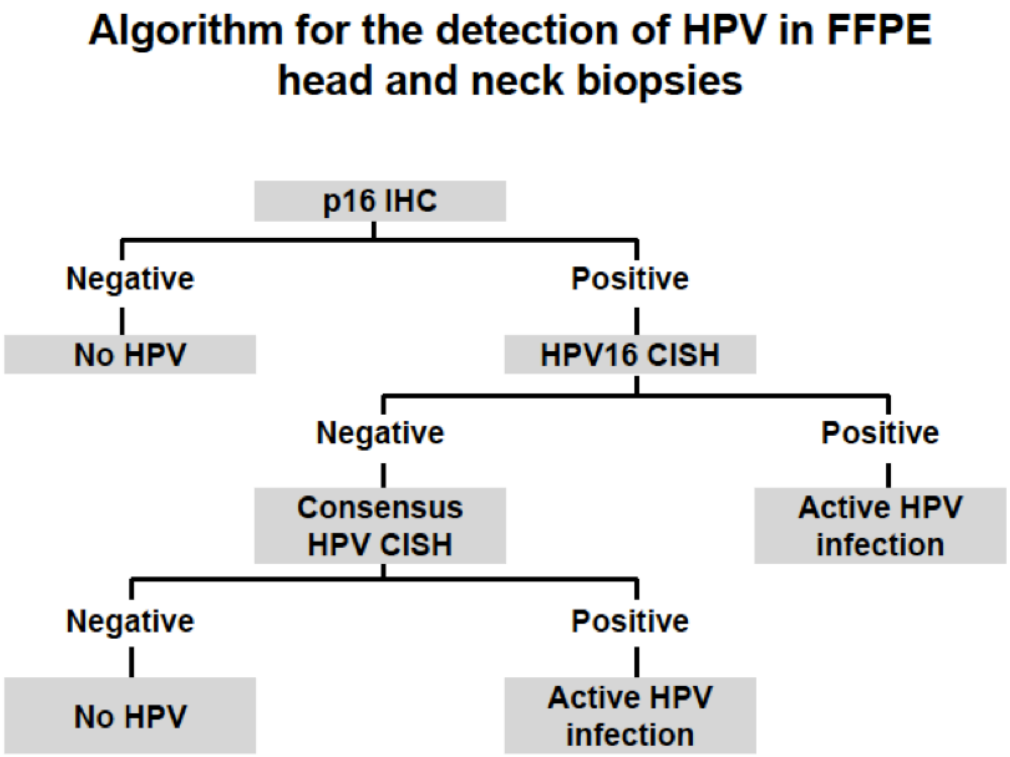


Figure 5: Algorithm for detection of HPV in FFPE head and neck biopsies (Modified from Robinson et al[119] as previously described by Westra[136])

Currently there is insufficient evidence to recommend one particular testing algorithm over another, however, the combination of p16 IHC with either an HPV

DNA directed PCR amplification technique or HPV DNA detection with signal amplification methods (ISH) have been suggested[125,131,135,137]. The basis upon which such recommendations have been made varies and only one study measures diagnostic tests against the gold standard for HPV-mediated malignancy, viral oncogene expression[125].

2.6 DNA METHYLATION IN HNSCC

Epigenetics refers to those heritable changes in gene expression that do not result from an alteration in the DNA sequence[138]. Four particular modifications or mechanisms are presently considered together under the umbrella term epigenetics; DNA methylation, covalent histone modification, nucleosome positional remodeling and microRNA. Although considered as discrete entities, they interact closely in order to impact upon gene expression[139].

Whilst contributing to the facilitation of appropriate gene expression in healthy cells, epigenetic control also contributes to dysregulated gene expression in a variety of disease states, including human malignancy[139,140].

DNA methylation occurs almost exclusively when a methyl group is added to the 5' position of cytosine rings that immediately precede guanine nucleotides in the linear DNA sequence (so called CpG dinucleotide or sites). CpGs are not distributed randomly throughout the genome and tend to cluster in regions called CpG islands regions of more than 200 bases which are particularly GC rich and have an observed

to expected ratio of CpGs of greater than 0.6[139]. CpG islands can be found within 60-70% of human gene promoters and, in contrast to the generally sparsely-distributed and hypermethylated CpGs in repetitive genomic sequences (intergenic regions and transposable elements), CpG islands are typically hypomethylated under normal conditions in order to facilitate gene expression[141].

Increases in DNA methylation are associated with chromatin remodeling and a subsequent reduction in transcriptional activity. In addition to the implications in normal biological processes, both specific and complimentary epigenetic alterations contribute to disease pathogenesis; in particular, epigenetic aberrations appear to be of increasing relevance in specific human malignancies[140,142].

DNA methylation in malignancy

Evidence of the frequent epigenetic aberrations apparent in human malignancy[142,143] raises the possibility of exploiting of these changes for clinical benefit; diagnostic (both early detection and definitive diagnosis), predictive and prognostic biomarker evaluation, therapeutic stratification and disease monitoring.

Two patterns of DNA methylation are specifically observed in malignancy; firstly, global or genome-wide hypomethylation, a large proportion of which is confined to repetitive DNA elements, and secondly gene specific hypermethylation of CpG island within the promoter regions of particular genes, which may in turn result in reduced gene expression[144,145].

Generation of global DNA methylation has been hypothesised to have evolved from initial host attempts to silence exogenous DNA from pathogens such as viruses[146]. Reduction in genome-wide methylation, or global hypomethylation, in malignancy is thought to increase genomic instability[147], loss of imprinting and activation of oncogenes[148] and as such alter clinical outcomes,[149-151] which is clearly of particular clinical relevance. These observations are supported from a mechanistic standpoint by the observation that global hypomethylation becomes more pronounced during neoplastic transformation from initial dysplastic lesions to invasive malignancy[148].

Attention towards site-specific hypermethylation in human cancers has focused on the CpG islands of promoter regions of genes, particularly those of tumour suppressor genes, which may become functionally silenced by elevation in methylation[148]. There is an extensive catalogue of such genes that have been demonstrated to be functionally inactivated in cancer by promoter methylation (Table 3) impacting on a variety of cellular pathways involved in oncogenesis; cell cycle control, apoptosis, DNA repair, cellular adhesion & invasion, angiogenesis[143]. Although not an exhaustive list, it does draw attention to the variability in the profiles of hypermethylated TSGs between different cancer types.

Gene	Function	Location	Tumour type	Consequences
MLH1	DNA mismatch repair	3p21.3	Colon, endometrium, stomach	Frameshift mutations
BRCA1	DNA repair, transcription	17q21	Breast, ovary	Double-strand breaks?
$p16^{INK4a}$	Cyclin-dependent kinase inhibitor	9p21	Multiple types	Entrance in cell cycle
$p14^{ARF}$	MDM2 inhibitor	9p21	Colon, stomach, kidney	Degradation of p53
$p15^{INK4b}$	Cyclin-dependent kinase inhibitor	9p21	Leukaemia	Entrance into cell cycle
MGMT	DNA repair of 06–alkyl-guanine	10q26	Multiple types	Mutations, chemosentivity
GSTP1	Conjugation to glutathione	11q13	Prostate, breast, kidney	Adduct accumulation?
p73	p53 homologue	1p36	Lymphoma	Unknown
LKB1/STK11	Serine–threonine kinase	19p13.3	Colon, breast, lung	Unknown
ER	Oestrogen receptor	6q25.1	Breast	Hormone insensitivity
PR	Progesterone receptor	11q22	Breast	Hormone insensitivity
AR	Androgen receptor	Xq11	Prostate	Hormone insensitivity
PRLR	Prolactin receptor	5p13–p12	Breast	Hormone insensitivity
TSHR	Thyroid-stimulating hormone receptor	14q31	Thyroid	Hormone insensitivity
RARβ2	Retinoic acid receptor-β2	3p24	Colon, lung, head and neck	Vitamin insensitivity?
CRBP1	Retinol-binding protein	3q21–q22	Colon, stomach, lymphoma	Vitamin insensitivity?
RASSF1A	Ras effector homologue	3p21.3	Multiple types	Unknown
NORE1A	Ras effector homologue	1q32	Lung	Unknown
VHL	Ubiquitin ligase component	3p25	Kidney, haemangioblastoma	Loss of hypoxic response?
Rb	Cell-cycle inhibitor	13q14	Retinoblastoma	Entrance into cell cycle
THBS1	Thrombospondin-1, Anti-angiogenic	15q15	Glioma	Neovascularization
CDH1	E cadherin, cell adhesion	16q22.1	Breast, stomach, Leukaemia	Dissemination
CDH13	H cadherin, cell adhesion	16q24	Breast, lung	Dissemination?
FAT	Cadherin, tumour suppressor	4q34–35	Colon	Dissemination?
HIC1	Transcription factor	17p13.3	Multiple types	Unknown
APC	Inhibitor of β-catenin	5q21	Aerodigestive tract	Activation β-catenin route
SFRP1	Secreted frizzled-related protein 1	8p12–p11	Colon	Activation Wnt signalling
DKK1	Extracellular Wnt inhibitor	10q11.2	Colon	Activation Wnt signalling
WIF1	Wnt inhibitory factor	12q14.3	Colon, lung	Activation Wnt signalling
COX2	Cyclooxygenase-2	1q25	Colon, stomach	Anti-inflammatory resistance?
SOCS1	Inhibitor of JAK–STAT pathway	16p13.13	Liver, mieloma	JAK2 activation
SOCS3	Inhibitor of JAK–STAT pathway	17q25	Lung	JAK2 activation
GATA4	Transcription factor	8p23–p22	Colon, stomach	Silencing of target genes
GATA5	Transcription factor	20q13	Colon, stomach	Silencing of target genes
ID4	Transcription factor	6p22–p21.3	Leukaemia, stomach	Unknown
SRBC	BRCA1-binding protein	1p15	Breast, lung	Unknown
SYK	Tyrosine kinase	9q22	Breast	Unknown
RIZ1	Histone/protein methyltransferase	1p36	Breast, liver	Aberrant gene expression?
DAPK	Pro-apoptotic	9q34.1	Lymphoma, lung, colon	Resistance to apoptosis
TMS1	Pro-apoptotic	16p11	Breast	Resistance to apoptosis
IGFBP3	Growth-factor-binding protein	7p14–p12	Lung, skin	Resistance to apoptosis
TPEF/HPP1	Transmembrane protein	2q33	Colon, bladder	Unknown
SLC5A8	Sodium transporter	12q23	Glioma, colon	Unknown
HOXA9	Homeobox protein	7p15–p14	Neuroblastoma	Unknown
EXT1	Heparan sulphate synthesis	8q24	Leukaemia, skin	Cellular detachment
Lamin A/C	Nuclear intermediate filament	1q21.2	Lymphoma, leukaemia	Unknown
WRN	DNA repair	8p12–p11.2	Colon, stomach, sarcoma	DNA breakage, chemosensitivity

Table 3: Selective list of gene previously reported to be silenced by promoter hypermethylation in malignancy

(Modified from Esteller[143])

Regulation of DNA methylation

Both in normal physiology and in pathological situations, establishment of de novo methylation and maintenance of the pre-existing methylation pattern is the function of the DNA methyltransferases (DNMTs) and to a lesser extent related proteins responsible for recruiting DNMTs to positions requiring their action, for example UHRF-1.

DNMT1 is predominantly responsible for maintenance of methylation patterns during both DNA replication and repair. It displays a marked preference (x30) for hemimethylated DNA by comparison to unmethylated form[152], yet its action is not exclusively maintenance as it also demonstrates de novo methylation capacity[153].

DNMT3a and 3b are primarily viewed as de novo methylators, in effect establishing the pattern of DNA methylation in embryonic development, however their role is not strictly confined and they too undertake an element of methylation maintenance[153].

The UHRF1 protein has recently been shown to play a supporting role in methylation maintenance through a strong affinity for binding to hemimethylated DNA and subsequent recruitment of DNMT1[154]. This has been evidenced by significant reductions in methylation levels in UHRF1 knock out mice[154]. Once bound to DNA, UHRF1 extrudes targeted cytosines from the double helix into an active site when upon a methyl group transfer takes place from DNMT1 via UHRF1 to the cytosine[153]. Additionally, a facilitative role for UHRF1 in DNMT3a and 3b methylation has also been proposed[153].

In reality, the role of DNMTs and their related proteins in DNA methylation is likely to be more collaborative in nature, where DNMTs act in concert with one and other rather than in exclusivity[153].

The significance of DNMTs in cancer biology lies with the implications they may have in silencing of tumour suppressor genes. Accumulating evidence indicates elevated levels of DNMTs in a variety of tumour types[155-157] and that this elevation frequently correlates with hypermethylation of key TSGs and statistically significant worse clinical outcomes[158-161].

In a similar fashion to DNMTs, UHRF1 has been shown to be a key epigenetic switch, which controls cell cycle progression in Non-Small Cell Lung Carcinoma through its ability to induce transcriptional silencing of tumour suppressor genes by maintaining their promoters in a state of hypermethylation[161].

Within HPV driven malignancy the role that DNMTs may play in viral methylation state has not been explored and it remains of particular interest given that the virus lacks its own methylation machinery and must therefore come under the influence of, or potentially direct, the hosts' methylation machinery.

DNA Methylation in HNSCC

As in other malignancies, DNA methylation in HNSCC acts through either global or site specific methylation change to induce both genomic instability and TSG repression respectively. Methylation therefore plays an important role in both tumour initiation and progression. HNSCC is not dissimilar to other malignancies in

that it is a heterogenous disease and therefore methylation states and implications vary according to a variety of factors such as tumour site and aetiology (viral or otherwise)[145].

Exploration of methylation changes has received increasing interest due, at least in part, to their potential as early predictive biomarkers in premalignant lesions or as diagnostic biomarkers within surrogates such as such as saliva[145]. Also of relevance is the changing aetiology of HNSCC brought about the impact of HPV-mediated oncogenesis in the oropharynx and the implications that this may have on the host methylome[162]. Early data suggests a picture of variation in the methylation marks seen in different tumours based on HPV status alone[144,163,164]. This variation extends to include HNSCC and in particular HPV positive OPSCC

Although a variety of genes, selected via both candidate gene approaches and genome wide association studies, have demonstrated variations in specific TSG methylation levels when comparing tumours and normal pairs, there remains considerable variation reported[145]. Tumour heterogeneity may influence this, however a failure to differentiate between the varying head and neck subsites, and, by inference, HPV status, in some studies almost certainly contributes to variability in results[138].

2.7 HPV16 INTEGRATION STATE IN MALIGNANCY

Following viral infection of the keratinocyte, HPV DNA is thought to remain in a circular episomal form within the mucosal basal cell layer. Typically, there is maintenance of 50-100 copies of the virus during this latent cellular infection and remains so until the virally infected cell progresses through routine cellular differentiation to reach the upper layers of the stratified squamous epithelium. In contrast to the permissive infection of squamous epithelium where, upon terminal cellular differentiation, the viral life cycle completes with shedding of encapsidated virions, the transforming oncogenic infection results in non-productive infection and progressive cellular abnormality[165]. The factors discriminating between permissive and transforming infections remain particularly unclear, however viral integration has been suggested as an important potential mechanism in this process[166].

In cancer samples, HPV can be found in episomal (extra-chromosomal) form, integrated into the host genome, or in a mixed form constituting variable proportions of both episomal and integrated. Clinical studies of HPV integration in cervical neoplasia show that proportions of viral integrant and episome vary significantly, however there is a consistent trend towards increasing frequency of the former as the severity of cervical lesion increases (from early dysplasia to invasive malignancy)[110,167-169]. This observation would suggest that integration is a key event in cervical carcinogenesis and that it may facilitate genome instability[5].

Several particular consequences of viral integration have been considered of marked importance in oncogenesis[5];

1. Viral E2 gene disruption
2. Altered E6/E7 expression and or stabilization
3. Insertional mutagenesis
4. Numerical and structural chromosomal alteration

Viral E2 gene Disruption

In the majority of cervical malignancies, disregulation of viral early gene function, in particular the oncogene transcriptional repressor E2[170], occurs through cleavage of the circular viral genome and insertion into the host DNA[6]. Cells in which this occurs are thought to hold a selective growth advantage[171] although conversely, the integration event inhibits the viruses natural life cycle by removing key early genes necessary for synthesis of an infectious virus[6]. The implied importance of a disrupted E2 gene in oncogenesis has been supported by observations from cell culture experiments in which reintroduction of an intact viral E2 gene to established cervical carcinoma cell lines induced growth arrest and senescence due to reactivation of the p53 and Rb pathways[172,173]

In addition to the uncontrolled expression of viral oncogenes, integration-related loss of E2 removes its inhibition of hTERT expression[112], potentiating the effect of E6

on telomerase activation[14] and therefore increasing opportunities for cellular immortalization (Figure 6).

Altered viral oncogene expression and stabilization

Depending upon the site of integration into the host genome, *cis*-acting host sequences may potentiate viral oncogene expression[174] and, further, those integrant derived E6 and E7 transcripts may be stabilized by co-transcribed cellular sequences, enhancing oncogenic potential[171,175] (Figure 6).

Insertion mutagenesis

It has been suggested that integration of the HPV genome may activate cellular oncogenes[176,177] or disrupt key tumour suppressor genes[178] such that an additional neoplastic selective pressure becomes apparent. A systematic review of genomic integration sites within cervical squamous intraepithelial lesions (SIL) and invasive cervical malignancy by Wentzensen et al.[110] failed to provide evidence of frequent gene specific integration. Indeed they demonstrated that integration is an apparently randomly distributed across the human genome albeit with a clear predilection for chromosomal fragile sites (CFSs).

To date there is insufficient evidence to substantiate a definitive or essential role for insertional mutagenesis in HPV-induced carcinogenesis. It has been speculated that the absence of a viral integrase indicates that integration is a chance occurrence, a theory supported by the lack of specific integration sites within the host genome[110].

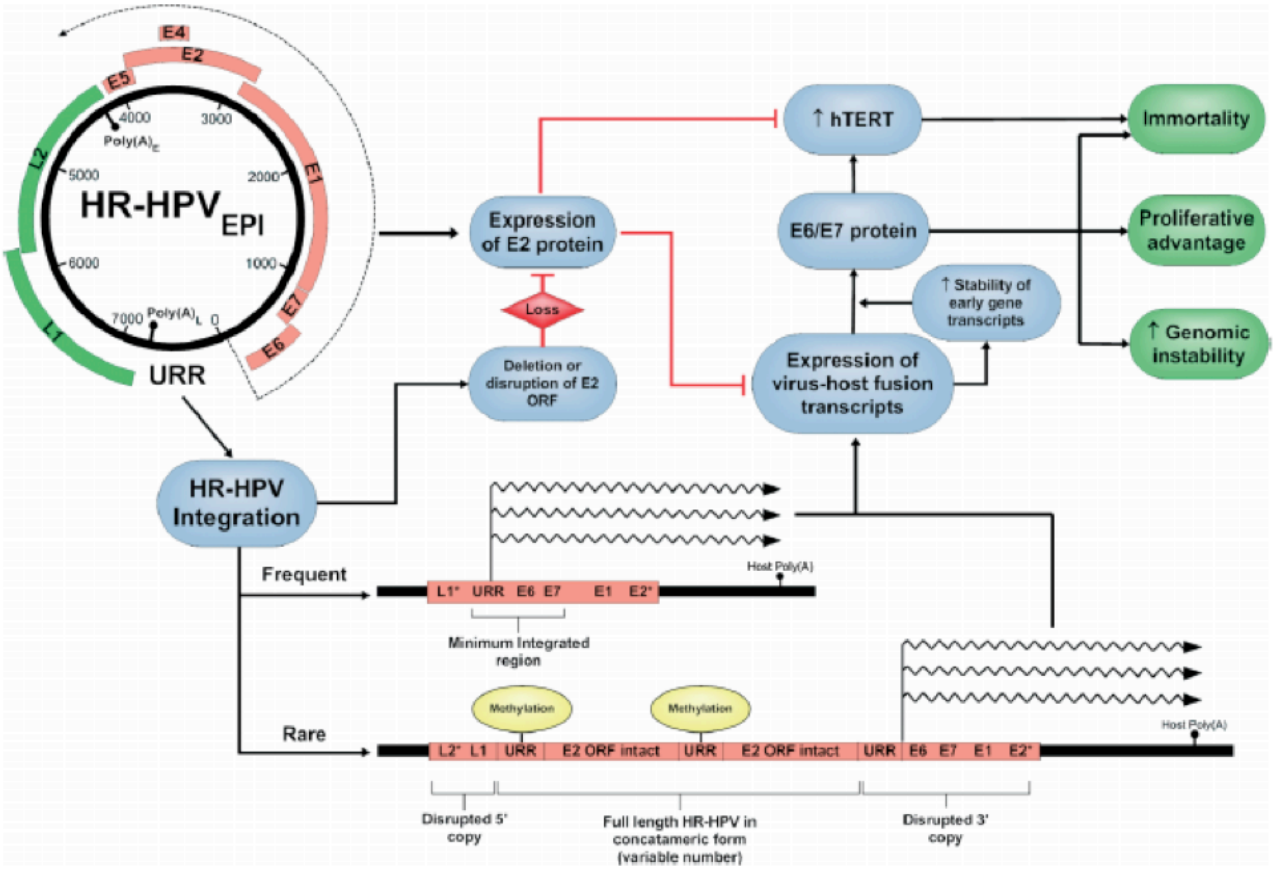


Figure 6: Molecular Consequences of Viral integration

Typically viral DNA insertion contains the viral oncogenes (E6 and E7) and the upstream regulator (URR) that is the start site for early gene transcription (dotted line). The blue text boxes represent the viral gene alterations and implications on host genes with subsequent implications for cellular disregulation (green boxes).
Viral integration can also dissociate the viral polyadenylation signal from early gene transcription, however leading to utilisation of host poly(A) signals once integrated, produce fusion transcripts with greater stability.
Less commonly, multiple viral integrants are inserted back to back, including intact E2 genes. Host methylation of the URR, limiting the functional significance of the multiple inserts is a speculative hypothesis for expression regulation[6,171] (Modified from Pett and Coleman[6]).

Integration in HPV positive HNSCC

Although the progressive impact of viral integration in cervical cancer has been described at length, the understanding within HPV positive HNSCC is somewhat different. Coupled with the obvious lack of a detectable analogous dysplastic lesion in the oropharynx, there seems to be a significantly different level of integrative events in the final invasive malignancy. In the cervix integration is evident in the vast majority of invasive lesions[5,167,168] whilst in OPSCC this level is markedly lower, ranging from 14 – 60%, depending on analytical method and tissue specimen type being analysed[96,179,180]. Explanations for this variation have yet to be advanced nor experimentally explored.

There are several techniques described to facilitate detection of HPV integration. They can be divided into those analyses that classify integration on the basis of implied evidence or those which allow direct detection, typically by means of sequencing techniques.

Implied evidence extends primarily from experimental data that suggests there is an integration-dependent disruption of the HPV E2 gene when any integration event occurs and hence detection of E2 status, integrity or expression informs integration state[181-183]. One weakness of such techniques surrounds the inference that detection of episomal viral DNA is absolute however such a finding belies the possibility of integrated virus “hidden” behind the noise/signal of episomal viral sequence.

Direct detection of integration differs in that fewer assumptions are made and evidence of integration is sought directly. Once again, several techniques have been described, ranging from Southern Blot analysis, with or without subsequent fragment sequencing, to the more recently described HPV specific techniques of Amplification of Papillomavirus Oncogene Transcripts PCR (APOT-PCR) and Detection of Integrated Papillomavirus Sequences PCR (DIPS-PCR) for detection of transcripts or genomic sequence respectively.

An ideal integration detection technique would provide direct evidence of viral cleavage position(s) within the circular viral DNA and accompanying host chromosomal integration position(s). Although a single viral copy could be all that is necessary. Additionally, the technique would allow detection of relative viral load, be applicable to clinical samples for biologically relevant analysis. Finally, it should involve a step that demonstrates active viral transcription, as this remains the prerequisite for defining a virally mediated malignancy as opposed to a potentially innocuous viral infection irrespective of copy number.

A novel approach to integration analysis is utilisation of next generation sequencing technologies with capture of HPV and host sequence at the position of viral cleavage/host insertion. To date, such an application of technique has yet to be adopted for integration analysis in HPV positive malignancy.

3 MATERIAL & METHODS

3.1 HUMAN TISSUE PROCUREMENT & CHARACTERISATION

Ethical Agreements for Clinical Sample & Data Collection

Fresh Frozen Tissue Samples

Tissue samples utilised in this research were sourced from patients treated at two sites within the Merseyside region, namely the Royal Liverpool & Broadgreen University Hospital Trust (1988 – 1996) and University Hospital Aintree NHS Foundation Trust (1992 – present). Tissue collection was undertaken following the granting of ethical approval by the South Sefton Research Ethics Committee (EC 47.01), the Liverpool (Adult) Research Ethics Committee (REC 07/Q1505/15) and North West 5 Research Ethics Committee (REC 09.H1010.54).
Utilisation of tissue resources collected under the latter two of these agreements was made possible following the granting of approval from the North West 3 – Liverpool East Research Ethics Committee (REC 10/H1002/53).

Tumours and adjacent normal uninvolved marginal tissue (where additionally available) was sourced from one hundred and eight (n=108) Oropharyngeal Squamous Cell Carcinomas (n=53 matched normal).

Formalin Fixed Paraffin Embedded Tissue Samples

Formalin fixed paraffin embedded (FFPE) tissue samples corresponding to the above fresh tissue samples were collected in parallel under the ethical agreements already

noted. Ninety-seven such samples were available for utilisation in TMA construction.

To quantify the potential contribution of stromal, inflammatory and adjacent, non-involved epithelial cells within the non-microdissected tissue samples a tumour cell burden analysis was performed by analysis of the tumour cell proportion within Haematoxylin and Eosin (H&E) stained, fixed tissue slides corresponding to the fresh frozen samples of 20 randomly selected cases.

Alternatively Sourced DNA & RNA Samples

Nucleic acids (DNA and RNA) extracted from 20 OPSCC at source by a collaborator conducting the PREDICT-TPF Trial (Primary Investigator; Dr B Lallemant, Faculty of Medicine Montpellier-Nîmes, CHU de Nîmes, France) was kindly offered for analysis. All nucleic acid extraction was undertaken from fresh frozen samples using the same techniques detailed below for locally sourced samples (page 63), with DNA and RNA quality assurance also following identical processes (page 67). Complimentary DNA synthesis was undertaken alongside all other OPSCC samples detailed in this research using the techniques detailed below (page 67).

DNA from Lymphoctyes derived from whole blood samples of healthy controls, sourced under ethical agreements granted to the Liverpool Lung Project & the Roy Castle Lung Cancer Foundation, was kindly provided by Dr George Nikolaidas.

3.2 TISSUE MICROARRAY (TMA) CONSTRUCTION & UTILISATION

TMA Construction

Tissue microarrays (TMAs) were constructed from formalin-fixed paraffin-embedded (FFPE) tissue blocks of OPSCC using a manual tissue arrayer (MTA-I, Beecher Instruments, USA), as previously described[184].

Recipient paraffin blocks were constructed from paraffin wax, melted in a wax bath, and subsequently poured into metal frame moulds, to reduce the incidence of air bubble entrapment that might subsequently affect TMA section integrity. Once set, blocks were inspected and any evident impurity or flaws in the block resulted in its rejection before use.

Array construction followed the procedure as detailed by Parsons et al[184]. Briefly, haematoxylin and eosin (H&E) stained sections were reviewed by a consultant Head & Neck histopathologist to identify areas of representative tumour and adjacent normal mucosa within each donor block. Triplicate tumour cores and solitary matched normal mucosal cores (height 4mm, diameter 0.6mm) were transferred from individual donor blocks to the recipient block employing a predetermined asymmetrical distribution. The asymmetrical core distribution was intended to ensure that the chance of confounding results due to positional staining artifacts were minimized for any individual case. Spacing between cores was 1.2 mm on the x-axis and 1.0 mm on the y-axis to increase core retention and minimise the chance of block fracture during construction.

Following completion of TMA construction, blocks were incubated at 40 °C overnight to facilitate bonding of the donor cores with the paraffin wax of the

recipient block. 5µm H&E stained sections of the TMAs were subsequently examined by a two consultant pathologists, blinded to the core tissue origin, to confirm accurate sampling of tumour bearing (or conversely normal) tissue.

TMA Utilisation

5 µm sections of recipient blocks were cut on a microtome and floated on a water bath at or below 42 °C before immediate transfer and mounting on 4% APES coated Superfrost Plus glass slides (Fisher Scientific, Leicestershire, UK). Slides were dried at 37 °C overnight before storage or usage.

3.3 p16 IMMUNOHISTOCHEMISTRY

$p16^{INK4A}$ Immunohistochemistry (p16 IHC) was carried out on 5 µm sections of TMA or whole mount blocks, prepared as detailed above. A proprietary kit (CINtec Histology; Roche mtm laboratories AG, Heidelberg, Germany) was utilised on a Ventana Benchmark autostainer (Ventana Medical Systems Inc, Tucson, AZ, USA). A tonsil squamous cell carcinoma with previously demonstrated high p16 expression was used as a positive control during slide staining. Omission of the proprietary primary antibody for p16 staining (E6H4™) was made for the purposes of negative control. p16 staining was scored as positive if there was strong diffuse nuclear and cytoplasmic staining present in greater than 70% of malignant cells[131]. All other staining patterns were scored as negative. Details of observer scoring and interpretation of test results are detailed in 3.6.

3.4 HIGH RISK HPV DNA IN SITU HYBRIDISATION

High Risk HPV DNA in situ hybridisation (HR HPV DNA ISH) was carried out using proprietary reagents (Inform III Family 16 Probe (B), Ventana Medical Systems Inc., Tucson, AZ, USA) on a Benchmark Autostainer (Ventana Medical Systems Inc.) applied to 5 µm sections of TMA or whole mount blocks, prepared as detailed above. The Inform III Family Probe (B) detects high risk HPV genotypes HPV16, -18, -31, -33, -35, -39, -45, -51, -52, -56, -58, -66. Three control samples were included with each slide; formalin fixed paraffin embedded samples of CaSki cells (HPV16 positive, approximately 870 copies per diploid genome, or approximately 1700 copies per near-tetraploid cell[185]), HeLa cells (HPV18 positive, 10-50 copies per cell) and C-33A (HPV negative control, Ventana Medical Systems Inc.). The HR HPV DNA ISH test was scored as positive if there was any blue reaction product colocalizing to the cell nucleus of malignant cells (either punctate or diffuse)[131]. Cases with specifically punctate nuclear staining cases were recorded in accordance with previously published reference to possibility of such staining pattern corresponding to genomic integration of viral DNA[186].

Details of observer scoring and interpretation of test results are detailed below.

3.5 HIGH RISK HPV RNA IN SITU HYBRIDISATION (RNAscope)

Detection of High Risk HPV E6/E7 mRNA on 5 µm sections of TMA sections, prepared as detailed above, or, where necessary, on 5 µm representative whole

mount sections, was performed using the High Risk HPV RNAscope kit (Advanced Cell Diagnostics, Inc., Hayward, CA, USA) as previously described[187] by the product manufacturers in their own research facility (Advanced Cell Diagnostics).

Briefly, 5 µm TMA sections were deparaffinised and pretreated with heat and protease before hybridisation with target-specific probes for the E6 and E7 genes of seven High Risk HPV genotypes (HPV-16, -18, -31, -33, -35, -52, -58). Ubiquitin C (*UBC*, a constitutively expressed endogenous gene) and the bacterial gene, *dapB* were used as positive and negative controls respectively.

Whole tissue sections for selected cases were stained for HR HPV RNA, *UBC* and *dapB* by a fully automated RNAscope assay (RNAscopeVS) using the Ventana Discovery XT slide autostaining system (Ventana Medical Systems Inc, USA).

Application of analysis to full sections of tumour specimens was deemed necessary where there was inter-observer discrepancy in the TMA analysis reporting process.

The *UBC* test was used to assess the presence of hybridisable RNA and was defined as adequate if there was strong staining in the majority of cells in the section. The *dapB* test was used to assess non-specific staining; only those cases that were negative or weakly stained were considered for HPV scoring. A positive HPV test result was defined as punctate staining that co-localised to the cytoplasm and/or nucleus of any of the malignant cells. Where staining was detected in the *dapB* negative control, a positive HPV result was only recorded if staining was at least twice as strong as the *dapB* test. With respect to HR HPV RNA ISH test controls, the UBC test was used to assess the presence of hybridisable RNA and was defined as adequate if there was strong staining in the majority of cells within the core or section.

3.6 HIGH RISK HPV TEST INTERPRETATION TECHNIQUES

The TMA section analysis was conducted for each test and involved assessment of individual cores, with assessor remaining blind to the core sample origin. Scoring was conduced by two consultant pathologists (Dr Max Robinson & Professor Philip Sloan, Newcastle University, UK) using a binary classification (positive vs. negative). Following collation of the independent staining interpretation, discordant scores were re-examined at a meeting between the pathologists to establish a consensus interpretation. In order to quality assure the results, cases that had discordant scores between the pathologists and/or variable scores between cores from the same tumour were additionally subjected to analysis of whole tumour sections. Test analysis for whole sections was identical to that of the TMA sections.

3.7 CELL LINES

All cell lines were grown and maintained in plastic culture dishes in media appropriate to their growth requirements at 37 °C in an atmosphere of 5% CO_2. Cells were split as required (usually 60-80% confluent) by detachment with trypsin (0.25% w/v)/EDTA (5mM) prewarmed to 37 °C. Briefly, the culture medium was removed from the flask and cells were washed with sterile PBS for 30 seconds before discarding from the flask. Trypsin was added and allowed to coat the cells before being incubated for 5 minutes at 37 °C. Detachment of cells was confirmed by tapping the flask (and incubation time increased if detachment not apparent) before neutralizing the effects of trypsin with serum containing media in an equal

volume. The cell population, suspended within media/trypsin was pipetted from the flask and centrifuged at 1200 rpm for 5 minutes. Supernatant was subsequently removed and the cell pellet used in onward processes.

Established Cell Lines

CaSki

The Human Caucasian cervical epidermoid carcinoma derived cell line was procured from ATCC (Middlesex, UK). CaSki cells contain 600-800 integrated copies of HPV16 per cell[185]. CaSki was maintained in RPMI 1640, 2nM Glutamine and 10% Foetal Bovine Serum.

SiHa

The Human Caucasian cervical squamous cell carcinoma derived cell line was procured from ATCC (Middlesex, UK). SiHa cells contain between 1 and 2 integrated copies of HPV16 per cell[188]. Cell growth is maintained in Eagle's Minimum Essential Medium, 2nM Glutamine and 10% Foetal Bovine Serum.

HeLa

The Human Negroid cervical epitheloid carcinoma derived cell line was procured from ATCC (Middlesex, UK). HeLa contains approximately 50 copies of HPV18 per cell. Cell growth was maintained in Eagle's Minimum Essential Medium, 2nM Glutamine and 10% Foetal Bovine Serum.

3.8 NUCLEIC ACID EXTRACTION

Fresh Frozen Tissue Samples

The AllPrep DNA/RNA Mini Kit (Qiagen®, Crawley, UK) was used to purify genomic DNA and total RNA simultaneously from fresh frozen tissue specimens which had been maintained at -80°C since time of surgical resection. Briefly, 2mm^3 portions of tumour tissue were divided from main tissue sample blocks within a class II biological safety cabinet with new sterile disposable consumables to avoid cross-contamination. Tissue samples were placed directly into 350 µl of Buffer RLT Plus (with 1:100 ß-mercaptoethanol) and homogenized within Precellys soft tissue homogenizing tubes containing 1.4mm ceramic (zirconium oxide) beads (Bertin Technologies, Montigny-le-Bretonneaux, France). Homogenization was undertaken in the Precellys® 24 using 2 x 20s cycles at 6500rpm (Bertin Technologies, Montigny-le-Bretonneaux, France). Samples were centrifuged at 10,000g for 3 minutes before transfer to the AllPrep DNA spin column for further centrifuging for 30s at 10,000g. The flow-through product was retained for RNA extraction as below. The samples were washed once with 500µl buffer AW1 and centrifuged at 10,000g before being washed again with 500µl buffer AW2. After centrifugation at 10,000g, 50µl of EB buffer was added to each sample and following incubation at room temperature for 1 min, DNA was recovered by centrifugation at 10,000g for 1 min. A subsequent elution of remaining DNA was made as above for retention of a "B-sample". Samples were subsequently stored at 4°C until required.

RNA extraction was undertaken using the flow-through product from AllPrep DNA spin column (above) according to the manufacturers protocol and including DNase digestion using a Qiagen RNase-Free DNase set (Qiagen, Crawley) to eliminate DNA carryover in the RNA preparations. Briefly, 350µl of 70% ethanol was mixed with the flow-through and transferred to an RNeasy spin column for centrifuging at 10,000g for 15s. The sample was washed with buffer RW1 before on-column DNase digestion with DNase I stock solution in Buffer RDD. The samples were incubated at room temperature (20-30°C) for 30 minutes before a further wash with Buffer RW1. The samples were centrifuged for 15s at 10,000g, washed with Buffer RPE and centrifuged repeatedly until the spin column was dry (2 x 2min at 10,000g). Purified RNA was eluted in 50uL of RNase-free water and stored at -80°C until required.

Cell Line Derived Samples

For the DNA extraction from cell lines the DNeasy kit (Spin column protocol) (Qiagen) was used. A maximum of $5x10^6$ cells were pelleted at 300 x g for 5 minutes and the pellet was re-suspended in 200 µl PBS. 20 µl proteinase K and 4 µl of RNase A (100 mg/ml) (Qiagen) were added, the lysate was then mixed by vortexing and incubated at room temperature for 10 minutes. Subsequently 200 µl Buffer AL were added and the lysate was mixed thoroughly by vortexing and incubated at 56°C for 10 min. 200 µl of ethanol (96–100%) were added to the sample which was mixed thoroughly by vortexing. The mixture was transferred into the DNeasy Mini spin column (which carries a silica based membrane) placed in a 2 ml collection tube,

and was centrifuged at 6000 x g for 1 min. 500 µl of Buffer AW1 were added, and the sample was centrifuged for 1 min at 6000 x g. 500 µl of Buffer AW2 were then added and the sample was centrifuged for 3 min at 20,000 x g to dry the DNeasy membrane. The DNeasy mini spin column was then placed in a 1.5 ml microcentrifuge tube and the DNA was recovered into 200 µl Buffer AE with centrifuging at 6000 x g for 1 min. DNA quality and quantity was assessed by spectrophotometry at 260/280 nm wavelength.

DNA & RNA from the HPV negative human bronchial epithelial cell line, HBEC-3KT, was kindly provided by Dr George Nikolaidis. These samples were utilised as an important HPV negative control and source of DNA and RNA for qPCR assay optimization.

Cell Line Verification Procedure

The identities of all cell lines utilised were confirmed following initial culturing and subsequent preparation of nucleic acids. Using the Cell ID System (Promega, Madison, WI, USA), confirmation of previously determined short tandem repeat (STR) loci profiles for each cell line was made. Briefly 2ng of sample DNA (or positive control K562 DNA) was combined with Cell ID Enzyme Mix (x5) and Primer Pair Mix (x10). Samples were transferred for thermal cycling as detailed in Table 4.

Temperature (°C)	Time	No of cycles
96	2 min	
94	30 sec	
60	30 sec	10
70	45 sec	
90	30 sec	
60	30 sec	22
70	45 sec	
60	30 min	

Table 4: Thermal Cycling Conditions for Cell ID System

Detection of amplified fragments in samples was conducted on an Applied Biosystems 3130 Genetic Analyser (Applied Biosystems) using GeneMapper 4.0 software. Briefly 1µL of PCR product was combined with 10µL of highly deionized formamide (HDF) and 1µL of the Internal Lane Standard (ILS 600) and denatured at 95°C for 3 minutes before returning to ice then immediately loading for capillary electrophoresis as detailed by the manufacturer. Comparison of STR loci results was made with those published by the cell line procurement source (ATCC, Middlesex, UK) and are detailed below in Table 5.

STR Locus	CaSki		SiHa		HeLa	
	Database	Actual	Database	Actual	Database	Actual
Amelogenin	X	X	X	X	X	X
CSF1PO	10	10	12	12	9,10	9,10
D13S317	8,12	8,12	11	11	12,13.3	12,13.3
D16S539	11,12	11,12	12	12	9,10	9,10
D5S818	13	13	9	9	11,12	11,12
D7S820	8,11	8,11	10	10	8,12	8,12
THO1	7	7	6,9	6,9	7	7
TPOX	8	8	8	8	8,12	8,12
vWA	17	17	14,17	14,17	16,18	16,18

Table 5: Reported and Actual STR Profiles for Cell Lines Utilised

3.9 NUCLEIC ACID QUANTIFICATION

The purified DNA & RNA samples were quantified by spectrophotometry using a NanoDrop 1000™ (NanoDrop Technologies, Thermo Fischer Scientific, Wilmington, DE, USA). Nucleic acid extraction was repeated where sample purity fell outside expected normal range, as determined by the ratio of absorption at 260 and 280nm ($A_{260:280}$). For DNA samples repeat extraction was conducted for $A_{260:280}$ less than 1.75 and for RNA samples when $A_{260:280}$ fell below 1.90.

3.10 COMPLIMENTARY DNA SYNTHESIS

Complimentary DNA synthesis (cDNA) was undertaken using the QuantiTect Reverse Transcription Kit (Qiagen®, Crawley, UK) according to the manufacturers protocol.

Briefly, following thawing on ice, 500-600ng of Total RNA was combined with 2 µL of gDNA wipe-out buffer and a variable volume of RNase free water to make a final volume of 14 µL. Samples were incubated at 42°C for 2 min then placed immediately on ice.

A reverse-transcription master mix of Quantiscript reverse transcriptase (1 µL per sample), Quantiscript RT buffer (4 µL per sample) and RT primer mix (1 µL per sample) was made on ice. 6 µL of the prepared master mix was combined with each sample, mixed by pipetting and incubated at 42 °C for 15 min with a final incubation at 95°C for 3 min to inactivate Quantiscript reverse transcriptase.

Each finished reverse-transcription reaction sample was diluted 1:5 with RNAse free water before storing at -20°C until required.

3.11 QUANTITATIVE REAL-TIME PCR ANALYSIS

For the detection of viral and host gene DNA presence and gene expression, quantitative real-time PCR assays were used. These constituted either custom designed assays (HPV16 E2, E6 & E7 and HPV 18 E6) or commercially available assays (HPV 33 E6) as detailed below. Analysis was undertaken for a cohort of 96 OPSCC and where available adjacent matched normal pairs (n=53).
Additionally, proprietary gene expression assays were utilised to determine differential gene expression levels based on HPV status for the key determinants of DNA methylation, DNA methyltransferases (DNMT-1, -3a & -3b) and UHRF1.

HR HPV qPCR Assays Design and Optimisation

The defined gold standard test for this research was HR HPV RNA qPCR which included assays for HPV16 E6 and E7 genes, HPV18 E6 gene and HPV33 E6 gene. The aggregation of results from these four assays was deemed appropriate to correctly classify HPV status in over 99% of HPV positive OPSCC in accordance with the published results of a systematic review of the role of HR HPV in 969 cases OPSCC[76].

HPV16 E2, E6 & E7, HPV18 E6 Assays

Using the Primer Express v2.0 Software (Applied Biosystems), primers and probes were designed for the HPV E2, E6, E7, and the HPV18 E6. Reference sequences utilised were as follows; Genebank NCBI Reference Sequence NC_001526.2 & AY262282.1 respectively. All primer sequences passed Basic Local Alignment Seach Tool (BLAST) analysis to ensure an absence of human genome target sequence homology.

Probes were FAM-labeled MGB Taqman probes synthesized by Applied Biosystems, whilst primers were synthesized by MWG (Ebersberg, Germany). Details of the primer and probe sequences are contained in Table 6.

Commercially available primers and a VIC-TAMRA labeled probe for the single-copy gene RNase P (Taqman RNase P Control Reagents, Applied Biosystems) were used as an endogenous reference in each multiplex reaction. This served to demonstrate availability and quantitative adequacy of detectable human genomic DNA within each reaction.

Target	Forward Sequence	Reverse Sequence	Probe Sequence
HPV16 E2	GATGGAGACATATGCAATACAATGC	CACAGTTACTGATGCTTCTTCACAAA	TACAAACTGGACACATATAT
HPV16 E6	CTGCGACGTGAGGTATATGACTTT	ACATACAGCATATGGATTCCCATCT	CTTTTCGGGATTTATGC
HPV16 E7	TTCGGTTGTGCGTACAAGC	AGTGTGCCCATTAACAGGTCTTC	CACGTAGACATTCGTACTT
HPV18 E6	AAACCGTTGAATCCAGCAGAA	GTCGTTCCTGTCGTGCTCG	TTGCAGCACGAATGG

Table 6: Primer and probe sequences for HPV-16 & -18 gene specific assays

Optimisation of HPV16 qPCR assays was conducted with CaSki and SiHa cell line derived DNA as positive controls and, in the case of CaSki, as an important threshold determiner. The positive control for HPV18 E6 qPCR assay optimization was HeLa derived DNA.

In addition to non-template controls, Hela was included as a negative control when CaSki was the target positive control and vice versa.

As Figure 7 demonstrates, the capacity to detect HPV viral sequence was preserved in concentrations as low as 1:10,000 of CaSki, whilst remaining reliably quantifiable to a level of 1:1,000 (Figure 8).

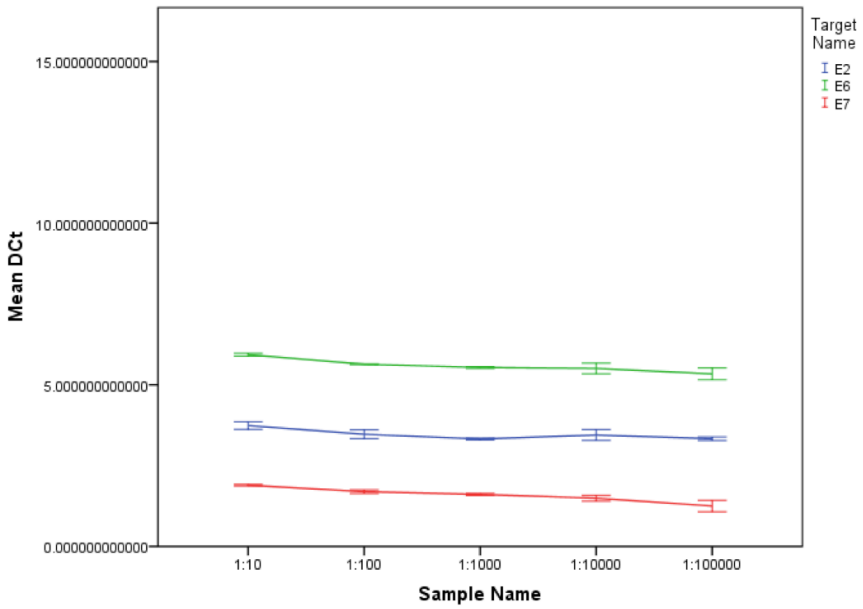


Figure 7: HPV16 Assay Detection Sensitivity

Serial dilution experiment of CaSki cDNA in ddH_2O to assess the reliability of the designed qPCR assay. The mean ΔCt for relevant HPV16 target (E2, E6 & E7) is represented on the Y axis while X-axis shows the dilution factor. The almost straight lines across dilutions demonstrate reliability over 5 logs, as required for valid qPCR assays. Error Bars: ± 1 S.E.

