UNIVERSITY OF DERBY

The prevalence and pathogenesis of Human Papillomavirus (HPV) in normal tonsil tissue

Mary-Anne Freckleton

A submission in partial fulfilment of the requirements of the University of Derby for the award of the degree of Doctor of Philosophy

College of Science and Engineering

Doctor of Philosophy 2025

CONTENTS

TITLE		PAGE
TABLE OF CONTENTS	…….....	**2-9**
LIST OF FIGURES	…….....	**10-12**
LIST OF TABLES	…….....	**13-14**
ABBREVIATIONS	…….....	**15-18**
COVID-19 STATEMENT	…….....	**19**
DECLARATION	…….....	**20**
DEDICATION	…….	**21**
ACKNOWLEDGEMENTS	…….	**22-23**
ABSTRACT	…….....	**24-25**
CHAPTER 1:	**INTRODUCTION…….....	**26-72**
1.0	Human Papillomaviruses – An Introduction……...	**26**
1.0.1.	HPV Genome Structure…….….	**28**
1.0.2.	The viral life cycle of HPV…….	**30**
1.0.3.	Viral attachment and cell entry……...	**30**
1.0.4.	Viral replication……..	**31**
1.0.5.	Viral gene expression…….	**34**
1.1.	HR-HPV – a virus with carcinogenic potential……..	**34**
1.1.1.	HPV-induced cervical carcinogenesis……	**35**
1.1.1.2.	Integrated or episomal viral forms…….	**35**
1.1.1.3.	The roles of E6, E7, and E5 in cervical carcinogenesis…….	**37**
1.1.2.	HPV-positive Oropharyngeal Cancer (HPV-OPC)…….	**40**
1.2.	The natural history of an oral HPV infection…….	**43**
1.2.1.	The prevalence of oral HPV within the non-cancerous oral cavity/oropharynx……...	**43**
1.2.2.	The pathogenesis of HPV within the oropharynx……..	**47**

1.2.2.1. HPV pathogenesis within OPC…….. **47**

1.2.2.2. HPV infection of the tonsillar crypts…… **49**

1.3. Biomarkers of, and treatments for, HPV-OPC……. **52**

1.3.1. HR-HPV and viral oncoprotein expression in HPV-OPC….. **52**

1.3.1.1. p53 is not mutated in most HPV-positive oropharyngeal malignancies…… **52**

1.3.2. $p16^{INK4A}$ expression as a surrogate marker for HPV within HNC…… **53**

1.3.3. microRNA (miRNA) expression profiles in HPV-OPC……. **54**

1.3.4. Oher biomarkers for HPV-HNC…….. **54**

1.3.5. HPV-HNC targeted treatments…… **56**

1.4. The risk factors for oral HPV infection…….. **57**

1.4.1. Gender and age…….. **57**

1.4.2. Sexual behaviours……. **59**

1.4.3. Lifestyle practices – alcohol consumption, smoking, and oral hygiene…… **61**

1.4.4. Ethnicity…….. **62**

1.4.5. HIV infection…….. **63**

1.5. Public health interventions…….. **65**

1.5.1. The Cervical Screening Programme…… **65**

1.5.2. The HPV vaccination programme…….. **66**

1.5.3. Education on the risk of HPV infection within the public…. **70**

1.6. Conclusion…….. **70**

1.7. Aims and objectives of this project **71**

CHAPTER 2: THE DEVELOPMENT OF A PROTOCOL TO EXTRACT DNA FROM FORMALIN-FIXED TONSIL TISSUE …….. 73-93

2.0. Introduction……. **73**

2.1. The impact of SARS-CoV-2 on sample collection……. **74**

2.2. The isolation of nucleic acid and protein from tonsil tissues.. **74**

2.2.1. TRIzol™ reagent for DNA extraction from formalin-fixed tonsil tissue…….. **75**

2.3. The AllPrep® DNA/RNA FFPE kit [Qiagen] for the extraction of DNA from formalin-fixed tonsil tissue……... **78**
2.3.1. Initial AllPrep® DNA/RNA FFPE kit trial…….. **78**
2.3.2. Optimisation of AllPrep® DNA/RNA FFPE kit extractions... **79**
2.3.3. Tonsil tissue storage……. **79**
2.3.4. Physical disruption and homogenisation of formalin-fixed tissue…… **80**
2.3.5. Final protocol modification for prolonged formalin-fixed tonsil tissue……. **81**
2.4. 24-hour formalin fixation period of tonsil tissues…….. **85**
2.5. Amendment to the HPV clinical study protocol……. **85**
2.6. Bead type…… **85**
2.7. The extraction of RNA from formalin-fixed tonsil tissue….. **88**
2.8. Discussion……... **92**
2.8.1. Conclusion……. **92**
CHAPTER 3 FINAL RESEARCH METHODOLOGY…….. 94-122
3.1. Meta-analysis of oral HPV prevalence in the healthy adult population…… **94**
3.1.1. Literature searching and selection of papers for analysis….. **94**
3.1.1.1. Study abstract and full-text screening…… **95**
3.1.2. Data extraction…….. **95**
3.1.3. Statistical analysis of oral HPV prevalence…….. **96**
3.2. NIHR-portfolio HPV clinical study…… **97**
3.2.1. Recruitment to the study of routine tonsillectomy patients **97**
3.2.2. Sample collection and specimen storage/fixation requirements…… **98**
3.2.3. Demographics and lifestyle characteristics of participants…. **98**
3.3. University of Birmingham Human Biomaterials Resource Centre (HBRC) tissues…… **99**
3.4. Sectioning of non-cancerous palatine tonsil tissue……. **99**
3.4.1. 10% NBF fixed palatine tonsils……. **99**
3.4.2. Fresh-frozen tonsil tissue…….. **101**