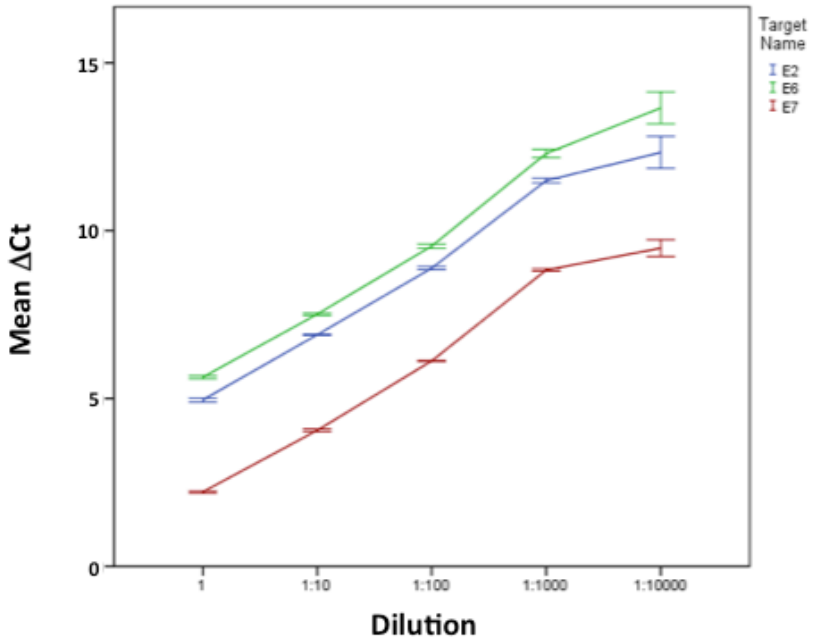


Figure 8: Establishment of the Linear Dynamic Range of the HPV16 Assay

Establishement of the Linear Dynamic Range of the HPV16 Assay. Mean ΔCt for relevant HPV16 target (E2, E6 & E7) of the HPV positive CaSki RNA in serial dilution with the HPV negative HBEC-3KT RNA with subsequent reverse transcription prior to qPCR assay. The results demonstrate a linear relationship of detection down to a dilution of 10^{-3} with HPV negative RNA.

Optimisation of each assay (HPV-16 and -18) lead to defined qPCR thermal conditions, as detailed in Table 7 and Table 8.

Each sample was run in duplicate with an ultimate reaction volume of 25µL consisting of 1x Taqman Gene Expression Master Mix (Applied Biosystems), 500nmol/l of relevant primer and 250nmol/L of appropriate probe, 1x endogenous reference primer/probe mix (VIC-TAMRA-labeled probe for single copy gene RNaseP for DNA qPCR (Taqman RNase P Control Reagents, Applied Biosystems) or Human VIC-MGB-labeled Actin ß (ACTB) primers and probe (Applied Biosystems, Carlsbad, CA; assay ID: 4352935E) for expression analysis) and 100ng of genomic DNA or cDNA respectively.

Step	Temperature (°C)	Time	No of cycles
UDG Incubation	50	2 min	
Activation	95	10 min	
Denaturation	95	15 sec	
Annealing/Extension	61	60 sec	45

Table 7: HPV16 E2, E6 & E7 qPCR Thermal Conditions

Step	Temperature (°C)	Time	No of cycles
UDG Incubation	50	2 min	
Activation	95	10 min	
Denaturation	95	15 sec	
Annealing/Extension	60	60 sec	45

Table 8: HPV18 E6 qPCR Thermal Conditions

Assays were run on an Applied Biosystems 7500 real time PCR machine. Reactions were set up in a duplicate volume and reactions split to final 25 µl volume duplicates. Reaction duplicates were run on the same PCR cycling machine in immediate time sequence.

HPV33 E6 Assay

Detection of HPV33 E6 expression was undertaken using a proprietary assay (Human papillomavirus 33 E6 gene, Genesig, Southampton, UK). qPCR Assay preparation and amplification was conducted in accordance with the manufacturers protocol. Briefly, in a reaction volume of 20µL, template cDNA, genesig 2x qPCR MasterMix and HPV33 Primer/Probe mix was amplified under the conditions detailed in Table 9.

A serial dilution of HPV33 positive control was prepared and run simultaneously for relative quantification of detected target.

Step	Temperature (°C)	Time	No of cycles
UDG Incubation	37	15 min	
Activation	95	15 min	
Denaturation	95	10 sec	
Annealing/Extension	60	60 sec	50

Table 9: Genesig HPV33 qPCR Thermal Conditions

DNA qPCR Assays

DNA qPCR Detection Threshold

The detection threshold for HPV positive status was set in accordance with the previously reported frequency of E6 gene copies per diploid genome for CaSki (869 copies)[185]. Assuming an HPV16 driven tumour is composed of a dominant clonal population of cells, we scored as positive those samples with ≥1 E6 gene copy/diploid genome. A sample was only deemed positive if the threshold was met in both of the duplicate runs.

Expression (mRNA qPCR) Assays

HPV Expression Assays

The HPV genome is intronless and, as such, primers and probes designed for amplification of DNA sequences can be utilised for expression analysis. Removal of

DNA, however, is therefore an absolute requirement to ensure detected amplification is from mRNA (cDNA) origin rather than from either residual DNA or contamination with DNA. To achieve this, DNase treatment occurred both in the RNA preparation from tissue and as an essential component of reverse transcriptase generation of cDNA.

Subsequent to the DNase treatment and reverse transcription, eight cDNA samples (selected at random) were subjected to microsatellite marker analysis of non-exonic areas of two human genes that would therefore be present in gDNA yet absent in RNA (or cDNA). Primers were selected from the LMS High Density Panel 5 set (Applied Biosystems); D9S161 (9p21.2) and D17S938 (17p13.2), and synthesized with fluorescent-labelled reverse primers (FAM). Microsatellite marker reactions were carried out in a multiplex manner using the Qiagen Multiplex Master Mix (Qiagen, UK), 50ng of cDNA and ddH20 to a volume of 20µL. The thermal conditions for the subsequent PCR are detailed below (Table 10).

Step	Temperature (^{o}C)	Time	No of cycles
Taq Activation	95	15 min	
Denaturation	94	30 sec	
Annealing	56	90 sec	30
Extension	72	60 sec	
Final extension	72	30 min	

Table 10: Microsatellite Marker Analysis PCR Thermal conditions

2µL of the resultant PCR products were dissolved in 10µL of Highly Deionized Formamide (HDF) with 0.5µL of the proprietary Genescan 600 LIZ Size Standard

(Applied Biosystems). For maximal sensitivity of detection of gDNA carrover, analysis was undertaken on a 3130 sequencer (Applied Biosystems).

DNMT1, 3A, 3B & UHRF1 Expression Assays

Determination of expression levels of DNA Methyltransferases 1, 3A and 3B (DNMT1, -3A and -3B) and the UHRF1 gene in tumour tissue samples, associated matched normal pairs and HPV positive cell lines, was undertaken using commercially available assays (Applied Biosystems). The relevant assays are detailed in Table 11.

GENE	Assay Id	DYE	Unigene Id	Amplicon length (bp)
DNMT1	Hs00154749_m1	FAM	Hs.202672	77
DNMT3A	Hs01027166_m1	FAM	Hs.515840	79
DNMT3B	Hs00171876_m1	FAM	Hs.713611	55
UHRF1	Hs00273589_m1	FAM	Hs.108106	105
ACTB	4326315E	VIC	Hs.520640	171

Table 11 Identification and additional information for proprietary gene expression assays DNMT1, -3A, -3B & UHRF1

The proprietary endogenous control, Human VIC-MGB-labeled Actin ß (ACTB) primers and probe (Applied Biosystems, Carlsbad, CA; assay ID: 4352935E), was utilised in each multiplex reaction. By comparison to target genes, ACTB had a greater amplicon length and as such its inclusion enabled use as an internal control for cDNA integrity.

Each final reaction volume of 20µl contained 10 µl of 2x TaqMan® Gene Expression Master Mix (Applied Biosystems), 900 nmol/L of each primer and 250 nmol/L probe,

1µL of endogenous control ACTB-VIC (Applied Biosystems) and approximately 100 ng of cDNA following the thermal cycling conditions detailed below in Table 12.

Step	Temperature (^{o}C)	Time	No of cycles
UDG Incubation	50	2 min	
Activation	95	10 min	
Denaturation	93	30 sec	
Annealing/Extension	60	45 sec	45

Table 12: DNMT & UHRF1 qPCR Thermal Conditions

Reactions were set up with duplicate volumes and subsequently split to reach final 20 µl reactions. Using an Applied Biosystems 7500 real time PCR machine reactions were run in immediate time sequence. Using the 7500 Software v2.0.1 (Applied Biosystems), qPCR data was analysed with post-assay expression levels expressed as relative quantification values (RQ, where RQ=$2^{(-\Delta\Delta Ct)}$)[189]. Average RQ values were determined from the duplicated runs for subsequent analysis.

3.12 BISULPHITE TREATMENT OF DNA

1 µg DNA was bisulphite treated using the EZ-96 DNA Methylation-Gold™ Kit, (Shallow-Well Format) (Zymo Research, Irvine, CA, USA) following the manufacturer's protocol.

Preparation of the CT Conversion Reagent was made by the addition of 9 ml of water, 500 µl of M-Dissolving Buffer, and 3 ml of M-Dilution buffer. It was then mixed at room temperature by constant shaking for 15 minutes. The M-Wash

Buffer was prepared by adding 144 ml of 100% ethanol to the 36 ml M-Wash Buffer concentrate.

Each sample containing 1 µg of DNA was made up to 20 µL with water before combining with 130 µl of the CT Conversion Reagent in a Conversion Plate. All samples were mixed by pipetting action before sealing the plate prior to thermal cycling as follows; 98°C for 10 minutes, 64°C for 2.5 hours. The samples were subsequently transferred to individual wells of the Silicon-A™ Binding Plate mounted on a Collection Plate together with 400 µl of M-Binding Buffer. The plate was centrifuged at 3,000 x *g* for 5 minutes and the flow-through discarded. 400 µl of M-Wash Buffer was added to each well and the plate re-centrifuged at 3,000 x *g* for 5 minutes. 200 µl of M-Desulphonation Buffer was then added to each well and incubated at room temperature (20 °C – 30 °C) for 20 minutes. Following incubation, the plate was centrifuged at 3,000 x *g* for 5 minutes and the pursuant flow-through discarded. 400 µl of M-Wash Buffer was added to each well of the plate and the plate centrifuged at 3,000 x *g* for 5 minutes with the flow-through discarded once more. A final 400 µl of M-Wash Buffer was added and the plate and centrifuged at 3,000 x *g* for 10 minutes.

The Silicon-A™ Binding Plate was then placed onto an Elution Plate and 50µl of M-Elution Buffer were added directly to each well. After 5 minutes incubation at room temperature (20 °C – 30 °C), it was centrifuged at 3,000 x *g* for 3 minutes to elute the DNA. Storage of samples following bisulphite treatment was at -20°C until use but not beyond one month. 2.5 µl of bisulphite treated DNA was used for each PCR reaction.

3.13 PYROSEQUENCING METHYLATION ANALYSIS

Pyrosequencing (PSQ) was utilised to detect and quantify variation in methylation levels within target sequences from clinical samples.

LINE-1 Assay

LINE-1 methylation status determination was undertaken using the previously described Pyrosequencing Methylation Assay (PMA)[190]. Briefly, this assay detects methylation of 6 CpGs contained within the CpG island of the 5' internal promoter. All primers were synthesized by MWG (Germany) in accordance with the previously published sequences[190] (Table 13).

Promoter	Forward primer (5'→3')	Reverse primer (5'→3')	Sequencing primer (5'→3')
LINE-1	BIO-TAGGGAGTGTTAGATAGTGG	AACTCCCTAACCCCTTAC	CAAATAAAA CAATACCTC

Table 13: LINE-1 Primer Sequences

(BIO: biotinylated primer)

PCR reaction set up was undertaken with the following components; Qiagen HotStarTaq Plus Master Mix Kit, 5µM biotinylated primer, 10 µM non-biotinylated primer and 2.5 µl (approximately 50ng) of bisulphite treated DNA. All reactions were setup in duplicate with sequencing runs conducted consecutively. An excess of non-biotinylated primer was included to reduce residual biotinylated primer that may otherwise impact on sequencing reaction fidality. The PCR thermal profile for LINE-1 is detailed in Table 14.

Step	Temperature (°C)	Time	No of cycles
Taq Activation	95	5 min	
Denaturation	94	30 sec	
Annealing	58	45 sec	40
Extension	72	45 sec	
Final extension	72	10 min	

Table 14: PCR conditions for LINE-1 Assay

The quality and quantity of resultant PCR product was confirmed by agarose gel (2%) electrophoresis and UV visualisation on a UVP VisionWorks LS instrument prior to progression to Pyrosequencing analysis.

For the pyrosequencing workflow, proprietary PyroGold reagents were used following the suppliers' protocol (Qiagen), with sequencing performed in the PSQ96MA pyrosequencer (Qiagen, Crawley, UK) instrument.

As above, duplicate PCR reactions for each sample were sequenced and the mean methylation levels recorded from the average of both sample runs (where positive results were apparent in both instances).

Briefly, PCR products were combined with binding buffer that contained both buffer and sepharose streptavidin beads. Once transferred to a 96 well plate the samples were agitated using a vortex plate at 350rpm for 10 minutes to allow for template binding to streptavidin beads. Bound template was then transferred, using a vacuum tool, to successive washes of 70% Ethanol for 10 seconds, 0.2M NaOH for 20 seconds and then proprietary wash buffer for a further 10 seconds. Templates, whilst still retained on the vacuum tool were transferred to annealing buffer that

contains sequencing primer. The samples were heated to 80°C for 2 minutes before being returned to room temperature for 2 minutes. Samples then entered the PSQ 96MA for sequencing.

HPV16 Gene Methylation Assays

To explore the effect of viral genome methylation in OPSCC, pyrosequencing methylation assays were designed for the most CpG rich regions of the E2 gene and two regions within the HPV16 long control region (LCR).

The reference HPV16 genome sequence utilised was Genebank NC_001526.2. All PMA primers were designed using the Assay Design Software (Qiagen) and synthesized by WMG (Germany). Primer sequences are listed below (Table 15).

Target	Forward primer (5'→3')	Reverse primer (5'→3')	Sequencing primer (5'→3')
HPV E2	GTGAAATTATTAGGTAGTATTTGG	BIO-CAACAACTTAATAATATAACAAAAA	GTGAAATTATTAGGTAGTA
LCR 1	BIO-ATTGTATTATGTGTAATTATTGAA	CCAAAAATATATACCTAACAAC	CCAAAAATATATACCTAACAAC
LCR 2	GTAAAATTGTATATGGGTGT	BIO- TAAAATATCTACTTTTATACTAACC	TAATTTATGTATAAAATTAAGG

Table 15: HPV16 Pyrosequencing Methylation Assay Primer Sequences

(BIO: biotinylated primer)

The LCR 2 PMA was designed to include five CpGs, of which four fell within the E2 protein binding sites 3 and 4 (E2BS3 and E2BS4). Within each binding site there were two corresponding CpGs. The relevant components of the LCR and PMA primer locations are graphically represented on Figure 9 (below).

PCR amplifications were performed using Qiagen Hotstar Plus Master Mix Kit, 5uM biotinylated primer, 10uM non-biotinylated primer and 2.5µL of bisulphite treated DNA. As before, excess non-biotinylated primer was included and PCR cycling was increased to 40 cycles to reduce carryover of unused biotinylated primer into the sequencing reaction.

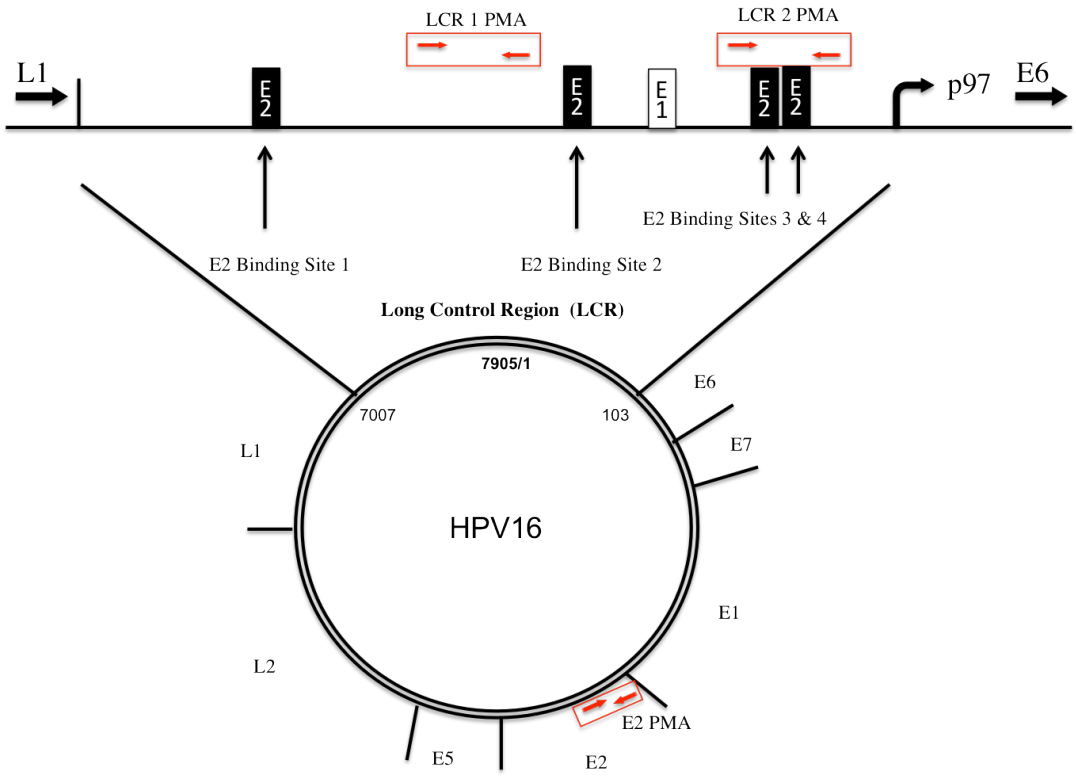


Figure 9: Schematic representation of HPV genome with detailed LCR and Pyrosequencing Methylation Assay sites.

Specific genes and LCR sequence start points are highlighted. The single E1 binding site and four E2 binding sites are highlighted. Also depicted are the PMA target locations. LCR 2 PMA analyses five CpG including four that fall within E2BS3 (nt 37 & 43) and E2BS4 (nts 52 & 58).
(Modified from Snellenberg et al[191])

Thermal profiles for E2, LCR 1 and LCR 2 reactions are detailed below in Table 16, Table 17 and Table 18 respectively.

Step	Temperature (°C)	Time	No of cycles
Taq Activation	95	5 min	
Denaturation	94	30 sec	
Annealing	48	40 sec	40
Extension	72	30 sec	
Final extension	72	10 min	

Table 16: HPV16 E2 PMA Thermal Conditions

Step	Temperature (°C)	Time	No of cycles
Taq Activation	95	5 min	
Denaturation	94	30 sec	
Annealing	46	60 sec	40
Extension	72	20 sec	
Final extension	72	10 min	

Table 17: HPV16 Long Control Region 1 PMA Conditions

Step	Temperature (°C)	Time	No of cycles
Taq Activation	95	5 min	
Denaturation	94	30 sec	
Annealing	46	35 sec	40
Extension	72	15 sec	
Final extension	72	10 min	

Table 18: HPV16 Long Control Region 2 PMA Conditions

Following thermal cycling, PCR products were run on a 2% agarose gel and visualised using UV visualisation on a UVP VisionWorks LS instrument to ensure sufficient product quality and quantity. As for LINE-1 PMA sequencing, products were sequenced in a 96MA Pyrosequencer (Qiagen) (as above, LINE-1 Assay).

3.14 METHYLATION MICROARRAY ANALYSIS

To analyse differences in gene promoter methylation between HPV positive and HPV negative OPSCC, the recently validated[192] Infinium HumanMethylation450 BeadChip microarray analysis (Illumina, San Diego, CA, USA) was utilised.

A separate cohort of 24 OPSCC was submitted for analysis of differential promoter sequence methylation status on the basis of HPV status. The Infinium platform is designed to provide analytical coverage of 96% of CpG islands and associated flanking island scores. Of the approximate 450,000 CpG sites included in the assay, just over 150,000 (30.9%) are deemed to be from within CpG islands. When considered in a functional setting, 41% (200,339) CpGs are in proximal gene promoter regions however at varying distances upstream of the relevant transcriptional start site.

Sample preparation and processing on the Infinium Methylation Assay was conducted by a third party technical provider (Gen-Probe, Manchester, UK). Briefly 500ng of gDNA was outsourced to the third party for bisulphite treatment, with subsequent utilisation of 4µL of the resultant bisulphite treated DNA according to the Infinium HD Methylation Assay protocol.

HPV16 status determination was undertaken for all samples using the HPV16 E6 mRNA qPCR assay as detailed above (2.5).

Bioinfomatic Analysis & Validation Pyrosequencing Methylation Assay Design

Raw data delivered from the third party were initially assessed for quality assurance. No outliers were detected. All data originating from the X chromosome were excluded from further analysis to exclude potential bias on the basis of gender.

With knowledge of the previously determined HPV status of the 24 samples (HPV positive n=6, HPV negative n=18), R statistical package, version 2.14.1 (Lucent Technologies, 2012), was utilised to perform Wilcoxon signed-rank and Wilcoxon rank sum tests. Where the average Beta score for any individual gene was greater than 0.2 or less than -0.2 and where there were 4 or more probes included on the array that fell within the promoter region for that specific gene, it was shortlisted for technical validation.

From the top ranked methylation variable positions, 14 genes met the above criteria (13 differentially hypermethylated in HPV positive tumours and 1 gene hypomethylated). Following mapping of probes to the relevant gene sequence, PMA primer design was undertaken using PyroMark Assay Design Software 2.0 (Qiagen, Crawley, UK). PMA assay design was unlikely to provide satisfactory PCR products capable of meeting the stringent requirements for pyrosequencing in the case of 6 genes. In the case of the 8 genes for which PMA design was successful, it was ensured that coverage of sequence containing a minimum of three CpGs corresponding to the probes was adhered to. Primer sequences (Table 19) and PCR conditions are listed below for each of the genes, along with details of the number

of CpGs covered by both the Infinium array probes and the PMA (Table 20 and Table 21).

As above, all target genes for validation had originally been demonstrated to be differentially hypermethylated in HPV positive OPSCC using the Infinium analysis.

Target Gene	Forward primer (5'→3')	Reverse primer (5'→3')	Sequencing primer (5'→3')
GalR1	GGGGTGAGGGTGGGATTA	BIO- CTCCTCCCCAAAATAACTATCC	GGTGAGGGTGGGATT
C12 orf42	BIO- AGTATTTTGTTGGGTTTTGG	CACAAAACAACCCCATTATA	AAACAACCCCATTATAATTA
HOXA7	BIO- ATTTTAGTAGTTTTTATAGGTGGT	AAACCTCTTACCCTTCCAT	CTTACCCTTCCATTCTAA
FLJ26850	GGTGGTTATTAGAGAATTGAAT	BIO- ACTCAATATAAAAATTCTCAAAAC	AGAGAATTGAATTTAGGAGG
SYN2	BIO- GGAAGGATAAGAGGTGTTAG	TTCCTCCTCACTACAAAATAT	CCTCACTACAAAATATTC
KCNA1	BIO- GGATTTGATTATTTTTAATGTG	AACTCTACTTCCCCTATAACC	ACTCTACTTCCCCTATA
SLCo4C1	AGTGTTTGGGTTTAAGGG	BIO- AAAATTCTCACCCCACAA	TGTTTGGGTTTAAGGGAG
CCNA1	GATAGAGTTGGGGTTTGGG	BIO- CAAAAACTCCTCTCCCCAC	AGAGTTGGGGTTTGGGA

Table 19: Infinium Technical Validation PMA Primer Sequences

(BIO: biotinylated primer)

Step	Temperature (^{o}C)	Time	No of cycles
Taq Activation	95	5 min	
Denaturation	94	30 sec	
Annealing	*	30 sec	40
Extension	72	30 sec	
Final extension	72	10 min	

Table 20: Infinium Technical Validation PMA Conditions

Target Gene	Annealing Temperature (°C)	CpG Inclusion	Number of Probes Covered
GalR1	60	5	4
C12 orf42	56	7	4
HOXA7	52	4	3
FLJ26850	52	4	3
SYN2	54	4	3
KCNA1	50	9	3
SLCo4C1	51	7	5
CCNA1	60	8	5

Table 21: Infinium PMA Gene Specific Annealing Temperatures & CpG Inclusions

For each gene, the number of CpGs is listed and the corresponding number of CpGs correlating to the Infinium array probes.

PMA PCR amplifications were performed using Qiagen Hotstar Plus Master Mix Kit, 5uM biotinylated primer, 10uM non-biotinylated primer and 2.5µL of bisulphite treated DNA. As before, excess non-biotinylated primer was included and PCR cycling was increased to 40 cycles to reduce carryover of unused biotinylated primer into the sequencing reaction.

Following thermal cycling, PCR products were run on a 2% agarose gel and visualised using UV visualisation on a UVP VisionWorks LS instrument to ensure sufficient product quality and quantity. Once more, as for LINE-1 PMA sequencing,

products were sequenced in a 96MA Pyrosequencer (Qiagen) (as above, LINE-1 Assay).

3.15 DNA ARTIFICIALLY METHYLATED CONTROLS

Leukocytes from a healthy individual previously shown not to harbor HPV16 DNA, were methylated in vitro with excess SssI methyltransferase (New England Biolabs) to generate completely methylated DNA, as per the manufacturers protocol. Briefly, 160 µM S-adenosylmethionine (SAM, 2µL) was combined with 10x NEBuffer 2 (2µL), 1µg of lymphocyte gDNA and SssI methylase (CpG Methyltransferase, 1µL) before mixing by pipetting.

The reaction volume was incubated for one hour at 37ºC before stopping the reaction by heating to 65ºC for 20 minutes.

Subsequently, a serial dilution of the artificially methylated DNA was made for the purposes of PMA calibration/reference methylation analysis.

Samples corresponding to unmethylated 5%, 10%, 20%, 40%, 80% and 100% methylated lymphocyte DNA were bisulphite treated as detailed in section 3.12.

3.16 HPV16 INTEGRATION ANALYSIS

Analysis of a cohort of HPV positive OPSCC (n=43) and control cell lines (CaSki & SiHa) was undertaken to determine the presence or otherwise of viral integration into the host genome. This analysis was undertaken using direct PCR based analysis of the E2 gene integrity.

Additionally, a pilot series of OPSCC sourced from the above cohort (n=9) and the control cell lines (CaSki & SiHa) were further interrogated using a recently described technique coupled with massively parallel sequencing.

E2 Gene Integrity Analysis

To determine the integrity of the HPV16 E2 gene, the previously modified and optimised approach described by Collins et al was employed[182]. The technique utilised sets of overlapping sequence-specific primers for the E2 gene (Figure 10). Determination of integration state relies on the assumption that integration occurs exclusively in the E2 gene and that failure of amplification of a component (or components) of the E2 gene implies integration. Conversely, detection of all components of the E2 gene by PCR amplification reflects a presumed episomal viral state.

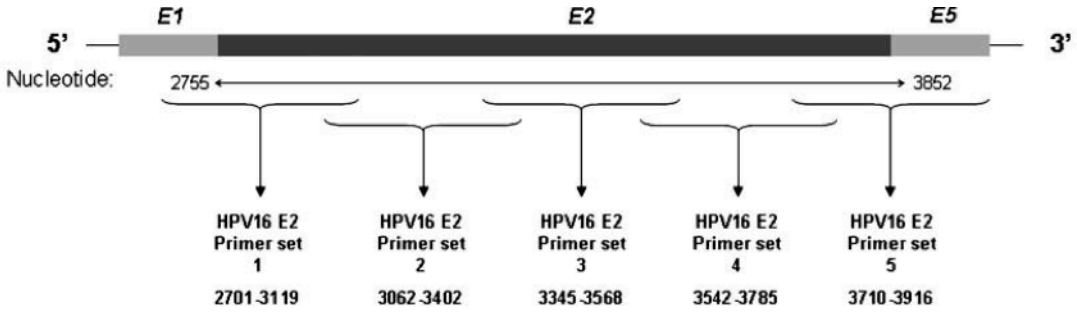


Figure 10: Schematic Diagram of HPV16 E2 Integrity Overlapping Primer Analysis

Location of primer sets detailed with respect to the E2 gene. Nucleotide numbers are according to the whole HPV16 genome. (Modified from Collins et al[182])

Briefly, 60ng of DNA samples from each case was amplified using Hotstart Mastermix with 0.4umol/L of the appropriate primer set. Thermal cycling conditions are detailed in Table 22 and Table 23, the only alteration from the conditions described by Collins et al. is a reduction in number of cycles from 60 to 40.

Step	Temperature (°C)	Time	No of cycles
Taq Activation	95	5 min	
Denaturation	95	30 sec	
Annealing	57	60 sec	40
Extension	72	120 sec	
Final extension	72	10 min	

Table 22: Thermal Cycling Conditions for HPV16 E2 Whole Gene

Step	Temperature (°C)	Time	No of cycles
Taq Activation	95	5 min	
Denaturation	95	30 sec	
Annealing	54	60 sec	40
Extension	72	120 sec	
Final extension	72	10 min	

Table 23: Thermal Cycling Conditions for HPV16 E2 Component Parts (P1 – P5)

Following thermal cycling, PCR products were run on a 2% agarose gel and visualised using UV visualisation on a UVP VisionWorks LS instrument to demonstrate product presence (or absence). Controls included DNA samples from CaSki and SiHa cell lines, which had previously been demonstrated to contain complete head-to-tail complete viral gene integrants[193] (hence positive control for all primer pairs) and a solitary integrant with loss of the E2 gene respectively[194] (integration positive control with expectation of primer set 2 amplicon failure). The negative controls were DNA derived from the known HPV16 negative cell line HBEC-3KT and DNA from the HPV negative OPSCC (sample No.11).

Next Generation Sequencing (NGS) Analysis

Prior to commencement of the project, options for both target sequence acquisition and sequencing were subject to collaborative discussion with the third party organization chosen to undertake sample preparation and sequencing; Centre for Genomics Research, University of Liverpool, Liverpool, UK.
Due to the previous success of Depledge et al[195], target capture and library preparation was undertaken using the previously validated SureSelectXT Target Enrichment System for extraction of the sequences of interest and generation of an Illumina Paired-End Sequencing Library (Agilent, Santa Clara, CA, USA)(Figure 11). Once more, selection of the platform, best suited to specifics of the project, was made in response to guidance provided by the third party collaborator and in keeping with project goals. Paired-end sequencing of all target sequences was completed using the HiSeq 2000 (Illumina, San Diego, USA)[195].

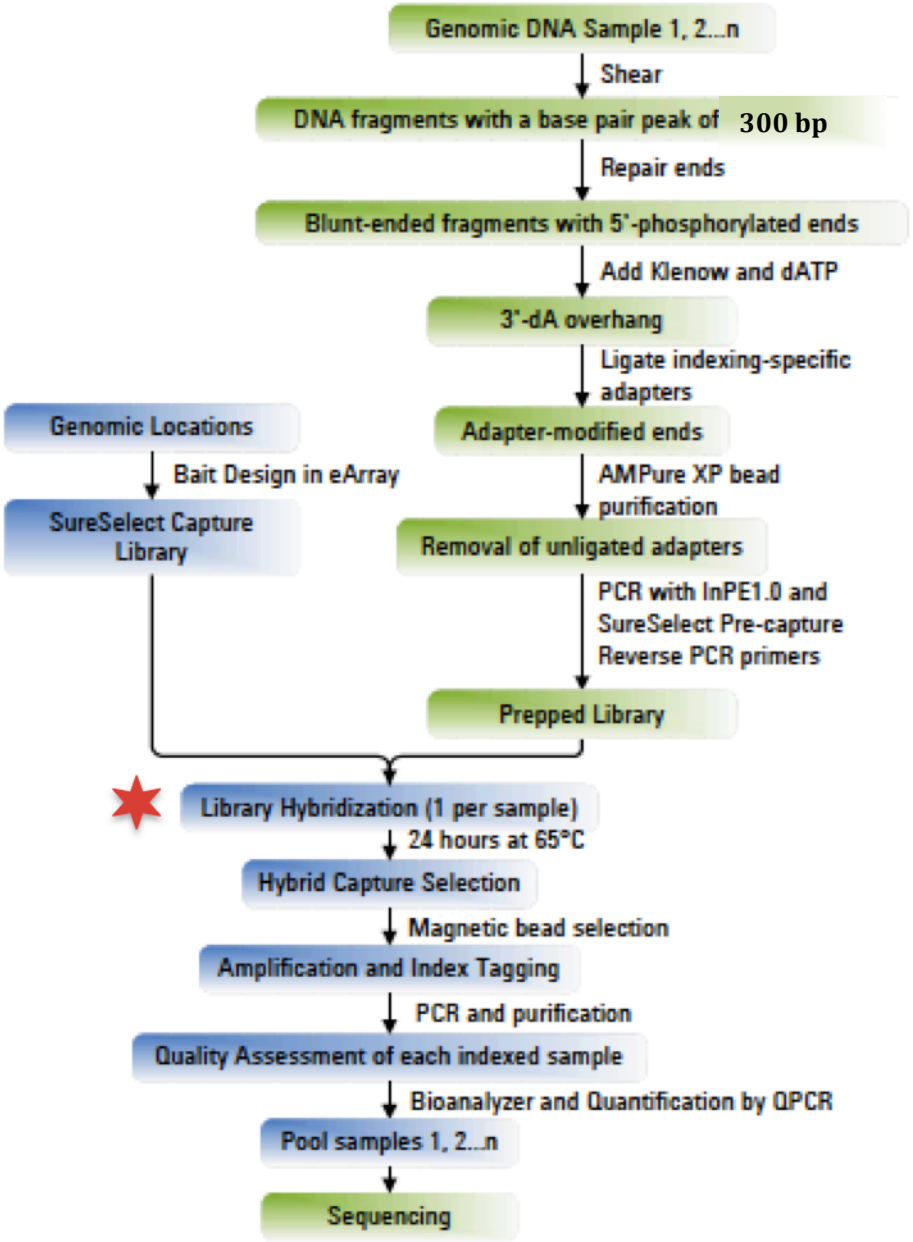


Figure 11: Overall sequencing sample preparation workflow

(Modified from Agilent SureSelect XT Protocol) * indicates correlation with hybridisation workflow (Figure 12).

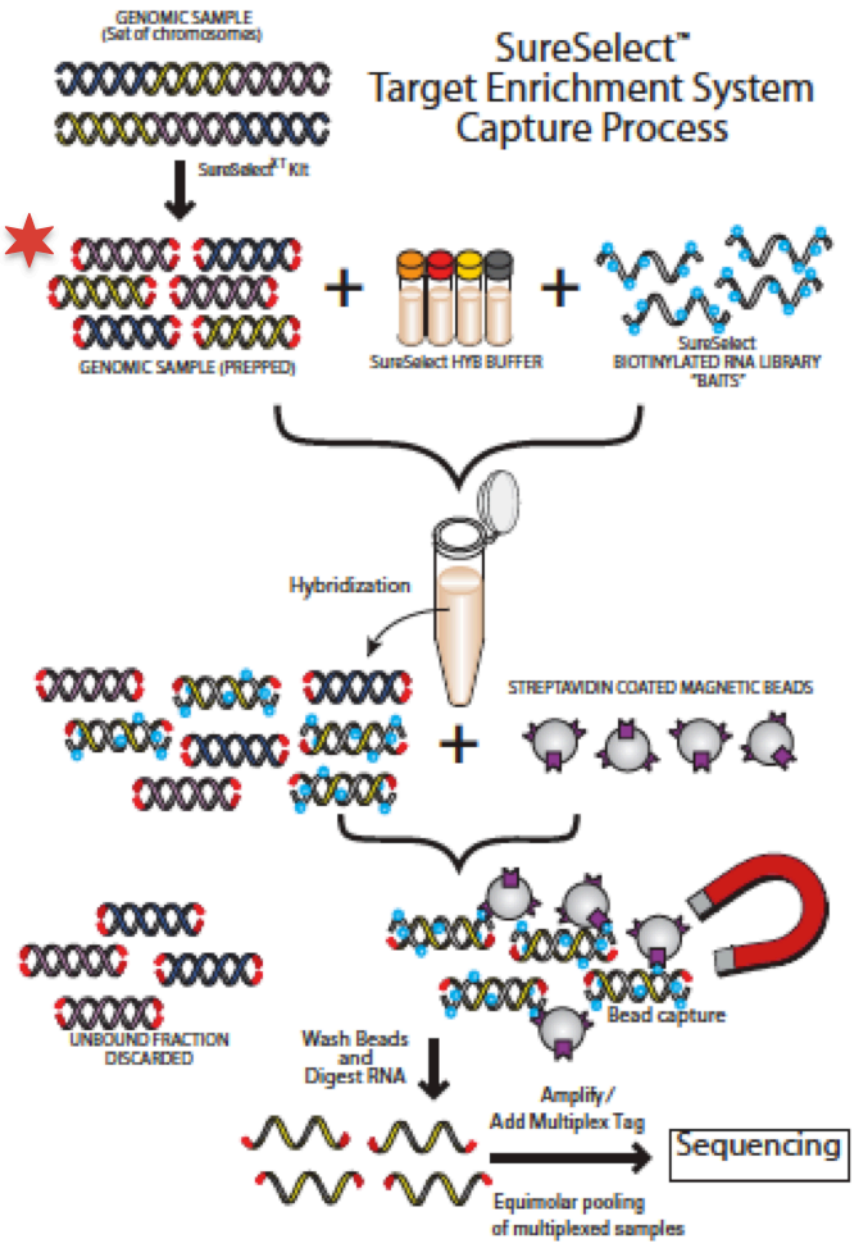


Figure 12: Sample Hybridisation Schematic

(Modified from Agilent SureSelect XT Protocol) * indicates input point of prepared and purified sample libraries.

Selection of cases for analysis was undertaken, ensuring adequate available DNA (3µg) and sample quality as detected by Nanodrop analysis ($A_{260}/_{280}$ and $A_{260}/_{230}$ rations ensured to be ≥1.8 and 1.9, respectively).

The sample preparation, hybridisation and sequencing were outsourced to a third party organisation; Centre for Genomics Research, University of Liverpool, Liverpool, UK. The workflow for sample preparation is graphically represented in Figure 11 and the simplified graphical representation of the target sequence hybridisation, portrayed in Figure 12.

Briefly, the protocol entailed shearing of 3ug of gDNA for each of 9 HPV16 positive OPSCC samples and 2 HPV16 positive cell lines (CaSki and SiHa) using the Covaris 300 programme to a target size of 300bp. The sheared and size-selected DNA was analysed on a DNA 1000 chip. Samples were compared to optimal DNA shearing profiles to ensure accurate shearing prior to proceeding to hybridisation (Figure 13).

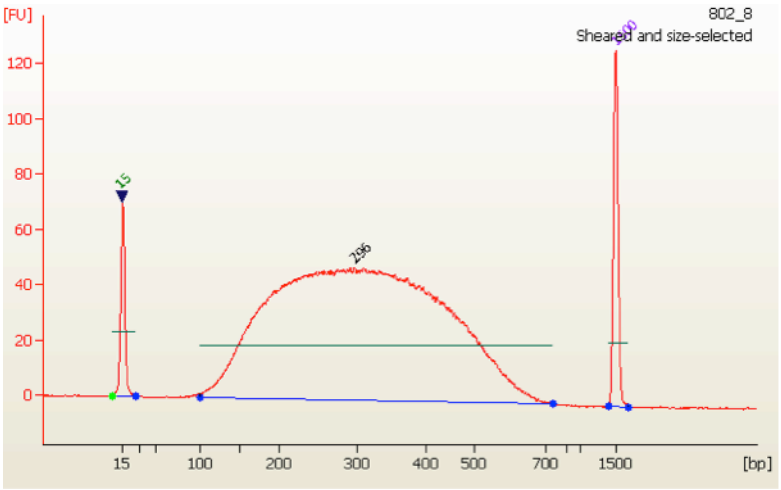


Figure 13: Optimal DNA shearing profile from Agilent 2100 Bioanalyzer electropherogram (12k chip)

Target fragment size 300bp. Peaks at extreme left (15bp) and extreme right of profile (1500bp) represent reference control fragments.

Following confirmation of satisfactory shearing profiles, samples underwent end repair, non-templated addition of 3'-A, adaptor ligation, hybridisation, enrichment PCR and related sample purification steps according to the SureSelect Illumina Paired-End Sequencing Library protocol (version 1.2, May 2011). The SureSelect

capture library or “baits” were customized for the HPV genome and the RNAse P human gene as follows. Overlapping 120-mer RNA baits allowing x5 coverage of the entire HPV16 genome was designed with the Agilent eArray software and then synthesized by Agilent Biotechnologies (NCBI Reference Sequence: NC_001526.2). Bait design paid additional attention to the circular nature of the genome to ensure coverage (x5) at the extremes of linearized text sequence, resulting in a total of 335 baits for the HPV16 genome.

Additionally, baits were designed and synthesized for the host gene, RNaseP and multiplexed with HPV16 baits. As before coverage was x5 for the 341bp RNase P gene (NCBI Reference Sequence: NC_000014.8). Inclusion of this gene was intended to serve two purposes; firstly, it would allow direct validation of the sequencing method with previously determined quantitative PCR results for each sample, and secondly allow calculation of relative HPV viral load between samples with RNaseP reads being the equilibrator for input DNA.

Sequencing was performed on the Illumina HiSeq platform in accordance with standard manufacturers protocols. Raw data management and bioinformatic analysis was provided by the third party. Bioinfomatic outputs were predetermined with the third party to ensure specific research targets and data were both realistically achievable and delivered to allow interpretation in keeping with the project aims. Specific reporting features were paired-end read origin (host or viral), mapping positioning, viral-host read analysis with specific interpretation of chimeric reads to report viral and host genomic break point/insertion locations and relative viral load.

4 EVALUATION OF DIAGNOSTIC TESTING IN OROPHARYNGEAL SQUAMOUS CELL CARCINOMA

4.1 INTRODUCTION

Human Papillomavirus-16 (HPV16) is the causative agent in a biologically distinct subset of oropharyngeal squamous cell carcinoma (OPSCC). HPV status has been demonstrated to be an important prognostic biomarker in OPSCC[98,100] with a hazard ratio for overall survival around 0.4 from systematic reviews of clinical trials[50] (2.3). In the design and introduction of new clinical trials, HPV16 status has become an essential inclusion or stratification parameter, highlighting the importance of accurate status determination. Understandably calls have been made to standardise the definitions and clarify the best test or combination of tests for accurate diagnosis[120]. Currently, a variety of detection methods are available[119], each with specific benefits and detractions (2.5). Additionally, considerable variation in sensitivity and specificity exists between the tests defining HPV status[118], such that the utility of some has been questioned.

4.2 DIAGNOSTIC TEST ANALYSIS AIMS & METHODS

Hypotheses

In keeping with previous evidence[125], detection of viral mRNA expression carried out by quantitative PCR (qPCR) techniques on fresh-frozen tissue samples can be considered the gold standard for HPV tumour diagnostics. Although invaluable in the research setting this gold standard has several logistical and practical difficulties that have ensured a reliance on alterative tests for routine clinical pathology services.

It is hypothesised that the tests currently used or advocated in routine clinical practice are less than optimal when directly compared to the gold standard.

Against this standard, this research aimed to determine the diagnostic and prognostic capacity of the frequently applied or advocated clinical tests (or combination tests); p16 IHC, HR HPV DNA ISH, combined p16 IHC/HR HPV DNA ISH, HPV16 DNA qPCR, and combined p16 IHC/HPV16 DNA qPCR. Such an application of a comprehensive diagnostic test panel to strictly classified OPSCC samples had not been conducted previously and sought to define a single clinical standard for HPV diagnostic testing in OPSCC.

In addition, the analysis aimed to provide data to clarify the role of HPV within a cohort of OPSCC patients from the United Kingdom, a region from which a the combination of both incidence and prognostic data had yet to be published. In doing so this would test the hypothesis that reported increases in HPV positive

OPSCC incidence, whilst being apparent, are overstated due to a lack of diagnostic stringency by virtual of suboptimal testing techniques.

Testing Methods & Statistical Analyses

The analysis cohort contained 108 OPSCC all of which had been strictly classified according to site[46,47] to ensure only tumours from the oropharynx were included (2.2). Samples originated from 3 distinct time periods (1988-1997, 2004-2007 & 2008-2009), allowing analysis of relative HPV positive tumour incidence over time. FFPE based analyses were conducted using triplicate fixed tissue tumour cores mounted on a tissue microarray (TMA) whilst DNA and RNAqPCR testing was conducted on relevant nucleic acid samples (DNA or cDNA) on a real time PCR platform as detailed previously in material and methods (3.11). Due to a hiatus in sample collection in the period from 1997-2004, cases/samples were not available for inclusion in the research project. This precluded analysis of a cohort of samples that spanned the entire period 1988 – 2009.

HPV status, as defined by the gold standard test, was only deemed positive where positive results were apparent in both of duplicate runs of RNA qPCR and as such a tumour would therefore be deemed as a reliably HPV- driven malignancy. The chi-squared and Kruskal–Wallis tests were used for comparison of demographic and tumor-specific features between periods of sample collection and HPV positive and -negative subgroups. Kaplan–Meier estimates for survival analysis and determination of testing sensitivity and specificity of the 7 alternative tests (p16 IHC; HR HPV ISH; DNA qPCR; and combined analysis tests: p16 IHC/HR HPV ISH; p16

IHC/DNA qPCR; DNA/ RNA qPCR; and p16 IHC/RNA qPCR) were carried out. The log-rank (Mantel–Cox) test was used for comparison between survival curves according to each of the detection methods.

Disease-specific survival was defined as death from or due to OPSCC, and overall survival was defined as death resulting from any cause. Both disease-specific survival and overall survival were calculated at 36 months follow-up beyond the date of initial diagnosis.

Tissue procurement was undertaken across an extended period (1988 – 2009), therefore, to ensure that era of collection did not impact upon detection rates, the quality of both DNA and RNA (cDNA) was assessed by a Kruskal–Wallis test of the cycle threshold (CT) of the relevant reference gene (RNase P for DNA qPCR and β Actin for RNA qPCR).

It is appreciated that oropharyngeal SCC typically contains an inflammatory cell infiltrate[196], particularly tumour infiltrating lymphocytes, and as such detection of HPV16 DNA/RNA in samples prepared from non-microdissected tissues may be reduced. Quantification of any potential reduction in tumour purity was undertaken through analysis of 20 randomly selected fixed tissue slides corresponding to the fresh frozen samples, from which DNA & RNA were extracted.

4.3 DIAGNOSTIC & PROGNOSTIC CAPACITY RESULTS

Cohort characteristics

The characteristics of the overall group and comparisons between samples derived from the three periods of collection are demonstrated in Table 24. Whilst comparison was made between the numbers of cases collected in each era and the clinical and demographic characteristics of those cases, no evidence of significant differences was detected. Tobacco consumption proved to be the sole exception in this instance, as there was a demonstrable increase in the proportion of non-smokers in latter years ($p=0.018$). As would be expected in a cohort of head and neck cancer patients, a male to female ratio of approximately 3:1 was consistently observed. The majority of tumours originated from within the tonsillar subsite however the contributions to the cohort from soft palate and base of tongue subsites remained meaningful.

	Time Period			Overall	Statistical Significance
	1988-1997	2004-2007	2008-2009	**1998-2009**	
Patient/Tumour Data					
No. of Patients	40 (37%)	37 (34%)	31(29%)	**108 (100%)**	
Age at Diagnosis (years)					
Mean	60.2	57.2	57.8	**58.5**	NS
Median	61.7	56.8	58.7	**58.6**	
Sex					
Female	8 (20%)	10 (27%)	7 (23%)	**25 (23%)**	NS
Male	32 (80%)	27 (73%)	24 (77%)	**83 (77%)**	
Tumour Site					
Tonsil	22 (55%)	17 (46%)	20 (64%)	**59 (55%)**	NS
Soft Palate	8 (20%)	7 (19%)	3 (10%)	**18 (17%)**	
Base of Tongue	7 (18%)	9 (24%)	4 (13%)	**20 (18%)**	
Oropharynx (not further spec.)	3 (7%)	4 (11%)	4 (13%)	**11 (10%)**	
Smoking					
Non-smoker	2 (7%)	12 (37%)	9 (31%)	**23 (25%)**	p=0.018
<20 pack-year history	16 (51%)	7 (22%)	7 (24%)	**30 (33%)**	
≥20 pack-year history	13 (42%)	13 (41%)	13 (45%)	**39 (42%)**	
Alcohol Consumption					
Non-drinker	4 (13%)	6 (19%)	4 (13%)	**14 (15%)**	NS
<28 Units/Week	12 (40%)	13 (42%)	15 (48%)	**40 (44%)**	
≥28 Units/Week	14 (47%)	12 (39%)	12 (39%)	**38 (41%)**	
Nodal Stage					
N0	9 (35%)	10 (28%)	8 (27%)	**27 (29%)**	NS
N1 (without ECS)	3 (11%)	5 (14%)	4 (13%)	**12 (13%)**	
N1 (with ECS) & N2/3	14 (54%)	21 (58%)	18 (60%)	**53 (58%)**	
Tissue Available					
Fresh frozen	36 (90%)	29 (78%)	30 (97%)	**95 (88%)**	NS
FFPE	31 (78%)	36 (97%)	30 (97%)	**97(90%)**	

Abbrevations: ECS, extracapsular; FFPE, formalin fixed parafin embedded

Table 24: Cohort Characteristics

Availability for testing, sample quality and consistency between repeats.

DNA and RNA qPCR

98/108 (91%) and 95/108 (88%) of samples were evaluable for HPV status determination by DNA and RNA qPCR respectively, and importantly with all samples providing analysable results from duplicate qPCR runs.

A Kruskal-Wallis test was applied to the mean CT of the relevant reference gene by year of sample collection (RNase P for DNA qPCR and β Actin for RNA qPCR). There was no evidence of statistically significant change in DNA quality as determined by mean CT for RNase P across the period of collection of samples (p=0.87) (Figure 14). Whilst there was an apparent difference in the mean CT for β Actin amongst cDNA samples (RNA) (p=0.01), the graphical representation below demonstrates variation to be in a non-linear fashion and therefore appears not to relate to increasing age of samples (Figure 15). It was concluded therefore that there was no conclusive evidence of DNA or RNA degradation over time that could influence HPV status determination.

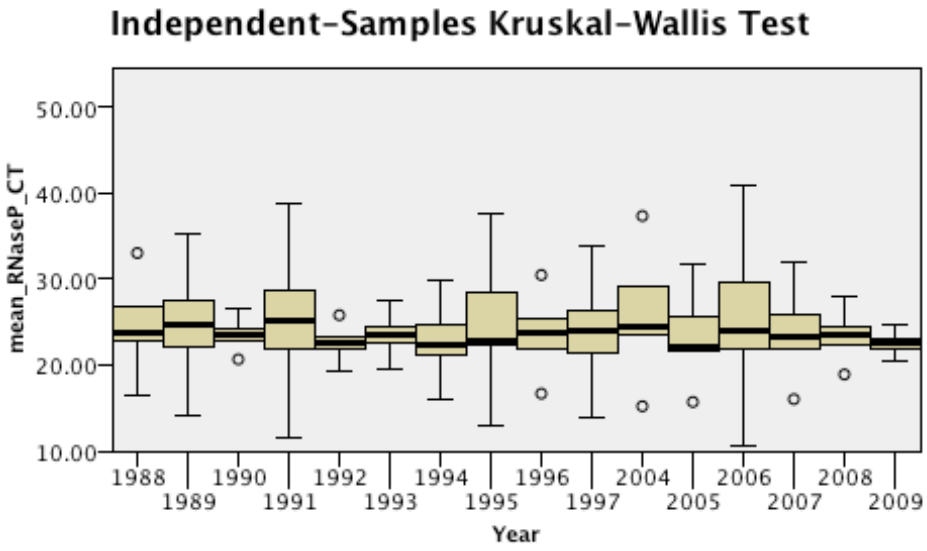


Figure 14: DNA Quality Over Time

(as reflected by mean CT of reference gene, RNase P for each sample) p=0.87

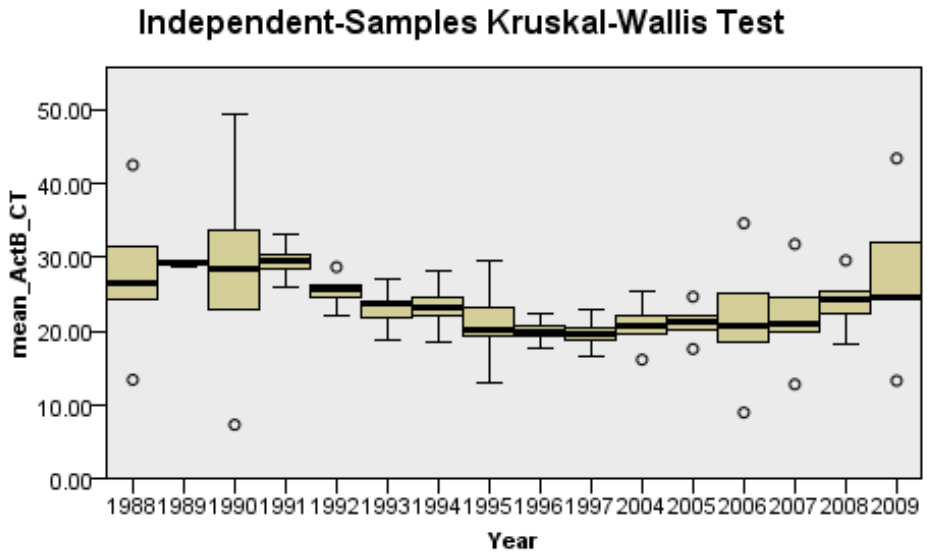


Figure 15: RNA Quality Over Time

(as reflected by mean CT of reference gene, β Actin for each sample). p=0.01

It is acknowledged that oropharyngeal SCC typically contains an inflammatory cell infiltrate, particularly tumour infiltrating lymphocytes, and as such detection of HPV16 DNA/RNA in samples prepared from non-microdissected tissues may be

reduced. Quantification of any potential reduction in tumour purity was undertaken through analysis of 20 randomly selected fixed tissue slides corresponding to the fresh frozen samples from which DNA & RNA were extracted. This analysis demonstrated a tumour cell proportion of greater than 50% in all cases, and greater than 2/3 in 16 (80%) samples. Given the sensitivity of the assays (as detailed previously, 3.11) failure to detect HPV, if present, would not be expected with a tumour cell proportion of 50% or more.

TMA: P16 IHC & HR HPV ISH

FFPE blocks were available for tissue microarray inclusion for 97/108 cases. p16 IHC and HR HPV ISH results were analysable from at least one or more representative tumour core for each case (97/97, 100%). Upon consideration of cases with positive staining results, complete consistency of p16 IHC and HR HPV ISH results between all tumour cores originating from the same FFPE block was seen in 36/41 (88%) and 20/29 (69%) cases respectively. A combined threshold of ≥2/3 core concordance for combined p16 IHC and HR HPV ISH was achieved by 97/97 (100%) cases.

In appreciation of the limitations of sampling which apply when undertaking TMA analysis, it was felt important to undertake additional p16 IHC of whole sections for 5 cases where a complete absence of staining in the TMA cores in the face of HPV positive tests. This internal control confirmed true negative scores for p16 IHC in each case by virtue of the presence of p16 IHC staining within adjacent normal tissue components such as follicular dendritic cells, tonsillar crypt epithelium and fibroblasts.

HPV status

The proportion of HPV positive cases within each time period or era, and the overall total, is expressed as a trend 1988-2009 in Table 25. The percentage of cases considered HPV positive as defined by RNA qPCR increased from 14% (5/36) within the era 1988-1997, to 57% (17/30) for cases originating from 2008-2009 (P=0.001). The overall tumour HPV positive rate for the sample cohort irrespective of era of tissue collection was 36% (34/95).

The increase in incidence remained statistically significant irrespective of the test used although the 2008/9 measures of HPV rates varied markedly between 52% for combined DNA/RNA qPCR and 77% for p16 IHC reflecting the variable sensitivity of tests.

In comparison to the defined gold standard test, the sensitivity of the seven tests, and combinations of tests, ranged from 88% to 97% and the specificity from 82% to 100% (Table 25).

With the exception of a single case (case 87,Table 27) all samples that were positive by RNA qPCR were also positive by DNA qPCR, however 8 DNA qPCR positive cases were negative by RNA qPCR, a feature reflected in the consequent reduction in specificity for DNA qPCR (87%).

3/95 (3%) of cases were positive for either HPV18 (1/95) or HPV33 (2/95) E6 expression. Of the latter, one of these cases demonstrated a multiple HPV infection with evidence of both HPV16 and HPV33 E6 gene expression. FFPE tissue was not available for this case, however the second HPV33 positive case did demonstrate a

positive result for HR HPV ISH yet this positive result occurred in the absence of p16 IHC staining.

The single case shown to be positive for HPV18 was p16 IHC/HR HPV ISH positive whilst negative by both HPV16 DNA and RNA qPCR.

A series of cases (n=5) were subjected to full section analysis of HR HPV ISH due to conflicting HPV status reporting between FFPE results and those of the gold standard. Full section analysis was undertaken to ensure sure conflict was due to the testing regime rather than the effects of tumour sampling.

1/5 cases was reclassified to exclude HR HPV ISH staining following the same criteria and as such brought fixed tissue and frozen tissue derived results into accord with one and other. The remaining 4/5 cases correlated with the TMA results. Of these one of the conflicting HR HPV ISH cases was positive for HPV33 by RNA qPCR, however 3/5 cases remain in conflict with the expression results for three HR HPV types (HPV-16, -18 & -33). Results (summarized in Table 25) were amended to reflect the findings of whole section analysis.

	Sensitivity	Specificity	Number of HPV +ve by Presentation Era			Total	Statistical Significance
HPV Status by Test	(compared to RNA qPCR)		1988-1997	2004-2007	2008-2009	1988-2009	p =
RNA qPCR "Gold Standard"			**5/36 (14%)**	**12/29 (41%)**	**17/30 (57%)**	**34/95 (36%)**	**0.001**
p16 IHC	94%	82%	6/31 (19%)	13/36 (36%)	23/30 (77%)	**42/97 (43%)**	<0.001
HR HPV ISH	88%	88%	4/31 (13%)	14/36 (39%)	18/30 (60%)	**36/97 (37%)**	0.001
Combined p16/HR HPV ISH	88%	90%	4/31 (13%)	12/36 (33%)	18/30 (60%)	**34/97 (35%)**	0.001
DNA qPCR	97%	87%	8/35 (23%)	15/33 (46%)	17/30 (57%)	**40/98 (41%)**	0.02
Combined p16/DNA qPCR	97%	94%	3/26 (12%)	13/32 (41%)	17/30 (57%)	**33/88 (38%)**	0.002
Combined p16/RNA qPCR	94%	100%	3/27 (11%)	12/28 (43%)	16/29 (55%)	**31/84 (37%)**	0.008
Combined DNA qPCR/RNA qPCR	94%	100%	3/35 (9%)	12/29 (41%)	15/29 (52%)	**30/93 (32%)**	0.001

Table 25: Tumour HPV Status determination Over Time and HPV test Diagnostic Stringency

Case Number	Patient & Tumour Specifics								Tissue Availabilty		Tissue Microarray (TMA) & quantitative PCR Results					
	Gender	Age at Presentation	Site	Smoking Hx	Alcohol Hx	Nodal Stage	Era	Postoperative Radiotherapy	Fresh Tissue	FFPE	p16 IHC	HR HPV ISH	HPV16 DNA qPCR	HPV16 RNA qPCR	HPV18 RNA qPCR	HPV33 RNA qPCR
1	M	49	Tonsil	1	1	0	1	0	1	0	-	-				
2	M	51.8	Soft Palate	1	2	2	1	1	1	1						
3	M	44.5	Tonsil	2	-	2	1	0	1	0	-	-				
4	M	71.8	Tonsil	1	1	-	1	1	1	0	-	-				
5	M	68.6	Tonsil	1	1	2	1	0	1	1						
6	M	45.8	Soft Palate	2	2	2	1	1	1	0	-	-				
7	F	62.4	Base of Tongue	1	1	0	1	0	1	0	-	-				
8	M	69.7	Base of Tongue	1	1	-	1	1	1	1						
9	M	51.1	Base of Tongue	1	2	-	1	0	1	1						
10	M	71.5	Tonsil	0	1	2	1	1	0	1	-	-	-	-	-	-
11	F	55.6	Tonsil	-	-	-	1	-	1	0	-	-				
12	M	54.6	Base of Tongue	-	-	-	1	-	1	1						
13	M	65.5	Base of Tongue	-	-	-	1	-	1	1						
14	M	69.6	Tonsil	1	0	-	1	-	0	1	-	-	-	-	-	-
15	M	39.6	Tonsil	-	-	-	1	-	1	1	-	-				
16	M	64.2	Base of Tongue	-	-	-	1	-	1	1						
17	M	87.9	Soft Palate	2	2	0	1	0	1	1						
18	F	69.8	Tonsil	1	0	0	1	0	0	1	-	-	-	-	-	-
19	M	73.4	Soft Palate	2	2	0	1	1	1	1						
20	M	86.7	Soft Palate	2	2	1	1	1	1	1						
21	M	41.8	Tonsil	1	-	-	1	-	1	1						
22	M	54.5	Tonsil	1	1	2	1	0	1	1						
23	M	71	Tonsil	2	2	2	1	1	0	1	-	-	-	-	-	-
24	M	61.7	Soft Palate	1	1	2	1	0	1	0						
25	M	62.2	Soft Palate	-	-	-	1	0	1	1	-	-				
26	M	51.7	Tonsil	1	1	2	1	0	1	1						
27	F	71.9	Tonsil	1	0	2	1	0	1	1						
28	M	50.6	Tonsil	2	2	2	1	0	1	1						
29	M	51.8	Tonsil	2	2	-	1	0	1	1	-	-				
30	F	57.2	Tonsil	2	2	1	1	0	1	0						
31	M	65.5	Tonsil	2	1	1	1	1	1	0	-	-				
32	F	73.7	Oropharynx	2	1	0	1	0	1	1						
33	M	71.7	Tonsil	2	0	0	1	0	1	1						
34	F	36.1	Tonsil	1	2	2	1	1	1	1						
35	M		Oropharynx	-	-	-	1	-	1	1						
36	M	72.6	Oropharynx	-	-	-	1	-	1	1						
37	M	58.7	Tonsil	2	2	0	1	1	1	1						
38	M	43.4	Tonsil	-	2	2	1	1	1	1						
39	F	56.9	Base of Tongue	0	1	0	1	1	1	1						
40	M	42.9	Soft Palate	1	2	2	1	0	1	1						
41	M	55.5	Tonsil	2	2	2	2	0	1	1						
42	M	78.3	Tonsil	0	-	2	2	0	1	1						
43	M	47.6	Base of Tongue	1	2	2	2	0	1	1						
44	F	60.4	Tonsil	1	1	2	2	0	1	1						
45	M	93.4	Tonsil	0	1	2	2	0	0	1	-	-	-	-	-	-
46	F	41.5	Tonsil	2	0	0	2	0	1	1						
47	M	22	Tonsil	-	-	-	2	-	1	1	-	-	-	-	-	-
48	M	45	Base of Tongue	2	2	2	2	0	1	1						
49	F	75.5	Soft Palate	0	0	0	2	0	1	1						
50	M	49.5	Base of Tongue	1	1	2	2	0	0	1	-	-	-	-	-	-
51	M	56.8	Tonsil	1	1	2	2	0	1	1						
52	M	56.5	Tonsil	0	1	2	2	0	1	1						
53	F	70.4	Tonsil	2	2	1	2	0	1	1						
54	M	57.4	Soft Palate	1	1	0	2	0	1	1						

Table 26: Patient & Tumour Individual Case Diagnostic Analysis

(Continued in Table 4)

Case Number	Patient & Tumour Specifics									Tissue Availabilty		Tissue Microarray (TMA) & quantitative PCR Results					
	Gender	Age at presentation	Site	Smoking Hx	Alcohol Hx	Nodal Stage	Era	Postoperative Radiotherapy		Fresh Tissue	FFPE	p16 IHC	HR HPV ISH	HPV16 DNA qPCR	HPV16 RNA qPCR	HPV18 RNA qPCR	HPV33 RNA qPCR
55	F	47.4	Base of Tongue	2	2	0	2	0		1	1						
56	M	55	Soft Palate	2	2	1	2	0		1	1			-	-	-	-
57	F	74.6	Soft Palate	2	0	2	2	0		0	1						
58	F	59.2	Base of Tongue	1	1	2	2	0		1	1						
59	M	52.4	Base of Tongue	2	2	2	2	0		0	1			-	-	-	-
60	M	50.9	Oropharynx	-	-	0	2	0		1	1			-	-	-	-
61	M	65.5	Tonsil	1	1	2	2	0		1	1						
62	F	63.4	Soft Palate	2	2	2	2	0		1	1						
63	M	53.7	Tonsil	0	0	2	2	0		1	1						
64	M	43	Tonsil	0	1	2	2	0		1	1						
65	M	47.9	Tonsil	2	2	0	2	0		1	1						
66	M	67.1	Tonsil	0	1	0	2	0		1	1						
67	M	32.8	Oropharynx	-	-	0	2	0		1	1						
68	M	53.1	Soft Palate	0	1	2	2	0		1	1						
69	F	65.2	Soft Palate	0	0	0	2	0		1	1						
70	M	59.1	Base of Tongue	-	-	2	2	-		1	1				-	-	-
71	M	62.5	Tonsil	0	1	2	2	0		1	1						
72	M	63	Tonsil	2	2	1	2	0		1	1						
73	M	70.4	Base of Tongue	0	1	1	2	1		1	1						
74	M	65.4	Base of Tongue	-	-	2	2	1		1	1						
75	M	50.9	Oropharynx	2	2	0	2	1		1	0	-	-				
76	M	58.9	Oropharynx	2	2	1	2	1		1	1		-				
77	F	43.6	Tonsil	0	0	2	2	1		1	1						
78	M	51.4	Tonsil	0	1	2	3	0		1	1			-			
79	M	52.1	Tonsil	0	1	1	3	1		1	1						
80	M	77.1	Tonsil	0	1	2	3	0		1	1						
81	M	74.6	Soft Palate	1	2		3	0		1	1						
82	M	65.3	Tonsil	2	2	2	3	1		1	1						
83	M	61.8	Tonsil	2	2	2	3	1		1	1						
84	M	48.8	Tonsil		1	0	3	1		1	1						
85	M	67.3	Tonsil	1	1	0	3	1		1	1						
86	M	58.4	Base of Tongue	0	1	2	3	1		1	1						
87	M	45.7	Base of Tongue	2	2	2	3	1		1	0						
88	M	49.1	Oropharynx	2	2	2	3	1		1	1	-	-				
89	M	59.5	Tonsil	0	1	2	3	1		1	1						
90	M	59.6	Oropharynx	2	2	2	3	1		1	1						
91	M	62.4	Tonsil	1	2	0	3	1		1	1						
92	M	67.6	Tonsil	-	1	0	3	1		1	1						
93	F	63.4	Soft Palate	2	1	2	3	1		1	1						
94	F	34.8	Tonsil	0	0	2	3	1		1	1						
95	F	61.5	Tonsil	2	0	2	3	1		1	1						
96	M	56.2	Tonsil	2	2	2	3	1		1	1						
97	F	60.6	Tonsil	1	1	1	3	1		1	1						
98	F	48.9	Soft Palate	2	2	1	3	0		1	1						
99	F	66.8	Oropharynx	2	2	0	3	1		1	1						
100	M	52.6	Tonsil	1	1	2	3	1		1	1						
101	F	57	Tonsil	1	1	2	3	1		1	1						
102	M	57.7	Tonsil	0	0	2	3	1		1	1						
103	M	54.6	Base of Tongue	0	1	0	3	1		1	1						
104	M	56.8	Tonsil	2	2	0	3	0		1	1						
105	M	58.7	Base of Tongue	1	2	2	3	1		1	1						
106	M	61.6	Tonsil	2	1	2	3	1		1	1						
107	M	73.5	Oropharynx	2	0	1	3	1		1	1				-		-
108	M	27.4	Tonsil	0	1	0	3	1		1	1						

Table 27: Patient & Tumour Individual Case Diagnostic Analysis

(Continuation of Table 3).

Implications of Diagnostic Thresholds

The diagnostic threshold for positive status was predetermined to be equivalent to a single HPV-16 E6 gene copy per diploid genome

(by reference to CaSki copy number, 869 integrated copies). As Figure 16 clearly demonstrates, HPV16 DNA qPCR provided a continuous variable without clear delineation around the threshold. In contrast, HR HPV RNA qPCR was unequivocal by classifying samples in a binary fashion. It is the variability around the threshold that has, at least in part, led to variation in specificity of DNA qPCR as a diagnostic test.

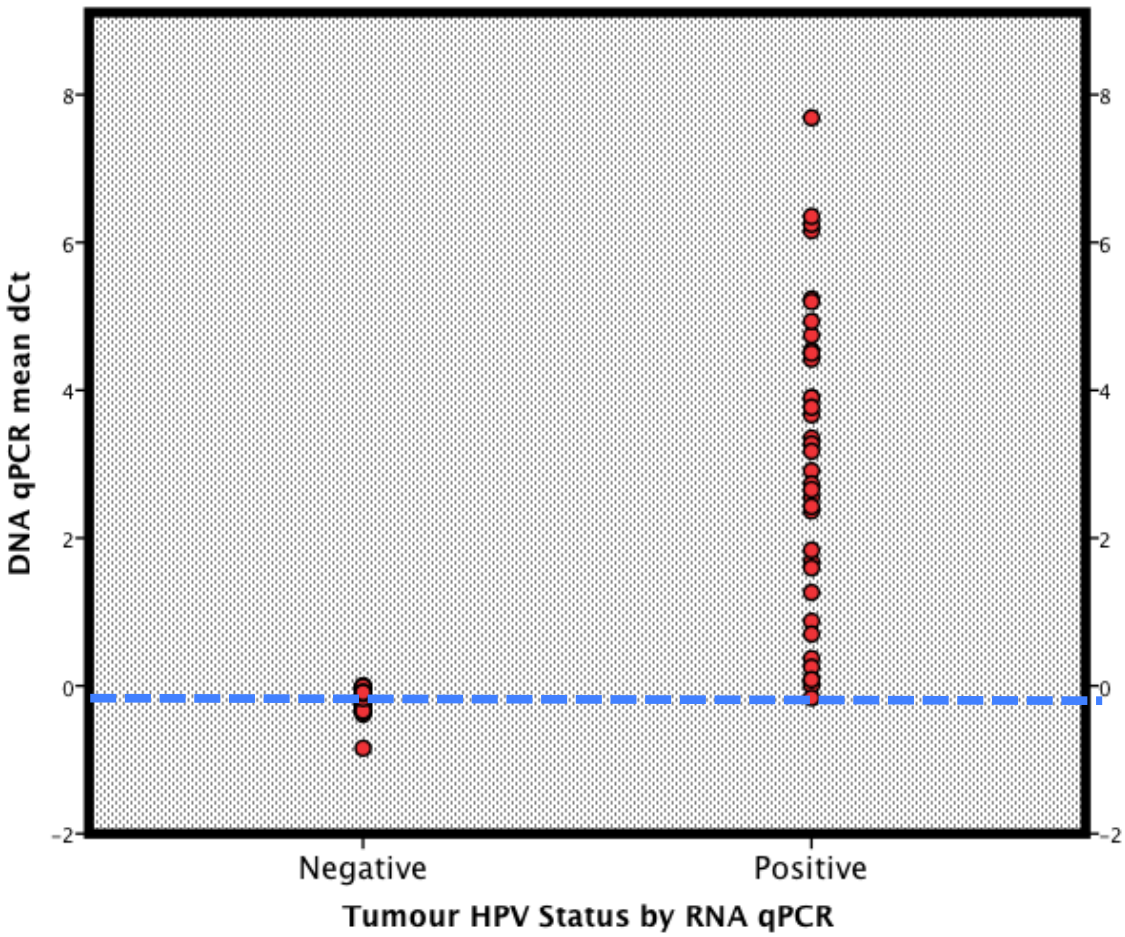


Figure 16: Plot of Individuals case HPV16 ΔCT DNA qPCR results by HPV status

Where HPV status is determined by RNA qPCR for HPV16 E6. Horizontal blue dash line represents the predetermined threshold for positive DNA qPCR status.

HPV16 Status in Tumours & Adjacent Normal Marginal Tissues

A subcohort of samples (n=53) had available matched normal tissue, resected following completion of tumour specimens intraoperatively and therefore beyond both the clinical macroscopic extent of disease and the traditional surgical clinical margins (typically 1cm or more for the majority of samples).

Table 28 depicts the relative HPV positive and negative proportions within the tumour and normal samples both in terms of HPV16 DNAqPCR (E6) and also RNAqPCR (E6 and or E7). The rate of coincident HPV positivity in both the primary tumour and its adjacent marginal specimen was 44% (11/25) when assessed using the gold standard test on fresh frozen samples. Evidence of E2 gene expression in the absence of viral oncogene expression was apparent in a single tumour sample (without any detectable viral expression in the matched normal tissue) and similarly E2 expression in normal tissue samples of two further cases was detected without E6 or E7 expression.

Diagnostic Test	**Tumour**			**Normal**		
	Positive	Negative	**Total**	Positive	Negative	**Total**
DNA E6	27 (51)	26 (49)	**53 (100)**	15 (28)	38 (72)	**53 (100)**
RNA E6+/orE7	25 (47)	29 (53)	**53 (100)**	11 (21)	42 (79)	**53 (100)**

Table 28: HPV Status in Matched Tumour and Adjacent Normals

Sample HPV status classified by DNAqPCR and RNAqPCR with percentage of cases in brackets.

HPV16 status vs. clinical characteristics

Individuals with HPV positive malignancy were younger than the HPV negative group (mean 53.3 vs. 60.8 yrs, p=0.003). The age of patients at the time of diagnosis conformed to a normal distribution (1-sample Kolmogorov-Smirnov test, p=0.997) and, significantly, the modest 7.5 yrs difference seen between mean ages exceeded the 6.8 yrs difference seen between the median ages. The other notable clinical characteristic correlating with HPV status was smoking history. Of the 82 patients for whom reliable smoking history could be determined, the non-smokers and those smoking <20 pack years were more common in the HPV positive group (Pearson's chi square, p=0.007). There were no significant differences between the groups when gender, tumour site, cervical lymph node stage or alcohol consumption were considered.

	HPV Status by combined RNA qPCR Analysis		Total	Statistical Significance
	Negative	Positive		p =
Patient/Tumour Data				
No. of Patients	61 (64%)	34 (36%)	**95 (100%)**	
Age at Diagnosis (years)				
Mean	60.8	53.3		0.003
S.E. of Mean	1.4	1.7		
Sex				
Female	16 (26%)	6 (18%)	**22 (23%)**	NS
Male	45 (74%)	28 (82%)	**73 (77%)**	
Tumour Site				
Tonsil	30 (49%)	22 (65%)	**52 (55%)**	
Soft Palate	13 (21%)	4 (12%)	**17 (18%)**	NS
Base of Tongue	11 (18%)	6 (18%)	**17 (18%)**	
Oropharynx (not further spec.)	7 (11%)	2 (6%)	**9 (9%)**	
Nodal Stage				
N0	15 (31%)	9 (28%)	**24 (30%)**	
N1 without ECS	8 (16%)	3 (9%)	**11 (13%)**	NS
N2/3 or N1 with ECS	26 (53%)	20 (63%)	**46 (57%)**	
Total	*49 (100%)*	*32 (100%)*	***81 (100%)***	
Smoking				
Non-smoker	8 (16%)	13 (42%)	**21 (26%)**	
<20 pack-year Hx	16 (31%)	11 (36%)	**27 (33%)**	0.007
≥20 pack-year Hx	27 (53%)	7 (23%)	**34 (42%)**	
Total	*51 (100%)*	*31 (100%)*	***82 (100%)***	
Alcohol Consumption				
Non-drinker	6 (14%)	5 (16%)	**11 (14%)**	
<28 Units/Week	14 (33%)	18 (56%)	**32 (43%)**	NS
≥28 Units/Week	23 (53%)	9 (28%)	**32 (43%)**	
Total	*43 (100%)*	*32 (100%)*	***75 (100%)***	

Table 29: Clinical Characteristics by HPV status

(As defined by RNA qPCR)

HPV testing methods as prognostic biomarkers: survival analysis

Kaplan-Meier survival curves, segregating cases by HPV status (as assigned by the gold standard RNA qPCR test) showed a significant prognostic benefit in both Overall Survival (OS, p= 0.003) and Disease Specific Survival (DSS, p=0.005) (Figure 17). Kaplan-Meier estimates of mean survival for the other tests are shown in Table 30. Although very similar to the gold standard RNA qPCR outcome measures, the

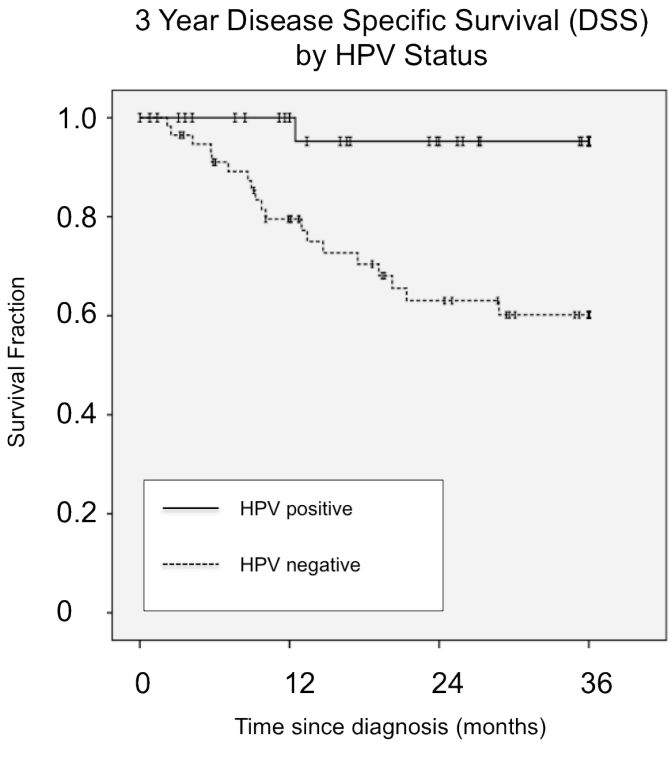
test combination demonstrating a prognostic benefit of greatest significance for both disease specific and overall survival was combined p16 IHC/ DNA qPCR (OS, p=0.002 and DSS, p=0.005). The least satisfactory tests in this regard were p16 IHC or HR HPV ISH, either alone or in combination. Although remaining statistically significant, the differences in OS (p=0.021, 0.011 & 0.016 respectively) vary by an order of magnitude by comparison with the gold standard. All tests using target amplification of DNA and RNA, performed relatively well in differentiating survival outcomes for both OS and DSS, although it is important to note that DNA qPCR lacked specificity (87%).

Disease Specific Survival		Mean (months)	SE	95% Confidence Interval		p value
				Lower	Upper	
RNA qPCR	HPV -ve	26.7	1.7	23.3	30.1	0.005
	HPV +ve	34.9	1.1	32.7	37.0	
p16 IHC	HPV -ve	27.2	1.9	23.6	30.9	0.018
	HPV +ve	33.5	1.4	30.7	36.2	
HR HPV ISH	HPV -ve	27.6	1.8	24.1	31.0	0.02
	HPV +ve	34.0	1.3	31.4	36.7	
p16 IHC/HR HPV ISH	HPV -ve	27.7	1.7	24.3	31.1	0.027
	HPV +ve	33.9	1.4	31.2	36.7	
DNA qPCR	HPV -ve	26.1	1.9	22.5	29.9	0.008
	HPV +ve	33.8	1.3	31.3	26.2	
DNA qPCR/p16 IHC	HPV -ve	26.7	1.9	23.0	30.4	0.005
	HPV +ve	34.9	1.0	32.9	37.0	
RNA qPCR/p16 IHC	HPV -ve	26.8	1.9	23.0	30.6	0.007
	HPV +ve	34.9	1.1	32.7	37.0	
Combined DNA/RNA qPCR	HPV -ve	26.7	1.7	23.3	30.1	0.006
	HPV +ve	34.8	1.1	32.6	37.1	

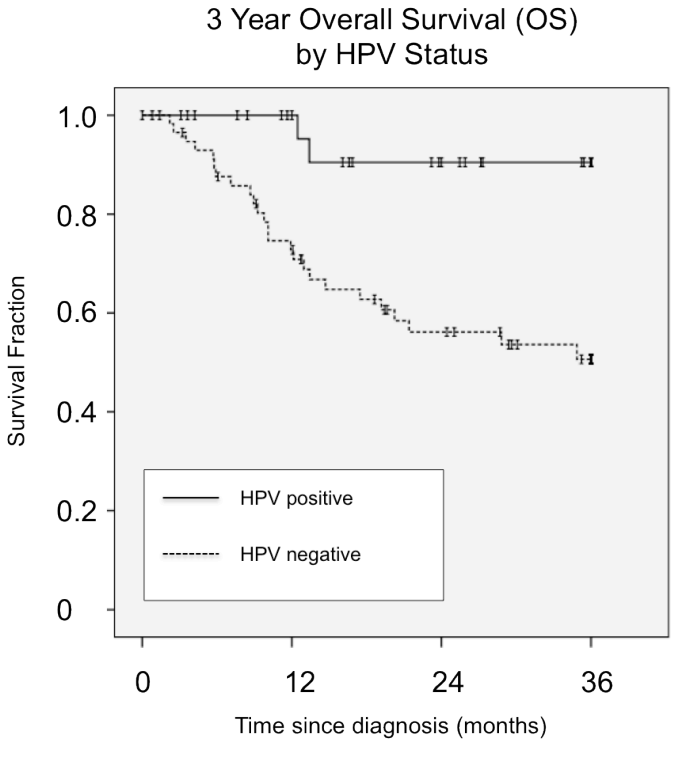
Overall Survival		Mean (months)	SE	95% Confidence Interval		p value
				Lower	Upper	
RNA qPCR	HPV -ve	24.7	1.8	21.2	28.2	0.003
	HPV +ve	33.8	1.5	30.9	36.7	
p16 IHC	HPV -ve	25.7	1.9	21.9	29.5	0.021
	HPV +ve	31.8	1.7	28.5	35.2	
HR HPV ISH	HPV -ve	25.7	1.8	22.2	29.3	0.011
	HPV +ve	33.0	1.6	29.8	36.2	
p16 IHC/HR HPV ISH	HPV -ve	25.9	1.8	22.4	29.5	0.016
	HPV +ve	32.9	1.7	29.5	36.2	
DNA qPCR	HPV -ve	24.4	1.9	20.7	28.0	0.007
	HPV +ve	32.1	1.6	28.9	35.3	
DNA qPCR/p16 IHC	HPV -ve	24.8	1.9	21.0	28.6	0.002
	HPV +ve	33.9	1.4	21.0	36.7	
RNA qPCR/p16 IHC	HPV -ve	24.8	2.0	21.0	28.7	0.003
	HPV +ve	33.8	1.5	30.9	36.7	
Combined DNA/RNA qPCR	HPV -ve	26.7	1.8	21.2	28.2	0.004
	HPV +ve	33.7	1.5	20.7	30.1	

Table 30: Kaplan-Meier estimates of survival at 36 Months by HPV status

(HPV status as defined by individual tests. Gold standard, RNA qPCR, highlighted)



p=0.005



p=0.003

Figure 17: 36 Month Kaplan-Meir Plots of Disease Specific Survival and Overall Survival by HPV status (HPV status determined by RNA qPCR, DSS (above), OS (Below))

4.4 DISCUSSION & IMPLICATIONS FOR CLINICAL PRACTICE

The results of this research provide the first evidence of relative incidence of HPV driven OPSCC, with associated prognostic implications, from a United Kingdom cohort.

Reflecting trends seen in other developed countries, the proportion of HPV mediated OPSCC cases has increased from 14% to 57% between 1998 and 2009 within the Liverpool cohort. Although there was variation in this rate depending on diagnostic test used (52 to 77%, 2008-9), these results show that there has been a substantial rise in relative incidence of HPV positive OPSCC. Although choice of test will influence the final HPV positive rate, these data are comparable to published rates from Western Europe and North America[3,51,76,90,197-200].

It remains to be seen whether the trend of increasing HPV positive OPSCC, and indeed OPSCC as a whole, continues, however speculative extrapolations of current rates suggest that by 2020 in a comparable US population, HPV-positive OPSCC cases alone will exceed those of cervical cancer for the first time[63]. Whether or not rates exceed their present state, the data generated here will provide more evidence in support of calls to strengthen primary and secondary prevention through prophylactic vaccination & education.

In keeping with the substantially improved outcomes, individuals with HPV positive OPSCC typically have improved locoregional control and reduced incidence of second primaries of the aerodigestive tract[91,201]. Evidence has previously suggested a link between genetic or epigenetic alterations within otherwise normal marginal tissues adjacent to HPV negative HNSCC and potential for disease recurrence[202,203].

In keeping with this it would be conceivable that the oncogenic drivers of HPV OPSCC, namely E6 and E7 expression, would be absent in marginal normal tissues. This however, did not appear to be the case, with almost half (44%) of normal samples matched to HPV positive OPSCC demonstrating viral oncogene expression. Clearly, inclusion of tumour cells within normal samples cannot be excluded in the absence of a full sectional histological analysis of the sample prior to nucleic acid extraction. Reliance on FFPE sourced material with microdissection may provide greater clarity; however there would therefore be a reliance on RNAqPCR results derived from FFPE rather than fresh frozen samples.

HPV analysis of OPSCC in clinical practice is becoming a fundamental requirement in the provision of both adequate prognostic information but also to facilitate entry into appropriately stratified clinical trials, including those investigating the potential to de-escalate the intensity of curative therapies. Paradoxically, there remains no 'international standard' for defining HPV related OPSCC in clinical practice[118]. Through the analysis of the currently applied diagnostic tests and test algorithms, these results highlight the compromises that have been necessary in terms of sensitivity and specificity, to achieve a test which can be used in a clinical setting. Van Houten et al[128] alluded to the inherent variability in the sensitivity of tests that may lead to potential overestimation of the role of HPV in OSPCC. Clearly the implications of reduced specificity for individuals recruited to trials aiming to de-escalate therapeutic intensity on the basis of HPV status could be critical.
p16 IHC was initially described as a surrogate for HPV status by Klussmann et. al.[132] and was later applied in OPSCC survival analysis by Lassen et al. [100] in their

description of HPV status within the Danish Head and Neck Cancer Group (DAHANCA) 5 Trial. When used in isolation, p16 IHC will identify tumours with excess p16 protein due to both the effects of viral E7 protein but also through, as yet unexplained, non HPV-mediated mechanisms[204]. In addition, the strength of stain and proportion of tumour specimen stained in a given FFPE sample will vary leading to ongoing debate surrounding the best p16 threshold and an ever-present potential for inaccurate classification[131,133,136].

The inclusion of HPV DNA testing (by PCR or ISH) in combination with p16 has been advocated to improve testing specificity, with such combined tests allowing classification into one of four groups[205] depending on a score for the two components. Robinson et. al.'s review[119] of HPV testing included a pooling of results from six studies examining 496 tumours using such a classification and found p16 positive/HPV negative in 5% of cases and p16 negative/HPV positive results in 8%. Based on our current series of 108 cases, the p16 IHC/HR HPV ISH classification demonstrates a p16 positive/HPV negative rate of 8% and p16 negative/HPV positive status in 2% of cases. Both of the p16 negative/HPV positive cases were negative by both DNA qPCR and RNA qPCR, however one sample was positive by HPV33 RNAqPCR analysis. Of particular interest however is the finding that RNA qPCR results highlight 2/97 cases (2%) that were p16 positive/HR HPV ISH negative. Such a finding of false negative results reflects reduced sensitivity for the combined p16 IHC/HR HPV ISH test in determining tumour HPV16 status. By comparison, combined p16 IHC/DNA qPCR showed 6/88 (7%) cases that were p16 positive/HPV negative, none of which demonstrated HPV16 E6 expression (RNA qPCR). The presence of HPV16 DNA was detected in 8 cases (20% of DNA qPCR positive

samples) where expression was not evident. Given the stringent efforts employed to avoid contamination at each step in this analysis, it is likely that this reflects detection of an innocent bystander viral infection in the absence of true virally mediated malignancy. Clearly the reduced specificity of DNA qPCR alone limits its utility in most settings. The threshold applied to DNAqPCR could however be questioned in this situation, yet the determined cut off was in keeping with a biological rationale for a truly HPV-driven malignancy[185] and indeed was more stringent than previous publications would suggest[130]. Although these results would suggest that the threshold is somewhat arbitrary, the level prescribed in this setting was made with a sound biological basis and indeed it would be hard to defend a threshold that falls below the level of a single viral copy per cell. The problem of threshold establishment is problematic and has been highlighted in the literature previoulsy[130] and it remains far from resolved leading one to question the use of DNA PCR based analysis in clinical practice. Once more, we recognise the implication that non-microdissected fresh frozen tissue samples may have on the potential for altered proportion of HPV to genomic DNA where tumour cell percentage is reduced. In an attempt to quantify any reduction in tumour cell percentage in fresh frozen samples, matched FFPE histopathological analysis was undertaken in a representative sample subcohort. This indicated tumour cell burden to be over 50% in all cases and greater the 2/3 in 80% of cases.

Detection of target signal for HPV in all samples, albeit below the threshold, was a curious finding which defied stringent attempts to exclude cross-contamination at all steps from sample procurement through to laboratory analysis (disposable instruments at time of tumour sectioning, fume cupboard isolation, separate

laboratory space, refrigeration and tools (pipettes etc). Indeed, at all times, non-template controls remained free from signal, making PCR setup errors unlikely. By its very nature PCR has potential for specificity weakness depending on the assay design (eg. sequence specificity or mispriming) or technique (eg. PCR cycle number for detection) however the possibility of a transcriptionally inactive remnant of viral DNA sequence (conceivably following HPV infection) should be considered. The results of the HPV16 E6 assay were paralleled by HPV16 E2 and E7 DNAqPCR assays lending some support to this explanation.

The Ventana Inform III HR HPV ISH probe detects twelve high risk HPV types including HPV16, 18 & 33. In this instance the ISH analysis defined a small subset of HPV negative cases as being positive, although by inclusion of a combined analysis with p16 IHC this group was reduced. It is possible that the HR HPV cocktail is detecting HPV types other than those tested using RNA qPCR, however contribution to OPSCC of other HR HPV types (beyond HPV16 & -18) in isolation is unlikely[65,76]. Unfortunately genotype specific probes are not available for diagnostic use in Europe due to licensing restrictions (Dako and Ventana Medical Systems Inc).
To date there is no published data that compares the genotype specific probes with cocktail probes however the probe utilised in this research has been shown to compare favorably when analysed against the detection of HPV16 E7/E6 mRNA[137].

If the calls for inclusion of OPSCC patients into appropriately designed and stratified clinical trials are to be met[50,120,206], then it is vital that accurate classification of HPV status be made prior to enrolment, and with a validated, clinically appropriate test.

This is now more than a theoretical problem, as several trials focussing towards de-escalation of treatment for individuals with HPV positive tumour and in early phase trials of HPV directed agents or immunotherapy, have recently opened. Our data suggests that HPV16 status determination with the p16 IHC/DNA qPCR combination test offers a valuable alternative to viral expression analysis on fresh tissue samples, and retains excellent prognostic value along with 97% sensitivity / 94% specificity. Such a test, although feasible, does carry with it logistical constraints for a routine pathology practice. The combination of p16 IHC/ HR HPV ISH is worthy of consideration as an alternative, consistent with the diagnostic algorithms suggested by Westra et. al. [136]. In our data, specificity for p16 IHC/ HR HPV ISH, albeit with a HPV high risk cocktail probe rather than a type specific probe, was acceptable (90%) but did come at the expense of sensitivity (88%) such that the test may be deemed undesirable.

Clinical trials in HNSCC frequently struggle to adequately recruit[207], and in those focussing within one anatomical sub-site, this difficulty may be exaggerated. In order to maximise sensitivity i.e. potential recruitment whilst maintaining specificity i.e. patient safety, the choice of satisfactory test is more limited. Faced with the potential "loss" of approximately 10% of eligible patients using tests such as p16 IHC/ HR HPV ISH, the benefits to sensitivity of employing DNA or RNA PCR assays appear to easily balance the additional logistic costs.

The contribution and clinical importance of high risk HPV subtypes other than HPV16, appears to be minimal by comparison with gynaecological and anogenital

malignancy[44]. Consequently, we feel that the use of "consensus" HPV PCR primers in HNSCC cases is difficult to justify, not least as this would merely confirm presence of viral DNA rather than the stronger burden of proof that viral oncogene expression bares when considering virally mediated malignancy. Kreimer et al[76] in their systematic review of prevalence and HPV type distribution in the head and neck found 86.7% of OPSCC were HPV16 positive whilst HPV18 and HPV33 positive cases, the subsequent largest percentage of types, accounted for only 1% each. Using viral oncogene expression, our findings are comparable; with HPV16 accounting for 94% of all HPV positive cases and HPV18 & 33 representing a small subset (3% and 6% respectively).

With respect to survival, this research reinforces previously reported favourable outcomes for individuals with HPV positive tumours[50,91,97,98], as demonstrated by both improved DSS and OS. It is apparent that, with the exception of p16 IHC or HR HPV ISH in isolation, most of the other assays available provide a reasonable prognostic guide.

Debate remains as to whether tobacco usage further stratifies HPV positive and negative malignancies into a third, middle, tier for prognostic value[91,208].

Due to the small numbers of non-smokers in this cohort, limiting meaningful analysis, the additional prognostic value of combining HPV16 and smoking history has not been assessed. We speculate that the addition of smoking data has added accuracy to some other published HPV typing methods that have inherent and proven inaccuracies (e.g. p16 IHC). In such a setting, non-smoking behaviour will

doubtless be strongly correlated with an HPV-16 positive category.

In conclusion, the presented data reflects a rigorous analysis of diagnostic tests, judging their value against the most clinically relevant demands of diagnostic accuracy and prognostic relevance. It is anticipated that the design of forthcoming clinical trials, aimed at both de-escalating therapy in HPV mediated OPSCC and, conversely, intensifying therapy for HPV negative cases, will be informed and guided by these results.

5 IS THE REFERENCE "GOLD STANDARD" TEST RESTRICTED TO FRESH FROZEN SAMPLES?

5.1 INTRODUCTION

Whilst there has been an increasing consensus surrounding the use of viral oncogene expression as a reference test, for reasons of RNA instability, such testing has both relied on the analysis of fresh frozen tissue, and upon utilisation of specialist research laboratory techniques. This has inevitably limited the potential translation of RNA-based tests into routine clinical diagnostics. For clinical utility therefore, HPV testing strategies have necessarily focused on formalin-fixed paraffin-embedded (FFPE) tissue (2.5), but this has been at the expense of reduced sensitivity and specificity for oncogenic HPV[125]. Further, the inclusion of multiple analytical stages to achieve an accurate and reliable HPV status is technically cumbersome, may produce discordant results across the different tests employed and inevitably increases costs.