3.5. DNA and RNA extraction from formalin-fixed tonsil tissues **101**

3.5.1. Homogenisation and lysis of 10% NBF fixed tonsil tissues.. **101**

3.5.2. AllPrep® DNA/RNA FFPE [Qiagen] for extraction of DNA from 10% NBF fixed tonsil tissue……... **102**

3.6. DNA isolation from fresh-frozen tonsil tissue…… **103**

3.6.1. Homogenisation and lysis of fresh-frozen tonsil tissue……. **104**

3.6.2. DNeasy® Blood & Tissue Kit [Qiagen] for DNA extraction **104**

3.7. DNA quantity and quality analysis and dilution for PCR assays…….. **105**

3.8. Beta-actin housekeeping gene PCR for DNA integrity analysis…… **105**

3.8.1. Beta-actin TaqMan Real-Time PCR screening…….. **105**

3.8.2. Beta-actin amplification analysis by standard endpoint PCR **108**

3.9. Gel electrophoresis of PCR amplimers…….. **108**

3.10. HPV-DNA qPCR consensus screening…….. **108**

3.10.1. Statistical analysis of qPCR data…… **109**

3.11. HPV genotyping by Sanger sequencing……. **112**

3.11.1. DNA purification from agarose gels…….. **112**

3.11.2. Sample preparation for Sanger sequencing…….. **112**

3.11.3. Data analysis of FASTA sequence data (NCBI BLAST)….. **113**

3.12. Genotyping of HPV-positive tonsil gDNA…… **113**

3.12.1. PCR amplification…….. **113**

3.12.2. Hybridisation of PCR products……. **113**

3.13. Tissue Sections and Staining……. **114**

3.14. Cell culture…… **115**

3.15. Tonsil tissue lysis and 8M urea protein extraction……. **115**

3.15.1. Preparation of tonsil tissue for lysis…… **115**

3.15.2. Tissue homogenisation and sonication…… **116**

3.15.3. Protein quantification…….. **117**

3.16. Tryptic protein digestion for SWATH-MS……. **117**

3.17. Proteomic analysis of non-cancerous palatine tonsils…… **120**

3.17.1. HPLC and SWATH-MS…….. **120**

3.17.2. Protein data searching – DIA-NN…….. **121**

3.17.3. DEA – Amica……... **121**

3.17.3.1. Ontological analysis…… **122**

3.17.3.2. Pathway enrichment analysis…….. **122**

CHAPTER 4: Meta-analyses of global oral HPV prevalence estimates amongst the adult population…….. 123-135

4.0. Introduction……. **123**

4.1. Selection of studies for meta-analysis……. **124**

4.1.1. Study classification and demographics of their cohorts for the final list of reports included in these meta-analyses…… **125**

4.2. Meta-analysis of global oral HPV prevalence estimates amongst the adult population……... **131**

4.2.1. Oral HPV infection in males is almost twice as prevalent than within females……. **132**

4.3. Discussion…… **134**

4.3.1. Conclusion……... **135**

CHAPTER 5: HPV PREVALENCE WITHIN THE NON-CANCEROUS TONSIL TISSUE OF ROUTINE TONSILLECTOMY PATIENTS……. 136-164

5.0. Introduction……. **136**

5.1. Final study sample…….. **137**

5.1.1. Beta-actin qPCR for DNA integrity analysis…….. **138**

5.1.2. Demographic information of the final sample…… **142**

5.2. Overall HPV prevalence in non-cancerous tonsil tissue by qPCR consensus screening……. **144**

5.2.1. Demographics and behaviours of individuals with HPV in their palatine tonsils…… **146**

5.3. The pathogenesis of HPV within non-cancerous tonsil tissue **147**

5.3.1. Viral type HPV-18 is most prevalent within non-cancerous tonsil tissue……. **151**

5.3.2. Histological analysis of HPV-positive and HPV-negative tonsil tissue……. **154**

5.3.2.1. $p16^{INK4A}$ – a true biomarker for HPV?..... **154**

5.3.2.2. CK7 – a marker of the tonsillar crypt epithelium in non-cancerous tonsil tissue……. **158**

5.4. Formalin fixation of tonsil tissues – HPV-negative or undetected HPV-DNA……. **161**

5.5. Discussion…… **162**

5.5.1. Conclusion……... **164**

CHAPTER 6 DIFFERENTIAL PROTEIN EXPRESSION IN NON-CANCEROUS PALATINE TONSILS WITH HPV-DNA INFECTION…….. 165-204

6.0. Introduction……. **165**

6.1. Final participant sample for protein extraction and SWATH-MS……... **166**

6.2. Quantification of protein from formalin-fixed versus fresh-frozen tonsil tissue……. **167**

6.3. SWATH-MS analysis of the proteome of non-cancerous tonsil tissue……. **167**

6.3.1. Proteins identified by SWATH data searching in DIA-NN.. **170**

6.4. Differential expression of HPV-positive and HPV-negative non-cancerous tonsil protein…… **170**

6.4.1. A panel of 222 proteins with significantly changed expression in HPV-positive versus HPV-negative tonsil tissues…….. **174**

6.4.2. Protein function of significantly changed protein hits in HPV-positive compared to HPV-negative non-cancerous tonsil tissue…… **176**

6.4.3. Classification of differentially expressed proteins in HPV-positive versus HPV-negative tonsil tissue……. **179**

6.4.4. Pathway enrichment analysis of differentially expressed proteins in HPV-positive tonsil tissue……. **183**

6.5. Analysis of HPV-positive males and HPV-positive females **190**

6.5.1. Differential protein expression of HPV-positive individuals based on their biological sex…… **194**

6.5.2. Ontological analysis of differential protein expression data of HPV-positive male versus HPV-positive female non-cancerous tonsil tissue……. **197**

6.5.3. Pathway enrichment analysis of differentially expression proteins between HPV-positive males and females……. **200**
6.6. Discussion…… **203**
6.6.1. Conclusion…….. **204**
CHAPTER 7: DISCUSSION…….. 205-229
7.0. Introduction……. **205**
7.1. Global burden of oral HPV in the healthy adult population… **205**
7.2. Oropharyngeal HPV prevalence in adult non-cancerous palatine tonsil tissues……. **206**
7.3. Experimental factors that influences oral HPV prevalence… **208**
7.3.1. Sample type for oral HPV prevalence investigation…… **208**
7.3.2. Viral types in oral HPV prevalence estimates……. **210**
7.3.2.2. HPV viral types within the non-cancerous oropharynx and oral cavity…… **211**
7.4. HPV pathogenesis in the non-cancerous oropharynx……. **212**
7.5. Demographics and lifestyle characteristics of individuals with HPV-positive tonsil tissue/s……. **213**
7.5.1. Gender, lifestyle practices, and oral HPV…… **214**
7.5.2. HPV-positivity and vaccination status……. **217**
7.6. HPV-negative or undetectable HPV in gDNA isolated from 10% fixed tonsil tissues…… **217**
7.7. Protein expression blueprint of HPV in the non-cancerous oropharynx……... **220**
7.7.1. The proteome of non-cancerous HPV-positive tonsil tissue… **220**
7.7.2. HPV-positive male versus HPV-positive female tonsil protein **223**
7.7.3. Classical markers for both HPV infection and HPV-OPC….. **225**
7.7.4. General limitations and future direction…….. **226**
7.8. Limitation of this study…… **227**
7.9. Conclusion…….. **227**
REFERENCES……. 230-266
APPENDICES
APPENDIX A: NIHR-portfolio HPV clinical study documentation……. **II**

APPENDIX B: University of Derby Ethical Approval……. **XXVI**

APPENDIX C: HBRC documentation……. **XXXIX**

APPENDIX D: A modified qPCR assay to detect HPV-DNA in non-cancerous tonsil tissue……. **XLIV**

APPENDIX E: Statistically significant protein hits within non-cancerous tonsil tissue by DEA…… **LXXV**

LIST OF FIGURES

FIGURE		PAGE
Figure 1.	The HPV-16 genome structure……..	**29**
Figure 2.	The lifecycle of HPV infection in cells of the squamous stratified epithelium……	**33**
Figure 3.	Interaction of HR-HPV oncoproteins, E6 and E7, with host cell proteins...	**39**
Figure 4.	Age-standardised cases of cervical, and male oropharyngeal cancer (OPC)	**41**
Figure 5.	Anatomy of the human oropharynx…….	**42**
Figure 6.	HPV infection of the tonsillar crypts within the human oropharynx……	**51**
Figure 7.	Oral HPV and its clinical relevance in HNC……	**64**
Figure 8.	Final tissue homogenisation protocol for the AllPrep® DNA/RNA FFPE kit……	**84**
Figure 9.	Beta-actin amplification in cDNA samples generated from formalin-fixed tonsil RNA……	**91**
Figure 10.	Sectioning of 10% NBF fixed palatine tonsil tissues for DNA, RNA, and protein extraction, and histological analysis…….	**100**
Figure 11.	Standard layout of a 96-well plate for beta-actin qPCR amplification analysis of non-cancerous tonsil gDNA……	**107**
Figure 12.	Standard layout of a 96-well plate for HPV-DNA qPCR consensus screening……	**111**
Figure 13.	PRISMA flowchart of report selection for meta-analysis of oral HPV prevalence estimates…….	**126**
Figure 14.	Oral HPV infection is almost twice as prevalent in males than females, with males over three times more likely to have an oral HR HPV-16 infection…….	**133**
Figure 15.	Amplification plot of the beta-actin gene in tonsillar gDNA and standard curve for copy number calculation……	**139**
Figure 16.	Analysis of standard PCR products for beta-actin amplification by gel electrophoresis……	**140**
Figure 17.	Amplification plot of housekeeping gene, beta-actin, within gDNA isolated from formalin-fixed tonsil tissues……	**141**
Figure 18.	Demographic information of the final sample included in our prospective HPV prevalence analysis..……	**143**
Figure 19.	Detection of oropharyngeal HPV infection in DNA from non-cancerous tonsil tissues by qPCR consensus screening……	**145**

Figure 20. HPV-positivity is mostly concordant in the left and right non-cancerous palatine tonsil tissues.…… **148**

Figure 21. HPV-positivity varies across different sections of the same tonsil tissue… **150**

Figure 22. H&E, p16, and CK7 staining of sections taken from the right tissue of representative participant tonsils…….. **159**

Figure 23. H&E, p16, and CK7 staining of sections taken from the left tissue of representative participant tonsils…….. **160**

Figure 24. TIC chromatograms of tonsil protein samples following SWATH-MS…. **169**

Figure 25. Quality control data generated by Amica software of differential abundance analysis of HPV-positive versus HPV-negative non-cancerous tonsillar protein…….. **173**

Figure 26. Volcano plot of the 5103 proteins quantified in Amica filtered on their log_2FC and p-value ($-log_{10}$)…….. **175**

Figure 27. Ontological classification of proteins differentially expressed in HPV-positive compared to HPV-negative tonsil tissue…… **182**

Figure 28. Respiratory electron transport pathway within the mitochondria is most significantly overrepresented within the statistically differentially expressed proteins of HPV-positive versus HPV-negative tonsil tissue…... **189**

Figure 29. Quality control data generated by Amica software of differential abundance analysis of protein from HPV-positive male versus HPV-positive female non-cancerous tonsil tissue…….. **193**