Limited recent evidence has suggested that a novel RNA-based chromogenic in situ hybridisation (ISH) technique (RNAscope, Advanced Cell Diagnostics, Hayward, CA) is capable of reliably detecting transcriptionally active genes, including High Risk HPV E6/E7 oncogenes, in FFPE tissue samples[209].

The RNA ISH techniques have been applied previously for detection of HPV transcripts in cervical neoplasia, genital condylomas and sinonasal papillomas, however the application of radioactively labelled probes and associated protracted

exposure times, or problems associated with variable RNA control detection limited or precluded routine clinical use[210-212].

This more recently described chromogenic RNA ISH technique has shown promising results in comparison to other FFPE-based HPV diagnostic tests both in a large (n=196) cohort of OPSCC[213] and on samples from head and neck subsites outwith the oropharynx[214] however, validation against the gold standard of fresh tissue derived mRNA qPCR has not been conducted in either of these clinical cohort studies. Although a test capable of detecting viral oncogene expression holds considerable promise, without formal validation against the previously described gold standard, clinical acceptance would be unlikely to be forthcoming.

High Risk HPV RNA In situ Hybridisation Aims

It was hypothesised that HR HPV ISH was capable of reaching a standard comparable to that displayed by the best performing test(s) currently applied in routine clinical practice, when measured against the current gold standard, evidence of viral oncogene expression derived from fresh tissue samples.

Test Methods, Interpretation & Analysis

Validation of the HR HPV RNAscope test was undertaken on a subcohort of TMA mounted FFPE cores from the previously described, well-characterised OPSCC cases (3.2), against data already generated for HR HPV expression (HPV16, -18 & -33 E6/E7 transcripts).

Test assessment for all tissue-based analyses was conducted independently, by two experienced Head & Neck Pathologists. Previously generated results for HR HPV DNA ISH, HPV16 DNA qPCR & the reference test, HPV16 E6/7, HPV18 E6 and HPV33 E6 expression, were correlated to relevant cases but not reinterpreted.

However p16 IHC status reanalysis was undertaken with the aim of allowing interpretation by two separate means, firstly using the threshold of strong and diffuse nuclear and cytoplasmic staining in ≥70% of the tumour [131] and also by means of the recently proposed and validated H score for p16 IHC[133]. The latter of these two scoring techniques reflects the cross-product (H score) of p16 staining intensity (scored from 0 to 3) and the percentage of tumour cells staining (from 0 to 100% in 5% increments), with an H score of >60 defined as p16 positive.

Statistical analysis was conducted in the same fashion as detailed previously (4.2)

5.2 RESULTS

Tissue sample quality and consistency

From the full cohort of OSPCC (4.3), matched FFPE tissue and correlating fresh tissue derived test results were available for 79 cases.

Interpretable results were available 79 cases, however one case had insufficient staining for ubiquitin C (positive control) and was therefore excluded from further analysis, leaving 78 of 79 (99%) cases for HPV analysis. Seventeen cases (22%) had discordant scores following TMA interpretation, due either to inter-observer variation or inter-core variation, and were subjected to further testing and

independent scoring using whole FFPE sections. A resultant Kappa scope of 0.948 (95% CI 0.88 – 1.0) for interobserver analysis of scoring was evident following complete analysis.

Cohort Characteristics

The entire cohort had a median follow up of 27 months (95% CI 27-37). The characteristics of the OPSCC cohort as a whole and subdivided by HPV status, defined by High Risk HPV qRTPCR, are shown in Table 31.

As would be expected from a subcohort of the original study cohort, these results are generally comparable.

The age of patients at diagnosis conformed to a normal distribution as signified by a one-sample Kolmogorov-Smirnov test (p=0.999). Individuals within the HPV positive group were statistically significantly younger than those in the HPV negative group (mean 54.2 vs. 61.3 years of age at diagnosis, p=0.003). Of the 69 cases for which reliable risk factor data was available, those individuals who were either non-smokers or who had smoked less than 20 pack-years were statistically more likely to have HPV positive OPSCC (p=0.004). Similarly, there was a trend towards lower alcohol exposure in the HPV positive group. There were no statistical differences between the groups by sex, tumour subsite or nodal category.

		HPV Status by HR HPV qRTPCR Analysis		Total	Statistical Significance
		Negative	Positive		p =
Patient/Tumour Data					
No. of Patients		45 (58%)	33 (42%)	**78 (100%)**	
Age at Diagnosis (years)					
Mean		61.3	54.2	**58.3**	0.003
S.E. of Mean		1.6	1.7	**1.2**	
Sex					
Female		12 (67%)	6 (33%)	**18 (23%)**	NS
Male		33 (55%)	27 (45%)	**60 (77%)**	
Tumour Site					
Tonsil		21 (49%)	22 (51%)	**43 (55%)**	
Soft Palate		9 (69%)	4 (31%)	**13 (17%)**	
Base of Tongue		8 (62%)	5 (38%)	**13 (17%)**	
Oropharynx (not further spec.)		7 (78%)	2 (22%)	**9 (11%)**	
	Total	45 (58%)	33 (42%)	**78 (100%)**	NS
Nodal Stage					
N0		13 (59%)	9 (41%)	**22 (31%)**	
N1 without ECS		6 (67%)	3 (33%)	**9 (13%)**	
N2/3 or N1 with ECS		20 (50%)	20 (50%)	**40 (56%)**	
	Total	*39 (55%)*	*32 (45%)*	***71 (100%)***	NS
Smoking					
Non-smoker		6 (32%)	13 (68%)	**19 (28%)**	
<20 pack-year Hx		9 (43%)	12 (57%)	**21 (31%)**	
≥20 pack-year Hx		21 (78%)	6 (22%)	**27 (41%)**	
	Total	*36 (54%)*	*31 (46%)*	***67 (100%)***	0.004
Alcohol Consumption					
Non-drinker		5 (50%)	5 (50%)	**10 (14%)**	
<28 Units/Week		11 (38%)	18 (62%)	**29 (42%)**	
≥28 Units/Week		21 (70%)	9 (30%)	**30 (44%)**	
	Total	*37 (54%)*	*32 (46%)*	***69 (100%)***	0.05

Table 31: HR HPV ISH Analysis Cohort Characteristics

High Risk HPV Detection in Normal Tissues

Of the seventy-nine FFPE cores, sampled from histologically normal mucosal tissue directly adjacent to related tumour resection specimens, seventy cores were available for analysis (70/79; 89%). Four of these cores displayed insufficient staining for ubiquitin C (positive control) and were therefore excluded from analysis. Within the remaining normal tissue cores there was no evidence of staining for HR HPV (0/66; 0%).

Oropharyngeal SCC Test Analysis

Photomicrographs of cases classified as HPV positive by High Risk HPV RNAscope are shown in Figure 18. The High Risk HPV RNAscope test had a sensitivity of 97% and a specificity of 93% against the reference test, with positive and negative predictive values of 91% and 98% respectively (Table 31).

Sensitivity values for other HPV tests when used as single tests were comparable to RNAscope; p16 IHC 97%, HR HPV ISH 94% and to a lesser extent HPV-16 DNA qPCR 91%, however, lower levels of specificity for oncogenic HPV were apparent for two of these tests; p16 IHC 82%, HPV-16 DNA qPCR 87%. Interpretation of more than one test per sample, in a diagnostic algorithm, appeared to improve specificity, but at the expense of sensitivity, exemplified by combined p16 IHC/HPV-16 DNA qPCR; sensitivity 91% and specificity 93%.

A comparison of the two p16 IHC scoring techniques revealed complete concordance with no difference in p16 status (positive or negative) either at the

level of individual TMA cores or specific tumour cases. As a result all reporting of p16 IHC testing performance can be considered to refer to either testing technique.

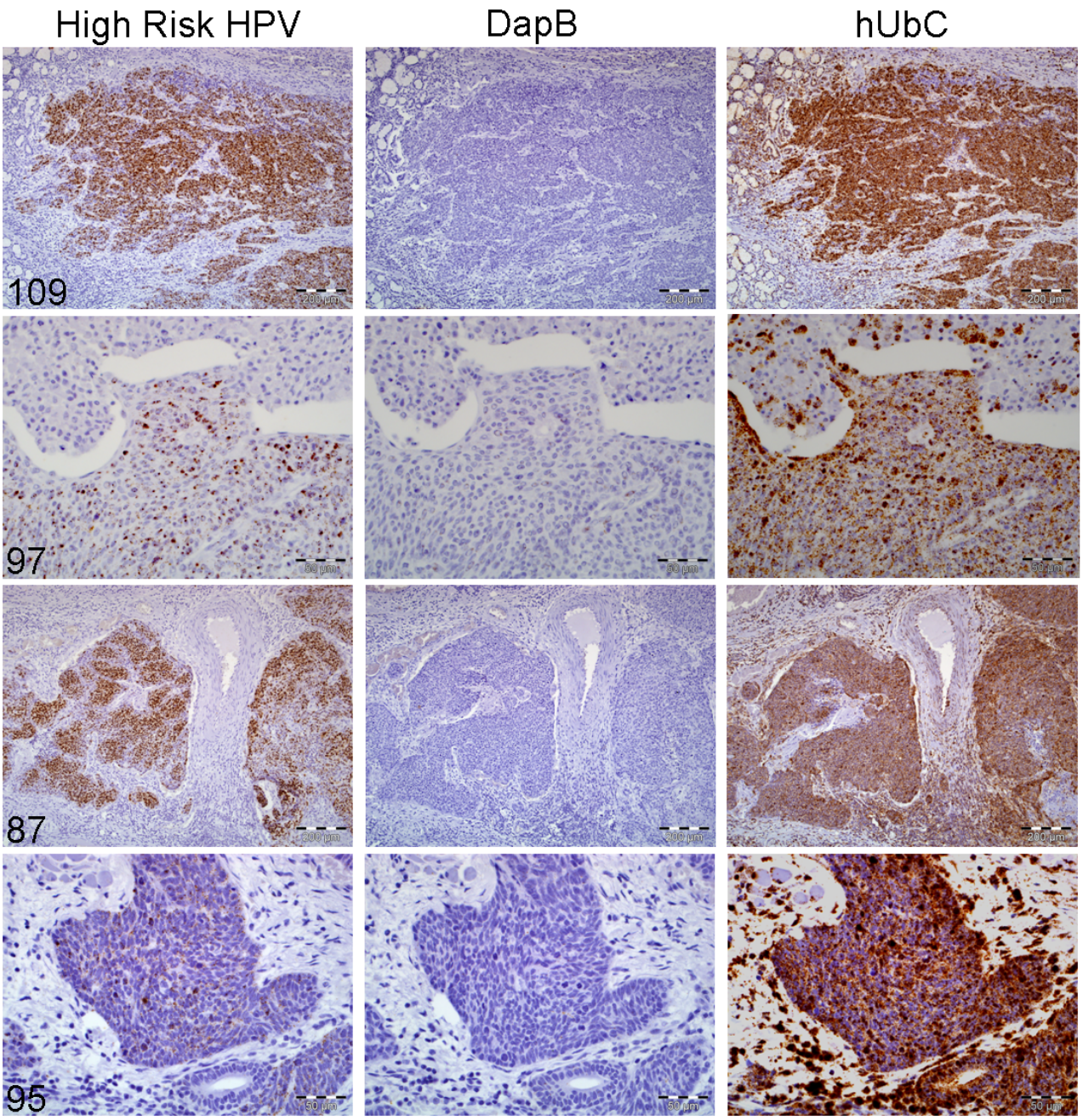
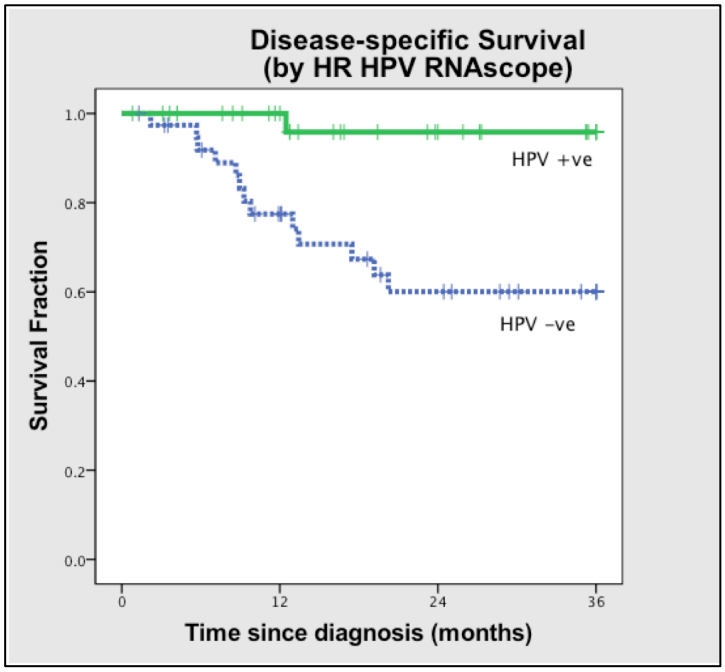


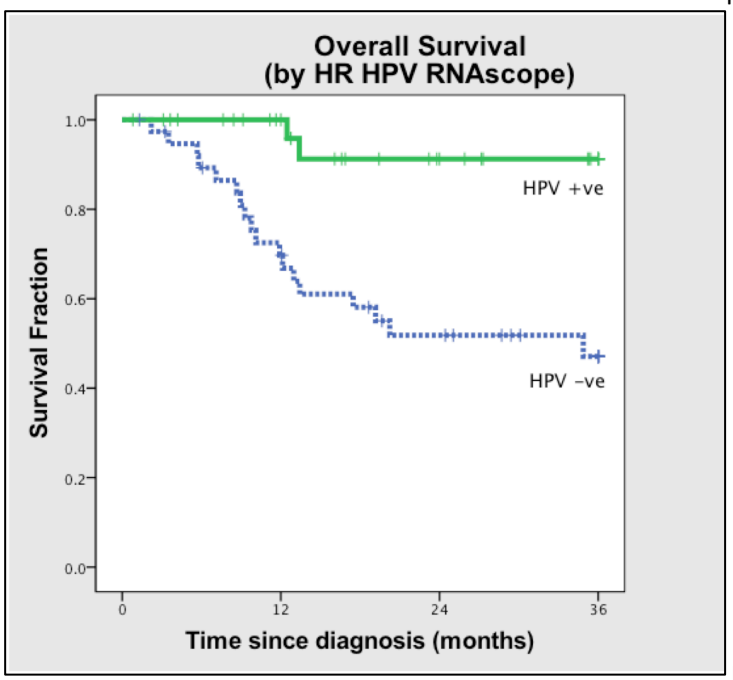
Figure 18: Photomicrographs of OPSCC cases stained using RNAscope

The cases demonstrate a range of positive staining patterns for high-risk HPV. Panels (left to right) represent the test result (HR HPV RNA ISH), negative control (*dapB*) and positive control (*UBC*). Scale bars are equivalent to 200 µm for cases 109 and 87, and 50 µm for cases 97 and 95
Cases 109 and 97 showed strong and moderate staining respectively, and contained HPV-16 E6/E7 mRNA by qRT–PCR. Case 87 showed strong staining and contained HPV-18 E6 mRNA by qRT–PCR. Case 95 (identified as case 101 on Table 27) showed weak staining and was negative for HPV-16 E6/E7 mRNA, HPV-18 E6 mRNA and HPV-33 E6 mRNA by qRT–PCR (false-positive result).

The Kaplan-Meier survival estimates illustrated in Figure 19 and detailed in Table 32 & 33 show the prognostic capacity of all HPV tests. High Risk HPV RNAscope displayed an encouraging capacity to discriminate survival, both in terms of overall survival (p=0.004) and disease specific survival (p=0.017), and this was comparable to the reference test (OS p=0.008, DSS p=0.025).



p=0.017



p=0.004

Figure 19: Kaplan-Meier Survival plots by HPV status as determined by HR HPV RNA ISH (RNAscope) Disease Specific Survival (DSS, above) & Overall Survival (OS, below)

Disease-specific survival		Three-year survival probability	95% CI	Logrank p value	Hazard Ratio	95% CI	p value
HR HPV RNAscope	**HPV -ve** **HPV +ve**	**0.6** **0.96**	**0.45 - 0.8** **0.88 - 1.0**	**0.0011**	**11.9**	**1.6 - 90.8**	**0.017**
HR HPV qRTPCR	**HPV -ve** **HPV +ve**	**0.62** **0.95**	**0.47 - 0.81** **0.87 - 1.0**	**0.0026**	**10.2**	**1.2 - 77.9**	**0.025**
p16 IHC	HPV -ve HPV +ve	0.61 0.9	0.46 - 0.82 0.81 - 1.0	0.0093	4.1	1.1 - 14.6	0.031
HR HPV ISH	HPV -ve HPV +ve	0.63 0.92	0.48 - 0.82 0.82 - 1.0	0.0074	5.2	1.2 - 23.3	0.031
p16 IHC/HR HPV ISH	HPV -ve HPV +ve	0.63 0.92	0.48 - 0.82 0.82 - 1.0	0.0074	5.2	1.2 - 23.3	0.031
DNA qPCR	HPV -ve HPV +ve	0.62 0.92	0.47 - 0.82 0.82 - 1.0	0.0035	5.5	1.2 - 24.6	0.025
p16 IHC/DNA qPCR	HPV -ve HPV +ve	0.62 0.95	0.47 - 0.81 0.87 - 1.0	0.0022	10.6	1.4 - 80.8	0.023

Table 32: Kaplan-Meier survival estimates of Disease Specific Survival with associated Hazard Ratios

Results segregated according to individual diagnostic tests or combination tests.

Overall Survival		Three-year survival probability	95% CI	Logrank p value	Hazard Ratio	95% CI	p value
HR HPV RNAscope	**HPV -ve** **HPV +ve**	**0.47** **0.91**	**0.33 - 0.68** **0.8 0 1.0**	**0.0002**	**8.3**	**1.9 - 35.9**	**0.004**
HR HPV qRTPCR	**HPV -ve** **HPV +ve**	**0.5** **0.91**	**0.35 - 0.7** **0.79 - 1.0**	**0.0007**	**7.1**	**1.7 - 30.8**	**0.008**
p16 IHC	HPV -ve HPV +ve	0.49 0.83	0.34 - 0.72 0.71 - 0.98	0.0044	3.3	1.2 - 9.1	0.02
HR HPV ISH	HPV -ve HPV +ve	0.5 0.88	0.35 - 0.71 0.75 - 1.0	0.0015	4.9	1.4 - 16.8	0.011
p16 IHC/HR HPV ISH	HPV -ve HPV +ve	0.5 0.88	0.35 - 071 0.75 - 1.0	0.0015	4.9	1.4 - 16.8	0.011
DNA qPCR	HPV -ve HPV +ve	0.5 0.85	0.34 - 0.72 0.72 - 1.0	0.0016	2.7	1.2 - 11.1	0.019
p16 IHC/DNA qPCR	HPV -ve HPV +ve	0.49 0.91	0.34 - 0.7 0.79 - 1.0	0.0005	7.4	1.7 - 32.0	0.007

Table 33: Kaplan-Meier survival estimates of Overall Survival with associated Hazard Ratios

Results segregated according to individual diagnostic tests or combination tests.

False positive and false negative reporting

High Risk HPV RNAscope conferred positive results for three cases (4%) where there was an absence of detectable HPV mRNA by qRTPCR. Corresponding test results for these cases indicated that they were also positive by p16 IHC, HR HPV DNA ISH, HPV-16 DNA qPCR, and consequently combinations of these tests.

High Risk HPV RNAscope classified as negative, one case (1%) which displayed evidence of HPV-16 transcripts by qRTPCR. The case was also classified as negative by p16 IHC, HR HPV DNA ISH, HPV-16 DNA qPCR, and consequently combinations of these tests.

Diagnostic Test	Sensitivity	95% CI	Specificity	95% CI	PPV	NPV
	(by comparison to "gold standard" HR-RNA qPCR)					
HR HPV RNAScope	**97%**	84-99%	**93%**	82-98%	**91%**	**98%**
p16 IHC	97%	85-99%	82%	67-91%	80%	97%
HR HPV DNA ISH	94%	80-98%	91%	79-96%	89%	95%
Combined p16/HR HPV DNA ISH	94%	80-98%	91%	79-96%	89%	95%
DNA qPCR	91%	76-97%	87%	74-94%	83%	93%
Combined p16/DNA qPCR	91%	76-97%	93%	82-98%	91%	93%

Table 34: Diagnostic capabilities of individual tests by comparison to HR HPV qPCR

5.3 DISCUSSION & IMPLICATIONS FOR CLINICAL PRACTICE

HPV analysis of OPSCC in clinical practice is swiftly becoming a fundamental requirement. Definitive HPV testing aims to provide both adequate prognostic information for patients and also facilitate entry into appropriately stratified clinical trials, including those investigating the potential to de-escalate the intensity of curative therapies. Paradoxically, there remains no 'international standard' for defining HPV related OPSCC in clinical practice and an adequately validated diagnostic standard for FFPE tissue has yet to be defined [118].

To demonstrate the efficacy of any test it must be appraised against a "reference" or "gold standard" test. As already detailed (2.4), in the context of HPV-driven malignancy, viral oncogenes expression is the prerequisite for carcinogenesis and its detection is therefore the most appropriate analytical standard[125]. Whilst acknowledging that HPV oncogene expression is only part of a complex process of altered molecular pathways in viral-driven cancer, it is against quantitative detection of transcriptionally active virus that the novel HPV test, High Risk HPV RNAscope, was measured.

High Risk HPV RNAscope has previously shown promising capability when compared to other HPV diagnostic tests [213,215] yet validation against an analytical standard had not been possible to date.

As has already been highlighted (2.3), previous evaluations of clinical outcomes in OPSCC based on HPV status have clearly demonstrated the survival advantage for individuals with HPV positive malignancy by comparison to their HPV negative counterparts [50,91,98,100,216,217]. In keeping with other clinically applicable tests, High

Risk HPV RNAscope has been shown to be capable of replicating this prognostic discrimination and indeed demonstrated a similar capacity to predict outcomes to the "gold standard".

It is however the high sensitivity (97%) and specificity (93%) of High Risk HPV RNAscope that offers considerable potential as a diagnostic test for HPV related OPSCC. This carries particular relevance given the fact that this diagnostic discrimination is achieved on FFPE tissue samples in a solo test format. The incorporation of control tests (*UBC* and *dapB*) on parallel sections enhances the quality control of test interpretation purposes.

The only other single test to have demonstrated comparable sensitivity in previous comparison to viral oncogene expression is p16 IHC (94-100%) [125,218]. The level of specificity (79-82%) demonstrated by p16 under the same conditions however, is considerably lower, due mostly to alternative, and as yet unexplained, elevations of p16 expression in HPV negative malignancy [204,215]. Diagnostic algorithms or combination tests (2.5) have been validated and as such are advocated to maximise diagnostic capability. In the context of these results however, it would appear that RNAscope is capable of outperforming combination tests by virtue of comparable specificity whilst displaying superior sensitivity.

Investigation of the potential differences between p16 IHC scoring techniques was undertaken by application of both the currently applied standard for p16 IHC analysis [131], strong and diffuse nuclear and cytoplasmic staining in >70% of the tumour, and the recently described *H* score analysis [133], derived from the cross product of staining intensity and proportion of tumour. Interestingly, both of the

p16 IHC scoring techniques gave identical results and therefore had no apparent bearing on either the sensitivity/specificity of p16 IHC or its prognostic capacity. It is appreciated that the modest cohort size may impact up on the ability to discern significant differences between scoring techniques however it would seem that neither test is capable of improving upon reduced levels of specificity previously reported[125], which remains the limiting factor of p16 IHC[† 219].

High Risk HPV RNAscope classified three cases as HPV positive in the absence of detectable HPV mRNA. It is conceivable that the samples might harbour other High Risk HPV genotypes not included in the reference test, which was restricted to the analysis of the three most common HPV genotypes isolated from OPSCC (HPV-16, -18, -33) or indeed potential heterogeneity for HPV within the tumour as suggested previously by Rietbergen et al[90]. It is interesting to note that one of the three cases had high levels of HPV-16 E2 expression detected by qRTPCR (6.2). E2 is a known transcriptional repressor of E6 and E7 genes, and its influence may have been sufficient to reduce E6 and E7 transcript levels below the detection threshold of qRTPCR whilst remaining within the detection range of High Risk HPV RNAscope. Alternatively, the mismatch between the qRTPCR result and the tests on FFPE raises the possibility of methodological flaws, despite the use of stringent experimental design and detection protocols to quality assure test results. A further possibility to explain discordant results between the novel test and the gold standard would be

[†] In contrast to several previous publications, Schlect et al were able produce specificity of 93% when comparing p16 IHC to fresh tissue derived mRNA from OPSCC. However, this research was undertaken utilising a p16 monoclonal antibody with markedly different performance profile, such that sensitivity was only 56%.

sample allocation errors such that FFPE cores and fresh frozen sourced sample results were for different samples. Attempts were made in stages of sample preparation and assay setup to avoid such an eventuality however it remains a possibility. Although not undertaken, and therefore potentially limiting the interpretation of these results, microsatellite marker analysis could have proven useful to discount such an eventuality.

The solitary HPV positive case reported as negative by High Risk HPV RNAscope, demonstrated an E6 and E7 expression level that was low by comparison with other samples, however it was not the lowest and remained well within the threshold for detection set prior to analysis of the samples. Interestingly, this case was similarly 'misclassified' by both p16 IHC and HR HPV DNA ISH. This therefore raises a possibility that fixation and processing parameters may have resulted in suboptimal preservation of the target molecules, however, given that the FFPE samples were all derived from the same diagnostic service, with storage in exactly the same conditions, this seems unlikely.

These results demonstrate that High Risk HPV RNAscope is capable of maintaining a high degree of accuracy against the most appropriate analytical gold standard and was the best discriminator of disease specific and overall survival. By comparison to the results presented in Chapter 4, RNAscope is capable of performing at a standard comparable to the best combination test p16 IHC/DNA qPCR whilst requiring only a single test which is confined to FFPE tissue resources alone.

Before adoption of HR HPV RNAscope into clinical practice could be formally advocated, this test requires mandatory approval as an in vitro diagnostic device

(IVD), however, the impending application and availability of High Risk HPV RNAscope to a widely available automated staining platform (Ventana Medical Systems Inc, USA) will facilitate standardisation of test conditions and reproducibility between laboratories. These features raise the possibility that High Risk HPV RNAscope could be developed to provide the "clinical standard" for assigning a diagnosis of HPV-related OPSCC.

6 EPIGENETIC REGULATION THROUGH DNA METHYLATION IN HPV MEDIATED OROPHARYNGEAL SQUAMOUS CELL CARCINOMA (OPSCC)

6.1 INTRODUCTION

Analysis of Epigenetic Regulation in HPV positive OPSCC Aims

Whilst virally mediated oncogenesis is relatively well understood, the influence that DNA methylation, and the key regulators of that methylation, have in this process remains unclear.

Viral oncogene expression has been considered to be under the repressive control of the early viral gene E2[220] and it has been believed therefore that loss of E2 expression, through mechanisms such as viral integration[171], is key to halting this repression with consequent potential for oncogenesis.

It was hypothesised that, within the HPV genome, methylation of either the E2 gene or its binding sites (E2BS) may be an alternative means of disruption of E2 gene expression or its downstream effects in the long control region (LCR), that would otherwise be seen following the process of integration. If this were true, E2 methylation would directly influence E2 expression levels and, as a consequence, result in an elevation of levels of the viral oncogenes, E6 and E7, normally under E2 repressive control. A similar resultant increase in E6 and E7 expression would be expected if methylation in the LCR, and more particularly the E2BSs, were to increase through restriction of E2 binding.

The impact of DNA methylation within the host genome was also explored in OPSCC samples. It was postulated that the differences in global DNA hypomethylation, previously reported between HPV positive and negative cells lines would also be evident in clinical tumour specimens. Similarly it was proposed that HPV positive malignancy holds a distinct DNA methylation profile within gene promoter sequences and that this may have biological relevance in terms of oncogenesis. Finally it was hypothesised that distinct variations in DNA methylation within virally mediated tumours are directly influenced by fundamental differences in the key drivers or regulators of methylation, DNA methyltransferases and UHRF1, when comparing HPV positive and negative OPSCC.

6.2 HPV E2 & LONG CONTROL REGION METHYLATION STATUS AND IMPLICATIONS FOR VIRAL ONCOGENE EXPRESSION

Analysis of Epigenetic Regulation in HPV positive OPSCC Methods

Pyrosequencing methylation assays (PMAs) were designed for the viral E2 gene and two CpGs within the long control region (LCR) of the HPV genome to determine methylation status. Within the target sequence of one of the LCR assays (LCR Region 2), lie the E2 binding sites E2BS3 and E2BS4. These sites have previously been shown to directly influence viral oncogene expression (E6 & E7) through E2 protein binding[221].

Pyrosequencing methylation assay (PMA) analysis was undertaken on a cohort of HPV positive OPSCC (n = 43) and the two HPV16 positive cell lines, CaSki and SiHa. qPCR was utilised to quantify viral oncogene RNA expression levels for the HPV-16 early genes E2, E6 and E7. RNA levels were expressed as relative quantification values ($RQ=2^{(-\Delta\Delta Ct)}$) with calibration against the HPV16 positive cervical cancer cell line CaSki. Mean RQ values for each of the three analysed genes were collated with mean CpG percentage methylation results. PMA methylation levels are calculated as an average of the methylation proportions at each individual CpG within the target sequence. The final methylation value reflects the mean of duplicated sample runs consecutively on the pyrosequencing platform. The threshold for scoring hypermethylated samples was conservatively set to 10%, which is higher than our previously established threshold[222,223], in order to ensure the biological relevance of detected methylation.

Non-parametric statistical tests were applied to determine correlation between detected methylation state and viral gene expression.

Epigenetic Regulation in HPV positive OPSCC Results

Analysis of the control cell lines is depicted below. In keeping with a previously described disruption of the viral E2 gene, SiHa showed no evidence of E2 expression and the PMA analysis of E2 methylation failed to amplify target sequence. CaSki expression was the reference against which all other samples were compared, and is therefore further relative expression analysis is precluded. In terms of viral methylation, CaSki showed high levels of methylation in both the E2 gene and also LCR regions. The proportion of tumour samples reaching the threshold for positive methylation is detailed in Table 36 for comparison with control cell lines

Sample	Average Methylation Analysis (PMA, %)			Viral Gene Expression (RNAqPCR)		
	E2	LCR Region 1	LCR Region 3	HPV16 E2 RQ	HPV16 E6 RQ	HPV16 E7 RQ
SiHa	NR	3	12	0	0.905	0.376
CaSki	84	67	94	1	1	1

Table 35: Control Sample Compiled Viral Expression and Methylation Analysis Results

All expression was relative with respect to CaSki (RQ). Viral methylation is recorded as a percentage. Abbreviations; PMA (Pyrosequencing Methylation Assay), NR (No result).

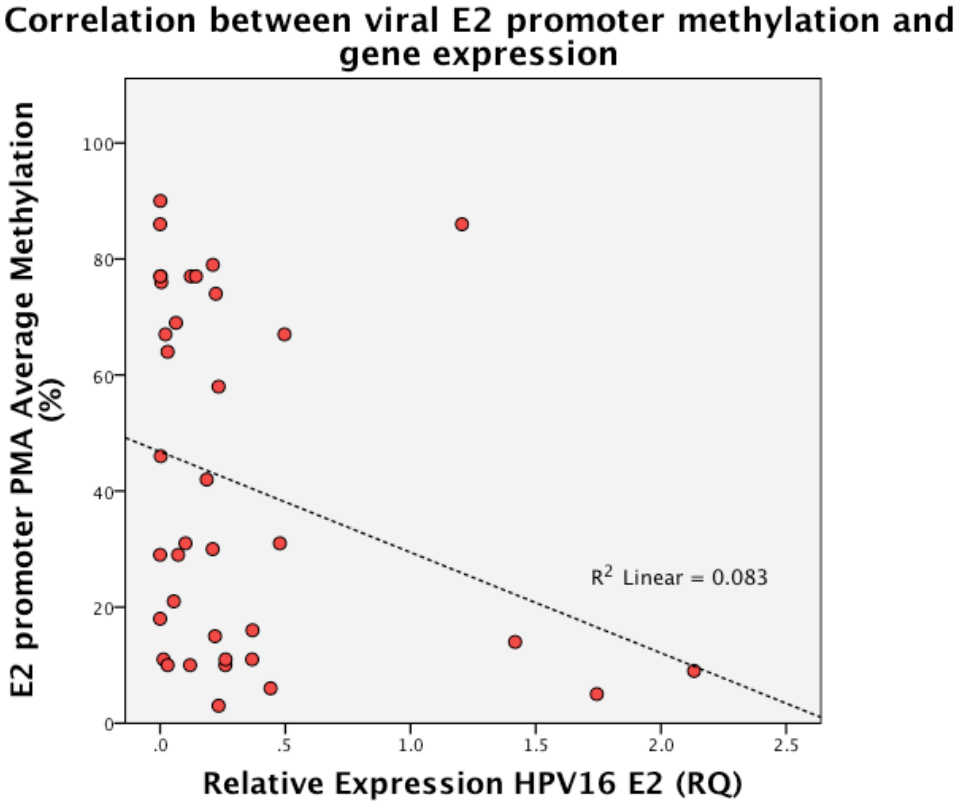
PMA Target	Proportion of Samples Methylation Positive
E2	33/37 (89%)
LCR Region 1	4/40 (10%)
LCR Region 2	11/41 (27%)

Table 36: Proportion of HPV samples displaying positive methylation at viral targets.

Positive methylation threshold set at ≥10%. Proportions refer to samples for which duplicate PMA results were available for analysis.

Within tumour-derived samples, there was a generalised lack of direct relationship between viral genome methylation and expression of key viral genes. Table 37 summarises the apparent relationships. E2 methylation and E2 expression demonstrated the relationship with greatest strength, however the correlation (negative) was only modest (correlation coefficient -0.362). There was no evidence of a relationship between methylation of either of the portions of the LCR (including E2BS3 or E2BS4) and the expression of E6 and E7 genes.

Figure 20 graphically represents the relative correlations between E2 gene methylation and expression (R^2=0.08), and LCR-2 and E6 expression (R^2=0.01) respectively.


Correlation between viral E2 promoter methylation and gene expression
E2 promoter PMA Average Methylation (%)
100
80
60
40
20
0
.0
.5
1.0
1.5
2.0
2.5
R2 Linear = 0.083
Relative Expression HPV16 E2 (RQ)

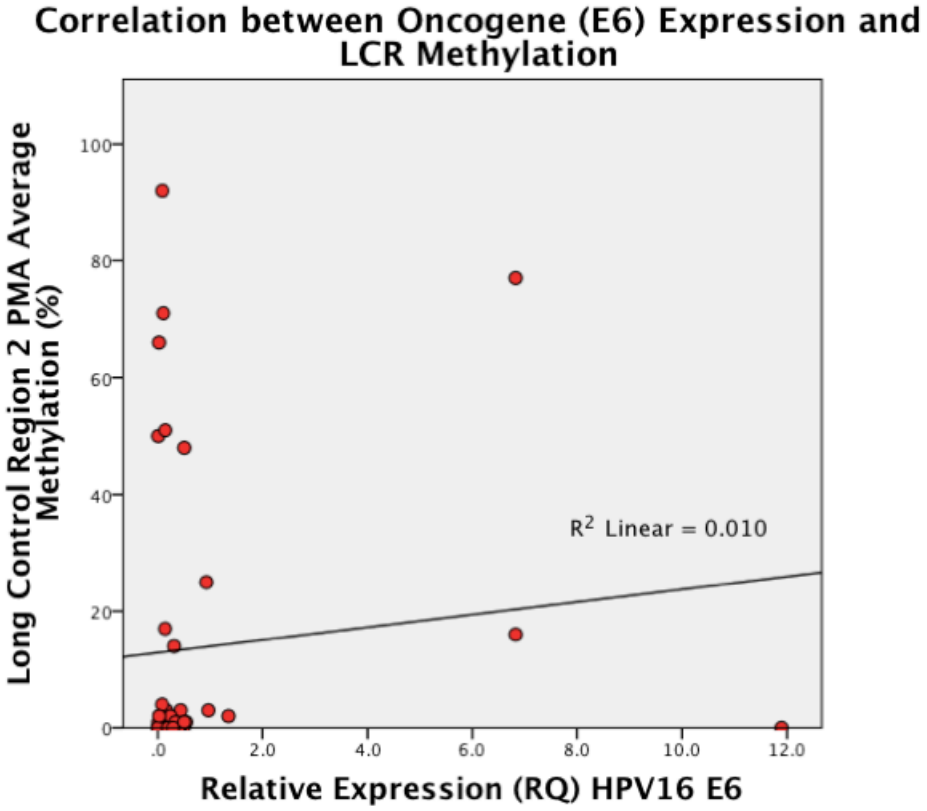

Correlation between Oncogene (E6) Expression and LCR Methylation
Long Control Region 2 PMA Average Methylation (%)
100
80
60
40
20
0
.0
2.0
4.0
6.0
8.0
10.0
12.0
R2 Linear = 0.010
Relative Expression (RQ) HPV16 E6

Figure 20: Correlation between methylation of viral regions and related viral gene expression

Top Scatterplot: Correlation between viral E2 gene promoter methylation and relative E2 gene expression. A negative weak correlation was observed (R^2=0.08)
Bottom Scatterplot: Correlation between viral long control region 2 (LCR) and HPV16 E6 gene expression. Long control region 2 includes the E2 binding sites E2BS3 and E2BS4. No evidence of correlation was apparent (R^2=0.01).

		Viral Gene Expression (RQ)		
		HPV16 E2	HPV16 E6	HPV16 E7
E2 (C^m)	Correlation Coefficient	**-0.362**	**-0.243**	**-0.147**
	Sig. (2-tailed)	**0.03**	**0.153**	**0.393**
	N	36	36	36
LCR Region 1 (C^m)	Correlation Coefficient	0.145	0.059	0.061
	Sig. (2-tailed)	0.385	0.725	0.714
	N	38	38	38
LCR Region 2 (C^m)	Correlation Coefficient	-0.046	0.071	-0.055
	Sig. (2-tailed)	0.780	0.668	0.738
	N	39	39	39

Table 37: Correlation between Viral methylation and Viral Gene Expression

Analysis also sought to determine the relationship between E2 and E6/E7. Whilst levels of E6 and E7 expression showed predictable strong positive correlation (Table 38; r=0.956, p<0.0001), a positive correlation was also seen between E2 expression and E6/E7 expression, contrary to expectation for a known transcriptional repressor (as demonstrated for E2 vs. E6 gene expression in Figure 21.

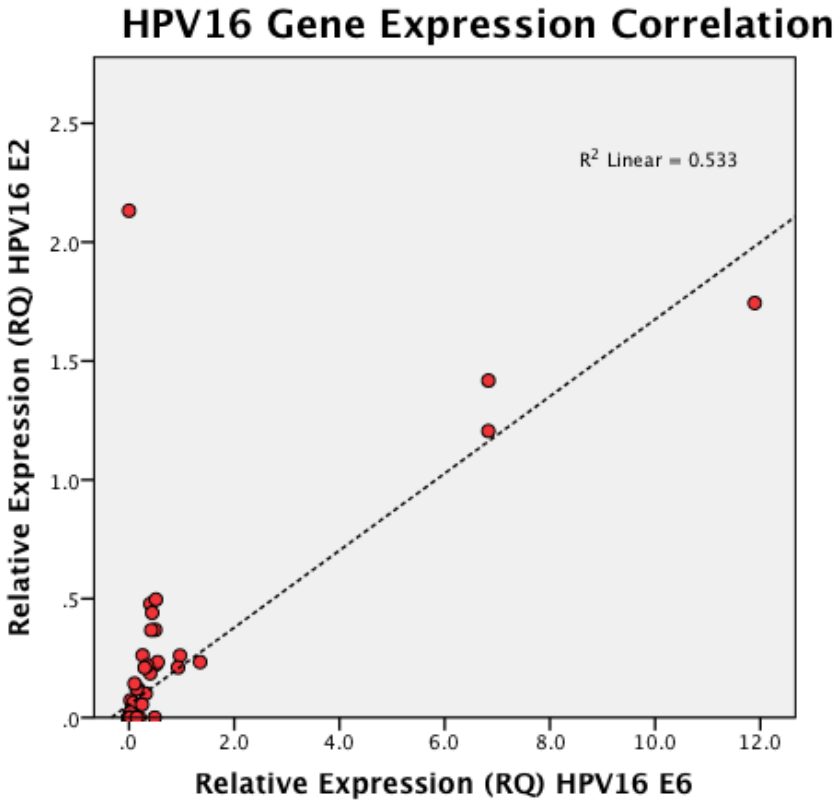


Figure 21: Scatter plot of correlation between HPV16 E2 and E6 gene expression (RQ)

Positive correlation is reflected in R^2=0.533

Target Gene		Viral Gene Expression (RQ)		
		HPV16 E2	HPV16 E6	HPV16 E7
HPV16 E2	Correlation Coefficient		**0.864**	**0.871**
	Sig. (2-tailed)		**<0.0001**	**<0.0001**
	N		103	103
HPV16 E6	Correlation Coefficient	0.864		0.956
	Sig. (2-tailed)	<0.0001		<0.0001
	N	103		103
HPV16 E7	Correlation Coefficient	0.871	0.956	
	Sig. (2-tailed)	<0.0001	<0.0001	
	N	103	103	

Table 38: Correlation between Viral E2 Expression and Viral Oncogene Expression (E6 & E7)

6.3 COMPARATIVE HOST GENOME-WIDE METHYLATION STATE IN HPV POSITIVE AND HPV NEGATIVE OPSCC

Genome-wide Methylation Analysis Methods

To quantify the extent of host genome-wide methylation in OPSCC clinical samples and, where available, their matched normal tissue samples, the LINE-1.2 retro-transposable element was used as a representative sequence region for analysis. Because LINE-1 retrotransponson constitutes a substantial proportion of the human genome (in excess of 17%) its methylation status has been suggested to be a robust and representative reflector of overall global methylation state.[224] LINE-1 methylation status analysis was undertaken using a previously published assay CpG[190] as detailed in (3.13), on a cohort of 65 OPSCC including 34 HPV positive cases and 31 HPV negative cases. In addition, matched normal samples for 25 OPSCC were analysed. A reference methylation score was determined by analysis of the variability of LINE-1 methylation in normal samples analysed simultaneously

Genome-wide Methylation Analysis Results

LINE-1 promoter was highly methylated, with a minimal degree of variability in normal oropharyngeal tissues (average normal sample LINE-1 methylation; 69.6% ± 2.8%, 2sd), however tumour derived samples displayed considerably greater variability in methylation (range 31.2% – 70.9%) and a “more” hypomethylated

LINE-1 promoter (average LINE-1 methylation; 62.0% ± 16%). Using a paired T-test, LINE-1 methylation was demonstrated to be statistically significantly higher in normal tissues when compared to tumours (p=0.008) (Figure 22).

When tumour samples were segregated on the basis of HPV16 status (as determined by RNAqPCR), HPV positive OPSCC had LINE-1 methylation profile, similar to mean normal tissue levels (average LINE-1 methylation; 64.2% ± 10.8%) whilst HPV negative OPSCC displayed significant global hypomethylation (average LINE-1 methylation; 55.3% ± 12.0%, paired t-test p<0.001) (Figure 23).

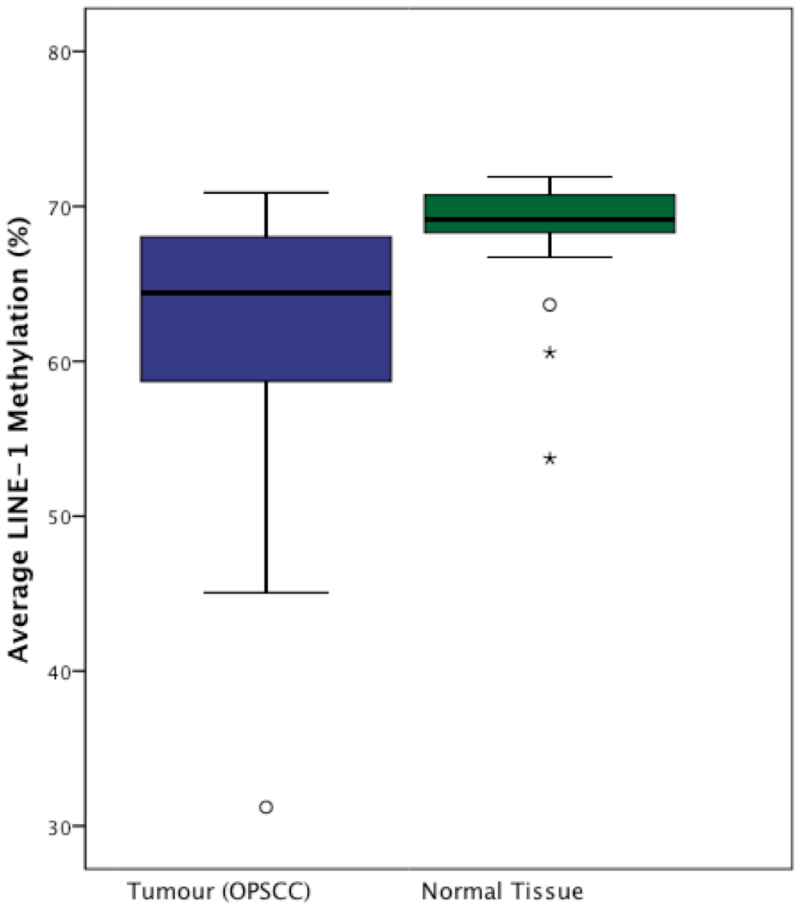


Figure 22: Boxplot representation of Average LINE-1 Methylation values in OPSCC and Adjacent Normal Tissues.

P=0.008

Tumours demonstrate significantly increased degree of hypomethylation. Average values are derived from duplicate sample runs of LINE-1 PMA.