Figure 30. 394 significantly changed protein hits in HPV-positive males compared to HPV-positive females when filtered by a log2FC ±1.5 and -log10 adj. p-value >2…….. **195**

Figure 31. Ontological classification of proteins differentially expressed in non-cancerous tonsil tissue of HPV-positive males versus HPV-positive females **199**

Figure 32. HRA approval from the Research Ethics Committee (REC) for the HPV clinical study…….… **V**

Figure 33. Letter of invitation for HPV clinical study (v5.3)…… **VIII**

Figure 34. Patient Information Sheet for HPV clinical study (v5.3)……. **XIII**

Figure 35. Consent form for HPV clinical study (v5.3)…… **XVI**

Figure 36. Questionnaire for HPV clinical study (v5.3)…… **XXV**

Figure 37. Ethical approval for this study from the University of Derby's College of Science and Engineering REC……. **XXVI**

Figure 38. Ethics application for this study from the University of Derby's College of Science and Engineering REC…….. **XXXVIII**

Figure 39. Confirmation of application to purchase HBRC biomaterials……. **XXXIX**

Figure 40. HRA approval from the REC for the HBRC…… **XLIII**

Figure 41. Melt curve analysis of products generated by qPCR using L1 consensus primers, MY09/MY11……... **XLIX**

Figure 42. Melt curve analysis of products generated by qPCR using L1 consensus primers, FAP59/FAP64……. **L**

Figure 43. Melt curve analysis of products generated by qPCR using L1 consensus primers, GP5+/GP6+……. **LI**

Figure 44. Analysis of qPCR products by gel electrophoresis…….. **LII**

Figure 45. Shift in melt curve alignment of tonsillar gDNA amplification products at lower final gDNA concentration; melt curve analysis and corresponding Sanger sequencing data……. **LVI**

Figure 46. Loss of amplification products for SiHa DNA control with reduced GP5+/GP6+ primer concentration……. **LX**

Figure 47. Shift in melt curve alignment of tonsillar gDNA amplification products at lower annealing temperature; melt curve analysis and corresponding Sanger sequencing data……. **LXIII**

Figure 48. Non-specific amplification of GP5+/GP6+ products from tonsillar gDNA with an annealing temperature of 48°C……. **LXVI**

Figure 49. Generation of increased amount of specific PCR product with limited non-specific GP5+/GP6+ primer binding at annealing temperature 52°C…….... **LXVII**

Figure 50. Melt curve analysis of products generated using GP5+/GP6+ consensus primers with different SYBR master mixes…… **LXX**

Figure 51. Analysis of PCR products using GP5+/GP6+ consensus primers with a variety of master mixes in a standard PCR approach by gel electrophoresis **LXXI**

LIST OF TABLES

TABLE		PAGE
Table 1.	HR- and LR-HPV viral types and associated diseases……..	**27**
Table 2.	Oral HPV prevalence reports by global region…….	**46**
Table 3.	Summary of concentration and integrity of DNA extracted using TRIzol™ reagent……	**77**
Table 4.	Summary of measured concentration and integrity of tonsil DNA extracted using the AllPrep® kit, with indicated protocol modifications……..	**83**
Table 5.	Summary of concentration and integrity of DNA isolated from 24-96 hour formalin-fixed tonsil tissue, using the AllPrep® DNA/RNA FFPE kit……...	**87**
Table 6.	Summary of the quantity and integrity of RNA isolated from formalin-fixed tonsil tissue……..	**90**
Table 7.	Oligonucleotides for PCR assays…….	**110**
Table 8.	Solutions for S-Trap™ Rapid Universal MS Sample Prep [Protifi] protein digestion…….	**119**
Table 9.	Studies included in meta-analysis of oral HPV prevalence in the healthy adult population…….……	**127**
Table 10.	Summary of demographics, lifestyle characteristics, and HPV-types detect for each HPV-positive individual……..	**153**
Table 11.	p16 scores of the right and left palatine tonsils of tissues considered HPV-positive by HPV-DNA qPCR consensus screening……..	**156**
Table 12.	p16 scores of the right and left palatine tonsils of fresh-frozen tissues considered HPV-negative by HPV-DNA qPCR consensus screening……	**157**
Table 13.	Total raw precursor numbers and proteins identified by DIA-NN searching of samples included in HPV-positive versus HPV-negative DEA…….	**171**
Table 14.	Top 10 significantly increased and decreased protein hits identified in HPV-positive tonsil tissue in comparison to HPV-negative tonsil tissue…….	**178**
Table 15.	Top 25 significantly enriched pathways of 222 protein hits identified by DEA of HPV-positive versus HPV-negative tonsil tissue……	**185**
Table 16.	Significantly enriched pathways within HPV-positive versus HPV-negative protein hits by FDR and their mapped proteins from each pathway……..	**188**
Table 17.	Total raw precursor numbers and proteins identified by DIA-NN searching of samples included in HPV-positive male versus HPV-positive female DEA	**191**

Table 18. Top six significantly increased and 10 decreased protein hits identified in HPV-positive male tonsil tissue compared to HPV-positive female tonsil tissue……. **196**

Table 19. 25 most enriched pathways within 394 significantly changed protein hits identified by DEA of HPV-positive male and HPV-positive female tonsil tissue…… **202**

Table 20. HPV L1 gene consensus primers…….. **XLVIII**

Table 21. Analysis of PCR amplification of known HPV-positive tonsil gDNA samples and melt curve peak alignment with HPV controls, across a DNA concentration gradient……. **LV**

Table 22. Analysis of PCR amplification of known HPV-positive tonsil gDNA samples and melt curve peak analysis across a GP5+/GP6+ concentration gradient, with two gDNA concentrations…… **LIX**

Table 23. Analysis of PCR amplification of known HPV-positive tonsil gDNA samples and melt curve peak alignment with HPV controls, across an annealing temperature gradient…… **LXII**

Table 24. All 222 significantly changed protein hits identified in HPV-positive compared to HPV-negative non-cancerous tonsil tissue…… **LXXV**

Table 25. All 394 significantly changed protein hits identified in HPV-positive male compared to HPV-positive female non-cancerous tonsil tissue…….. **LXXIX**

Abbreviations

(ds)DNA – Double Stranded Deoxyribonucleic Acid

ACN – Acetonitrile

ACTB – Beta-actin

ADME – Absorption, distribution, metabolism, and excretion

BEB – Back Extraction Buffer

BID – BH3-interacting domain agonist

BP – Base Pair

C7 – Complement Component C7

CBP – CREB (cAMP response element) binding protein

CDC – The Centres for Disease Control and Prevention

CDK - Cyclin-dependent Kinase

cDNA – Complementary DNA

CIN – Cervical Intraepithelial Neoplasia

CK7 – Cytokeratin 7

cNLS – Classical Nuclear Localisation Signal

CRH – Calderdale Royal Hospital

CRN – Clinical Research Network

CV – Coefficient of Variation

DEA – Differential Expression Analysis

DEA – Differential Expression Analysis

DFS – Disease-free survival

DIA-NN – Data Independent Acquisition Neural Networks

DIA – Data Independent Acquisition

DTT – Dithiothreitol

E6AP – E6 Associated Protein

EGFR – Epidermal Growth Factor Receptor

EJC – Exon junction complex

ER – Endoplasmic reticulum

EtOH – Ethanol

FA – Formic acid

FBS – Foetal Bovine Serum

FC – Fold-change

FDR – False Discovery Rate

FFPE – Formalin-fixed paraffin-embedded

FGFR – Fibroblast Growth Factor Receptor

gDNA – Genomic DNA

H&E – Haematoxylin and eosin

hADA3 – Human Alteration/Deficiency in Activation 3

HAT – Histone acetyltransferase

HBRC – Human Biomaterials Resource Centre

HFKs – Human Foreskin Keratinocytes

HIV – Human Immunodeficiency Virus

HNC – Head and Neck Cancer

hnRNP - Heterogeneous Nuclear Ribonucleoprotein

HNSCC – Head and Neck Squamous Cell Carcinoma

HPLC – High Performance Liquid Chromatography

HPV – Human Papillomavirus

HPV-HNC – HPV-positive HNC

HPV-OPC – HPV-positive OPC

HR-HPV or HR – High-risk Human Papillomavirus or High-risk

HSPG – Heparan Sulphate Proteoglycan

hTERT – Human Telomerase Reverse Transcriptase

IDA – Iodoacetamide

IHC – Immunohistochemistry

IQR – Interquartile Range

ISH – *In situ* hybridisation

JCVI – The Joint Committee on Vaccination and Immunisation

Ki-67 – Kiel-67

KLK8 – Kallikrein related peptidase 8

LBC – Liquid Based Cytology

LCR – Long Control Region

LR-HPV or LR – Low-risk Human Papillomavirus or Low-risk

m/z – mass:charge

MAC – Membrane Attack Complex

MeOH – Methanol

MHC – Major Histocompatibility Complex

miRNA – Micro RNA

mRNA – messenger RNA

MS – Mass Spectrometer

MS – Milliseconds

MSM – Men who have sex with men

MSW – Men who have sex with women

NBF – Neutral Buffered Formalin

NEB – Nuclear Envelope Breakdown

NHANES – National Health and Nutrition Examination Survey

NHS – National Health Service

NHSCSP – National Health Service Cervical Screening Programme

NIHR – National Institute for Health and Care Research

NMD – Nonsense-mediated Decay

NOK – Normal Oral Keratinocyte

OPC – Oropharyngeal Cancer

OPSCC – Oropharyngeal Squamous Cell Carcinoma

OS – Overall survival

OSCC – Oral Squamous Cell Carcinoma

PANTHER – Protein Analysis THrough Evolutionary Relationships

Pap Test – Papanicolaou Test

PBS – Phosphate Buffered Saline

PC – Proprotein Convertase

PCA – Principal Component Analysis

PCR – Polymerase Chain Reaction

PD-1 – Programmed cell death protein-1

PD-L1 – Programmed cell death ligand

PEDF – Pigment derived-epithelium factor

PHE – Public Health England

pHR-HPV – Probable High-Risk HPV

PIK3CA - phosphatidyl inositol-3 kinase

PIS – Patient Information Sheet

PLWH – People Living With HIV

poly(A) - Polyadenylation

pRB – Retinoblastoma protein

PRH – Princess Royal Hospital

PTM – Post-translational modification

qPCR – Quantitative Polymerase Chain Reaction

qRT-PCR – Quantitative Reverse Transcriptase Polymerase Chain Reaction

RCF – Relative Centrifugal Force

RINe – RNA Integrity Number Equivalent

RNA – Ribonucleic Acid

rRNA – Ribosomal ribonucleic acid

RRP – Recurrent Respiratory Papillomatosis

RT – Room temperature

SARS-CoV-2 – Severe acute respiratory syndrome-Coronavirus-2

SCJ – Squamocolumnar Junction

STI – Sexually transmitted infection

SWATH-MS – Sequential Window Acquisition of all Theoretical Mass Spectra

TEAB – Triethylammonium bicarbonate buffer

TEM – Tetraspanin-enriched microdomains

TIC – Total ion current

TZ – Transformation zone

UHDB – University Hospitals of Derby and Burton

UHL – University Hospital Leicester

UHN – University Hospital Nottingham

UIN – Unique Identifier Number

UK – United Kingdom

UPLC – Ultra Performance Liquid Chromatography

URR – Upstream Regulatory Region

US – United States

WHO – World Health Organisation

COVID-19 STATEMENT:

This project was significantly impacted by Covid-19. The project started in January 2021, but full access to the laboratory was not allowed until May 2021. The work relied heavily on tonsil tissues collected from routine surgeries; however, these were paused due to the pandemic and no elective tonsillectomies were performed between January 2021-November 2021. When these surgeries did resume, tonsils were fixed in formalin to enable the safety of researchers.