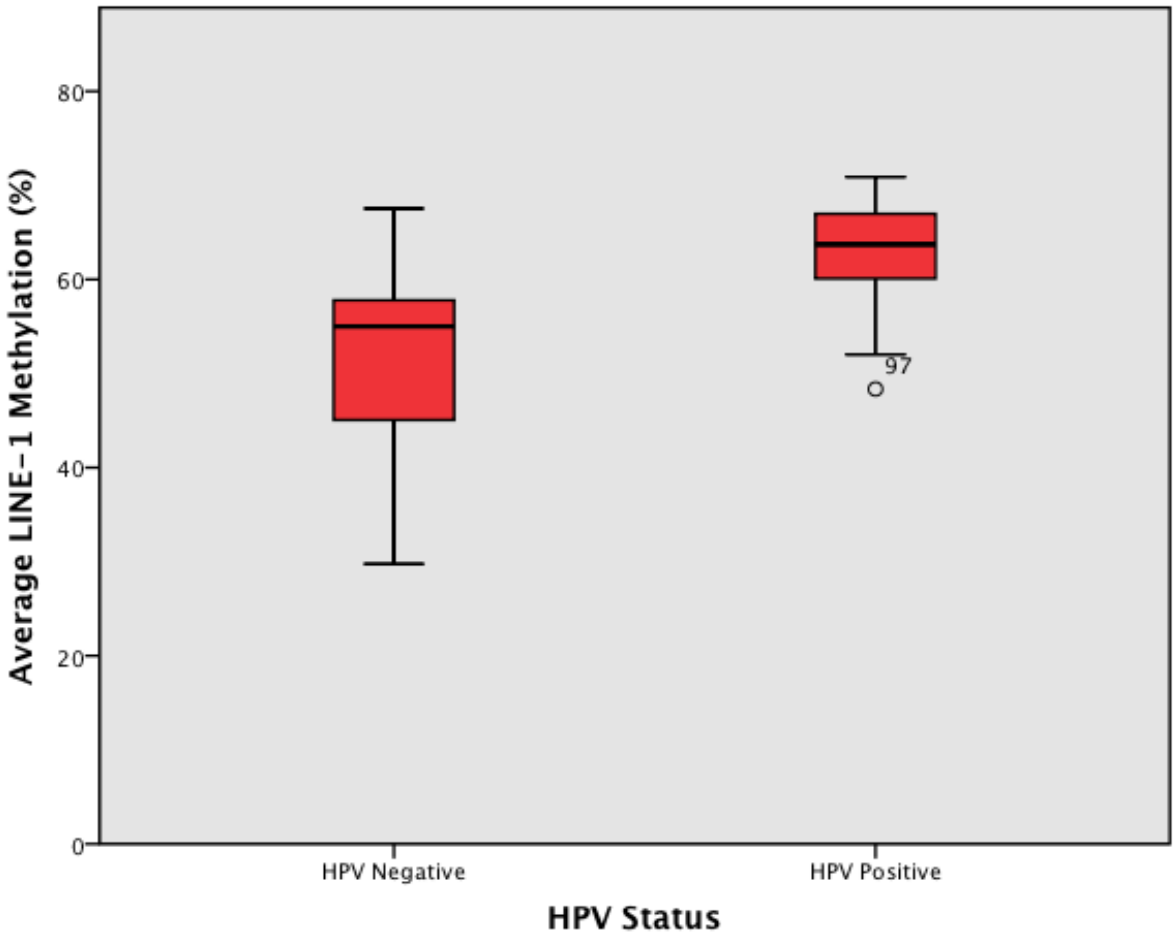


Figure 23: Boxplot representations of Average LINE-1 Methylation values in OPSCC stratified by HPV status.

p<0.0001

HPV positive tumours demonstrate significantly increased global hypomethylation (p<0.0001). Average values are derived from duplicate sample runs of LINE-1 PMA.

Exposure to tobacco smoking has previously been shown to strongly influence levels of global hypomethylation in head and neck cancers[225], therefore a further analysis of OPSCC tumour samples was undertaken to exclude smoking or any other demographic/clinicopathologcial feature as confounding factors. No significant association, however, was found between either smoking history or any other demographic/clinicopathological feature (not shown).

6.4 DETECTION OF DIFFERENTIAL HOST GENE PROMOTER METHYLATION IN HPV POSTIVE & HPV NEGATIVE OPSCC

Detection of Promoter Methylation State in OPSCC: Methods

A cohort of 24 OPSCC cases, not previously examined, was analysed using the Infinium HumanMethylation450 Beadarray previously detailed in Chapter 3 (3.14). An initial shortlist of potential candidate genes, differentially methylated in OPSCC on the basis of HPV status, was established before pyrosequencing methylation assay design and technical validation. Candidate genes were selected on the basis of statistical and pragmatic means as detailed previously (3.14).

Host Tumour Suppressor Gene promoter sequence hypermethylation in HPV Positive & HPV Negative OPSCC: Results

Differential promoter hypo- and hypermethylation was apparent in the methylation array raw results following initial statistical analysis. Following ranking of promoter methylation variable positions stratified by HPV status, 31 genes with apparent hypermethylation and 13 genes differentially hypomethylated in HPV positive samples were highlighted. A subsequent screen to exclude candidate genes with less than 4 probes per gene promoter confirmed a shortlist of 14 genes, one of which was hypomethylated and the remainder hypermethylated in HPV positive disease.

From this initial shortlist, feasible pyrosequencing methylation assays (PMA) design was possible for eight genes. An explanation of the current evidence for potential

roles of all genes meeting the inclusion criteria, including the eight genes suitable for PMA validation, is presented in Table 39.

Three genes included (C12 orf42, SYN2 and FLJ26850) have not previously been reported to have a role or potential role in either malignancy or virally mediated disease.

Target Gene	Evidence for role in viral infection or oncogenesis	Ref
	PROMOTER HYPOMETHYLATION IN HPV POSITIVE OPSCC	
DERL-3	Required for degredation of misfolded proteins, upregulated by unfolded protein response. Prolonged activation of UPR implicated in sensitivity to platinum chemotherapy	Ma et al.[226]
	PROMOTER HYPERMETHYLATION IN HPV POSITIVE OPSCC	
CTNND2	Delta-Catenin (CTNND2) is a cadherin-associated protein involved in cell adhesion. Overexpressed in cancer and associated with progression in malignancy	Lu et al.[227]
GALR1	Epigenetic inactivation evidence in HNSCC suggesting a role as a TSG	Misawa et al[228]
	Putative regulator of resistance to chemotherapy in Colorectal cancer. Silencing induced apoptosis and synergistically enhanced effects of 5-FU or oxaliplatin	Stevenson et al[229]
C12 orf42	-	-
HOXA7	Frequently methylation target in early stage lung cancer. Putative diagnostic biomarker in lung cancer.	Rauch et al[230]
	Associated with a malignant phenotype in meningioma. Putative diagnostic biomarker for malignant behavior	Di Vinci et al[231]
FLJ26850	-	-
SYN2	-	-
KCNA1	Identified as key regulator of oncogene-induced senescence (OIS). Down regulation induces OIS escape.	Lallet-Daher et al[232]
LMO3	Transcription factor that interacts with p53 to regulate function. Hypermethylation and reduced expression in lung cancer	Kwon 2012 et al.[233]
MYOCD	Recognised promoter of smooth muscle cellular differentiation, overexpressed in sarcomas.	Perot et al.[234]

SLCo4C1	Organic cation/anion transporters known to affect platinum uptake and clearance. Proposed association in HNSCC	Ziliak et al[235]
STK32B	Gene product necessary for KRAS-dependent cell lines. Proposed target for small molecule inhibitors.	Babij et al.[236]
RPS6KA2	Putative tumour suppressor gene in Ovarian Malignancy	Bignone et al.[237]
CCNA1	Component of PI3 kinase pathway. Improved disease-free survival in HNSCC	Tan et al[238]
	Evidence of hypermethylation in HNSCC using other methylation array platforms – Illumina GoldenGate & illumina infinium HumanMethylation27 beadarray	Jithesh et al[239] Sator[163]

Table 39: Evidence for role of Target Genes in Oncogenesis (HPV related or otherwise)

Highlighted genes (red) were suitable for custom PMA design and subsequent validation

Analysis of HPV positive methylation controls provided internal validation of the pyrosequencing assays employed. A representative plot, with line of best fit and correlation coefficient (C12orf42), is demonstrated in Figure 24 and the complete validation gene panel correlation coefficients are detailed in Table 40.

The correlation coefficients were such that it was concluded that detection of methylation values by PMA was valid for each of the respective assays.

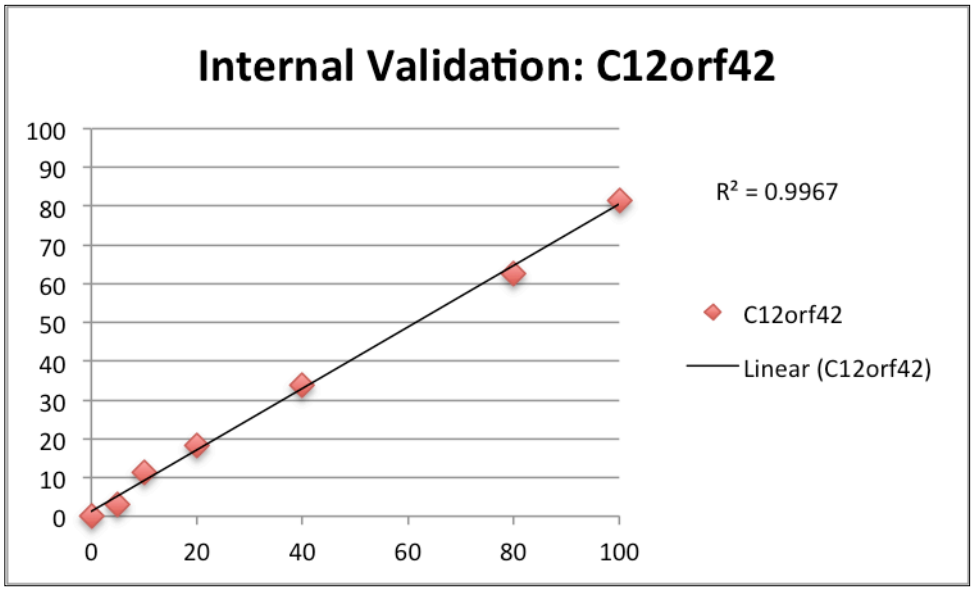


Figure 24: Representation Internal Validation Plot for C12orf42 Pyrosequencing Methylation Assay. Detection of artificially methylated DNA control samples

X axis: Artificially Methylated DNA Standards (%); y axis: PMA average methylation detection (%). Line of best fit and correlation coefficient (R^2).

Target Gene	Correlation coefficient (R^2=)
C12orf42	0.9967
CCNA1	0.97106
FLJ26850	0.97228
GARL1	0.98187
HOXA7	0.99855
KCNA1	0.94874
SLCO4Cl	0.99051
SYN2	0.98823

Table 40: PMA Internal validation analysis - Correlation Coefficients for Individual PMAs using methylated control DNA samples.

Raw data compiled from average gene probe methylation from the Infinium array is tabulated adjacent to technical validation results from individual gene PMA analysis (**Figure 25**). Results are conditionally formatted to highlight the differential methylation between HPV negative tumours (above table division) and HPV positive tumours (below table division). The table demonstrates the hypermethylation of target gene promoters in HPV positive tumour samples by comparison to the HPV negative samples.

Additionally, comparison of methylation values for each test was performed using Spearman's correlation test for non-parametric values. A positive correlation was seen for each of the gene assays as listed in Table 41, albeit the strength of correlation was only convincing for C12 orf42. Remaining correlations were weak to moderate (seven gene targets).

Graphic representations of correlation scatter plots for two of the best correlating target genes (FLJ26850 and C12 orf42) are made in Figure 26.

		C12 orf42		CCNA1		FLJ26850		GALR1		HOXA7		KCNA1		SLCO4CI		SYN2		
Sample ID	HPV16 Status	Infinium Array	PMA Validation	Infinium Array	PMA Validation	Infinium Array	PMA Validation	Infinium Array	PMA Validation	Infinium Array	PMA Validation	Infinium Array	PMA Validation	Infinium Array	PMA Validation	Infinium Array	PMA Validation	
3	0	6	9	19	3	11	21	16	5	7	3	10	4	10	1	13	21	
4	0	8	0	14	7	12	7	31	0	10	1	8	12	7	2	23		
5	0	10	1	10	8	17	9	7	5	59	5	13	3	8	1	49	20	
6	0	7	19	11	28	14	4	21	60	9	32	8		36	1	22		
8	0	9	23	24	6	18	6	29	19	13	3	11	11	9	1	29	27	
9	0	45	46	84	50	38	4	72	0	61	2	5	0	72	0	74	27	
10	0	17	26	12	32	28	35	11	62	8	8	7		6		23	2	
11	0	7	0	12	6	13	6	15	21	8	5	7	2	6	0	17	18	
12	0	7	0	13	13	12	8	25	21	44	4	8	5	7	0	20	18	
14	0	7	0	13	3	19	8	20	11	10	3	12	10	7	22	25	14	
15	0	5	0	7	2	9	3	8	0	6	1	5	0	5	1	7	0	
16	0	35	36	44	68	12	25	32	72	33	50	9	0	37	66	20	65	
17	0	31	1	56	2	51	11	58	0	14	51	13	4	68	2	9	32	
18	0	40	0	14	5	15	5	29	10	12	31	12	3	8	1	38	9	
19	0	16	6	18	9	27	29	18	7	7	2	5	1	7	2	15	8	
20	0	7	0	16	3	12	8	17	11	33	2	7	4	34	0	15	11	
23	0	10	1	59	2	24	7	34	6	9	2	35	1	49	0	50	9	Legend
24	0	60	2	77	47	11	15	6	21	6	3	5		5	35	6	8	0
1	1	54	26	56	38	47	34	60	41	58	29	48		59	32	56	47	10
2	1	41		41	60	26	71	41		37	69	31		35	67	39	46	20
7	1	51	33	76	41	73	30	52	62	72	52	53		72	53	79	47	40
13	1	43	17	54	25	46	16	48	28	44	20	31		43	20	52	26	60
21	1	44	34	50	27	24	35	32	75	33	0	33	47	42	43	43	45	80
22	1	42	38	47	48	41	27	60	23	34	48	27		50	35	40	45	100

Figure 25: Comparative Methylation Analysis: Average Methylation in promoter sequence of eight differentially methylated genes.
Technical Validation Raw Data comparing Infinium HumanMethylation450 Beadchip Array data and Pyrosequencing Methylation Assay results for 8 gene promoters determined to be differentially methylated on the basis of HPV status. Individual values reflect the average methylation (either between probes for Array data or between duplicate runs for PMA. Samples 1, 2, 7, 13, 21 & 22 are HPV positive whilst the remaining cases are HPV negative. Conditional formatting has been applied to highlight differences in average methylation (Legend: far right)

	Correlation between Infinium (450K) Array and PMA Methylation Results							
	C12orf42	**CCNA1**	**FLJ26850**	**GARL1**	**HOXA7**	**KCNA1**	**SLCO4Cl**	**SYN2**
Correlation Coefficient	0.649	0.353	0.442	0.125	0.563	0.445	0.249	0.484
Sig. (2-tailed)	0.001	0.091	0.031	0.570	0.081	0.084	0.251	0.022
N	23	24	24	23	24	16	23	22

Table 41: Spearman's Correlation Coefficients for Methylation Status (%) between Infinium HumanMethylation450 BeadChip Array and Individual Gene Pyrosequencing Assay Analysis by target gene.

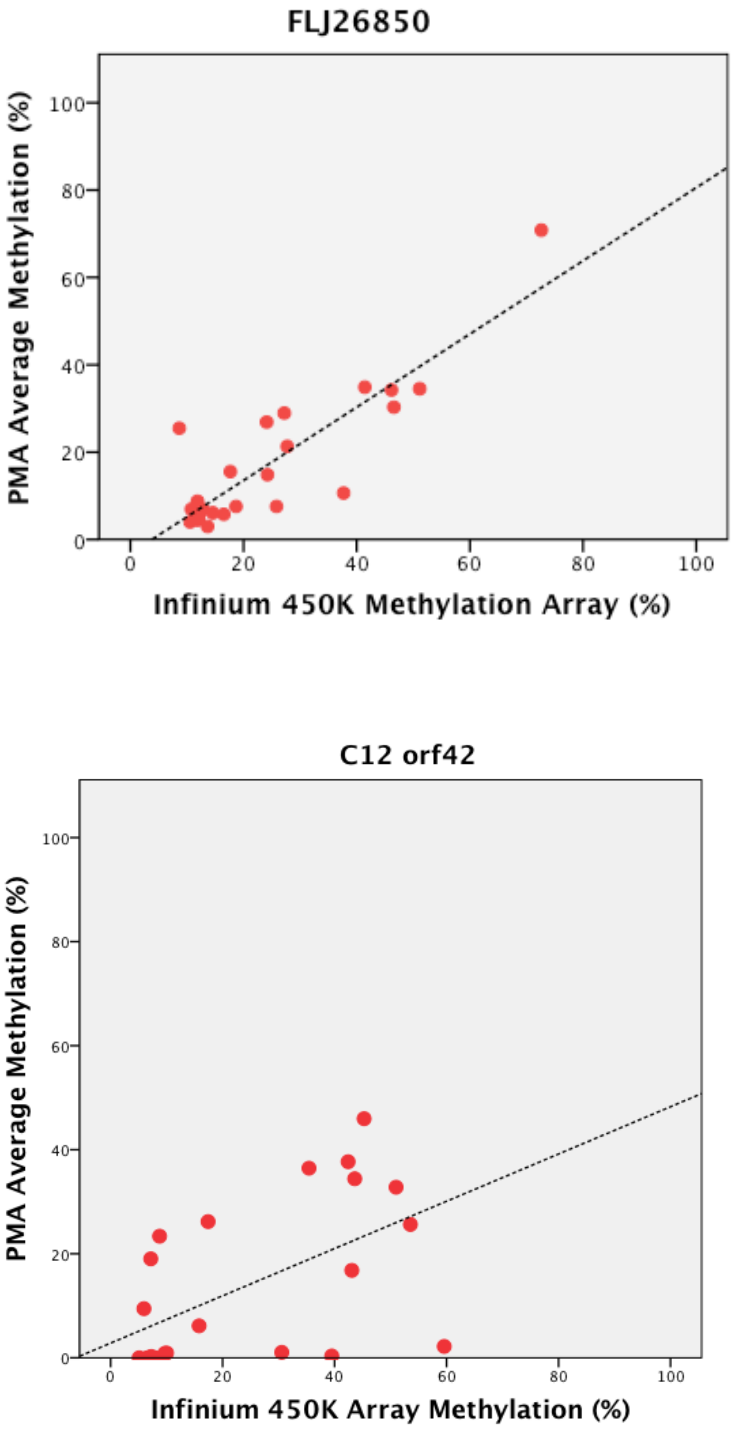


Figure 26: Scatter plots depicting correlation between Infinium Array methylation results and Pyrosequencing Methylation Array average methylation values.

FLJ26850 (above) & C12 orf42 (below) demonstrated correlation coefficients of 0.442 (p=0.031) and 0.649 (p=0.001) respectively (line of best fit identified, dotted line). Methylation values represent the average of contributory probe on the Infinium array for each gene and the average values of duplicate PMA runs.

6.5 EXPRESSION OF KEY REGULATORS OF METHYLATION STATE IN HPV POSITIVE HNSCC

To determine the significance of the regulators of DNA methylation in OPSCC stratified by HPV status, the expression of DNMTs and UHRF was determined within the cohort of oropharynx SCC samples (n= 103), and, where available, their corresponding matched normal pairs (n=53)(Table 42).

HPV status for each case had already been established using HPV mRNA (E6 & E7) expression as detailed previously (3.11).

It was hypothesized that different expression levels would be apparent depending upon HPV status and this would in turn correlate with a global methylation state that was fundamentally different.

	Tumour Cohort		**T/N Pairs**	
Tumour Site	Number	HPV +ve proportion	Matched Normal	HPV +ve proportion
Oropharynx	103	46/103 (45%)	53	11/53 (21%)

Table 42: Regulation of DNA Methylation Analysis Cohort

The mRNA expression levels of DNMTs (1, 3a & 3b) and UHRF1 mRNA were determined for each sample using quantitative PCR (qPCR). In addition, for a proportion of cases, matched normal sample derived RNA was available for concurrent analysis. Normalisation of expression levels was conducted by reference to the average expression level from the pooled normal sample cohort (calibrator) to determine a relative expression level. This relative expression level is expressed

as a relative quantification value (RQ). Statistical comparison of expression levels between tumours and their adjacent normal tissue samples was undertaken using a non-parametric test for paired analysis (Wilcoxon's).

DNA Methyltransferase & UHRF1 Expression Results

Variation in expression of DNMTs or UHRF1 between tumours and adjacent normal tissue within the 53 paired samples demonstrated a highly significant increased UHRF1 expression within tumours (p<0.0001) (Figure 27) and similarly increased levels of DNMT3b expression (p=0.01).

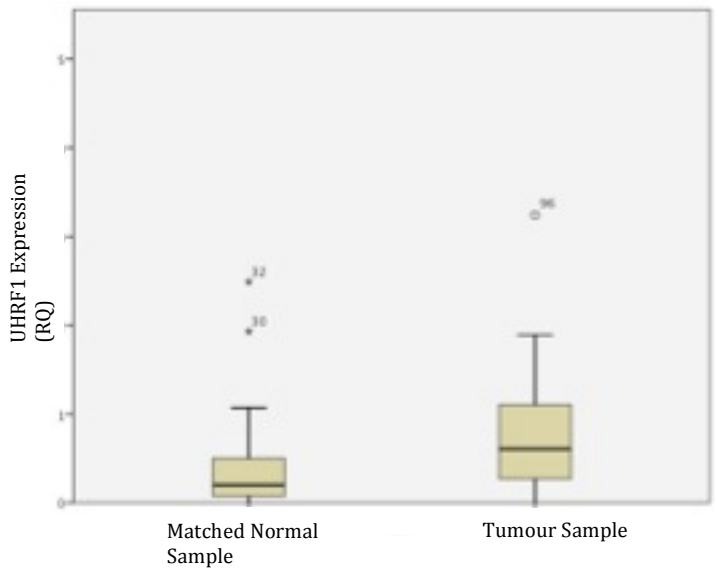


Figure 27: Comparative expression for UHRF1 between paired OPSCC tumour and adjacent normal tissues.

Outliers (normal =1, tumours =5) excluded from above figure.

A bivariate analysis of the detected expression levels in the cohort of 103 tumour samples alone demonstrated moderate correlation between DNMT1 and UHRF1, and to a lesser extent between DNMT1 and DNMT3a (Table 43).

Target Gene	DNMT1	DNMT3A	DNMT3B
UHRF1	0.731*	0.467	0.471
	<0.001§	<0.001	<0.001
DNMT1		0.660	0.457
		<0.001	<0.001
DNMT3A			0.295
			0.002
DNMT3B			

Table 43: Correlation in levels of expression of DNMTs/UHRF1 within OPSCC tumour samples.

* Spearman's Correlation Coefficient; § p value

Further analysis of correlation in DNMT and UHRF1 expression within tumours was made following stratification on the basis of HPV status. HPV negative tumours displayed a greater degree of correlation for each of the four genes than was seen for the whole cohort analysis. This was not the case for HPV positive tumours where correlations were consistently weaker than for the whole cohort (Table 44).

Target Gene	DNMT1		DNMT3A		DNMT3B	
	HPV Negative	HPV Positive	HPV Negative	HPV Positive	HPV Negative	HPV Positive
UHRF1	0.840*	0.609	0.615	0.245	0.508	0.464
	<0.001§	<0.001	<0.001	0.101	<0.001	0.001
DNMT1			0.709	0.533	0.484	0.474
			<0.001	<0.001	<0.001	0.001
DNMT3A					0.285	0.296
					0.032	0.048
DNMT3B						

Table 44: Correlation in levels of expression of DNMTs/UHRF1 within OPSCC tumour samples stratified by HPV status.

* Spearman's Correlation Coefficient; § p value

When comparison was made of independent gene expression levels within OPSCC tumours stratified on the basis of HPV status, DNMT3b expression was significantly lower in HPV positive tumours (Mann-Whitney test, p=0.007). Although a trend towards lower expression of the other analysed genes (DNMT1, -3a and UHRF1) in HPV positive tumours was apparent, evidence of altered expression on the basis of HPV status was not statistically significant.

Further analysis of related clinical outcomes measures for individuals cases on the basis of DNMT3b expression demonstrated that, at 36 months following diagnosis, both disease specific and overall survival were significantly worse for individuals whose tumours overexpressed DNMT3b (DSS; p=0.028, OS p=0.04)(Figure 28).

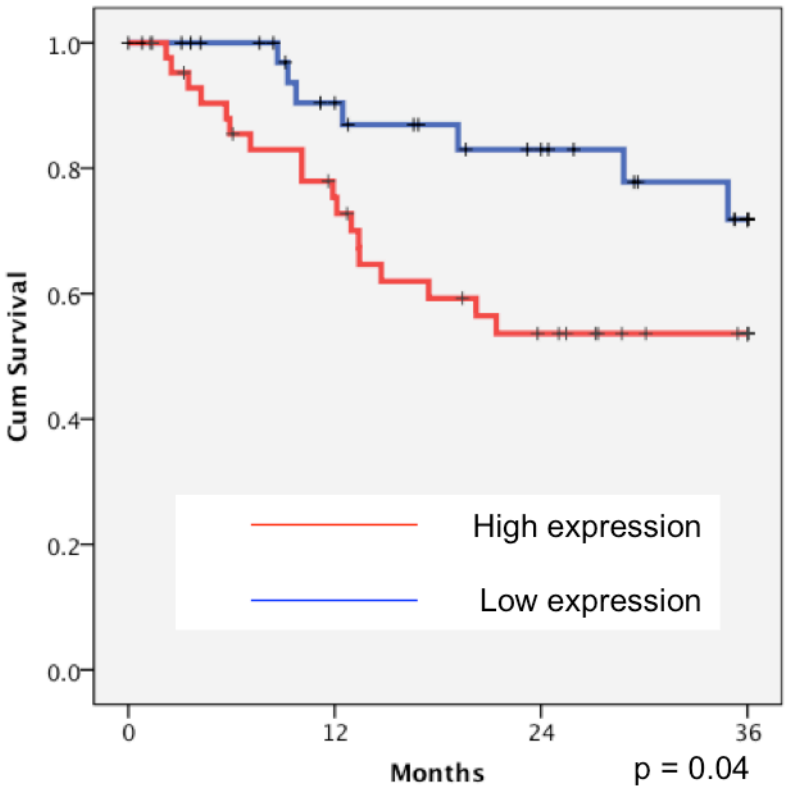


Figure 28: Kaplan Meier Estimates of Overall Survival at 36 Months on the basis of DNMT3b expression.

DNMT3b expression was classified as either High or Low with respect to the median DNMT3b RQ value (RQ=0.65). High expression levels correlated with a significantly worse overall survival (p=0.04).

HPV lacks intrinsic viral DNA methylation machinery, therefore it follows that induction of altered methylation state requires recruitment or opportunistic utilisation of the host methylation machinery. To determine if levels of DNMTs or UHRF1 expression influenced viral methylation state (E2 promoter or LCR methylation) non-parametric analysis was undertaken. Analysis showed no evidence of any correlation between the genes responsible for DNA methylation and the methylation state of the individual viral targets included.

Similarly, analysis of correlations between DNMTs/UHRF1 expression and global methylation within tumours stratified by HPV status was made using a Mann-Whitney test however there was no statistical correlation apparent.

6.6 DISCUSSION

Viral Methylome

The understanding that the HPV transcriptional repressor gene, E2, plays a restraining influence in oncogenesis has been widely held, particularly in consideration of HPV mediated cervical malignancy[240]. On this basis, the immediate implication of its disruption by an integration process is release of the viral oncogenes (E6 and E7) from transcriptional repression. Such a concept fails to explain the reported instances of tumours harboring intact virus in an episomal form presumably without integration[5].

Support for the concept of E2 disruption playing a fundamental role in HPV-mediated oncogenesis comes from *in vitro* evidence of the growth inhibitory effect of a reintroduction of E2 gene in HeLa (HPV18 positive) cells with a coincident fall in E6 and E7 expression[111,220]. In HPV positive OPSCC, integration has been described as a critical and necessary step[78] yet analyses of viral state in OPSCC suggests the proportion of integrated cases may be as low as 50%[179], leading to the possibility of other means to bring about E2 disruption; for that reason, the role of viral methylation in both the E2 promoter and LCR was analysed.

Pyrosequencing assays proved to be a successful technique for the detection of the methylation levels of both the E2 gene (37/43, 86%) and two distinct regions within

the LCR (40/43, 93%; 41/43, 95%). Although there was a trend towards negative correlation between E2 gene promoter methylation and expression, it was only modest (correlation coefficient = -0.362, p=0.03) in nature. Interestingly however, was the finding that viral oncogene expression (E6 & E7) appeared not to be influenced by LCR methylation and, further, that it showed a strong positive correlation to E2 expression (E6; 0.864, p<0.001 & E7; 0.871, p<0.001). This would suggest that in established cases of HPV positive OPSCC, E2 is either a positive driving force for oncogene expression or, perhaps more plausibly, that E6 and E7 expression occurs autonomous to E2 expression, and perhaps elevated levels of E2 expression are an unsuccessful negative feedback mechanism and that a separate event, such as E2 binding site blockade, negates its repressive influence.
Also of note in this analysis, is the frequency with which E2 expression was detected in HPV positive OPSCC (35/43, 81%). The results from the SiHa cell line demonstrate a complete loss of detectable E2 expression in keeping with previous findings and a well-documented disruption to the E2 open reading frame[194]. Although evidence of E2 expression does not necessary exclude viral integration, it does imply the E2 ORF is intact in each of these cases, whether it be integrated, episomal or a combination of both forms.

As mentioned above, the assays employed have demonstrated excellent capability to detect viral methylation state. In support of the capability of the PMA analysis is the comparable viral methylome results for CaSki when considered alongside those reported by previously[241,242]. Using PMA analysis our results showed high levels of methylation detected at all sites for CaSki, but particularly in the E2 promoter and

the LCR region 2 (immediately adjacent to the p97 promoter). Using bisulphite sequencing of the entire HPV16 methylome in CaSki samples, both Fernandez et al.[242] and Park et al.[241] gave an equivalent depiction at these sites ($E2^{m}$>50% & LCR>80%[242] and LCR 100%[241] respectively). Fernandez et al. went on to demonstrate that in 75% (6/8) of cervical cancer samples, methylation specific PCR demonstrated the region encompassing E2BS3 and E2BS4 to be methylated. By comparison, and from a larger series of clinical samples, our results showed 11/41 (27%) of HPV positive OPSCC to be definitively methylated (average methylation greater than 10%). This would add some weight to a theory of induced E2 binding inhibition due to CpG methylation and, therefore, an alternative to E2 disruption through integration. However, our analysis failed to demonstrate any correlation between CpG methylation encompassing E2BS3 and E2BS4 and E2 expression.

Without a means to accurately determine the integration state of analysed target sequences, it remains unclear whether integration is a prerequisite step, as suggested by Park et al.[241], in rendering the viral genome susceptible to the acquisition of an altered methylation profile in sites such as the E2BS. In addition to a validated quantifiable integration test, it would also be necessary to source clinical samples derived from individuals “at risk” of progression from transient HPV infection to transformative infection; as yet, such a cohort remains difficult to define let alone recruit.

Genome-wide Host Methylation State

The characterisation of genome-wide methylation state by LINE-1 PMA has been described previously in several tumour types, including HNSCC[144,148,163,190]. It is apparent that genome-wide hypomethylation is a hallmark of all human cancers[225], yet when considering differences in HNSCC global DNA methylation, both *in vitro*[163] and *in vivo*[144] analyses have demonstrated that HPV negative tumours display more frequent genome wide hypomethylation than HPV positive counterparts. Our findings in a larger cohort of OPSCC clearly support this, with demonstration of increased hypomethylation of tumours of the oropharynx as a whole group when compared with adjacent uninvolved margins. Additionally it was seen that stratification on the basis of HPV status highlighted HPV positive tumours to have global methylation levels more comparable to normal than HPV negative disease (69.6% vs 64.2% vs 55.3% for normal, HPV positive and HPV negative respectively). Richards et al[144]. reported similar findings with respect to maintenance of a more normal LINE-1 methylation profile in HPV positive tumours and raised the question if the this reflects an exuberate cellular methylation response to HPV infection. In terms of evidence of altered expression of the key genes for DNA methylation regulation (DNMTs and UHRF1) in concert with correlating changes in LINE-1 methylation, our data was unable to provide further support for this theory.

It has been surmised that the raised global methylation levels in HPV positive disease are a result of host defense mechanism attempting to silence viral gene expression yet mechanistic studies exploring this have not been forthcoming[144]. Given the statistically significant reduction in smoking history amongst individuals with HPV positive OPSCC and previous evidence of its impact on global methylation state[225], we explored smoking exposure as alternative explanation for reduced hypomethylation in HPV positive tumours in an attempt to exclude its confounding effect. There was no evidence to support smoking or any other demographic or pathological feature as confounding variables in LINE-1 methylation analysis.

Host Gene Promoter Methylation in OPSCC

Methylation microarray analyses depicting differences in promoter methylation between HPV positive and negative clinical and cell line derived samples have been conducted using several platforms previously, with increasing depth as the technology has evolved. Applying the Illumina GoldenGate Methylation Array (1505 CpG loci) to a series of Oral SCC (OSCC) Jithesh et al.[239] found significant hypermethylation for HPV positive lesions across nine probes, two of which (CCNA1 and CTSL1) were also highlighted in an analysis of HPV positive and negative OPSCC using the Infinium HumanMethylation27 Beadchip array (over 27,000 CpGs across almost 15,000 genes)[163].

Interestingly, using a third generation methylation microarray, Infinium HumanMethylation450 Beadchip array, CCNA1 (cyclin-A1) was also identified as being hypermethylated in HPV positive OPSCC.

In a further cohort of HNSCC analysed with the GoldenGate array, CCNA1 was once more highlighted as being hypermethylated in HPV positive cases. In this instance the findings were also validated using a PMA assay providing corroboration[243].

CCNA1 is an important promoting modulator of the cyclin-dependent kinase pathway and as such play an integral part in cell cycle progression from S phase to G2.

It is interesting to speculate on the role that hypermethylation of CCNA1 might play in HPV positive OPSCC. In contrast to HPV negative OPSCC, the majority of HPV postive tumours display a p53 wild-type profile[244]. It is plausible in HPV positive disease that epigenetic downregulation of CCNA1 may be a necessary step in viral carcinogenesis, which is not readily apparent in HPV negative HNSCC.

Counterintuitive to such a premise, is evidence suggesting that HPV positive HNSCC consistently displays CCNA1 protein overexpression[245]. It remains conceivable that these two events; CCNA-1 hypermethylation and CCNA-1 protein overexpression, are not mutually exclusive events and that protein overexpression facilitates cell cycle deregulation whilst virally induced promoter methylation abrogates tumour suppressive influences of the gene.

GARL1, G-protein-coupled Galanin Receptor-1, is also a gene of particular interest in HNSCC and more particularly HPV positive malignancy. As a group, G-protein-coupled-receptors (GPCRs) have emerged as key regulators of tumour growth and

metastasis[246], more specifically, GARL1 has been highlighted previously as a putative tumour suppressor gene (TSG). Misawa et al.[228] demonstrated, in HNSCC cell lines, that GARL1 hypermethylation lead to reduction in gene expression and, by analysis of methylation specific PCR (MSP), that positive methylation corresponded to worse survival in HNSCC tumours. Although HPV stratification had not been conducted in this cohort, the clinicopathological features (increased T stage and poor survival) could be used as construed as surrogates of HPV negative disease. By contrast, we found significant differential hypermethylation in HPV positive OPSCC that had been strictly defined by site of origin and by HPV status. Misawa et al. included 11/38 (29%) OPSCC in their analysis but there was no attempt to analyse them separately to oral, hypopharyngeal, sinonasal and laryngeal tumours and as such has significant potential to confound results[228]. DNA viruses, such as Epstein-Barr Virus (EBV) and Kaposi sarcoma-associated herpesvirus (KSHV), have been shown to exploit host GCRPs or possess ORFs that encode viral-GCRPs for selective growth and metastatic advantage[246]. In contrast, the evidence of hypermethylation of GARL1 in HPV positive OPSCC we found, may reflect a host response to viral oncogenesis or a lack of deregulation of this gene within HPV positive tumours and, in keeping with findings of increased sensitivity to chemotherapy (5-fluorouricil & platinum based agents) in colorectal cancer[229], could correlate with known responsiveness of HPV positive OPSCC to chemotherapy based treatments[247].

Genome-wide Assay Validity

Although differential methylation is seen using both the 450K array and PMA analysis for each of the candidate genes, strong correlation of quantitative comparison between methods is not apparent and could be seen as a limitation of this analysis. The lack of quantitative correlation is perhaps unsurprising given the differences in techniques and in individual targets; 450K probes analyse solitary CpGs whilst corresponding PMA target sequence covers a number of probe CpGs and additional adjacent CpGs (for example, CCNA1 PMA included 8 CpGs for analysis, of which 5 had been individually analysed on the 450K platform).

In keeping with other authors[192,248], our validation of 450K results support the application of this array but as yet do not allow complete exploration of the changes in gene promoter methylation as the cohort analysis is underpowered. However, a biological validation on a large separate and distinct cohort may allow greater interpretation. Any further validation would seek to elucidate and better define predictive biomarkers in the transition to viral induced malignancy.

Regulators of Methylation in HPV positive OPSCC

DNMTs are the only recognised enzymes capable of methyltransferase activity, although contributory elements, such as UHRF1 can act as facilitators though DNMT1 recruitment to hemimethylated targets[154]. This analysis sought to define the quantitative differences in expression of the key elements of the cellular

methylation machinery between HPV positive and negative OPSCC and more importantly correlate findings in HPV positive tumours to the viral methylome analysis.

In accordance with findings in a wide variety of human malignancies[156], we demonstrated that OPSCC overexpressed all DNMTs, in particular UHRF1 and DNMT3b, when compared to their normal adjacent tissues. Underlining the divergent mechanisms of HPV positive and negative OPSCC, Sator et al.[163] made similar observations when analysing DNMT3a expression alone in HPV positive cell lines and in a comparative gene expression array, Martinez et al.[249] found DNMT1 to be overexpressed in HPV positive HNSCC by comparison to normal tissues.

It is recognised that the pRb/E2F pathway regulates expression of UHRF1 and DNMT1 in order to replicate DNA methylation profiles during the transition from S to G2 phases of the cell cycle[250]. We demonstrated that there was a strong positive correlation between DNMT1 and UHRF1 expression in OPSCC however interestingly, the strength of correlation was greatest for HPV negative tumours. Daskalos et al.[161] had previously shown in primary lung SCC, a tumour with similar aetiology to HPV negative HNSCC, that overexpression of UHRF1 was a critical feature responsible for maintenance of TSG hypermethylation. Such findings would support our observation in HPV negative disease.

Within HPV positive tumours, viral E7, through its impact on pRb and the pRb/E2F complex, liberates E2F that in turn acts as a transcriptional activator of both UHRF1 and DNMT1. Our results did not demonstrate a collaborative expression profile and, when compared to HPV negative tumours, HPV positive disease had a trend towards lower expression of all DNMTs and UHRF1. This trend was most profound

for DNMT3b (p=0.007) and when outcomes for individual cases was analysed based the level of DNMT3b expression, a clear positive prognostic impact was seen for low expressing tumours. Due to the profound survival benefit of HPV positive status it was not possible to exclude this as a confounding variable, however evidence from diffuse large B-cell lymphoma[251], acute myeloid leukaemia[252] and non-small cell lung cancer[161] would support a survival advantage for low DNMT3b expressing tumours. In HNSCC, irrespective of levels of expression, DNMT3b polymorphisms have also been shown to covey a negative survival impact[253,254]. In both instances, tumours were not subjected to HPV testing however the demographic features would be suggestive of HPV negative disease. It would be of interest to determine the relevance or proportional occurrence of DNMT3b polymorphisms in HPV positive and negative OPSCC.

Although HPV positive and negative OPSCC consistently display fundamentally different genetic and epigenetic profiles, we were unable to define consistent mechanistic links between the DNA methylation regulatory genes and either genome-wide methylation (LINE-1) or specific targets within the viral genome (E2 promoter, LCR sites including E2BSs).

7 VIRAL INTEGRATION STATE IN HPV POSITIVE OPSCC

7.1 INTRODUCTION

Integration Analysis Aims

The understanding and subsequent implications of viral integration in high risk HPV lesions come primarily from analyses of cell lines or clinical samples derived from cervical dysplasia or neoplasia[6,110,174]. Integration appears to represent a critical step in the progression to invasive disease as implied by the increasing frequency with which it observed from early cervical intraepithelial neoplasia (CIN) through carcinoma in situ to cervical cancer[255].

It was hypothesised that integration plays a critical role in the disruption of the E2 gene and represents an obligatory step in oncogenesis within HPV positive OPSCC and as such could demonstrably affect detectable levels of E2 gene expression.

Although potentially advantageous for viral DNA persistence, it was postulated that integration events within cellular DNA in HPV OPSCC are random in nature rather than targeted to specific cellular genes, in keeping with findings from cervical cancer.

Integration analysis aimed to detect the presence of E2 gene integrity and correlate this with E2 gene expression.

Additionally, on the premise that the HPV circular genome cleavage position might fell beyond the limits of the E2 gene we aimed to test a recently validated hybridisation capture and sequencing techniques to detect these positions and the related host cellular insertion points in a small of cohort of HPV positive OPSCC

samples. This analysis also aimed to determine, in clinical OPSCC samples, whether detection of integration events might have a role in positive selection bias during oncogenesis (insertional mutagenesis events).

7.2 VIRAL E2 GENE INTEGRITY ANALYSIS

E2 Integrity Assay - Methods

The integrity of the viral E2 gene was analysed in cell line samples (CaSki, SiHa and HBEC-3KT) and 44 OPSCC samples, including 43 that had previously been demonstrated to harbor HPV16 DNA (4), using overlapping endpoint PCR primers for E2 ORF as previously described[182].

The positive control utilised for disrupted E2 gene was SiHa and for intact E2 gene was CaSki. The negative controls were DNA derived from the known HPV16 negative cell line HBEC-3KT and DNA from the HPV negative OPSCC (sample No.11). Duplicate reactions were run for all samples and where equivocal or contradictory results were apparent, a triplicate reaction was run under identical conditions as a discriminator.

E2 Integrity Assay - Results

From the cohort of 43 HPV16 DNA positive cases E2 gene integrity results were available for all cases analysed. For 38/43 samples, all component parts of the E2 gene (E2 primer pairs 1-5) and the E2 whole gene product were apparent by gel electrophoresis. For 5/43 (12%) samples, PCR products were not visualised for the

whole E2 gene amplification (Table 45) and in each instance, one or more of the component parts (p1-p5) also failed to amplify product.

The positive control CaSki displayed PCR product for each PCR E2 component and the whole gene whilst for SiHa, the E2 whole gene and P2 primer set (coverage nt 3086-3388) demonstrated no evidence of product.

The negative control tumour and HPV16 negative cell line DNA samples demonstrated no visualised product for any assay, whole gene or component.

Of the 5 OPSCC samples with disrupted E2 genes, two also failed to have detectable levels of viral E2 gene expression (Table 45). As expected SiHa also showed no evidence of E2 expression.

	Viral Gene Expression (RQ)			E2 Gene Integrity						Integration analysis
Sample Number	E2	E6	E7	E2W	E2 P1	E2 P2	E2 P3	E2 P4	E2 P5	E2 Gene Integrity
CaSki	1	1	1							Episomal
SiHa	0	0.9049705	0.0947423							Integrated
13	0	0.0007309	0.0007884							Episomal
34	0	0.1459979	0.0685027							Episomal
106	0	0.4878131	0.4103934							Integrated
95	0.0003023	0.1935964	0.2297027							Integrated
105	0.0633823	0.0834319	0.0713859							Episomal
88	0.2108659	0.9311542	0.2648071							Integrated
108	1.4176697	6.8316684	5.5763587							Episomal
87	1.7438271	0	45.256418							Episomal
101	2.132007	0	0							Episomal
2	0	0	0							Episomal
29	0	0	0							Episomal
9	0	0	0							Episomal
24	0	0	0							Episomal
86	0	0	0							Episomal
73	0	0	0							Episomal
74	0	0	0							Episomal
43	0	0.141985	0.0669817							Integrated
78	1.219E-05	0.0457227	0.0672174							Episomal
94	0.000561	0.0006288	0.000668							Episomal
15	0.0008446	0.0248418	0.0099905							Episomal
56	0.0009473	0.088393	0.0905531							Episomal
52	0.002159	0.0017562	0.002059							Episomal
64	0.0027079	0.1240011	0.1204835							Episomal
79	0.0064853	0.0291312	0.0345321							Integrated
54	0.0137445	0.0122932	0.0196							Episomal
92	0.0210234	0	0.0075874							Episomal
68	0.0298181	0.0442212	0.1784394							Episomal
63	0.0299735	0.0654582	0.0490217							Episomal
91	0.0546892	0.246214	0.1142239							Episomal
58	0.0721929	0.0300186	0.1314877							Episomal
66	0.1014794	0.3165565	0.3471346							Episomal
102	0.1199433	0.1467459	0.0968757							Episomal
51	0.1226845	0.1678393	0.4207485							Episomal
40	0.1431897	0.1056965	0.0700243							Episomal
85	0.1861135	0.4027405	0.2244408							Episomal
22	0.2099776	0.2963961	0.1613667							Episomal
104	0.2617149	0.2578441	0.1539956							Episomal
103	0.3677299	0.4235508	0.331257							Episomal
75	0.4408241	0.4394722	0.3069655							Episomal
97	0.4964587	0.5106907	0.368357							Episomal
89	1.2053655	6.8301286	1.7082364							Episomal
46										Episomal
77	0.2192208	0.3467968	0.0886204							Episomal
11	0	0	0							Negative
HBEC-3KT	0	0	0							Negative

Table 45: Relative **v**iral gene expression (RQ) and E2 gene Integrity Analysis for HPV16 DNA positive OPSCC.

Gene expression data is relative to CaSki, HPV16 positive cervical cancer cell line expression (grey band represents no available data).
E2 gene integrity analysis schematic for results of end-point PCR analysis of E2 whole gene assay (E2W) and overlapping sequence components of the E2 gene (E2P1 – P5). Red signifies visualised PCR product of expected size in duplicate reactions. White signifies no PCR product.
Integration Analysis: Reflects the inferred integration state of HPV16 in the clinical sample; evidence of all components of the E2 gene and the E2 whole gene PCR suggests episomal viral state whilst loss of one or more component of E2 and the E2 whole gene is inferred to suggest integrated virus.

Cell Line samples (CaSki & SiHa) are underlined by red-hashed lines, HPV negative OPSCC tumour (No. 11) and cell line (HBEC-KT) samples are underlined by blue-hashed lines.

7.3 SPECIFIC VIRAL CAPTURE & NEXT GENERATION SEQUENCING FOR DETERMINATION OF VIRAL CLEAVAGE & HOST INSERTION POSITIONS

A series of 9 HPV16 DNA positive OPSCC and the HPV16 positive cervical cancer cell lines, CaSki and SiHa, were analysed using a recently described target sequence (HPV16) hybridisation, amplification and deep sequencing technique[195].

With an appreciation of the potential weaknesses of inferred integration assays such as the E2 integrity assay, detailed above and discussed more fully below, this analysis sought to better define the incidence of viral integration in clinical samples, and where integration was apparent, the cleavage point within the circular viral genome. Through capture of the viral sequences abutting integration sites in the host genome it was intended to also be able to determine the "recipient" sites within the host and quantify whether integration appears as a sporadic event or a specific driver through insertional mutagenesis.

Viral sequence capture and sequencing – Results

The 11 samples, including 9 clinical OPSCC DNA samples and 2 reference/control cell lines, all provided results amenable to analysis results following library preparation, sequence capture, amplification and sequencing.

All samples were multiplexed and sequenced using an Illumina HiSeq platform, yielding between $8x10^6$ – $9.5x10^6$ 100bp paired-end reads per sample analysed

(Appendix I), although following removal of duplicate reads to declare unique reads only, the number fell by between 44 – 66% (Appendix II).

When considering total paired-end reads (Appendix I) with duplicates included, reads which mapped to human genomic sequence in both paired ends represented a significant proportion of reads (31.23 – 92.71%) with the exception of CaSki, for which human-human paired reads represented only 4.82% of the total reads. Concordance in terms of proximity of the paired reads to one and other in the human genome (mapping sites <500bp apart) was typically seen, however a small proportion of reads mapped beyond proximity of 500bp (Human:Human mixed pairs 0.25 – 7.32%).

Paired end reads that contained viral sequence demonstrated substantial variation between samples. The CaSki cell line sample demonstrated a high frequency (91.24%) of reads containing one or more pair with viral sequence, however 5 tumour samples demonstrated less than 5% of paired end reads with detectable viral sequence. Concordance within the viral read pairs (mapping viral sites within <500bp of paired reads) was proportionate to the overall viral reads (86-94%) with the exception of one sample (Sample ID 13) that showed greater variation (60%) albeit from a significantly reduced number of viral reads.

Hybrid reads, which were composed of viral reads in one paired-end read and human sequence in the second were seen infrequently, greatest in a tumour sample (Sample ID 34) and the CaSki cell line sample (3.52%) and least frequently in the remaining tumour samples (range <0.001% - 1.92%).

RNaseP reads were consistent across the tumour samples, however CaSki demonstrated a substantially reduced total, and proportionate, read for RNaseP

pairs (Appendix I – bottom panel). Across the cohort of samples between 3 - 6% of RNaseP reads lacked concordance (over 500bp between paired reads).

Comparative integration results (Table 46) highlight the presence of detectable viral integration in both reference cell lines and all clinical samples. As detailed previously, the E2 integrity assay inferred integration to be present in only 3/9 samples, yet NGS analysis detects integration in all 9 clinical samples and the two cell lines. Generalised peak viral cleavage position, collated from Human:Viral mixed pair reads, demonstrates variability in viral break points.

More detailed analysis of individual samples (Appendix III-XIII) shows that viral cleavage, for the most part, is pan-genomic, with the exception of short viral regions in SiHa (E2 gene; Appendix IV), and in clinical samples No. 13 (E7, E2, L1, L2 and LCR; Appendix V), No. 106 (E2 – E5; Appendix VII) and No. 105 (L2; Addendix IX) where no viral reads are detected in mixed paired end reads (including human sequence in the corresponding paired read).

In the cases of the four samples that had undetectable E2 expression (3 clinical samples and SiHa cell line), a peak E2 viral cleavage position was not apparent. However within these same samples, absent representation of viral reads from a portion of the E2 gene was apparent in three of the four instances, SiHa and samples No. 13 and 106 (as above) suggesting the possibility of deleted sequence. For samples where E2 expression was apparent (including sample No. 101 which had exclusive high E2 expression without detectable viral oncogene expression), E2 was fully represented in matched human:viral reads.

Co-localisation of paired end viral DNA reads to the human genome (Appendix III-XIII; top images) demonstrated specific integration positions in 5 samples (SiHa and tumour samples 106, 95, 88 and 108; Appendix VII, VIII, X and XI respectively). Not only were these apparent in the Human/HPV mixed pairs read but also detected in chimaeric reads and were analysised for specific integration site/nucleotide (Table 46 & Appendix V, VII, VIII, X & XI - bottom schematics).

For the remaining samples, CaSki, and clinical samples Nos. 13, 34, 105, 87 & 101), co-localisation to the human genome was non-specific or pan-genomic in nature (Table 46 & Appendix III, V, VI, IX, XII & XIII). Interestingly, this group included all samples that had low frequency integrants (sample Nos. 13, 105, 87 & 101) as demonstrated in the relatively low chimaeric read frequency (<250 total chimaeric reads).

Chimaeric read interrogation revealed peak human insertion positions in each sample. Due to low chimaeric read number, samples deemed to be low frequency integrant samples were not further analysed for human insertion position. It is apparent in remaining samples that integration position is not conserved between samples with no evidence of repetition in the cohort analysed. Further, sites appear to be primarily within repetitive elements of the human genome rather than specific genes (Table 46). Specific gene disruption of GPR1, G-couple-protein receptor 1, was noted in sample No. 95 (Appendix VIII).

There was no apparent relationship between viral cleavage, host insertion or read proportions and the extent of viral oncogene expression (E6 & E7).

	Viral DNA qPCR	Viral Gene Expression (RQ)			Integration analysis		Chimaeric Host Integration Location Data					Prominence of Viral Cleavage Site (NGS)							
Sample Name or Number	E6	E2	E6	E7	E2 Gene Integrity	NGS Cohort	Host Peak Insertion Point (nt.)	Genomic Map	Gene	Gene Function	Correlated Viral Cleavage	E6	E7	E1	E2/E4	E5	L2	L1	LCR
CaSki	9.056	1	1	1	Episomal	Integrated	20:26257343	Ch20p11.1	-	satellite repeat	E6								
SiHa	0.884	0	0.905	0.0947	Integrated	Integrated	13:74087563	Ch13q22.1	-	Long intergenic non-protein coding RNA	E2								
13	-0.162	0	0.0007	0.0008	Episomal	Integrated	low frequency integrant												
34	6.158	0	0.146	0.0685	Episomal	Integrated	8:128638753	Ch8q24.21	PVT1 & SINE repeat element	non coding protein	E1, L1 & LCR								
106	1.837	0	0.4878	0.4104	Integrated	Integrated	1:223486447	Ch1q41	SusD4/LTR	Sushi domain-containing protein 4	E1 & L2								
95	0.700	0.0003	0.1936	0.2297	Integrated	Integrated	2:207070838	Ch2q33.3	GPR1	G-protein-coupled Receptor1	E1								
105	4.932	0.0634	0.0834	0.0714	Episomal	Integrated	low frequency integrant												
88	3.769	0.2109	0.9312	0.2648	Integrated	Integrated	15:90872202	Ch15q26.1	SINE Repeat element	-	E1								
108	4.502	1.4177	6.8317	5.5764	Episomal	Integrated	16:17039652	Ch16p12.3	SINE repeat element	-	E1 & E2								
87	-0.091	1.7438	0	45.256	Episomal	Integrated	low frequency integrant												
101	2.552	2.132	0	0	Episomal	Integrated	low frequency integrant												

Table 46: Compiled results for samples analysed using next generation sequencing (NGS)

For each tumour or cell line control sample, details of the viral E6 gene DNA qPCR results (ΔCT) and RNAqPCR (ΔΔCT, ref sample CaSki) results for gene expression of viral E2, E6 and E7 are included.
Inferred E2 integration results and overall NGS integration state is listed for each sample.
Representative host insertion position (table centre) is detailed from viral/host chimaeric reads. Peak host insertion position (nucleotide position, cytogenetic location) and where apparent human gene with known gene function. Corresponding viral cleavage position is also noted.
Schematic diagram of viral genome (far right of table), with relative viral open reading frame size, highlights peak viral cleavage positions (blue fill) in each sample.
Cell Line samples (CaSki & SiHa) are listed above red dividing line, OPSCC tumour samples are below red dividing line

7.4 DISCUSSION

Detection of viral integration – E2 gene integrity

Disruption of the E2 gene has been held as synonymous with integrated viral DNA and, by implication, the loss of E2 is subsequently a critical step in viral mediated oncogenesis. Collins et al.[182] demonstrated in cervical dysplastic lesions and invasive cervical disease that an assay capable of detecting the integrity of E2, will highlight progressive increases in inferred integration of viral DNA as severity of disease advances.

In HPV16 positive OPSCC, this assay classified 12% of cases as having a disrupted E2 ORF, suggestive of integrated viral DNA. The positive control for disrupted E2 gene, SiHa cell line DNA, gave results consistent with previous findings of disruption of the E2 gene between nucleotides 3132 – 3384 as reported by Baker et al.[194] and corroborated by Collins et al.[182] in their description of the application of this assay. This supports the capacity of the test to detect a disrupted E2 gene, at least when all copies in the sample are similarly disrupted.

As will be discussed further below, detection of integration using the sensitive sequencing technique employed, categorised all samples as having integrated viral DNA, effecting E2 and other ORFs, and therefore the conflicting results question the capacity of the E2 integrity test to detect all cases of integration and hence its efficacy.

Admittedly, no precursor lesion was available for inclusion in this analysis and therefore the temporal element to this assay's utility is lost, however the test also

fails to recognise the potential for viral cleavage events beyond the boundaries of the E2 gene and, in common with many indirect integration analyses, fails to "see past" episomal viral DNA when detecting integration. This situation is clearly demonstrated in the cervical cancer cell line, CaSki, which is known to contain multiple viral integrants[185] and yet registers intact E2 components and whole E2 gene, therefore interpreted as episomal.

E2 expression was absent in five samples with evident E6 and/or E7 expression (SiHa and 4 OPSCC derived samples), yet it would appear that complete E2 gene disruption can only be implicated in 3 of these samples, of which SiHa is one. Alterative means of disruption of the viral E2 ORF are possible, including down regulation through methylation of the gene promoter, although our results noted previously (6.2) do not support such a finding. Similarly the nature of end point PCR assays may have lead to misinterpretation of the biological situation through assay detection issues.

Next Generation Sequencing as an Analytical Tool

Using target capture technology and high throughput sequencing we were able to produce total paired-end reads of approximately 8×10^6 from HPV16 positive tumour samples. This is slightly lower than that described by Depledge et al. ($4.8 \times 10^7 – 7.2 \times 10^7$) when analysing larger viral target sequences in from clinical preparations. Duplicate reads accounted for 34 -56% of paired-end reads following initial analysis. It is recognised that duplicate reads can occur as a result of PCR

amplification, however independent sequencing fragments are also likely to generate identical reads by chance. In this analysis, since the region of target interest is small and resultant depth of sequencing in greatly increased, it is to be expected a limited number of exact sequences would be replicated resulting in numerous independent reads which start and end at the same position. As can be seen in Appendix I and Appendix II, duplicate reads follow a similar proportionate distribution to the overall reads. For these reasons it was therefore decided to include duplicate reads in the mapping statistics for all subsequent analysis.

The proportion of paired-end reads which map to Human sequence is surprisingly high in tumour samples when compared to CaSki, which is know to have a high number of integrants per genome (>800)[185]. In the initial description of this capture and sequencing technique, Depledge et al[195] was able to generate an average of 80% (range 34-99%) on target reads, however in situations where target abundance was low this fell to 18-20%. Our findings for CaSki demonstrate over 90% paired-end reads containing one or more read with viral sequence however OPSCC tumour samples had between 0.02% - 61.66% (mean 15.5%). Low natural abundance of target reads in complete genomic samples could explain this, at least in part. If viral copies are very low, then following complete uptake of viral sequence the baits may capture targets with reduced homology non-specifically.

Alternative explanations generally point to less-than-stringent washes applied post hybridisation however the techniques employed do not differ from previous descriptions nor the manufacturers ideals and it is felt that this is therefore an less likely explanation. Also, inclusion of RNaseP baits in the process could impact on genomic read frequency and a proportion of the 120mer baits had homology of

other human sequence of up to 20bp. This level, however, sat well within the confines of the manufacturers recommendations (<40bp homology) and is felt likely to have only had a minor contribution to increased human paired-end read frequency.

Analysis of paired-end reads that include viral sequence suggest the presence of integration in all samples, cell line or tumour derived. Such a finding would be expected in CaSki and SiHa, however in this instance all tumour samples demonstrated sequencing features such as human:viral mixed pairs and chimaeric reads (single reads containing viral and host sequence), consistent with integration. Although integration may be ubiquitous in HPV positive OPSCC, sample No 13 must be viewed with caution given the very low viral read number, yet on balance, it does retains a high HPV concordant:mixed pair ratio (0.23) despite the overall read number.

In keeping with all other integration analysis techniques, it is still not possible to quantify the proportion virus that is episomal and that which is integrated for any given sample. In support of the analysis presented, the disruptive peaks for SiHa demonstrate concordance with the previously described range for peak cleavage (nt 3134 – 3384) both published[182,188,194] and indicated in the E2 integrity results above. Similarly, CaSki integration findings are in keeping with previous results of multiple tandem head to tail nondisrupted and fragmented disrupted reads[188], although the interpretation is difficult given the high number of integrant events, the apparent lack of consistency in cleavage position and the disparate chimaeric reads.

Viral Cleavage Position Detection

This analysis provides evidence that cleavage of the viral genome in the integration process is pan-genomic in nature. Both Human:Viral mixed reads, and chimaeric reads have highlighted the positions of cleavage broadly across the viral genome. Although peaks of cleavage frequency, particularly in the E1 and E2 genes, are witnessed, short conserved viral sequence that does not appear in to cleave is also apparent, particularly within E2 (in SiHa and in 2 tumour samples) but also elsewhere in the genome.

On the evidence of disparate cleavage position, it is difficult to support the concept of viral E2 cleavage as being a driving event in oncogenesis. This is further refuted by no apparent correlation between integration and E2 expression and the, otherwise counterintuitive, positive correlation witnessed in E6 and E7 expression (6.2).

It is difficult to speculate on the significance of preserved regions of virus, not involved in mixed human:virus paired reads given their frequency in this small cohort of samples.

The samples that provided pangenomic cleavage positions fell into two groups; those with low viral integrant frequency (<250 chimaeric reads) and those with high integration as a feature (>250 chimaeric reads). Interestingly, the former of these two groups included one sample that demonstrated high E2 expression alone (in the absence of detectable E6 or E7 (sample No. 101, Appendix XIII) despite detectable viral reads from the E6 and E7 ORF. Such a tumour would normally be

classified as being HPV negative using the gold standard definition (viral oncogene expression from fresh frozen tissue samples) and as such could merely be a transient infection, albeit with low levels of integration present. Review of RNAscope (HR HPV RNAISH) data for the corresponding FFPE cores however, highlighted the sample to be clearly positive for viral E6 and E7 expression.

Host Integration Position

From the overall reads and the chimaeric reads that dictate exact nucleotide insertion positions, insertion is almost exclusively in repetitive elements rather than specific genes. This does question the theory of insertional mutagenesis bearing a substantial or obligatory stage in oncogenesis. These findings are from a small cohort however but are supported by a previous systematic review of viral insertion sites in HPV positive cell lines, dysplastic epithelial cervical lesions and invasive cervical cancer[110]. From over 190 previously reported loci of integration in HPV positive samples, integration appeared random in nature but with a predilection for reported common fragile sites (CFS). Two sites in particular show frequent integration hits and have been termed “hot spots” for integration[256]. From the samples analysed here, the cell line SiHa fits within a previously reported HPV related CFS (13q22.1, FRA13B)[257] as expected. Further, tumour sample No. 106 displayed the highest frequency of chimaeric reads within the CFS FRA1H[110]. Given the scale of different integrant locations there remains considerable scope to further correlate known and putative CFS with integration loci.

It is not yet apparent if integration at CSF occurs in an opportunistic fashion or whether there is a selective advantage however there have been calls made to better define CSFs in the context of HR HPV integration[6].

In a comparative analysis of integration sites and genomic alterations, using an array-CGH analysis, Peter et al.[174] demonstrated that at particular genomic integration sites, short amplifications occurred in multiple tumour samples within the test cohort. In our analysis, Sample No. 34 demonstrated peak chimaeric integration detection at the loci identified by Peter et al. (8q24.21) as the most frequently detected recurrent amplification point. Amplifications detected by Peters et al were modest in nature and they speculated that it was the integrated viral origin of replication that directed later amplification through co-amplification of viral and host sequences. Despite being an observation in separate cohorts, by different means, it is unclear whether this observation has any fundamental impact on the process of virally mediated tumourgenesis.

Evidence based support for the concept of insertional mutagenesis was not readily apparent when considering the peak integration positions detected in OPSCC in our analysis. The previously mentioned sample No. 34 was shown to have peak chimaeric reads with the chromosomal band 8q24.21, a region of the MYC gene. This has been seen previously, in particular in association with HPV18 mediated cervical malignancy and has raised interest given the implications MYC holds for universally up regulating gene expression[6].

Sample No. 95 demonstrated a large frequency of integration into the G-protein-coupled receptor-1 (GPR1). Part of the G-protein-coupled receptor superfamily, this gene has not featured in previous integration sites nor does it have reported links to malignancy, virally mediated or otherwise.

Although selection bias cannot directly be excluded within these results, it is apparent that the view of viral integration occurring in a significant minority of OPSCC should be reconsidered. Although this represents a modest cohort of samples, integration appears to be ubiquitous and the disruption to viral sequence and host alike does not fit to a single prescription. Expansion to a larger cohort using sensitive techniques such as those employed here would add strength to the initial findings. An important step beyond observational findings will be corroboration via alterative techniques such as quantitative PCR. Should specific genes, such as MYC, appear more frequently on a list of insertion sites, qPCR validation might prove a useful tool to differentiate incidental event from oncogenic driver.

8 DISCUSSION & FUTURE DIRECTIONS

HPV Diagnostics in HNSCC – Prognostic Biomarker & Disease Stratification

Detection of HPV, or surrogate markers of its presence, has swiftly become a fundamental requirement in the clinical management of OPSCC within the United Kingdom and the majority of Western health care systems.

Although therapeutic regimes do not as yet differ on the basis of HPV status, testing provides the basis for guiding discussions surrounding prognosis in the clinical setting.

We have demonstrated that the well-established increase in incidence of OPSCC in United Kingdom Head and Neck Oncology practice is coincident and comparable to the increase in relative incidence of HPV positive OPSCC, such that almost two thirds of OPSCC are now likely to harbor transcriptionally active HPV16. Such a change has increasingly been described as an epidemic of HPV positive OPSCC and has drawn the attention of public health officials responsible for health planning and vaccination.

Although we have shown that the clinical testing regimes available currently are universally capable as prognostic biomarkers, their diagnostic capacity is highly variable, in part due to a necessary reliance on formalin fixed tissue specimens. Where sensitivity is sought, for example in the application of p16 IHC, our results demonstrate it to be at the expense of specificity. Combination tests offer opportunities to improve specificity but in doing so compromise sensitivity. For so long as treatment does not depend on HPV status, each of the alternative test will remain fit for purpose, however this will no longer remain acceptable should clinical

practice evolve towards de-escalation of therapeutic intensity for HPV positive disease, or indeed the converse, intensification for individuals with locally advanced HPV negative disease. Both in this setting and in the clinical trials that are necessary to inform this practice, greater diagnostic stringency will be essential.

The research detailed in this thesis will form the basis for future applications to investigate the prognostic significance of HPV positive malignancy beyond the confines of the oropharynx.

Although, only briefly mentioned, data from small cohorts of tumours in other head and neck subsites (oral cavity, larynx and hypopharynx) suggest that only a small proportion, typically 5-10%, are HPV positive (Oral 4/102 (4%), Larynx 3/91 (3.2%), Hypopharynx 2/28 (7.1%), Overall 9/221 (4.1%); unpublished data kindly provided by Mr Nav Upile). The size of these cohorts has precluded meaningful analysis of outcome measures to determine if HPV status outwith the oropharynx has prognostic significance. Large multicentre bio-banks are established and with sound diagnostic rigor could be interrogated to answer this question.

We also intend to explore further, the relationship between smoking history and HPV status with a view to better defining any apparent survival difference between HPV positive smokers and HPV positive non-smokers. It remains to be seen whether strict anatomical site classification coupled to stringent HPV status determination will detract from previous findings of disease specific prognostic implications for the HPV positive smoker.

It is important to reflect upon the contradictory findings of definitive HPV16 oncogene expression in a significant minority of fresh frozen adjacent marginal

tissues (paired to HPV positive tumours), whilst examination of FFPE samples using RNAscope demonstrated no such findings. Interpretation has considerable importance as it is assumed that, in contrast to carcinogen-induced HPV negative malignancy, HPV positive tumours lack field change; a potential contributory factor in improved local recurrence free survival. Although tumour infiltration of normal samples is difficult to completely refute without formal histopathological review prior to analysis, our experience in such circumstances would consider this to be unlikely. A more convincing explanation would be “peri-resection” contamination of harvested samples as would be the case for any surgically resected specimen. Although tumour is not confined to the normal marginal tissue (harvested over 10mm from the macroscopic tumour margin), contaminated fluids, blood or saliva will inevitably abut all tissues. The high sensitivity of the analytical tests therefore will ensure positive results despite the probability of no active transcription in the normal tissue. FFPE samples by contrast will have any “peri-resection” contamination washed away or denatured in the process of fixation, resulting in the observed negative results for HR HPV in any normal FFPE tissue core. Alternatively, HPV positive non-tumour marginal tissue may occur in a proportion of HPV positive OSPCC. This would, by inference however, call into question the sensitivity of all other FFPE tests applied here (p16, DNA ISH and RNA ISH), as each scored normal tissues as negative in all instances.

Whilst a definitive answer cannot be made here based on the evidence available, on balance of probability and for sound scientific reasons the former explanation is remains considerably more plausible.

Virally mediated OPSCC has both distinct behavior and prognosis hence, as mentioned above, clinical trials specifically targeting HPV positive OPSCC are now being developed. The accuracy of trial stratification for individuals on the basis of the HPV status of their malignancy is of paramount importance and immediate clinical impact. The consequences of inaccurate stratification in the setting of de-intensification trials could result in a cohort of individuals receiving sub therapeutic treatment with potentially devastating clinical implications. Our results have shown, when compared to the gold standard test, the novel RNA in situ hybridisation test, High Risk HPV RNAscope, has diagnostic stringency better than any other single test and comparable to tests that would otherwise not be viable in a clinical setting. The utility of this test on FFPE samples, coupled with excellent levels of interobserver concordance imply that it could be applied in the clinical trial setting once through the necessary national and international in vitro diagnostic regulatory frameworks. It is anticipated that our experience with this diagnostic tool would leave us well placed to engage with such future works.

Biomarkers for HPV positive OPSCC

The necessity for a biomarker capable of discriminating between a persistent high-risk HPV infection and one with the potential to become transformative in the oropharynx exists. On the evidence provided in this research it appears that neither genome-wide methylation analysis nor viral methylome characteristics can be exploited to this end. We were able to build upon the growing evidence that HPV positive OPSCC is biologically distinct from HPV negative tumours but also, we were

able to prove the hypothesis that HPV positive disease has a distinct epigenetic profile. Using clinical samples we have corroborated findings initially generated in HPV positive cell lines and through array-based techniques and target-specific assays, shown fundamental differences that warrant further investigation. The results generated when comparing promoter methylation in HPV positive and negative OPSCC with the Infinium HumanMethylation450 BeadArray were subjected to technical validation through pyrosequencing assays yet expansion to a larger cohort to OPSCC remains necessary for biological validation of targets. It remains to be seen whether a panel of differentially methylated genes would assist biomarker development prediction of the potentially transformative viral infection.

Similarly, it is unclear whether the developing understanding of viral integration state in HPV positive OPSCC can be exploited to improve either diagnostics or therapeutic intervention. Without a dysplastic lesion analogous to that seen in the cervix and utilised to great effect in national cervical screening programmes, early diagnosis in HPV positive OPSCC is limited to clinically evident tumours or subsequent metastatic disease. If integration was shown to be a critical step in HPV positive OPSCC oncogenesis then it would follow that investment of research energy and resources could be directed towards investigating further, saliva based screening tools aimed at viral integration detection. Such advances would, by necessity, need to differentiate between latent and fully selected transcriptionally active integrants, conceivably through downstream effects on cellular function by E6 and E7. Sample recruitment from individuals with latent infection is needed and

would be invaluable as a reference to compare with samples derived from individuals with HPV mediated cancer.

We have however generated data that would suggest that viral integration is an inescapable eventuality for HPV mediated OPSCC. Additionally, we have brought into question the significance of the viral E2 gene following the establishment of malignancy. Our data appears in indicate that viral cleavage events span the HPV genome rather than concentrating on the E2 ORF. Indeed, expression of the E2 gene, at whatever level, seems to be an irrelevance once oncogenesis is established, supported by a finding that E6 and E7 expression is not under the transcriptional repression of E2, at least in invasive disease.

It is interesting to speculate whether HPV mediated malignancy conforms more to a theory of hit and run oncogenesis rather than persistent viral oncogene dependent malignancy. Certainly, the role that E2 plays in this process is less convincing.

Data generated with deep sequencing technology supports previous findings of random human integration sites although with a slight preponderance towards common fragile sites. The implications of viral integration in terms of insertional mutagenesis are certainly an interesting avenue for future investigation, both in terms of expansion of the number of tumour samples subjected to techniques such as those employed here, but also through validation of findings at particular loci such as MYC, a transcriptional regulator known to be constitutively expressed in many cancers.

In conclusion, this research has provided important clarification surrounding the relative utility of HPV diagnostic tests in OPSCC whilst validating a novel test that offers considerable promise when addressing the question of de-escalation of therapeutic intensity in HPV positive OPSCC.

Although new understandings of viral integration occurrence have been put forward in this thesis, uncertainly remains when considering potential for biomarkers for the progression to transforming viral infection.

9 PUBLICATIONS SUPPORTING THIS THESIS

Evaluation of human papilloma virus diagnostic testing in oropharyngeal squamous cell carcinoma: sensitivity, specificity, and prognostic discrimination.
Schache AG, Liloglou T, Risk JM, Filia A, Jones TM, Sheard J, Woolgar JA, Helliwell TR, Triantafyllou A, Robinson M, Sloan P, Harvey-Woodworth C, Sisson D, Shaw RJ.
Clin Cancer Res. 2011 Oct 1;17(19):6262-71. doi: 10.1158/1078-0432.CCR-11-0388.

Validation of a novel diagnostic standard in HPV-positive oropharyngeal squamous cell carcinoma.
Schache AG, Liloglou T, Risk JM, Jones TM, Ma XJ, Wang H, Bui S, Luo Y, Sloan P, Shaw RJ, Robinson M.
Br J Cancer. 2013 Apr 2;108(6):1332-9. doi: 10.1038/bjc.2013.63. Epub 2013 Feb 14.

Associated Publications

HPV specific testing: a requirement for oropharyngeal squamous cell carcinoma patients.
Robinson M, **Schache A**, Sloan P, Thavaraj S.
Head Neck Pathol. 2012 Jul;6 Suppl 1:S83-90. doi: 10.1007/s12105-012-0370-7. Epub 2012 Jul 3. Review.

Squamous cell carcinoma of the head and neck outside the oropharynx is rarely human papillomavirus related.
Upile N, Shaw RJ, Jones TM, Goodyear P, Liloglou T, Risk JM, Boyd MA, Sheard J, Sloan P, Robinson M, **Schache AG**
Oral Oncology (Under review August 2013)

10 APPENDICES

Sample ID	Total read pairs	Human within chromosome pairs	%	Human concordant pairs	%	Within HPV pairs	%
CaSki	9,511,505	458,537	4.82%	430,276	4.52%	8,678,529	91.24%
SiHa	8,386,165	7,377,237	87.97%	6,938,964	82.74%	364,594	4.35%
13	8,045,978	7,459,666	92.71%	7,081,186	88.01%	1,466	0.02%
105	8,817,575	5,389,265	61.12%	5,091,711	57.75%	2,797,628	31.73%
101	8,329,130	6,910,317	82.97%	6,590,243	79.12%	850,771	10.21%
106	8,841,535	7,880,586	89.13%	7,416,594	83.88%	221,956	2.51%
34	8,622,022	2,692,502	31.23%	2,533,755	29.39%	5,326,167	61.77%
108	9,639,841	6,508,255	67.51%	6,146,477	63.76%	2,560,347	26.56%
88	7,978,406	6,440,880	80.73%	6,087,751	76.30%	969,477	12.15%
95	8,335,501	7,621,955	91.44%	7,182,714	86.17%	100,866	1.21%
87	8,844,930	7,694,502	86.99%	7,290,046	82.42%	404,606	4.57%

Sample ID	Total read pairs	HPV concordant pairs	%	Human/HPV mixed pairs	%	Human:Human mixed pairs	%
CaSki	9,511,505	8,220,424	86.43%	335,243	3.52%	22,635	0.24%
SiHa	8,386,165	333,691	3.98%	28,012	0.33%	526,532	6.28%
13	8,045,978	883	0.01%	259	0.00%	520,963	6.47%
105	8,817,575	2,523,107	28.61%	168,960	1.92%	402,077	4.56%
101	8,329,130	798,060	9.58%	43,371	0.52%	460,877	5.53%
106	8,841,535	191,711	2.17%	19,271	0.22%	647,556	7.32%
34	8,622,022	4,880,479	56.60%	362,182	4.20%	208,732	2.42%
108	9,639,841	2,375,437	24.64%	141,848	1.47%	359,279	3.73%
88	7,978,406	894,543	11.21%	68,215	0.85%	435,048	5.45%
95	8,335,501	91,657	1.10%	6,789	0.08%	544,232	6.53%
87	8,844,930	377,091	4.26%	11,599	0.13%	335,302	3.79%

Sample ID	Total read pairs	RNaseP pairs	%	RNaseP mixed pair	%
CaSki	9,511,505	474	0.00%	15	0.00%
SiHa	8,386,165	12,271	0.15%	547	0.01%
13	8,045,978	12,808	0.16%	636	0.01%
105	8,817,575	11,072	0.13%	670	0.01%
101	8,329,130	11,505	0.14%	591	0.01%
106	8,841,535	14,774	0.17%	748	0.01%
34	8,622,022	5,643	0.07%	311	0.00%
108	9,639,841	12,157	0.13%	510	0.01%
88	7,978,406	12,990	0.16%	592	0.01%
95	8,335,501	13,163	0.16%	709	0.01%
87	8,844,930	8,569	0.10%	320	0.00%

Appendix I: Mapping Statistics for individual samples, including duplicate reads where apparent.

Human within chromosome pairs: Paired end reads mapping to human sequence at one or both ends. **Human Concordant Pairs**: Concordant Human paired-end reads, mapping within 500bp. **Within HPV Pairs:** Paired end reads both mapping to HPV sequence. **HPV Concordant Pairs:** Concordant Viral paired-end reads, mapping within 500bp. **Human:Human Mixed Pairs:** Paired end reads mapping to human sequence in excess of 500bp from paired reads. **RNaseP Pairs:** Paired end reads mapping to the 341bp RNaseP gene sequence. **RNaseP mixed Pair:** Paired end reads mapping to viral sequence or human sequence in excess of 500bp from opposing paired read

Sample ID	Total	Human within chromosome pairs	%	Human concordant pairs	%	Within HPV pairs	%
CaSki	3,222,204	437,504	13.58%	410,786	12.75%	2,431,180	75.45%
SiHa	4,361,982	3,804,007	87.21%	3,537,080	81.09%	117,570	2.70%
13	4,113,314	3,719,781	90.43%	3,492,696	84.91%	920	0.02%
105	4,565,928	3,259,749	71.39%	3,057,252	66.96%	872,978	19.12%
101	4,673,491	3,978,570	85.13%	3,770,027	80.67%	293,774	6.29%
106	4,686,207	4,115,781	87.83%	3,829,754	81.72%	69,605	1.49%
34	3,974,592	1,977,242	49.75%	1,854,511	46.66%	1,567,167	39.43%
108	5,046,944	3,880,349	76.89%	3,636,877	72.06%	772,022	15.30%
88	4,307,501	3,604,753	83.69%	3,375,994	78.37%	309,955	7.20%
95	3,959,364	3,543,148	89.49%	3,295,467	83.23%	27,935	0.71%
87	4,311,858	3,588,687	83.23%	3,356,769	77.85%	117,751	2.73%

Sample ID	Total	HPV concordant pairs	%	Human/HPV mixed pairs	%	Human:Human mixed pairs	%
CaSki	3,222,204	2,184,997	67.81%	315,043	9.78%	21,916	0.68%
SiHa	4,361,982	104,428	2.39%	11,981	0.27%	338,634	7.76%
13	4,113,314	616	0.01%	246	0.01%	328,743	7.99%
105	4,565,928	755,624	16.55%	86,256	1.89%	287,300	6.29%
101	4,673,491	271,627	5.81%	20,978	0.45%	316,375	6.77%
106	4,686,207	57,373	1.22%	7,894	0.17%	420,761	8.98%
34	3,974,592	1,374,755	34.59%	229,069	5.76%	168,675	4.24%
108	5,046,944	694,729	13.77%	68,514	1.36%	255,947	5.07%
88	4,307,501	278,471	6.46%	30,557	0.71%	297,450	6.91%
95	3,959,364	24,868	0.63%	2,689	0.07%	323,933	8.18%
87	4,311,858	106,870	2.48%	4,938	0.11%	201,561	4.67%

Sample ID	Total	RNaseP pairs	%	RNaseP mixed pair	%
CaSki	3,222,204	438	0.01%	15	0.00%
SiHa	4,361,982	4,592	0.11%	260	0.01%
13	4,113,314	4,598	0.11%	294	0.01%
105	4,565,928	4,876	0.11%	342	0.01%
101	4,673,491	4,865	0.10%	295	0.01%
106	4,686,207	5,242	0.11%	353	0.01%
34	3,974,592	3,331	0.08%	225	0.01%
108	5,046,944	5,313	0.11%	282	0.01%
88	4,307,501	5,132	0.12%	278	0.01%
95	3,959,364	4,369	0.11%	310	0.01%
87	4,311,858	3,021	0.07%	142	0.00%

Appendix II: Mapping Statistics for individual samples, excluding duplicate reads.

Human within chromosome pairs: Paired end reads mapping to human sequence at one or both ends. **Human Concordant Pairs**: Concordant Human paired-end reads, mapping within 500bp. **Within HPV Pairs:** Paired end reads both mapping to HPV sequence. **HPV Concordant Pairs:** Concordant Viral paired-end reads, mapping within 500bp. **Human:Human Mixed Pairs:** Paired end reads mapping to human sequence in excess of 500bp from paired reads. **RNaseP Pairs:** Paired end reads mapping to the 341bp RNaseP gene sequence. **RNaseP mixed Pair:** Paired end reads mapping to viral sequence or human sequence in excess of 500bp from opposing paired read.

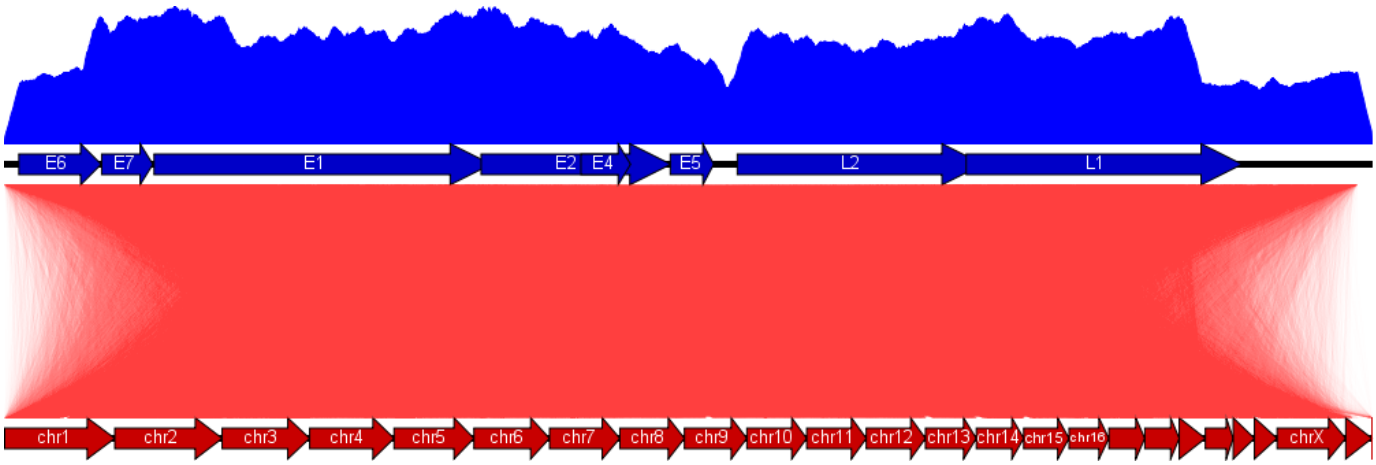
Legend for Appendices III – XIII: For each of the individual appendices

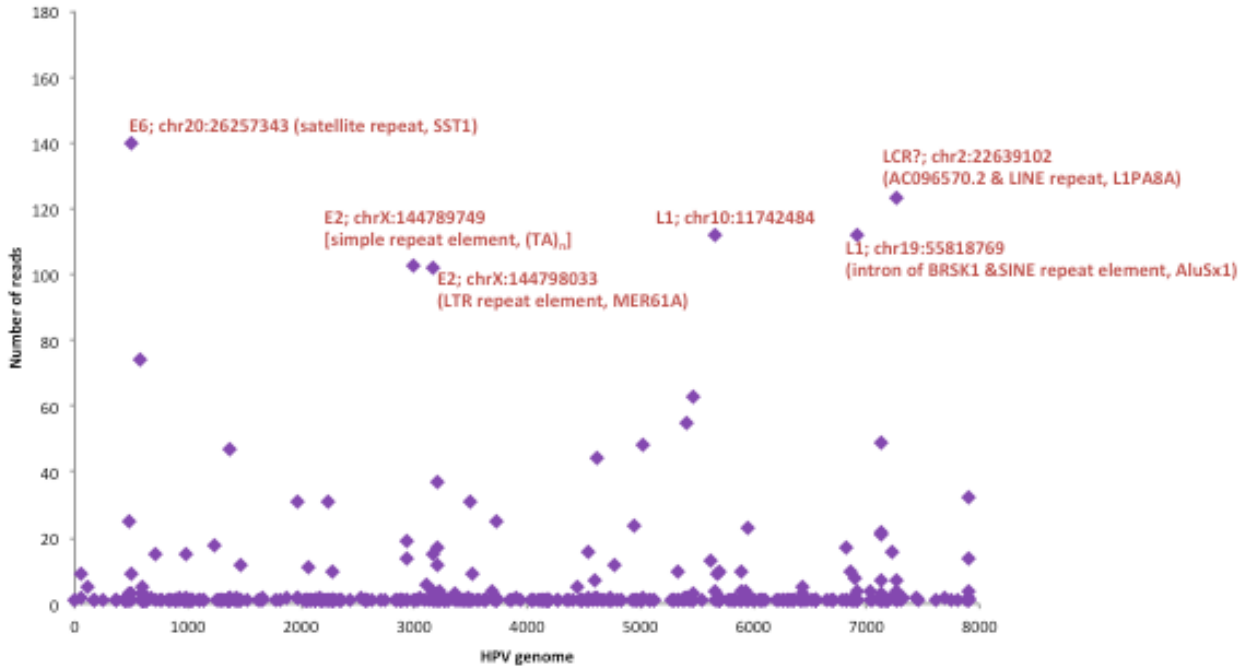
Top figure: Human-HPV Mixed Paired-end Reads mapped to respective genome l(HPV and Human) locations.
Blue read amplitude represents frequency of viral sequence read for specific viral genome location (respective HPV open reading frame annotated in blue arrows, with exception of E4 ORF which falls within E2 ORF). Proximal and distal HPV genome (no arrow) represents the HPV long control region. Red localizing lines delineate location of corresponding paired read to chromosomal position (each red annotated arrow reflects the respective chromosome).
For high frequency specific chromosomal integration sites the viral gene and human chromosome are listed with chromosomal co-ordinates and gene annotations where apparent (source: UCSC genome browser & ensembl).

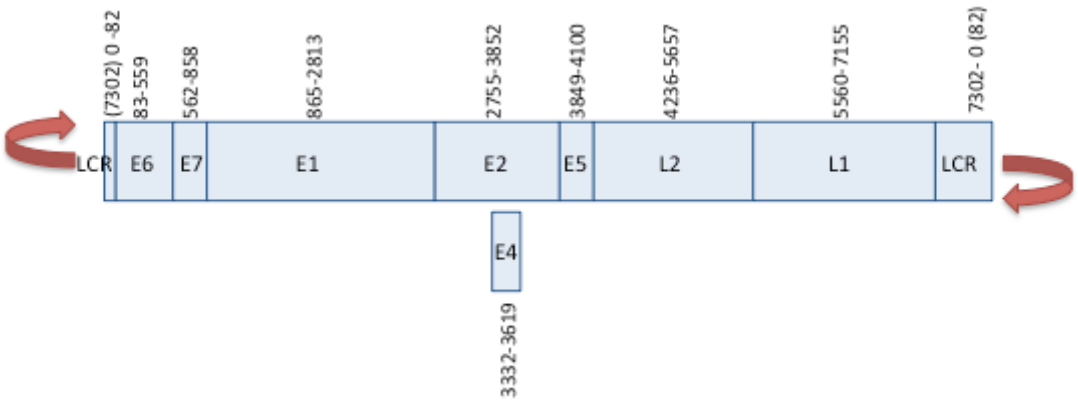
Middle Dot Plot: Graphic representation of viral cleavage points (breakpoints) within chimaeric reads distributed across the 7904bp circular genome (x axis). Frequency of cleavage point read is plotted on the y axis.
Chimaeric reads represent those reads for which both virus and human sequence are included in the same single read, thus demonstrating the exact point of transition between the sequences of differing origin (viral and human).

Bottom Schematic: Representation of the HPV16 genome highlighting the viral open reading frames and long control region (LCR) with respective nucleotide sequence span. Figures in parenthesis represents either start or end point for LCR which spans the 0/7904 circular start/end point for the HPV16 genome.

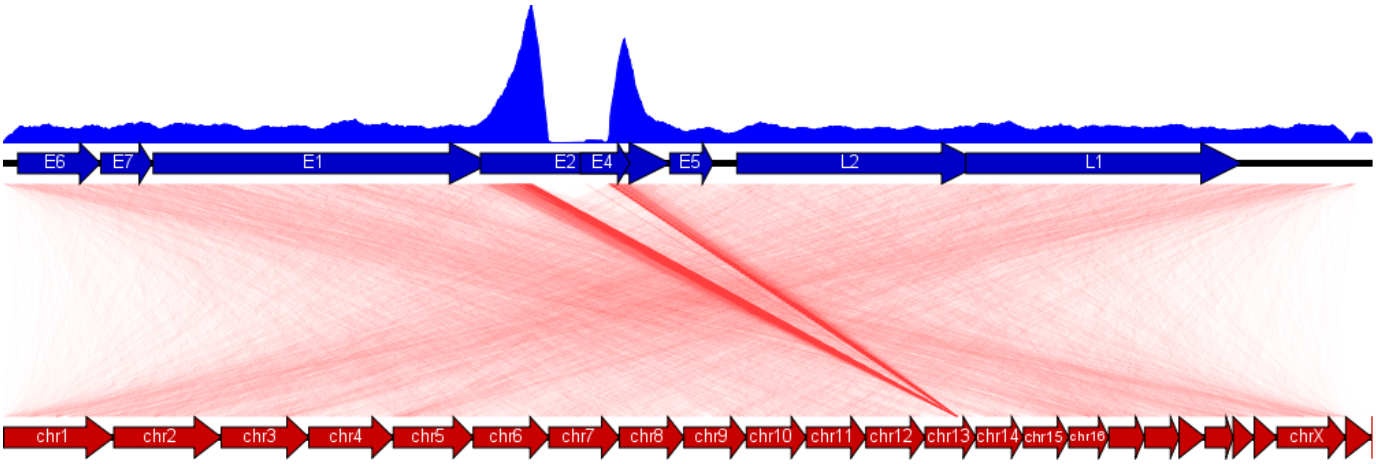
Appendix III: Viral & Host Sequencing Read Representations for the HPV16 positive Cell Line CaSki

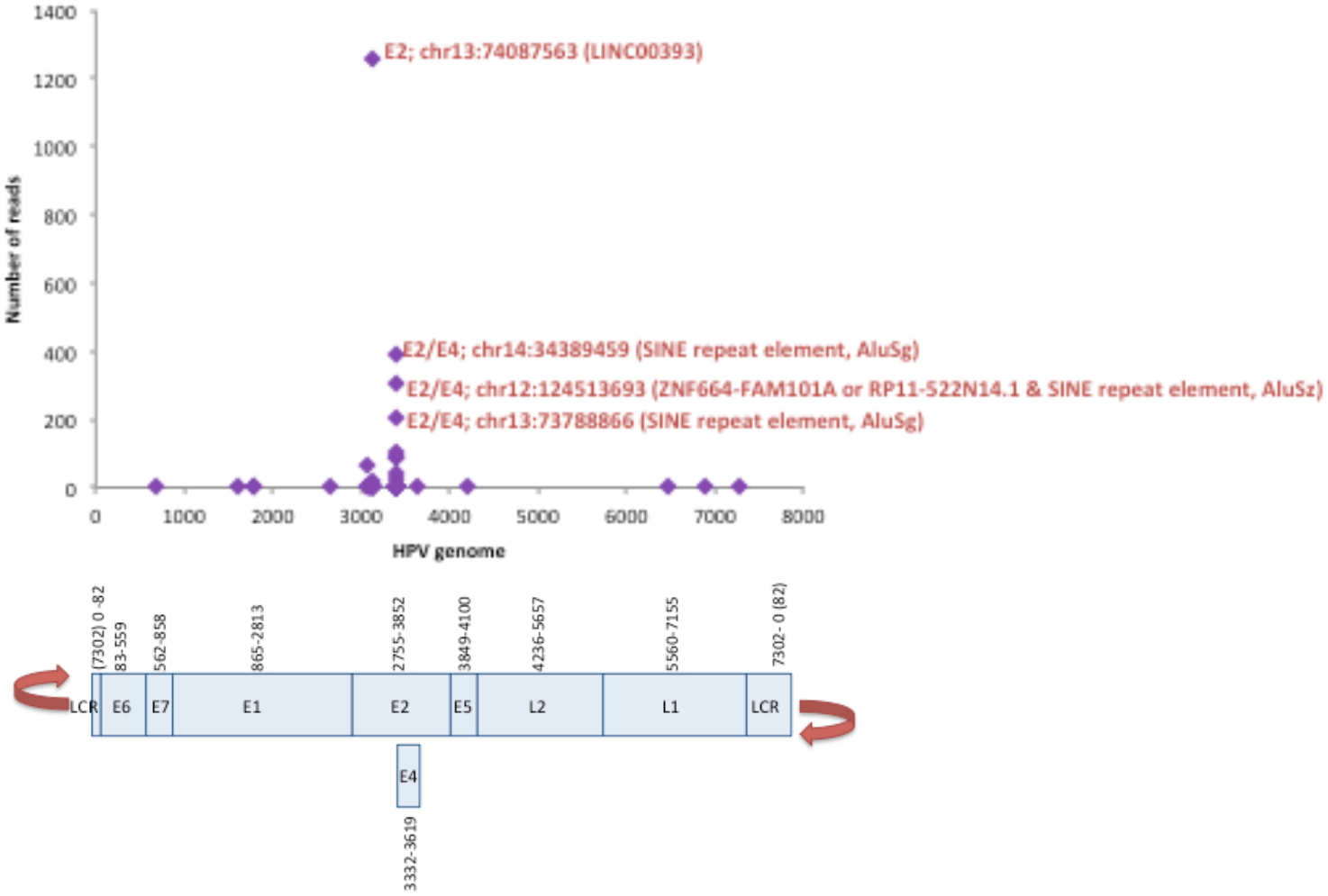




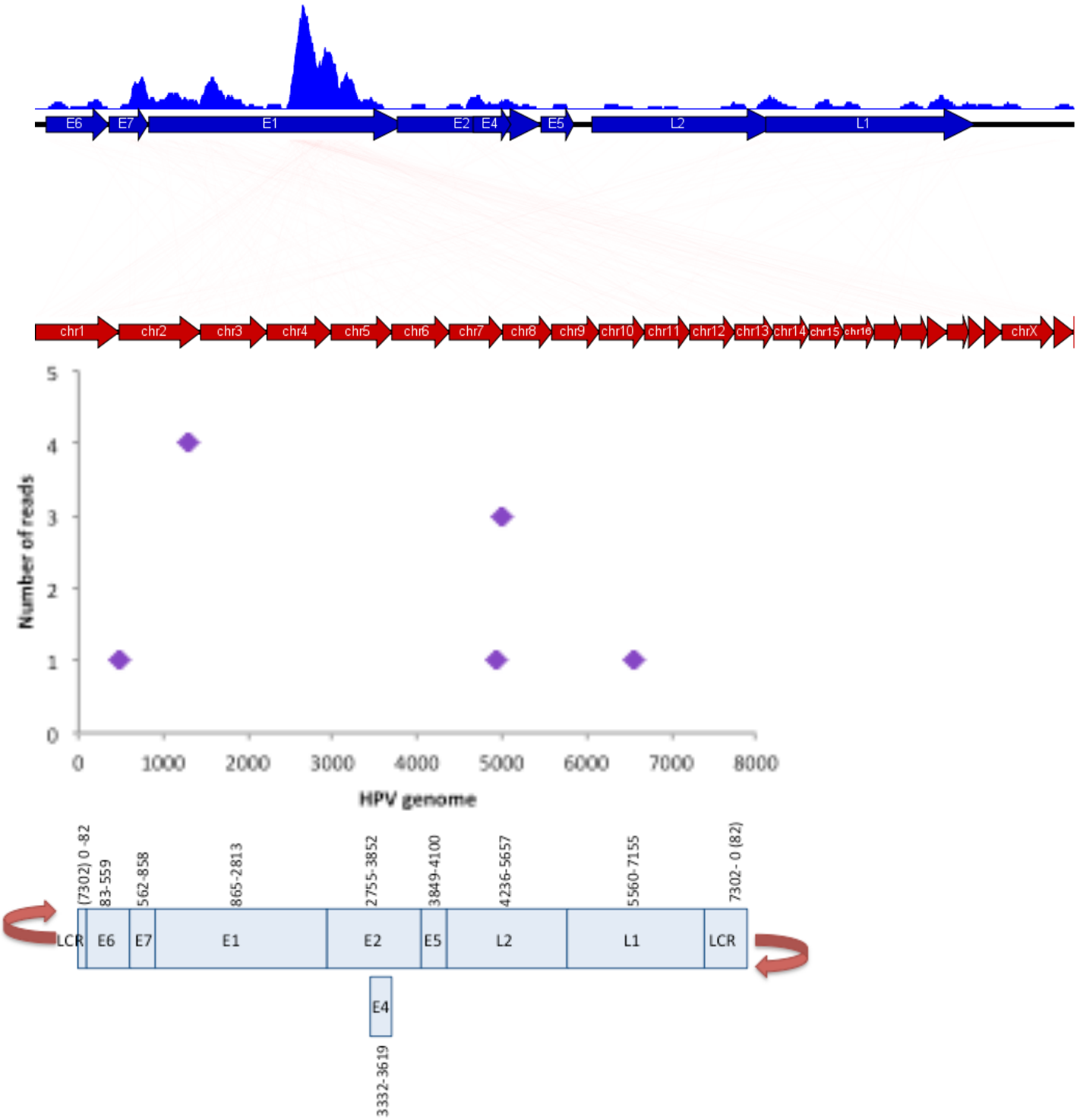


Appendix IV: Viral & Host Sequencing Read Representations for the HPV16 positive Cell Line SiHa