Declaration:

I hereby declare that the research presented within this thesis was conducted by me, and that the analysis and writing up of these data is my own. Where others have made a contribution to the data presented, this is acknowledged where appropriate.

This project has been ethically approved by the University of Derby's Ethics Committee (2021 – reference: ETH2021-2440; Appendix B) and by the South West NHS REC (June 2016 – reference: 16/SW/009, IRAS ID: 194857; Appendix A), with additional tissues purchased from Human Biomaterials Resource Centre, University of Birmingham (approval 21-374; Appendix C).

Dedication:

For Momma and Dadda,

I know you would both be so proud.

This is for you.

Acknowledgements:

Over the past 4 years of this Ph.D., I have been lucky enough to have worked with, and been supported by, several people (too many to thank individually). There are however individuals who have been instrumental to this work, and to me personally throughout the years I have spent completing this Ph.D., who I would like to acknowledge here.

Firstly, I would like to thank my Ph.D. supervisor, Dr Elizabeth Marsh. Without you, none of this would have been possible. Elizabeth, thank you for giving me this opportunity back in 2020 and believing in me all the way. Your kind, calm, and motivating nature has been constant throughout this journey, and I could not have done it without you. You always have encouraged me and reminded me that I can do it – thank you will never be enough, but I truly do thank you from the bottom of my heart.

Also, to my supervisory team – Dr Laurice Fretwell and Dr Thomas Illingworth. Thank you for your continuous support, I appreciate everything that you have done throughout these past 4 years to make my Ph.D. experience so enjoyable. You have both been amazing and I feel incredibly lucky to have worked with and had such a wonderful supervisory team.

As part of my Ph.D. project, I had the pleasure of working with external collaborators at NTU. To Dr David Boocock and Dr Clare Coveney, thank you for welcoming me into your lab, I really enjoyed learning from you both. The early mornings and late nights in the lab were worth it in the end! Also, to Dr James Hind, thank you for your statistical expertise and work on the oral HPV meta-analyses, it has been greatly appreciated. I would also like to thank all those involved in the HPV clinical study at all of the participating hospitals; in particular, Mr Bindy Sahota and Dr Amandeep Mann at Royal Derby Hospital whose contributions were so significant to the progress of this work.

I would also like to thank Dr Rebecca Mallender. Thank you for always being there – we started our time at Derby together and I am thankful to have made a friend in you. Also, to all those I have had the pleasure of sharing the office and laboratory with over the past 4 years. Priya, Asha, Marie, Jake, Jaya, Agnieszka, Nevin (and many more!), I will miss working with you all.

Thanks too goes to Oracle Head and Neck Cancer, UK and the Rosetrees Trust who funded this research.

To all my friends: Jade, Megan, Gemma, Jessica, Sheridan, Josh, and Lauren – you're all the best and I love you all. A special thank you though goes to my best friends; Jade – thank you for giving me the confidence to embark on this Ph.D. journey and for all the supportive words of encouragement. The memories we've had during the past 4 years (with the fab 4 – plus Henry and Tom) have been crucial in keeping me going and to getting me over the finish line. Megan – you have been my best friend for as long as I can remember, thank you for literally everything. Especially for all the curries and catch-up walks that have helped me get through the entire Ph.D. but particularly my final year. Gemma – you have been there right from the beginning of my University journey back in Salisbury Hall. Thank you for everything, not only your kind words during my Ph.D. but throughout the undergraduate days too. You have always believed in me and you have always been at the end of the phone when I have needed encouragement to keep going.

Finally, to my family, Mum, Dad, Sally and Lucy. It is impossible to put into words how thankful I am to you all. Mum and Dad, you are the most supportive, yet never pushy, parents. Your true belief in me has enabled me to complete this Ph.D.; thank you for everything. Sally and Lucy – you were my first and forever best friends. Knowing that I have you all (and the dogs of course – Billy, Coco, and Nugget) has been the ultimate motivation to keep going – this is for you all.

ABSTRACT

Background: Human papillomavirus (HPV)-positive oropharyngeal cancers (OPCs) of the base of the tongue and tonsils are increasing. The knowledge of the natural history of the virus within the oropharynx is limited, with prospective studies of oral HPV pathogenesis sparse. The prevalence of oral HPV within healthy individuals has been reported to be as high as ~80%. The risk of infection is linked to alcohol consumption, smoking, age, and certain sexual behaviours. It is men that are at greater risk of oral HPV infection and disease.

Objectives: This study aimed to provide an updated estimate of global oral HPV prevalence within healthy adults; to determine the prevalence of HPV within the non-cancerous palatine tonsil tissues of adults and scrutinise the pathogenesis of the virus; and, to examine the tonsil proteome of HPV-positive individuals with comparison to HPV-negative tissues.

Methods: A systematic review and meta-analysis of literature of adult oral HPV prevalence was performed. We established a clinical study on the NIHR-portfolio. Non-cancerous tonsils, a tonsillar swab, and a lifestyle questionnaire were collected from adults undergoing routine tonsillectomy surgery within five UK hospitals. We developed a protocol to isolate DNA from formalin-fixed (10% NBF) tonsil tissue for molecular analysis. A modified HPV-DNA qPCR screen was used to determine tonsillar HPV-status followed by HPV genotyping. HPV-positive tissues have been characterised for p16 and CK7 histologically. Protein biomarkers within tonsil tissues were analysed using a non-biased technique, SWATH mass spectrometry.

Results: We provide an up-to-date estimate of the global burden of overall oral HPV prevalence by meta-analysis of 5.044% (95% CI 4.897-5.193%); interestingly, men were twice as likely to have oral HPV than women. In the clinical study, we describe the prevalence and viral pathogenesis of HPV within the non-cancerous tonsils of our prospective sample. We have determined a 10.2% oropharyngeal HPV prevalence (N=12/118; comprising 4 males, 7 females, and 1 unknown individual). Proportionally, more men than women were HPV-positive, but this was not significant (p=0.49). We found HPV-status to be mainly concordant in both the left and right tonsil tissues; however, positivity varies across the tissue. Three viral types were detected within these tissues (HPV-18, -16, -73); HPV-18 was most prevalent within these individuals (N=10/12, 83.3%). HPV-16 was only detected in a co-infection with HPV-18. Upregulated proteins in HPV-positive tissues included those with immune function

(p<0.01), and functionally, the respiratory electron transport chain was the most enriched pathway. Analysis of dysregulated proteins within HPV-positive males versus HPV-positive females revealed enriched pathways to include rRNA processing and associated functions (adj. p<0.001).

Conclusions: Globally, oral HPV prevalence is a greater burden to men than women. Oropharyngeal HPV is prevalent within adult routine tonsillectomy patients in the UK. We describe changes in protein expression based on HPV-positivity in non-cancerous tonsil tissue, that require future validation to provide a 'signature' of HPV-positivity. Future multivariate analysis of HPV prevalence, risk factors, and protein expression, would allow identification of populations at risk of HPV-mediated disease who would benefit from intervention, and monitoring for HPV-induced cellular changes.

CHAPTER 1: Introduction

1.0. Human Papillomaviruses – An Introduction

Part of the Papillomaviridae family, Human Papillomaviruses (HPVs) are a group of circular double-stranded DNA (dsDNA) viruses of ~55 nm in diameter, with a genome of approximately 8 kb (Zheng and Baker, 2006, Graham, 2010, Bernard et al., 2010). Often sexually transmitted, HPVs infect, and then replicate within the nucleus of the basal cells of the squamous stratified epithelium (Edwards et al., 2013). The viral genome generally encodes for 8 proteins that possess roles within the infectious life cycle of the virus. These include 6 regulatory proteins (E1, E2, E4, E5, E6, E7) and 2 structural proteins (L1 and L2) (Zheng and Baker, 2006, Graham, 2010). At present, over 200 viral types have been identified that each have an L1 sequence of at least 10% difference (Bernard et al., 2010, Doorbar et al., 2015, Karolinska Institutet, 2023).

HPVs are broadly categorised as low-risk (LR) (including viral types HPV-6 and -11) (Zheng and Baker, 2006), and high-risk (HR) (HPV-16, -18, -31, -33, -34, -35, -39, -45, -51, -52, -56, -58, -59, -66, -68, -70, -73, -83) (Burd, 2003, Muñoz et al., 2003). The vast majority of infections are cleared by the immune system; however, persistent infection with LR HPV-6 and -11, can in some instances, cause benign lesions such as condyloma acuminata, these are anogenital warts specifically caused by HPV (Zheng and Baker, 2006), and LR types can also cause recurrent respiratory papillomatosis (RRP) (Fortes et al., 2017) and common warts and plantar warts (Doorbar et al., 2015). The HR-HPV types can cause certain cancers, including cervical, anal, penile, vulval, vaginal, and some cancers of the head and neck (HNC) (Szymonowicz and Chen, 2020) (Table 1). HPV-16, -18, -31, -33, -35, -45, -51, -52, -56, -58, and -59 have been recognised as class I carcinogens by the World Health Organisation (WHO) since 2012 (World Health Organization, 2024d). HPV is most well known as the causative agent of cervical cancer; the incidence and prevalence of which has declined over the past 30 years in the United Kingdom (UK) and other developed countries due to screening strategies and vaccination (Lechner et al., 2022, Falcaro et al., 2024), whereas incidence remains high in other parts of the world (Hull et al., 2020, Singh et al., 2023).

Table 1. HPV types and associated diseases. HPV viral types, their risk category, and associated diseases are summarised.

HPV type/s	HR-HPV or LR-HPV	Disease/s
HPV-16 and HPV-18	HR-HPVs	• Cervical cancer = ~99% HPV-positive • OPCs and other HNCs = >90% caused by HPV-16 • Anogenital cancers
HPV-31, -33, -34, -35, -39, -45, -51, -52, -56, -58, -59, -66, -68, -70, -73, -83	HR-HPVs	• Also associated with cancers (including cervical, HNC, anogenital)
HPV-6, HPV-11	LR-HPVs	• Condyloma acuminata • Recurrent respiratory papillomatosis (RRP)
HPV-1, -2, -4, -27, -57	N/A	• Common warts (i.e., on hands and verruca vulgaris on feet)

1.0.1. HPV Genome Structure

The HPV genome is divided into three regions; early, late, and non-protein encoding long control region (LCR; otherwise referred to as the upstream regulatory region (URR)). The early region contains 6 genes (E1, E2, E4, E5, E6, E7), and the late region encodes L1 and L2 (Zheng and Baker, 2006). The roles of the encoded proteins within the viral life cycle are discussed later. The early and late regions are flanked by polyadenylation (poly(A)) sites, whose differential modification are fundamental to the viral transcription programme (Figure 1).