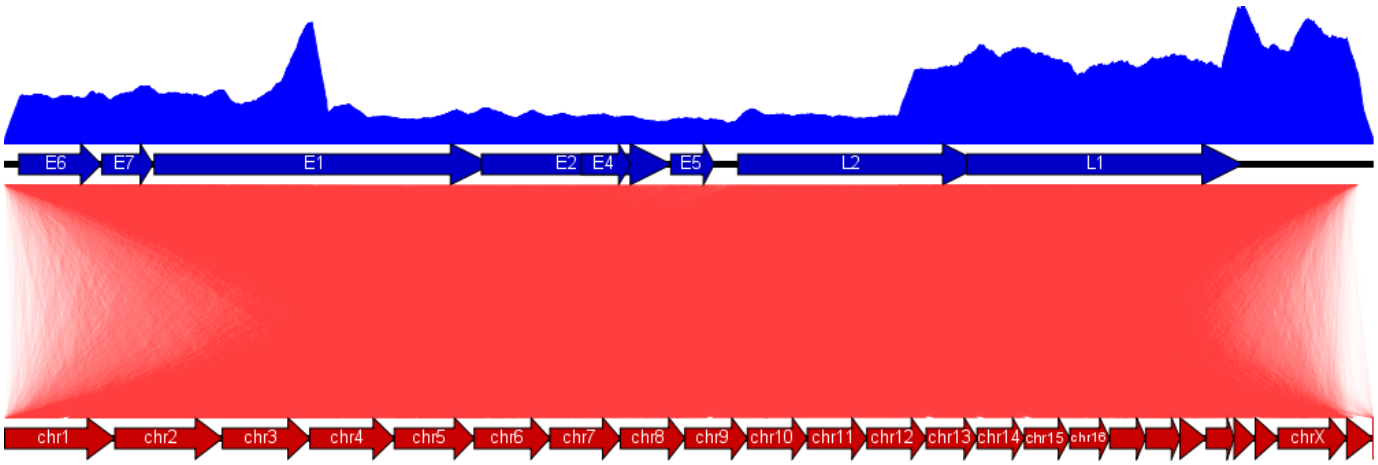


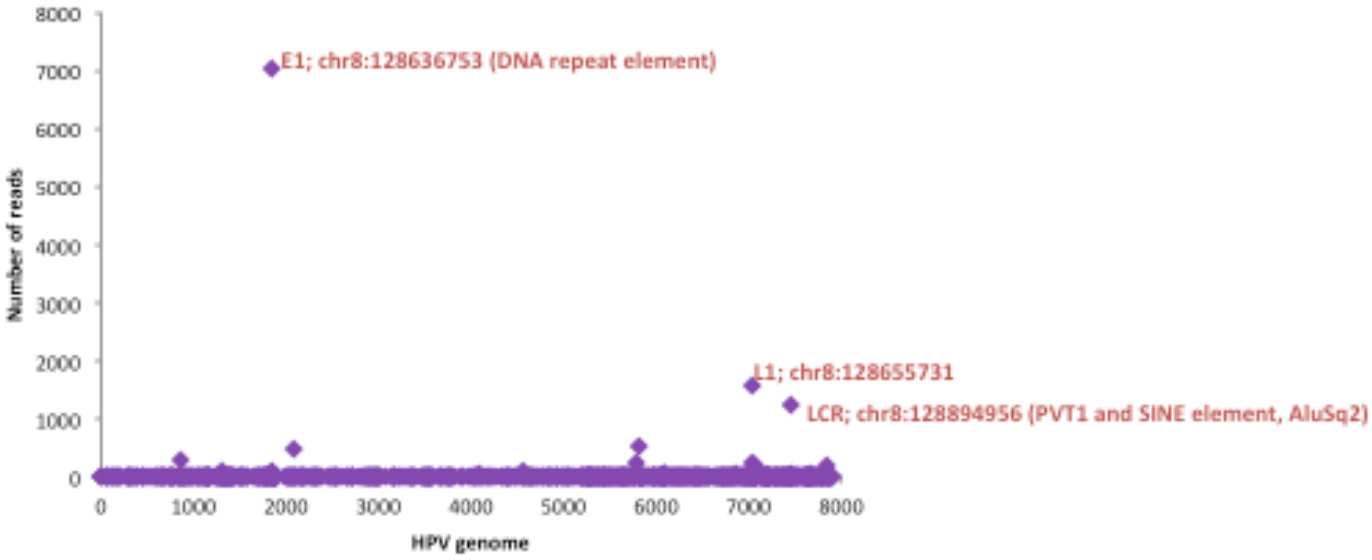


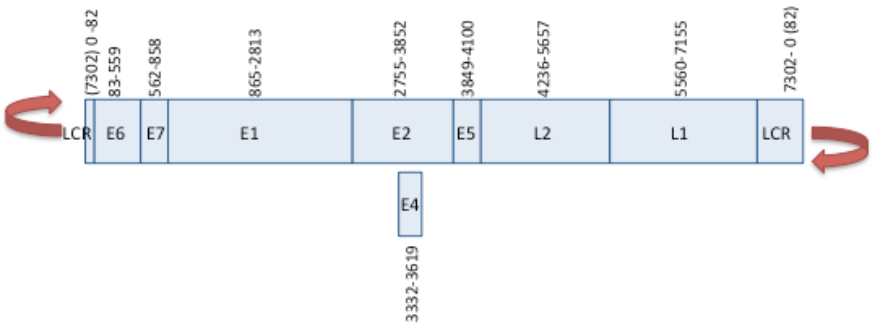
Appendix V: Viral & Host Sequencing Read Representations for the HPV16 positive OPSCC sample No. 13 (311T)



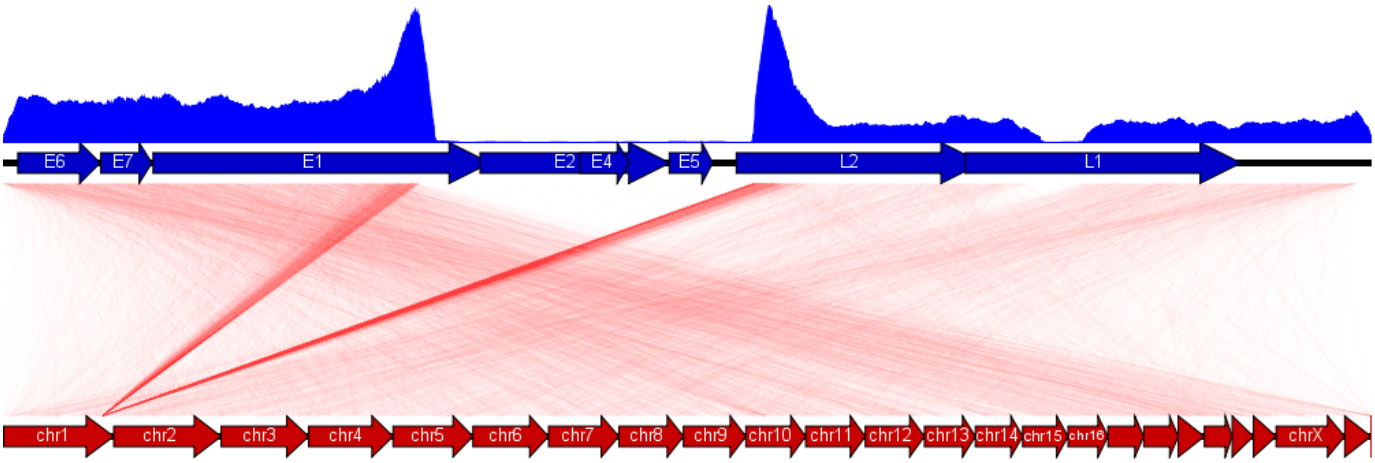
Appendix VI: Viral & Host Sequencing Read Representations for the HPV16 positive OPSCC sample No. 34 (427T)

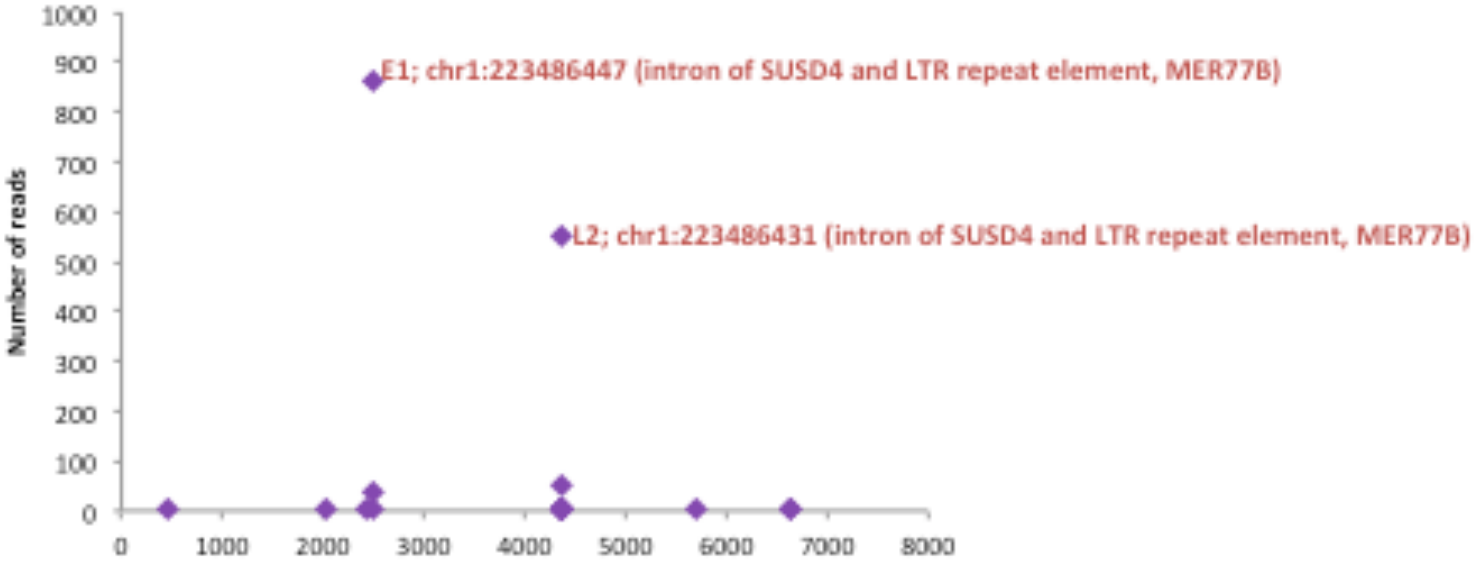


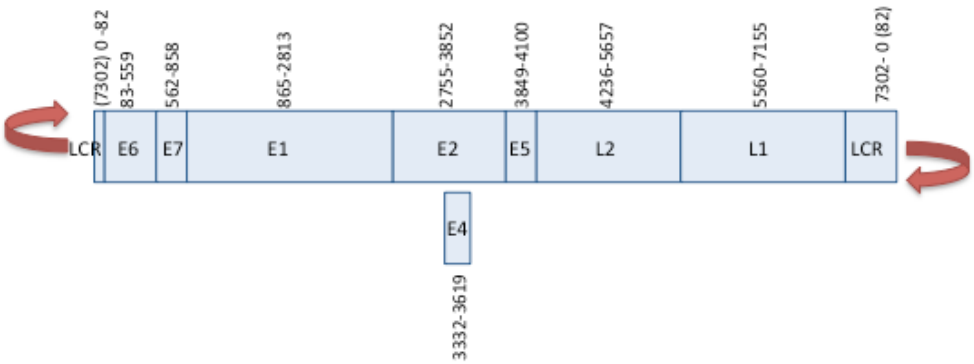




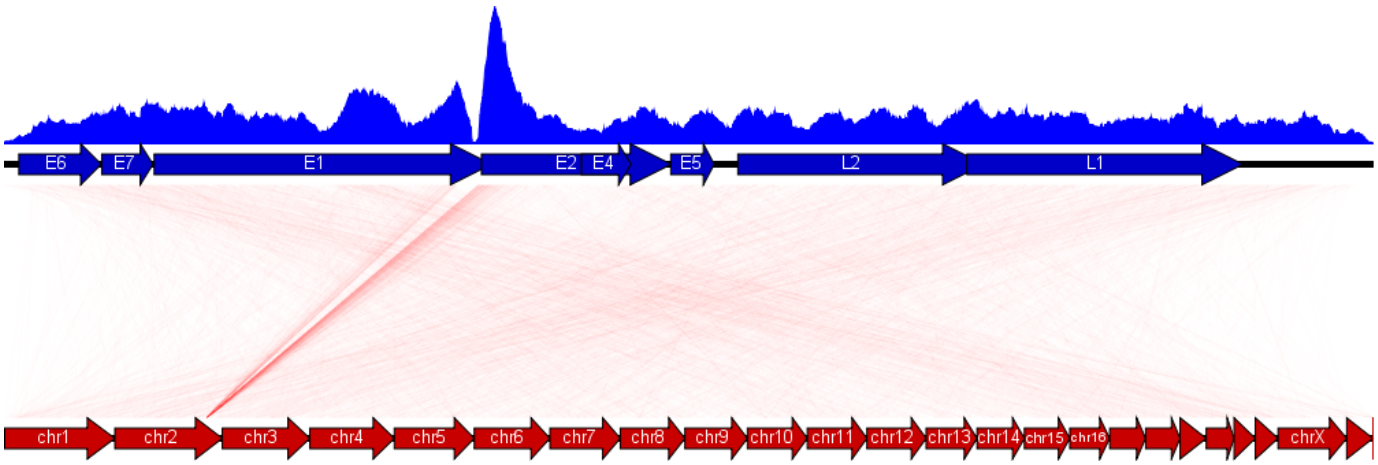
Appendix VII: Viral & Host Sequencing Read Representations for the HPV16 positive OPSCC sample No. 106 (045-09T)

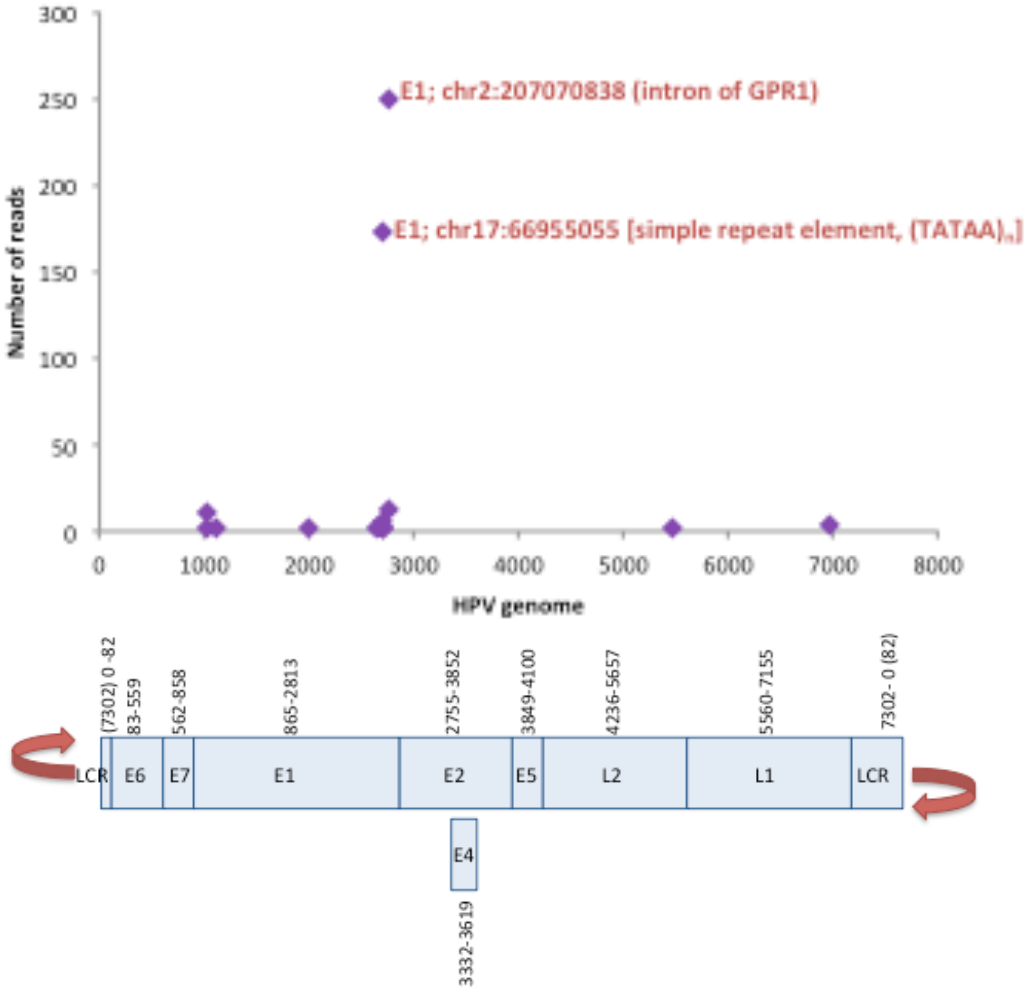




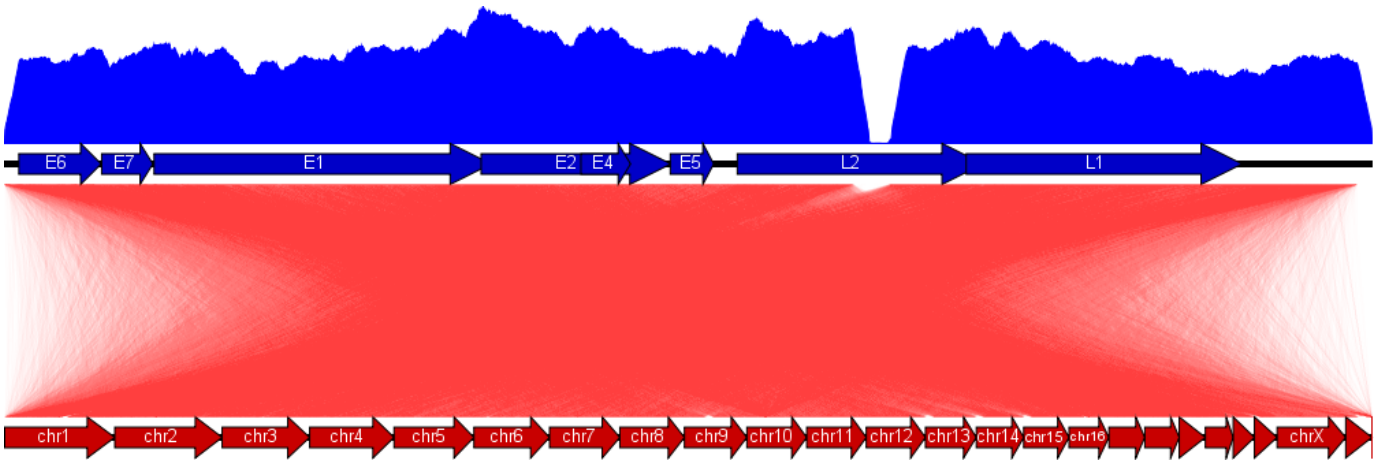


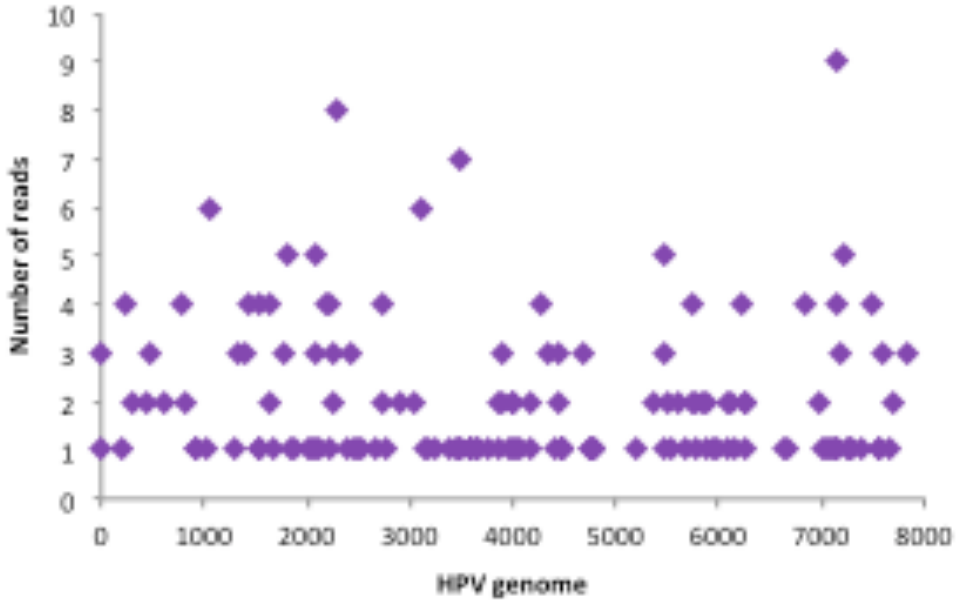
Appendix VIII: Viral & Host Sequencing Read Representations for the HPV16 positive OPSCC sample No. 95 (270-08T)

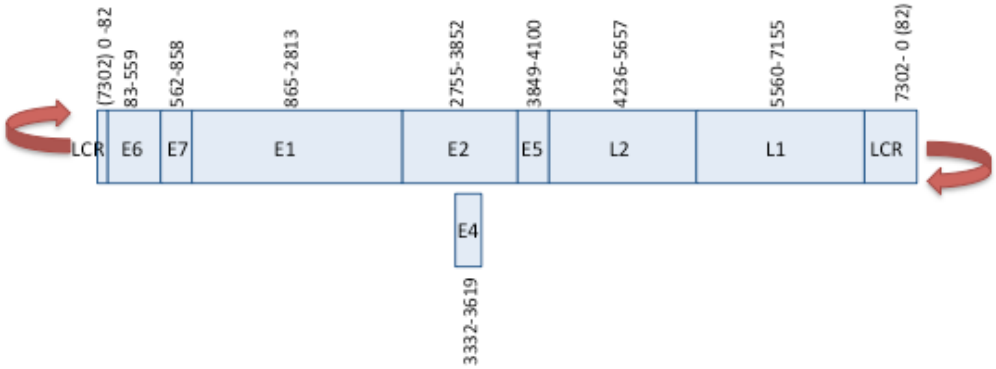




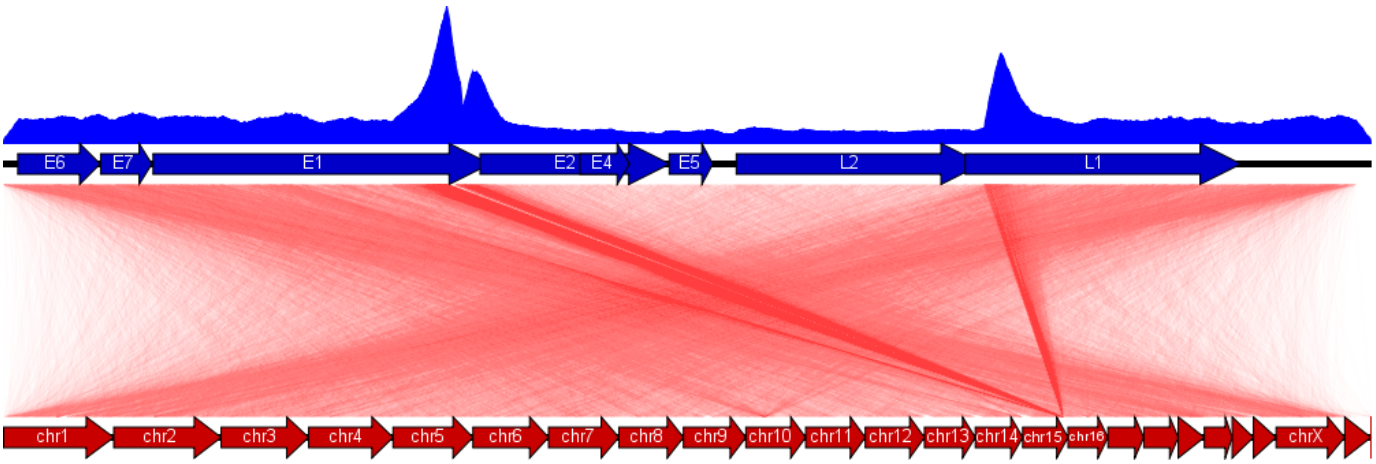
Appendix IX: Viral & Host Sequencing Read Representations for the HPV16 positive OPSCC samples No. 105 (043-09T)

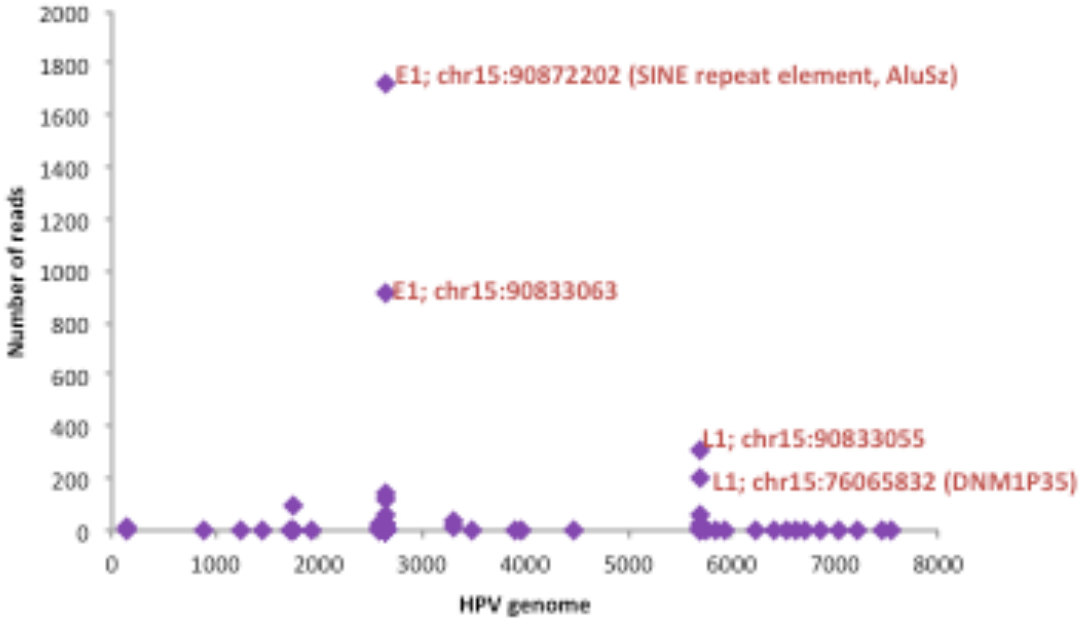


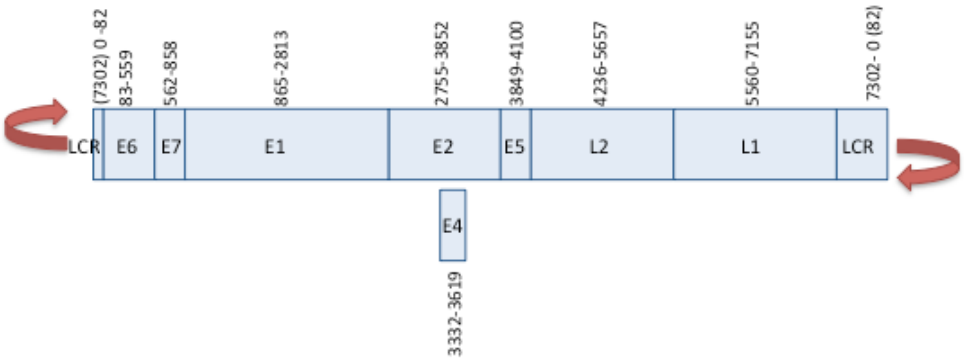




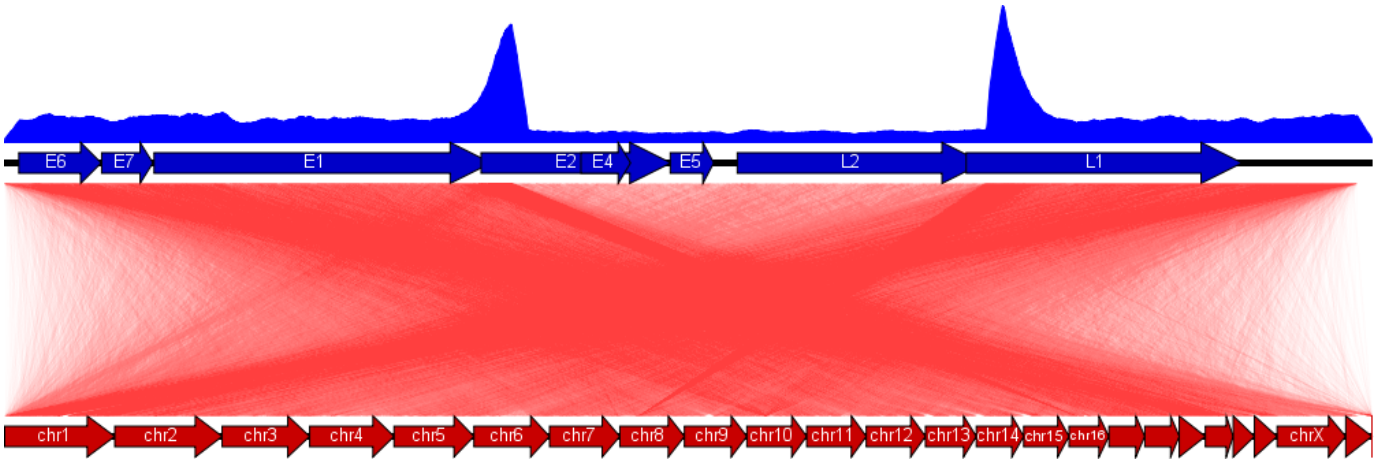
Appendix X: Viral & Host Sequencing Read Representations for the HPV16 positive OPSCC sample No. 88 (077-08T)

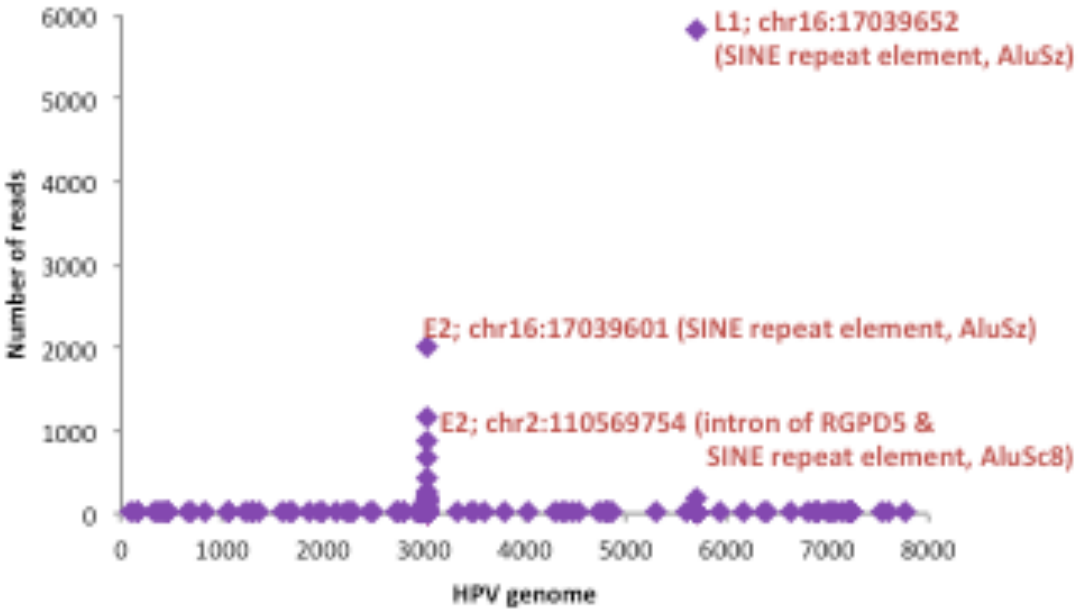


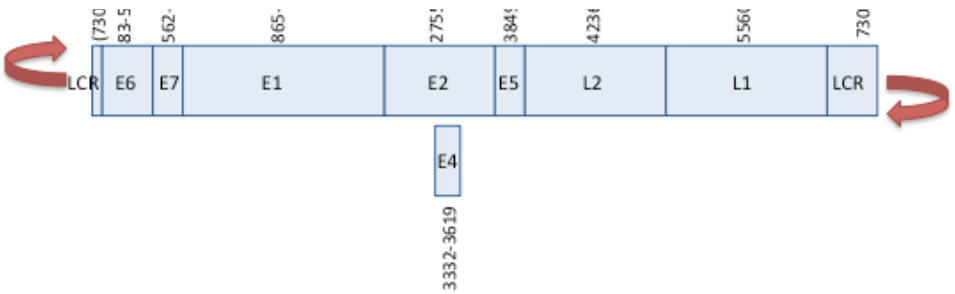




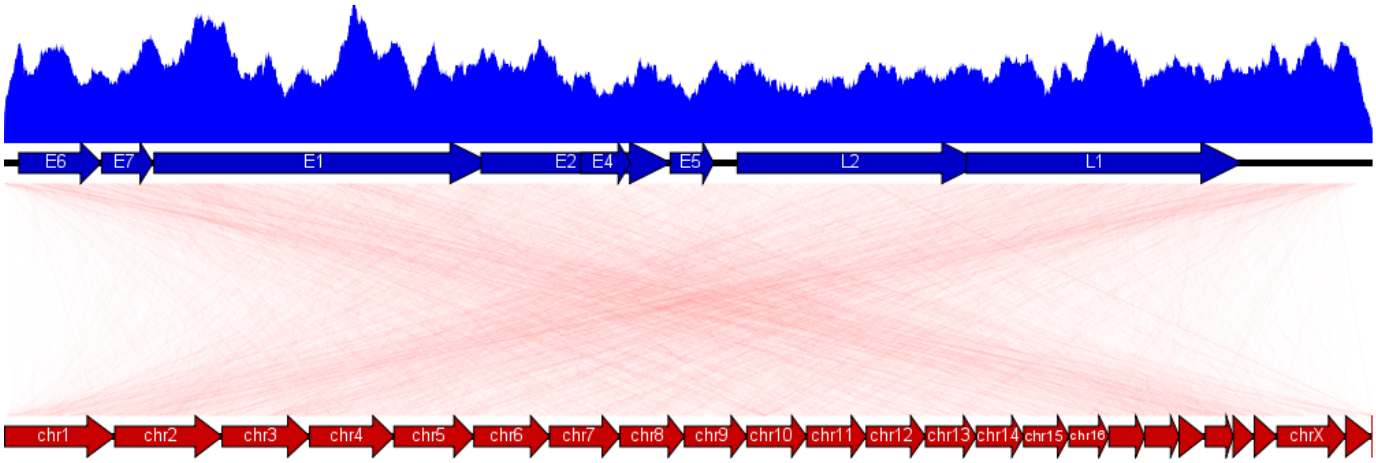
Appendix XI: Viral & Host Sequencing Read Representations for the HPV16 positive OPSCC sample No. 108 (239-08T)

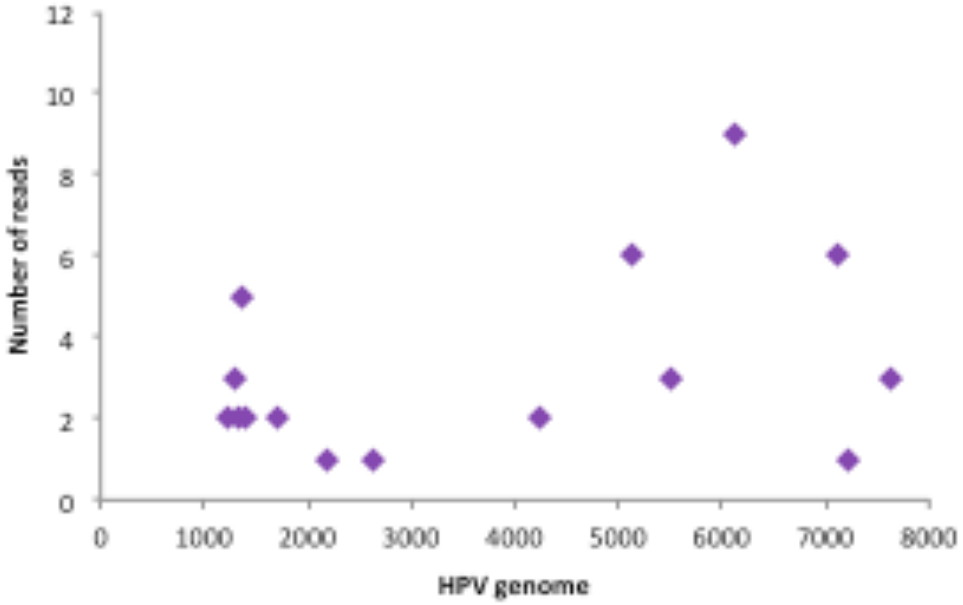


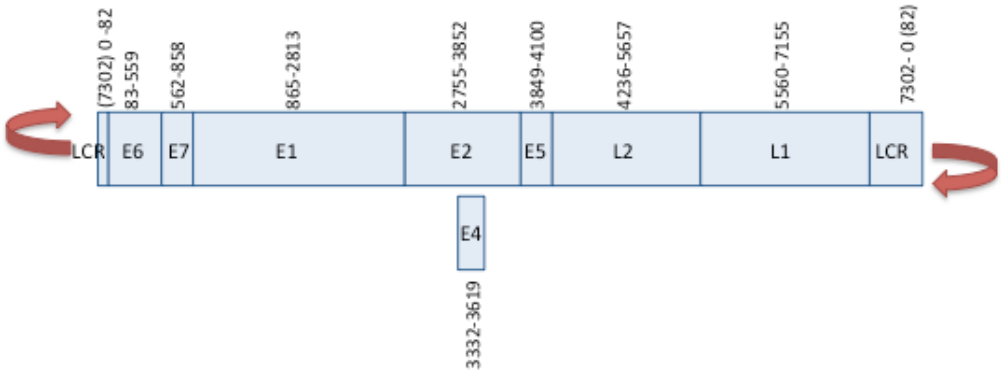




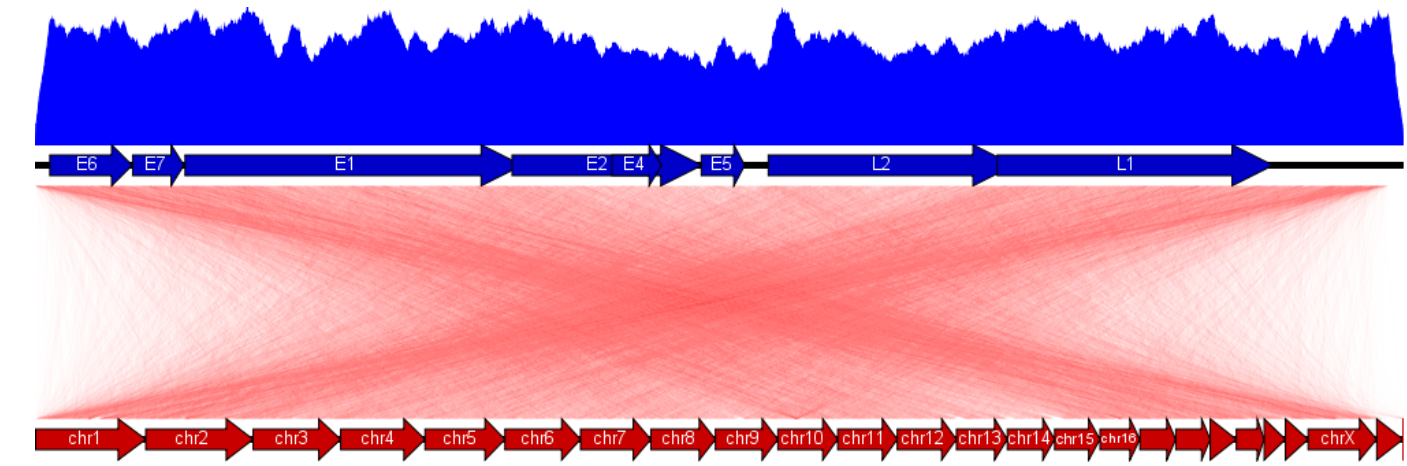
Appendix XII: Viral & Host Sequencing Read Representations for the HPV16 positive OPSCC sample No. 87 (075-08T)

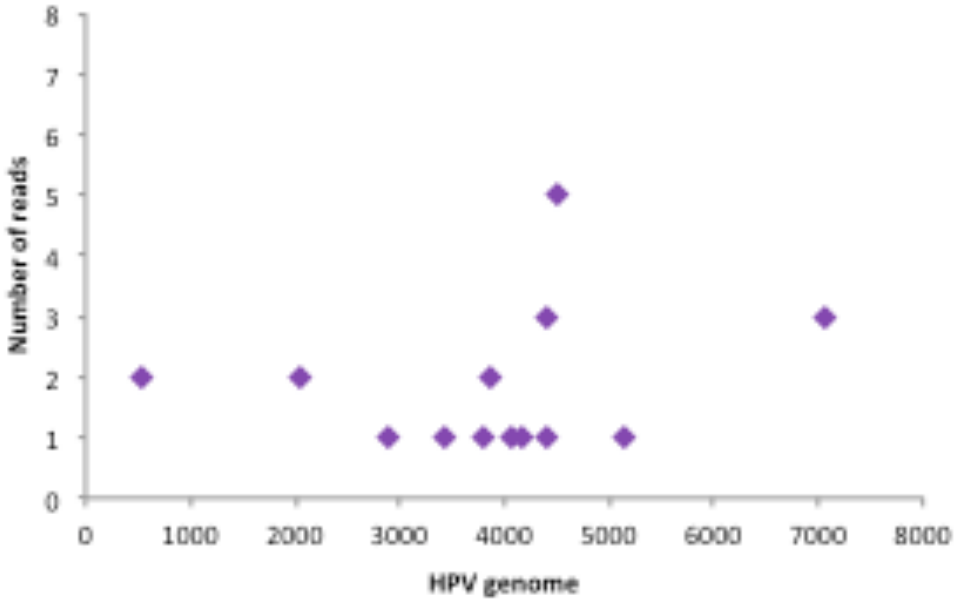


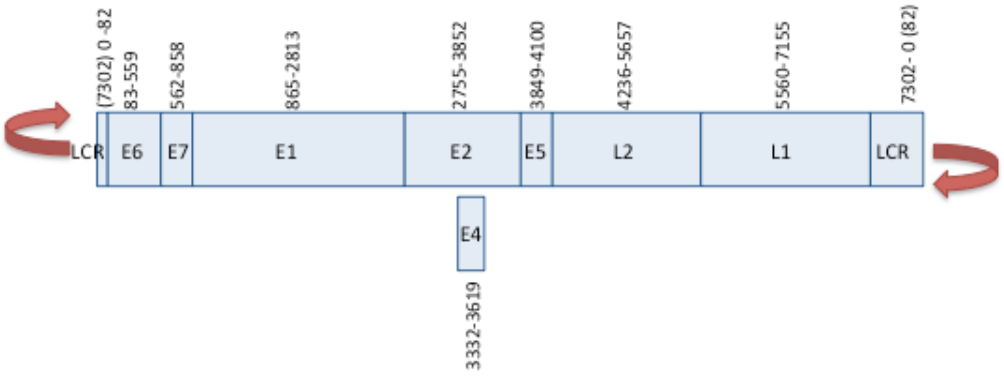




Appendix XIII: Viral & Host Sequencing Read Representations for the HPV16 positive OPSCC sample No. 101 (035-09T)







11 REFERENCES

1. Conway DI, Stockton DL, Warnakulasuriya KA, et al: Incidence of oral and oropharyngeal cancer in United Kingdom (1990-1999) -- recent trends and regional variation. Oral Oncol 42:586-92, 2006
2. Price G, Roche M, Crowther R, et al: Profile of Head and Neck Cancers in England - Incidence, Mortality and Survival. Oxford, United Kingdom, 2010
3. Marur S, D'Souza G, Westra WH, et al: HPV-associated head and neck cancer: a virus-related cancer epidemic. The lancet oncology 11:781-9, 2010
4. Chung CH, Gillison ML: Human papillomavirus in head and neck cancer: its role in pathogenesis and clinical implications. Clinical cancer research : an official journal of the American Association for Cancer Research 15:6758-62, 2009
5. Vinokurova S, Wentzensen N, Kraus I, et al: Type-dependent integration frequency of human papillomavirus genomes in cervical lesions. Cancer Res 68:307-13, 2008
6. Pett M, Coleman N: Integration of high-risk human papillomavirus: a key event in cervical carcinogenesis? J Pathol 212:356-67, 2007
7. Clarke MA, Wentzensen N, Mirabello L, et al: Human papillomavirus DNA methylation as a potential biomarker for cervical cancer. Cancer Epidemiol Biomarkers Prev 21:2125-37, 2012
8. de Villiers EM, Fauquet C, Broker TR, et al: Classification of papillomaviruses. Virology 324:17-27, 2004
9. Rautava J, Kuuskoski J, Syrjanen K, et al: HPV genotypes and their prognostic significance in head and neck squamous cell carcinomas. Journal of clinical virology : the official publication of the Pan American Society for Clinical Virology 53:116-20, 2012
10. Upile N: HPV associated Oropharyngeal Squamous Cell Carcinoma (OPSCC). The Otolaryngologist 4:57-61, 2011
11. Miller DL, Puricelli MD, Stack MS: Virology and molecular pathogenesis of HPV (human papillomavirus)-associated oropharyngeal squamous cell carcinoma. The Biochemical journal 443:339-53, 2012
12. Bernard HU: The clinical importance of the nomenclature, evolution and taxonomy of human papillomaviruses. Journal of clinical virology : the official publication of the Pan American Society for Clinical Virology 32 Suppl 1:S1-6, 2005
13. Nindl I, Gottschling M, Stockfleth E: Human papillomaviruses and non-melanoma skin cancer: basic virology and clinical manifestations. Disease markers 23:247-59, 2007

14. Ghittoni R, Accardi R, Hasan U, et al: The biological properties of E6 and E7 oncoproteins from human papillomaviruses. Virus genes 40:1-13, 2010

15. Hegde RS: The papillomavirus E2 proteins: structure, function, and biology. Annu Rev Biophys Biomol Struct 31:343-60, 2002

16. Rogers A, Waltke M, Angeletti PC: Evolutionary variation of papillomavirus E2 protein and E2 binding sites. Virology journal 8:379, 2011

17. Schwarz E, Freese UK, Gissmann L, et al: Structure and transcription of human papillomavirus sequences in cervical carcinoma cells. Nature 314:111-4, 1985

18. Buck CB, Cheng N, Thompson CD, et al: Arrangement of L2 within the papillomavirus capsid. Journal of virology 82:5190-7, 2008

19. Schiller JT, Day PM, Kines RC: Current understanding of the mechanism of HPV infection. Gynecologic oncology 118:S12-7, 2010

20. Roberts JN, Buck CB, Thompson CD, et al: Genital transmission of HPV in a mouse model is potentiated by nonoxynol-9 and inhibited by carrageenan. Nature medicine 13:857-61, 2007

21. Kajiji S, Tamura RN, Quaranta V: A novel integrin (alpha E beta 4) from human epithelial cells suggests a fourth family of integrin adhesion receptors. The EMBO journal 8:673-80, 1989

22. Evander M, Frazer IH, Payne E, et al: Identification of the alpha6 integrin as a candidate receptor for papillomaviruses. Journal of virology 71:2449-56, 1997

23. Horvath CA, Boulet GA, Renoux VM, et al: Mechanisms of cell entry by human papillomaviruses: an overview. Virology journal 7:11, 2010

24. Day PM, Lowy DR, Schiller JT: Papillomaviruses infect cells via a clathrin-dependent pathway. Virology 307:1-11, 2003

25. Bedell MA, Hudson JB, Golub TR, et al: Amplification of human papillomavirus genomes in vitro is dependent on epithelial differentiation. Journal of virology 65:2254-60, 1991

26. Stanley MA, Pett MR, Coleman N: HPV: from infection to cancer. Biochemical Society transactions 35:1456-60, 2007

27. Middleton K, Peh W, Southern S, et al: Organization of human papillomavirus productive cycle during neoplastic progression provides a basis for selection of diagnostic markers. Journal of virology 77:10186-201, 2003

28. Doorbar J: The papillomavirus life cycle. Journal of clinical virology : the official publication of the Pan American Society for Clinical Virology 32 Suppl 1:S7-15, 2005

29. Kanodia S, Fahey LM, Kast WM: Mechanisms used by human papillomaviruses to escape the host immune response. Curr Cancer Drug Targets 7:79-89, 2007

30. Petry KU, Scheffel D, Bode U, et al: Cellular immunodeficiency enhances the progression of human papillomavirus-associated cervical lesions. Int J Cancer 57:836-40, 1994

31. Venuti A, Paolini F, Nasir L, et al: Papillomavirus E5: the smallest oncoprotein with many functions. Mol Cancer 10:140, 2011

32. Kademani D: Oral cancer. Mayo Clinic proceedings. Mayo Clinic 82:878-87, 2007

33. Hayat MJ, Howlader N, Reichman ME, et al: Cancer statistics, trends, and multiple primary cancer analyses from the Surveillance, Epidemiology, and End Results (SEER) Program. Oncologist 12:20-37, 2007

34. Warnakulasuriya S: Global epidemiology of oral and oropharyngeal cancer. Oral oncology 45:309-16, 2009

35. Johnson NW, Jayasekara P, Amarasinghe AA: Squamous cell carcinoma and precursor lesions of the oral cavity: epidemiology and aetiology. Periodontol 2000 57:19-37, 2011

36. Rogers SN, Brown JS, Woolgar JA, et al: Survival following primary surgery for oral cancer. Oral oncology 45:201-11, 2009

37. Mehanna H, West CM, Nutting C, et al: Head and neck cancer--Part 2: Treatment and prognostic factors. BMJ 341:c4690, 2010

38. Chaturvedi AK: Epidemiology and Clinical Aspects of HPV in Head and Neck Cancers. Head Neck Pathol 6 Suppl 1:16-24, 2012

39. Boyle PaL, B: World Cancer Report 2008. Lyon, France, IARC Press, 2008

40. Meurman JH, Uittamo J: Oral micro-organisms in the etiology of cancer. Acta Odontol Scand 66:321-6, 2008

41. Syrjanen SM, Syrjanen KJ, Happonen RP: Human papillomavirus (HPV) DNA sequences in oral precancerous lesions and squamous cell carcinoma demonstrated by in situ hybridization. J Oral Pathol 17:273-8, 1988

42. Palefsky JM, Silverman S, Jr., Abdel-Salaam M, et al: Association between proliferative verrucous leukoplakia and infection with human papillomavirus type 16. J Oral Pathol Med 24:193-7, 1995

43. Syrjanen K, Vayrynen M, Castren O, et al: Morphological and immunohistochemical evidence of human papilloma virus (HPV) involvement in the dysplastic lesions of the uterine cervix. Int J Gynaecol Obstet 21:261-9, 1983

44. Bouvard V, Baan R, Straif K, et al: A review of human carcinogens--Part B: biological agents. Lancet Oncol 10:321-2, 2009

45. Scully C, Field JK, Tanzawa H: Genetic aberrations in oral or head and neck squamous cell carcinoma (SCCHN): 1. Carcinogen metabolism, DNA repair and cell cycle control. Oral Oncol 36:256-63, 2000

46. Sobin L: TNM Classification of Malignant Tumors (ed 7th). Oxford, Wiley-Blackwell, 2009

47. World Health Organization.: International statistical classification of diseases and related health problems (ed 10th revision, 2nd edition.). Geneva, World Health Organization, 2004

48. Moore KL, Persaud TVN, Torchia MG: Before we are born : essentials of embryology and birth defects (ed 7th). Philadelphia, PA, Saunders/Elsevier, 2008

49. Moore KL, Dalley AF, Agur AMR: Clinically oriented anatomy (ed 6th). Philadelphia, Wolters Kluwer/Lippincott Williams & Wilkins, 2010

50. Shaw R, Robinson M: The increasing clinical relevance of human papillomavirus type 16 (HPV-16) infection in oropharyngeal cancer. Br J Oral Maxillofac Surg, 2010

51. D'Souza G, Kreimer AR, Viscidi R, et al: Case-control study of human papillomavirus and oropharyngeal cancer. N Engl J Med 356:1944-56, 2007

52. Hobbs CG, Sterne JA, Bailey M, et al: Human papillomavirus and head and neck cancer: a systematic review and meta-analysis. Clinical otolaryngology : official journal of ENT-UK ; official journal of Netherlands Society for Oto-Rhino-Laryngology & Cervico-Facial Surgery 31:259-66, 2006

53. Klussmann JP, Weissenborn SJ, Wieland U, et al: Human papillomavirus-positive tonsillar carcinomas: a different tumor entity? Med Microbiol Immunol 192:129-32, 2003

54. Perry ME: The specialised structure of crypt epithelium in the human palatine tonsil and its functional significance. J Anat 185 (Pt 1):111-27, 1994

55. Evans M, Powell NG: The changing aetiology of head and neck cancer: the role of human papillomavirus. Clinical oncology 22:538-46, 2010

56. Schache AG, Lieger O, Rogers P, et al: Predictors of swallowing outcome in patients treated with surgery and radiotherapy for advanced oral and oropharyngeal cancer. Oral Oncol 45:803-8, 2009

57. Nicoletti G, Soutar DS, Jackson MS, et al: Chewing and swallowing after surgical treatment for oral cancer: functional evaluation in 196 selected cases. Plast Reconstr Surg 114:329-38, 2004

58. de Martel C, Ferlay J, Franceschi S, et al: Global burden of cancers attributable to infections in 2008: a review and synthetic analysis. Lancet Oncol 13:607-15, 2012

59. Chesson HW, Blandford JM, Gift TL, et al: The estimated direct medical cost of sexually transmitted diseases among American youth, 2000. Perspect Sex Reprod Health 36:11-9, 2004

60. Parkin DM, Bray F: Chapter 2: The burden of HPV-related cancers. Vaccine 24 Suppl 3:S3/11-25, 2006

61. Daling JR, Madeleine MM, Johnson LG, et al: Human papillomavirus, smoking, and sexual practices in the etiology of anal cancer. Cancer 101:270-80, 2004

62. Munoz N, Castellsague X, de Gonzalez AB, et al: Chapter 1: HPV in the etiology of human cancer. Vaccine 24 Suppl 3:S3/1-10, 2006

63. Chaturvedi AK, Engels EA, Pfeiffer RM, et al: Human Papillomavirus and Rising Oropharyngeal Cancer Incidence in the United States. Journal of clinical oncology : official journal of the American Society of Clinical Oncology, 2011

64. Campisi G, Giovannelli L: Controversies surrounding human papilloma virus infection, head & neck vs oral cancer, implications for prophylaxis and treatment. Head & neck oncology 1:8, 2009

65. Herrero R, Castellsague X, Pawlita M, et al: Human papillomavirus and oral cancer: the International Agency for Research on Cancer multicenter study. J Natl Cancer Inst 95:1772-83, 2003

66. Dubina M, Goldenberg G: Viral-associated nonmelanoma skin cancers: a review. Am J Dermatopathol 31:561-73, 2009

67. Hariri S, Unger ER, Sternberg M, et al: Prevalence of genital human papillomavirus among females in the United States, the National Health And Nutrition Examination Survey, 2003-2006. J Infect Dis 204:566-73, 2011

68. Schiffman M, Wentzensen N, Wacholder S, et al: Human papillomavirus testing in the prevention of cervical cancer. J Natl Cancer Inst 103:368-83, 2011

69. Burchell AN, Winer RL, de Sanjose S, et al: Chapter 6: Epidemiology and transmission dynamics of genital HPV infection. Vaccine 24 Suppl 3:S3/52-61, 2006

70. zur Hausen H: Papillomaviruses and cancer: from basic studies to clinical application. Nature reviews. Cancer 2:342-50, 2002

71. Gillison ML, Broutian T, Pickard RK, et al: Prevalence of oral HPV infection in the United States, 2009-2010. JAMA 307:693-703, 2012

72. Hernandez BY, Wilkens LR, Zhu X, et al: Transmission of human papillomavirus in heterosexual couples. Emerg Infect Dis 14:888-94, 2008

73. Fakhry C, Sugar E, D'Souza G, et al: Two-week versus six-month sampling interval in a short-term natural history study of oral HPV infection in an HIV-positive cohort. PLoS One 5:e11918, 2010

74. D'Souza G, Fakhry C, Sugar EA, et al: Six-month natural history of oral versus cervical human papillomavirus infection. Int J Cancer 121:143-50, 2007

75. Heck JE, Berthiller J, Vaccarella S, et al: Sexual behaviours and the risk of head and neck cancers: a pooled analysis in the International Head and Neck Cancer Epidemiology (INHANCE) consortium. Int J Epidemiol 39:166-81, 2010

76. Kreimer AR, Clifford GM, Boyle P, et al: Human papillomavirus types in head and neck squamous cell carcinomas worldwide: a systematic review. Cancer epidemiology, biomarkers & prevention : a publication of the

American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology 14:467-75, 2005

77. Gillison ML, Koch WM, Capone RB, et al: Evidence for a causal association between human papillomavirus and a subset of head and neck cancers. Journal of the National Cancer Institute 92:709-20, 2000

78. Begum S, Cao D, Gillison M, et al: Tissue distribution of human papillomavirus 16 DNA integration in patients with tonsillar carcinoma. Clin Cancer Res 11:5694-9, 2005

79. Kreimer AR, Clifford GM, Snijders PJ, et al: HPV16 semiquantitative viral load and serologic biomarkers in oral and oropharyngeal squamous cell carcinomas. Int J Cancer 115:329-32, 2005

80. Braakhuis BJ, Snijders PJ, Keune WJ, et al: Genetic patterns in head and neck cancers that contain or lack transcriptionally active human papillomavirus. J Natl Cancer Inst 96:998-1006, 2004

81. Ribeiro KB, Levi JE, Pawlita M, et al: Low human papillomavirus prevalence in head and neck cancer: results from two large case-control studies in high-incidence regions. Int J Epidemiol 40:489-502, 2011

82. Ramqvist T, Dalianis T: An epidemic of oropharyngeal squamous cell carcinoma (OSCC) due to human papillomavirus (HPV) infection and aspects of treatment and prevention. Anticancer research 31:1515-9, 2011

83. Blomberg M, Nielsen A, Munk C, et al: Trends in head and neck cancer incidence in Denmark, 1978-2007: focus on human papillomavirus associated sites. Int J Cancer 129:733-41, 2011

84. Hocking JS, Stein A, Conway EL, et al: Head and neck cancer in Australia between 1982 and 2005 show increasing incidence of potentially HPV-associated oropharyngeal cancers. Br J Cancer 104:886-91, 2011

85. Ioka A, Tsukuma H, Ajiki W, et al: Trends in head and neck cancer incidence in Japan during 1965-1999. Jpn J Clin Oncol 35:45-7, 2005

86. Junor EJ, Kerr GR, Brewster DH: Oropharyngeal cancer. Fastest increasing cancer in Scotland, especially in men. BMJ 340:c2512, 2010

87. Mork J, Moller B, Dahl T, et al: Time trends in pharyngeal cancer incidence in Norway 1981-2005: a subsite analysis based on a reabstraction and recoding of registered cases. Cancer Causes Control 21:1397-405, 2010

88. Reddy VM, Cundall-Curry D, Bridger MW: Trends in the incidence rates of tonsil and base of tongue cancer in England, 1985-2006. Ann R Coll Surg Engl 92:655-9, 2010

89. Syrjanen S: HPV infections and tonsillar carcinoma. Journal of clinical pathology 57:449-55, 2004

90. Rietbergen MM, Leemans CR, Bloemena E, et al: Increasing prevalence rates of HPV attributable oropharyngeal squamous cell carcinomas in the Netherlands as assessed by a validated test algorithm. Int J Cancer 132:1565-71, 2013

91. Ang KK, Harris J, Wheeler R, et al: Human papillomavirus and survival of patients with oropharyngeal cancer. N Engl J Med 363:24-35, 2010
92. Gillison ML, Koch WM, Capone RB, et al: Evidence for a causal association between human papillomavirus and a subset of head and neck cancers. J Natl Cancer Inst 92:709-20, 2000
93. Goldenberg D, Begum S, Westra WH, et al: Cystic lymph node metastasis in patients with head and neck cancer: An HPV-associated phenomenon. Head Neck 30:898-903, 2008
94. Fakhry C, Gillison ML: Clinical implications of human papillomavirus in head and neck cancers. J Clin Oncol 24:2606-11, 2006
95. D'Souza G, Zhang HH, D'Souza WD, et al: Moderate predictive value of demographic and behavioral characteristics for a diagnosis of HPV16-positive and HPV16-negative head and neck cancer. Oral Oncol 46:100-4, 2010
96. Mellin H, Friesland S, Lewensohn R, et al: Human papillomavirus (HPV) DNA in tonsillar cancer: clinical correlates, risk of relapse, and survival. Int J Cancer 89:300-4, 2000
97. Ragin CC, Taioli E: Survival of squamous cell carcinoma of the head and neck in relation to human papillomavirus infection: review and meta-analysis. Int J Cancer 121:1813-20, 2007
98. Fakhry C, Westra WH, Li S, et al: Improved survival of patients with human papillomavirus-positive head and neck squamous cell carcinoma in a prospective clinical trial. J Natl Cancer Inst 100:261-9, 2008
99. Licitra L, Perrone F, Bossi P, et al: High-risk human papillomavirus affects prognosis in patients with surgically treated oropharyngeal squamous cell carcinoma. J Clin Oncol 24:5630-6, 2006
100. Lassen P, Eriksen JG, Hamilton-Dutoit S, et al: Effect of HPV-associated p16INK4A expression on response to radiotherapy and survival in squamous cell carcinoma of the head and neck. J Clin Oncol 27:1992-8, 2009
101. Kumar B, Cordell KG, Lee JS, et al: EGFR, p16, HPV Titer, Bcl-xL and p53, sex, and smoking as indicators of response to therapy and survival in oropharyngeal cancer. J Clin Oncol 26:3128-37, 2008
102. Spanos WC, Nowicki P, Lee DW, et al: Immune response during therapy with cisplatin or radiation for human papillomavirus-related head and neck cancer. Arch Otolaryngol Head Neck Surg 135:1137-46, 2009
103. Hwang ES, Nottoli T, Dimaio D: The HPV16 E5 protein: expression, detection, and stable complex formation with transmembrane proteins in COS cells. Virology 211:227-33, 1995
104. Scheffner M, Werness BA, Huibregtse JM, et al: The E6 oncoprotein encoded by human papillomavirus types 16 and 18 promotes the degradation of p53. Cell 63:1129-36, 1990
105. Tommasino M, Accardi R, Caldeira S, et al: The role of TP53 in Cervical carcinogenesis. Human mutation 21:307-12, 2003

106. Howie HL, Katzenellenbogen RA, Galloway DA: Papillomavirus E6 proteins. Virology 384:324-34, 2009

107. McMurray HR, Nguyen D, Westbrook TF, et al: Biology of human papillomaviruses. International journal of experimental pathology 82:15-33, 2001

108. Huh K, Zhou X, Hayakawa H, et al: Human papillomavirus type 16 E7 oncoprotein associates with the cullin 2 ubiquitin ligase complex, which contributes to degradation of the retinoblastoma tumor suppressor. Journal of virology 81:9737-47, 2007

109. Khleif SN, DeGregori J, Yee CL, et al: Inhibition of cyclin D-CDK4/CDK6 activity is associated with an E2F-mediated induction of cyclin kinase inhibitor activity. Proc Natl Acad Sci U S A 93:4350-4, 1996

110. Wentzensen N, Vinokurova S, von Knebel Doeberitz M: Systematic review of genomic integration sites of human papillomavirus genomes in epithelial dysplasia and invasive cancer of the female lower genital tract. Cancer Res 64:3878-84, 2004

111. Magaldi TG, Almstead LL, Bellone S, et al: Primary human cervical carcinoma cells require human papillomavirus E6 and E7 expression for ongoing proliferation. Virology 422:114-24, 2012

112. Dowhanick JJ, McBride AA, Howley PM: Suppression of cellular proliferation by the papillomavirus E2 protein. Journal of virology 69:7791-9, 1995

113. Bellanger S, Tan CL, Xue YZ, et al: Tumor suppressor or oncogene? A critical role of the human papillomavirus (HPV) E2 protein in cervical cancer progression. Am J Cancer Res 1:373-389, 2011

114. Brandwein-Gensler M, Teixeira MS, Lewis CM, et al: Oral squamous cell carcinoma: histologic risk assessment, but not margin status, is strongly predictive of local disease-free and overall survival. Am J Surg Pathol 29:167-78, 2005

115. Jung AC, Guihard S, Krugell S, et al: CD8-alpha T-Cell infiltration in human papillomavirus-related oropharyngeal carcinoma correlates with improved patient prognosis. Int J Cancer, 2012

116. Wansom D, Light E, Worden F, et al: Correlation of cellular immunity with human papillomavirus 16 status and outcome in patients with advanced oropharyngeal cancer. Arch Otolaryngol Head Neck Surg 136:1267-73, 2010

117. Syrjanen S, Lodi G, von Bultzingslowen I, et al: Human papillomaviruses in oral carcinoma and oral potentially malignant disorders: a systematic review. Oral Dis 17 Suppl 1:58-72, 2011

118. Braakhuis BJ, Brakenhoff RH, Meijer CJ, et al: Human papilloma virus in head and neck cancer: the need for a standardised assay to assess the full clinical importance. Eur J Cancer 45:2935-9, 2009

119. Robinson M, Sloan P, Shaw R: Refining the diagnosis of oropharyngeal squamous cell carcinoma using human papillomavirus testing. Oral Oncol 46:492-6, 2010

120. Mehanna H, Jones TM, Gregoire V, et al: Oropharyngeal carcinoma related to human papillomavirus. BMJ 340:c1439, 2010

121. Scicchitano MS, Dalmas DA, Bertiaux MA, et al: Preliminary comparison of quantity, quality, and microarray performance of RNA extracted from formalin-fixed, paraffin-embedded, and unfixed frozen tissue samples. The journal of histochemistry and cytochemistry : official journal of the Histochemistry Society 54:1229-37, 2006

122. Srinivasan M, Sedmak D, Jewell S: Effect of fixatives and tissue processing on the content and integrity of nucleic acids. Am J Pathol 161:1961-71, 2002

123. Hall GL, Kademani D, Risk JM, et al: Tissue banking in head and neck cancer. Oral Oncol 44:109-15, 2008

124. Evers DL, He J, Kim YH, et al: Paraffin embedding contributes to RNA aggregation, reduced RNA yield, and low RNA quality. J Mol Diagn 13:687-94, 2011

125. Smeets SJ, Hesselink AT, Speel EJ, et al: A novel algorithm for reliable detection of human papillomavirus in paraffin embedded head and neck cancer specimen. Int J Cancer 121:2465-72, 2007

126. Leemans CR, Braakhuis BJ, Brakenhoff RH: The molecular biology of head and neck cancer. Nature reviews. Cancer 11:9-22, 2011

127. Wiest T, Schwarz E, Enders C, et al: Involvement of intact HPV16 E6/E7 gene expression in head and neck cancers with unaltered p53 status and perturbed pRb cell cycle control. Oncogene 21:1510-7, 2002

128. van Houten VM, Snijders PJ, van den Brekel MW, et al: Biological evidence that human papillomaviruses are etiologically involved in a subgroup of head and neck squamous cell carcinomas. Int J Cancer 93:232-5, 2001

129. Kroupis C, Vourlidis N: Human papilloma virus (HPV) molecular diagnostics. Clinical chemistry and laboratory medicine : CCLM / FESCC 49:1783-99, 2011

130. Ha PK, Pai SI, Westra WH, et al: Real-time quantitative PCR demonstrates low prevalence of human papillomavirus type 16 in premalignant and malignant lesions of the oral cavity. Clin Cancer Res 8:1203-9, 2002

131. Singhi AD, Westra WH: Comparison of human papillomavirus in situ hybridization and p16 immunohistochemistry in the detection of human papillomavirus-associated head and neck cancer based on a prospective clinical experience. Cancer 116:2166-73, 2010

132. Klussmann JP, Gultekin E, Weissenborn SJ, et al: Expression of p16 protein identifies a distinct entity of tonsillar carcinomas associated with human papillomavirus. Am J Pathol 162:747-53, 2003

133. Jordan RC, Lingen MW, Perez-Ordonez B, et al: Validation of Methods for Oropharyngeal Cancer HPV Status Determination in US Cooperative Group Trials. Am J Surg Pathol 36:945-954, 2012

134. Zhao N, Ang MK, Yin XY, et al: Different cellular p16(INK4a) localisation may signal different survival outcomes in head and neck cancer. Br J Cancer 107:482-90, 2012

135. Thavaraj S, Stokes A, Guerra E, et al: Evaluation of human papillomavirus testing for squamous cell carcinoma of the tonsil in clinical practice. Journal of clinical pathology 64:308-12, 2011

136. Westra WH: The changing face of head and neck cancer in the 21st century: the impact of HPV on the epidemiology and pathology of oral cancer. Head Neck Pathol 3:78-81, 2009

137. Perrone F, Gloghini A, Cortelazzi B, et al: Isolating p16-positive/HPV-negative oropharyngeal cancer: an effort worth making. Am J Surg Pathol 35:774-7; author reply 777-8, 2011

138. Shaw R: The epigenetics of oral cancer. Int J Oral Maxillofac Surg 35:101-8, 2006

139. Portela A, Esteller M: Epigenetic modifications and human disease. Nat Biotechnol 28:1057-68, 2010

140. Esteller M: Epigenetic changes in cancer. F1000 Biol Rep 3:9, 2011

141. Rodriguez-Paredes M, Esteller M: Cancer epigenetics reaches mainstream oncology. Nature medicine 17:330-9, 2011

142. Jones PA, Baylin SB: The epigenomics of cancer. Cell 128:683-92, 2007

143. Esteller M: Cancer epigenomics: DNA methylomes and histone-modification maps. Nat Rev Genet 8:286-98, 2007

144. Richards KL, Zhang B, Baggerly KA, et al: Genome-wide hypomethylation in head and neck cancer is more pronounced in HPV-negative tumors and is associated with genomic instability. PLoS One 4:e4941, 2009

145. Demokan S, Dalay N: Role of DNA methylation in head and neck cancer. Clin Epigenetics 2:123-50, 2011

146. Yoder JA, Walsh CP, Bestor TH: Cytosine methylation and the ecology of intragenomic parasites. Trends Genet 13:335-40, 1997

147. Matsuzaki K, Deng G, Tanaka H, et al: The relationship between global methylation level, loss of heterozygosity, and microsatellite instability in sporadic colorectal cancer. Clin Cancer Res 11:8564-9, 2005

148. Esteller M: Epigenetics in cancer. N Engl J Med 358:1148-59, 2008

149. Hsiung DT, Marsit CJ, Houseman EA, et al: Global DNA methylation level in whole blood as a biomarker in head and neck squamous cell carcinoma. Cancer Epidemiol Biomarkers Prev 16:108-14, 2007

150. Pattamadilok J, Huapai N, Rattanatanyong P, et al: LINE-1 hypomethylation level as a potential prognostic factor for epithelial ovarian cancer. Int J Gynecol Cancer 18:711-7, 2008

151. Tangkijvanich P, Hourpai N, Rattanatanyong P, et al: Serum LINE-1 hypomethylation as a potential prognostic marker for hepatocellular carcinoma. Clin Chim Acta 379:127-33, 2007

152. Yoder JA, Soman NS, Verdine GL, et al: DNA (cytosine-5)-methyltransferases in mouse cells and tissues. Studies with a mechanism-based probe. Journal of molecular biology 270:385-95, 1997

153. Jones PA, Liang G: Rethinking how DNA methylation patterns are maintained. Nat Rev Genet 10:805-11, 2009

154. Bostick M, Kim JK, Esteve PO, et al: UHRF1 plays a role in maintaining DNA methylation in mammalian cells. Science 317:1760-4, 2007

155. Girault I, Tozlu S, Lidereau R, et al: Expression analysis of DNA methyltransferases 1, 3A, and 3B in sporadic breast carcinomas. Clin Cancer Res 9:4415-22, 2003

156. Robertson KD: DNA methylation, methyltransferases, and cancer. Oncogene 20:3139-55, 2001

157. Ehrlich M, Woods CB, Yu MC, et al: Quantitative analysis of associations between DNA hypermethylation, hypomethylation, and DNMT RNA levels in ovarian tumors. Oncogene 25:2636-45, 2006

158. Kim H, Kwon YM, Kim JS, et al: Elevated mRNA levels of DNA methyltransferase-1 as an independent prognostic factor in primary nonsmall cell lung cancer. Cancer 107:1042-9, 2006

159. Park HJ, Yu E, Shim YH: DNA methyltransferase expression and DNA hypermethylation in human hepatocellular carcinoma. Cancer Lett 233:271-8, 2006

160. Lin RK, Hsu HS, Chang JW, et al: Alteration of DNA methyltransferases contributes to 5'CpG methylation and poor prognosis in lung cancer. Lung cancer 55:205-13, 2007

161. Daskalos A, Oleksiewicz U, Filia A, et al: UHRF1-mediated tumor suppressor gene inactivation in nonsmall cell lung cancer. Cancer 117:1027-37, 2011

162. Fernandez AF, Esteller M: Viral epigenomes in human tumorigenesis. Oncogene 29:1405-20, 2010

163. Sartor MA, Dolinoy DC, Jones TR, et al: Genome-wide methylation and expression differences in HPV(+) and HPV(-) squamous cell carcinoma cell lines are consistent with divergent mechanisms of carcinogenesis. Epigenetics 6:777-87, 2011

164. Biron VL, Mohamed A, Hendzel MJ, et al: Epigenetic differences between human papillomavirus-positive and -negative oropharyngeal squamous cell carcinomas. J Otolaryngol Head Neck Surg 41 Suppl 1:S65-70, 2012

165. Munger K, Baldwin A, Edwards KM, et al: Mechanisms of human papillomavirus-induced oncogenesis. Journal of virology 78:11451-60, 2004

166. zur Hausen H: Human papillomaviruses in the pathogenesis of anogenital cancer. Virology 184:9-13, 1991

167. Klaes R, Woerner SM, Ridder R, et al: Detection of high-risk cervical intraepithelial neoplasia and cervical cancer by amplification of transcripts derived from integrated papillomavirus oncogenes. Cancer Res 59:6132-6, 1999

168. Hopman AH, Smedts F, Dignef W, et al: Transition of high-grade cervical intraepithelial neoplasia to micro-invasive carcinoma is characterized by integration of HPV 16/18 and numerical chromosome abnormalities. J Pathol 202:23-33, 2004

169. Kalantari M, Blennow E, Hagmar B, et al: Physical state of HPV16 and chromosomal mapping of the integrated form in cervical carcinomas. Diagn Mol Pathol 10:46-54, 2001

170. Choo KB, Pan CC, Han SH: Integration of human papillomavirus type 16 into cellular DNA of cervical carcinoma: preferential deletion of the E2 gene and invariable retention of the long control region and the E6/E7 open reading frames. Virology 161:259-61, 1987

171. Jeon S, Allen-Hoffmann BL, Lambert PF: Integration of human papillomavirus type 16 into the human genome correlates with a selective growth advantage of cells. Journal of virology 69:2989-97, 1995

172. Goodwin EC, DiMaio D: Repression of human papillomavirus oncogenes in HeLa cervical carcinoma cells causes the orderly reactivation of dormant tumor suppressor pathways. Proc Natl Acad Sci U S A 97:12513-8, 2000

173. Goodwin EC, Yang E, Lee CJ, et al: Rapid induction of senescence in human cervical carcinoma cells. Proc Natl Acad Sci U S A 97:10978-83, 2000

174. Peter M, Stransky N, Couturier J, et al: Frequent genomic structural alterations at HPV insertion sites in cervical carcinoma. J Pathol 221:320-30, 2010

175. Jeon S, Lambert PF: 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, 1995

176. Ferber MJ, Thorland EC, Brink AA, et al: Preferential integration of human papillomavirus type 18 near the c-myc locus in cervical carcinoma. Oncogene 22:7233-42, 2003

177. Peter M, Rosty C, Couturier J, et al: MYC activation associated with the integration of HPV DNA at the MYC locus in genital tumors. Oncogene 25:5985-93, 2006

178. Reuter S, Bartelmann M, Vogt M, et al: APM-1, a novel human gene, identified by aberrant co-transcription with papillomavirus oncogenes in a cervical carcinoma cell line, encodes a BTB/POZ-zinc finger protein with growth inhibitory activity. The EMBO journal 17:215-22, 1998

179. Koskinen WJ, Chen RW, Leivo I, et al: Prevalence and physical status of human papillomavirus in squamous cell carcinomas of the head and neck. Int J Cancer 107:401-6, 2003

180. Snijders PJ, Scholes AG, Hart CA, et al: Prevalence of mucosotropic human papillomaviruses in squamous-cell carcinoma of the head and neck. Int J Cancer 66:464-9, 1996

181. Canadas MP, Darwich L, Sirera G, et al: New molecular method for the detection of human papillomavirus type 16 integration. Clin Microbiol Infect 16:836-42, 2010

182. Collins SI, Constandinou-Williams C, Wen K, et al: Disruption of the E2 gene is a common and early event in the natural history of cervical human papillomavirus infection: a longitudinal cohort study. Cancer Res 69:3828-32, 2009

183. Evans MF, Cooper K: Human papillomavirus integration: detection by in situ hybridization and potential clinical application. J Pathol 202:1-4, 2004

184. Parsons MWG, H: How to make tissue microarrays. Diagnostic Histopathology 15:142-150, 2009

185. Roberts I, Ng G, Foster N, et al: Critical evaluation of HPV16 gene copy number quantification by SYBR green PCR. BMC Biotechnol 8:57, 2008

186. De Marchi Triglia R, Metze K, Zeferino LC, et al: HPV in situ hybridization signal patterns as a marker for cervical intraepithelial neoplasia progression. Gynecologic oncology 112:114-8, 2009

187. Ukpo OC, Flanagan JJ, Ma XJ, et al: High-risk human papillomavirus E6/E7 mRNA detection by a novel in situ hybridization assay strongly correlates with p16 expression and patient outcomes in oropharyngeal squamous cell carcinoma. The American journal of surgical pathology 35:1343-50, 2011

188. Meissner JD: Nucleotide sequences and further characterization of human papillomavirus DNA present in the CaSki, SiHa and HeLa cervical carcinoma cell lines. The Journal of general virology 80 (Pt 7):1725-33, 1999

189. Livak KJ, Schmittgen TD: Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods 25:402-8, 2001

190. Daskalos A, Nikolaidis G, Xinarianos G, et al: Hypomethylation of retrotransposable elements correlates with genomic instability in non-small cell lung cancer. Int J Cancer 124:81-7, 2009

191. Snellenberg S, Schutze DM, Claassen-Kramer D, et al: Methylation status of the E2 binding sites of HPV16 in cervical lesions determined with the Luminex(R) xMAP system. Virology 422:357-65, 2012

192. Sandoval J, Heyn H, Moran S, et al: Validation of a DNA methylation microarray for 450,000 CpG sites in the human genome. Epigenetics 6:692-702, 2011

193. Graham DA, Herrington CS: HPV-16 E2 gene disruption and sequence variation in CIN 3 lesions and invasive squamous cell carcinomas of the cervix: relation to numerical chromosome abnormalities. Mol Pathol 53:201-6, 2000

194. Baker CC, Phelps WC, Lindgren V, et al: Structural and transcriptional analysis of human papillomavirus type 16 sequences in cervical carcinoma cell lines. Journal of virology 61:962-71, 1987

195. Depledge DP, Palser AL, Watson SJ, et al: Specific capture and whole-genome sequencing of viruses from clinical samples. PLoS One 6:e27805, 2011

196. Wansom D, Light E, Thomas D, et al: Infiltrating lymphocytes and human papillomavirus-16--associated oropharyngeal cancer. Laryngoscope 122:121-7, 2012

197. Fischer CA, Zlobec I, Green E, et al: Is the improved prognosis of p16 positive oropharyngeal squamous cell carcinoma dependent of the treatment modality? Int J Cancer 126:1256-62, 2010

198. Klussmann JP, Preuss SF, Speel EJ: [Human papillomavirus and cancer of the oropharynx. Molecular interaction and clinical implications]. HNO 57:113-22, 2009

199. Charfi L, Jouffroy T, de Cremoux P, et al: Two types of squamous cell carcinoma of the palatine tonsil characterized by distinct etiology, molecular features and outcome. Cancer Lett 260:72-8, 2008

200. Nasman A, Attner P, Hammarstedt L, et al: Incidence of human papillomavirus (HPV) positive tonsillar carcinoma in Stockholm, Sweden: an epidemic of viral-induced carcinoma? Int J Cancer 125:362-6, 2009

201. Morris LG, Sikora AG, Patel SG, et al: Second primary cancers after an index head and neck cancer: subsite-specific trends in the era of human papillomavirus-associated oropharyngeal cancer. J Clin Oncol 29:739-46, 2011

202. Shaw RJ, Hobkirk AJ, Nikolaidis G, et al: Molecular Staging of Surgical Margins in Oral Squamous Cell Carcinoma Using Promoter Methylation of p16(INK4A), Cytoglobin, E-cadherin, and TMEFF2. Annals of surgical oncology 20:2796-802, 2013

203. Tabor MP, Brakenhoff RH, Ruijter-Schippers HJ, et al: Genetically altered fields as origin of locally recurrent head and neck cancer: a retrospective study. Clin Cancer Res 10:3607-13, 2004

204. Harris SL, Thorne LB, Seaman WT, et al: Association of p16(INK4a) overexpression with improved outcomes in young patients with squamous cell cancers of the oral tongue. Head Neck, 2010

205. Weinberger PM, Yu Z, Haffty BG, et al: Molecular classification identifies a subset of human papillomavirus--associated oropharyngeal cancers with favorable prognosis. J Clin Oncol 24:736-47, 2006

206. Worden FP, Ha H: Controversies in the management of oropharynx cancer. J Natl Compr Canc Netw 6:707-14, 2008

207. Kaur G, Hutchison I, Mehanna H, et al: Barriers to recruitment for surgical trials in head and neck oncology: a survey of trial investigators. BMJ Open 3, 2013

208. Evans M, Newcombe R, Fiander A, et al: Human Papillomavirus-associated oropharyngeal cancer: an observational study of diagnosis, prevalence and prognosis in a UK population. BMC cancer 13:220, 2013

209. Wang F, Flanagan J, Su N, et al: RNAscope: a novel in situ RNA analysis platform for formalin-fixed, paraffin-embedded tissues. The Journal of molecular diagnostics : JMD 14:22-9, 2012

210. Fu YS, Hoover L, Franklin M, et al: Human papillomavirus identified by nucleic acid hybridization in concomitant nasal and genital papillomas. Laryngoscope 102:1014-9, 1992

211. Stoler MH, Broker TR: In situ hybridization detection of human papillomavirus DNAs and messenger RNAs in genital condylomas and a cervical carcinoma. Hum Pathol 17:1250-8, 1986

212. van den Brule AJ, Cromme FV, Snijders PJ, et al: Nonradioactive RNA in situ hybridization detection of human papillomavirus 16-E7 transcripts in squamous cell carcinomas of the uterine cervix using confocal laser scan microscopy. Am J Pathol 139:1037-45, 1991

213. Ukpo OC, Flanagan JJ, Ma XJ, et al: High-risk human papillomavirus E6/E7 mRNA detection by a novel in situ hybridization assay strongly correlates with p16 expression and patient outcomes in oropharyngeal squamous cell carcinoma. Am J Surg Pathol 35:1343-50, 2011

214. Lewis JS, Jr., Ukpo OC, Ma XJ, et al: Transcriptionally-active high-risk human papillomavirus is rare in oral cavity and laryngeal/hypopharyngeal squamous cell carcinomas--a tissue microarray study utilizing E6/E7 mRNA in situ hybridization. Histopathology 60:982-91, 2012

215. Bishop JA, Ma XJ, Wang H, et al: Detection of Transcriptionally Active High-risk HPV in Patients With Head and Neck Squamous Cell Carcinoma as Visualized by a Novel E6/E7 mRNA In Situ Hybridization Method. Am J Surg Pathol, 2012

216. Licitra L, Perrone F, Bossi P, et al: High-risk human papillomavirus affects prognosis in patients with surgically treated oropharyngeal squamous cell carcinoma. Journal of clinical oncology : official journal of the American Society of Clinical Oncology 24:5630-6, 2006

217. Lindquist D, Romanitan M, Hammarstedt L, et al: Human papillomavirus is a favourable prognostic factor in tonsillar cancer and its oncogenic role is supported by the expression of E6 and E7. Molecular oncology 1:350-5, 2007

218. Schache AG, Liloglou T, Risk JM, et al: Evaluation of human papilloma virus diagnostic testing in oropharyngeal squamous cell carcinoma: sensitivity, specificity, and prognostic discrimination. Clinical cancer research : an official journal of the American Association for Cancer Research 17:6262-71, 2011

219. Schlecht NF, Brandwein-Gensler M, Nuovo GJ, et al: A comparison of clinically utilized human papillomavirus detection methods in head and neck cancer. Modern pathology : an official journal of the United States and Canadian Academy of Pathology, Inc 24:1295-305, 2011

220. Wells SI, Francis DA, Karpova AY, et al: Papillomavirus E2 induces senescence in HPV-positive cells via pRB- and p21(CIP)-dependent pathways. The EMBO journal 19:5762-71, 2000

221. Kim K, Garner-Hamrick PA, Fisher C, et al: Methylation patterns of papillomavirus DNA, its influence on E2 function, and implications in viral infection. Journal of virology 77:12450-9, 2003

222. Hall GL, Shaw RJ, Field EA, et al: p16 Promoter methylation is a potential predictor of malignant transformation in oral epithelial dysplasia. Cancer Epidemiol Biomarkers Prev 17:2174-9, 2008

223. Shaw RJ, Liloglou T, Rogers SN, et al: Promoter methylation of P16, RARbeta, E-cadherin, cyclin A1 and cytoglobin in oral cancer: quantitative evaluation using pyrosequencing. Br J Cancer 94:561-8, 2006

224. Cordaux R, Batzer MA: The impact of retrotransposons on human genome evolution. Nat Rev Genet 10:691-703, 2009

225. Furniss CS, Marsit CJ, Houseman EA, et al: Line region hypomethylation is associated with lifestyle and differs by human papillomavirus status in head and neck squamous cell carcinomas. Cancer Epidemiol Biomarkers Prev 17:966-71, 2008

226. Ma Y, Hendershot LM: The role of the unfolded protein response in tumour development: friend or foe? Nature reviews. Cancer 4:966-77, 2004

227. Lu Q: delta-Catenin dysregulation in cancer: interactions with E-cadherin and beyond. J Pathol 222:119-23, 2010

228. Misawa K, Ueda Y, Kanazawa T, et al: Epigenetic inactivation of galanin receptor 1 in head and neck cancer. Clin Cancer Res 14:7604-13, 2008

229. Stevenson L, Allen WL, Turkington R, et al: Identification of galanin and its receptor GalR1 as novel determinants of resistance to chemotherapy and potential biomarkers in colorectal cancer. Clin Cancer Res 18:5412-26, 2012

230. Rauch T, Wang Z, Zhang X, et al: Homeobox gene methylation in lung cancer studied by genome-wide analysis with a microarray-based methylated CpG island recovery assay. Proc Natl Acad Sci U S A 104:5527-32, 2007

231. Di Vinci A, Brigati C, Casciano I, et al: HOXA7, 9, and 10 are methylation targets associated with aggressive behavior in meningiomas. Transl Res 160:355-62, 2012

232. Lallet-Daher H, Wiel C, Gitenay D, et al: Potassium channel KCNA1 modulates oncogene-induced senescence and transformation. Cancer Res, 2013

233. Kwon YJ, Lee SJ, Koh JS, et al: Genome-wide analysis of DNA methylation and the gene expression change in lung cancer. Journal of thoracic oncology : official publication of the International Association for the Study of Lung Cancer 7:20-33, 2012

234. Perot G, Derre J, Coindre JM, et al: Strong smooth muscle differentiation is dependent on myocardin gene amplification in most human retroperitoneal leiomyosarcomas. Cancer Res 69:2269-78, 2009

235. Ziliak D, O'Donnell PH, Im HK, et al: Germline polymorphisms discovered via a cell-based, genome-wide approach predict platinum response in head and neck cancers. Transl Res 157:265-72, 2011

236. Babij C, Zhang Y, Kurzeja RJ, et al: STK33 kinase activity is nonessential in KRAS-dependent cancer cells. Cancer Res 71:5818-26, 2011

237. Bignone PA, Lee KY, Liu Y, et al: RPS6KA2, a putative tumour suppressor gene at 6q27 in sporadic epithelial ovarian cancer. Oncogene 26:683-700, 2007

238. Tan HK, Saulnier P, Auperin A, et al: Quantitative methylation analyses of resection margins predict local recurrences and disease-specific deaths in patients with head and neck squamous cell carcinomas. Br J Cancer 99:357-63, 2008

239. Jithesh PV, Risk JM, Schache AG, et al: The epigenetic landscape of oral squamous cell carcinoma. Br J Cancer 108:370-9, 2013

240. Thierry F: Transcriptional regulation of the papillomavirus oncogenes by cellular and viral transcription factors in cervical carcinoma. Virology 384:375-9, 2009

241. Park IS, Chang X, Loyo M, et al: Characterization of the methylation patterns in human papillomavirus type 16 viral DNA in head and neck cancers. Cancer Prev Res (Phila) 4:207-17, 2011

242. Fernandez AF, Rosales C, Lopez-Nieva P, et al: The dynamic DNA methylomes of double-stranded DNA viruses associated with human cancer. Genome Res 19:438-51, 2009

243. Colacino JA, Dolinoy DC, Duffy SA, et al: Comprehensive analysis of DNA methylation in head and neck squamous cell carcinoma indicates differences by survival and clinicopathologic characteristics. PLoS One 8:e54742, 2013

244. Stransky N, Egloff AM, Tward AD, et al: The mutational landscape of head and neck squamous cell carcinoma. Science 333:1157-60, 2011

245. Weiss D, Koopmann M, Basel T, et al: Cyclin A1 shows age-related expression in benign tonsils, HPV16-dependent overexpression in HNSCC and predicts lower recurrence rate in HNSCC independently of HPV16. BMC cancer 12:259, 2012

246. Dorsam RT, Gutkind JS: G-protein-coupled receptors and cancer. Nature reviews. Cancer 7:79-94, 2007

247. Worden FP, Kumar B, Lee JS, et al: Chemoselection as a strategy for organ preservation in advanced oropharynx cancer: response and survival positively associated with HPV16 copy number. J Clin Oncol 26:3138-46, 2008

248. Lechner M, Fenton T, West J, et al: Identification and functional validation of HPV-mediated hypermethylation in head and neck squamous cell carcinoma. Genome Med 5:15, 2013

249. Martinez I, Wang J, Hobson KF, et al: Identification of differentially expressed genes in HPV-positive and HPV-negative oropharyngeal squamous cell carcinomas. Eur J Cancer 43:415-32, 2007

250. Vogel MC, Papadopoulos T, Muller-Hermelink HK, et al: Intracellular distribution of DNA methyltransferase during the cell cycle. FEBS Lett 236:9-13, 1988

251. Amara K, Ziadi S, Hachana M, et al: DNA methyltransferase DNMT3b protein overexpression as a prognostic factor in patients with diffuse large B-cell lymphomas. Cancer Sci 101:1722-30, 2010

252. Hayette S, Thomas X, Jallades L, et al: High DNA methyltransferase DNMT3B levels: a poor prognostic marker in acute myeloid leukemia. PLoS One 7:e51527, 2012

253. Wang L, Rodriguez M, Kim ES, et al: A novel C/T polymorphism in the core promoter of human de novo cytosine DNA methyltransferase 3B6 is associated with prognosis in head and neck cancer. Int J Oncol 25:993-9, 2004

254. Liu Z, Wang L, Wang LE, et al: Polymorphisms of the DNMT3B gene and risk of squamous cell carcinoma of the head and neck: a case-control study. Cancer Lett 268:158-65, 2008

255. Woodman CB, Collins SI, Young LS: The natural history of cervical HPV infection: unresolved issues. Nature reviews. Cancer 7:11-22, 2007

256. Lazo PA: The molecular genetics of cervical carcinoma. Br J Cancer 80:2008-18, 1999

257. Thorland EC, Myers SL, Gostout BS, et al: Common fragile sites are preferential targets for HPV16 integrations in cervical tumors. Oncogene 22:1225-37, 2003