The LCR varies in size, dependent on viral type, and occupies 7-11% of the genome. Although not encoding for any functional proteins, the LCR is necessary for the regulation of viral gene expression. It contains the origin of replication, indicated by an E1 protein binding site; indeed, the E1 protein is pivotal to HPV replication (Egawa et al., 2012). A number of transcription factors can bind within the LCR enhancer, and co-ordinate viral gene expression at early (p97 in HPV-16, p105 in HPV-18), and late (p670 in HPV-16, p811 in HPV-18) promoters (Burley et al., 2020). The number of major promoters differs between HPV types (Yu et al., 2022). Viral protein E2 is also essential to transcription regulation, owing to its ability to bind four sequences within the LCR to manipulate viral expression, and its involvement in recruitment of transcription factors/proteins to the viral genome coordinate appropriate gene expression (McBride, 2013, Burley et al., 2020).

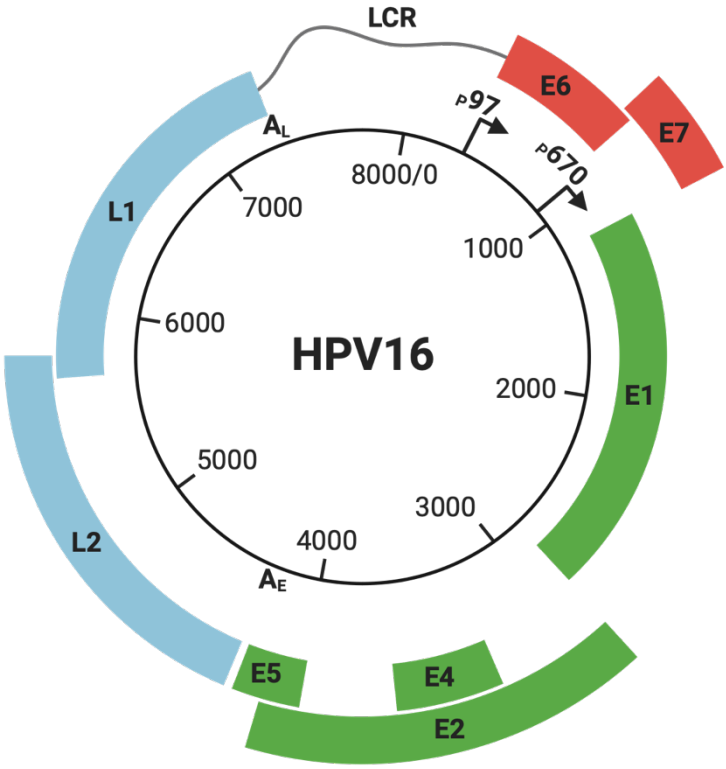


Figure 1. The HPV-16 genome structure. The circular double-stranded HR HPV-16 genome is divided into the early, late, and long control region (LCR). The early region contains 6 genes (E1, E2, E4, E5, E6, and E7). The late region contains genes L1 and L2. The regions are flanked by two polyadenylation sites (A_E and A_L). In the case of HPV-16, transcription is controlled from two promoters, the early p97, and late p670 promoters. Image was created using BioRender.com.

1.0.2. The viral life cycle of HPV

The HPV life cycle is closely co-ordinated with the innate terminal differentiation of its target cell, the keratinocyte (Graham, 2017b). HPVs access the basal lamina through micro-wounds or abrasions within the multi-layered stratified epithelium (Schneider, 1993, Doorbar, 2005, Doorbar et al., 2012, Ozbun, 2019). Here, infectious virions establish a novel infection within differentiating, mitotically active cells (Ozbun, 2019). The viral pathogenesis within the cervix is well understood, and the majority of current knowledge is within the context of HPV-induced cervical carcinogenesis. In the case of the cervix, it is the single-layered cells of the squamocolumnar junction (SCJ) between the ecto- and endocervix in which HR viral types establish productive and non-productive infections. Indeed, unlike infection of the basal cells that are not usually accessible without an abrasion of the cell layers, the single celled natured of the SCJ give direct access of the virus to their target cell (Yang et al., 2015). These cells are able to become the basal layer of the transformation zone (TZ) (Doorbar, 2005, Graham, 2017a).

1.0.3. Viral attachment and cell entry

HPVs target heparan sulphate proteoglycans (HSPGs) on the epithelial surface, or the basement membrane (Johnson et al., 2009, Kines et al., 2009, Schiller et al., 2010). A conserved region of the carboxy-terminus of the viral capsid protein, L1 that contains residues lysine-278, -356, and -361 mediates this interaction (Knappe et al., 2007). Upon binding HSPG, L1 then undergoes a conformational change that allows for cleavage of the minor capsid protein, L2, by the proprotein convertase (PC), furin (Richards et al., 2006, Kines et al., 2009). This cleavage step appears to be one that is imperative to infection, as L2 cleavage alone can bypass the need for HSPG attachment (Day et al., 2008). In addition to this, other changes to L1 that impact viral internalisation include L1 cleavage by kallikrein related peptidase (KLK8) after binding to HSPGs (Cerqueira et al., 2015), and/or a cyclophilin induced conformational change (Bienkowska-Haba et al., 2009), and these mechanisms can promote the important L2 protein exposure and processing. However, the role of cyclophilins has been questioned, with evidence that L2 cleavage can occur in their absence (Bronnimann et al., 2016). The essential cleavage of L2 is followed by transfer of the virion to an unknown uptake receptor, most probably not a HSPG, to allow viral entry to its target cell (Selinka et al., 2007).

Viral entry is facilitated by an endocytic mechanism, similar to the process of macropinocytosis (Schelhaas et al., 2012). Originally believed to be clathrin- or caveolin-dependent (Day et al.,

2003, Bousarghin et al., 2003), more recent work has suggested a role for tetraspanin-enriched microdomains (TEMs) (Spoden et al., 2008, Schelhaas et al., 2012, Spoden et al., 2013). HPV-16 has been shown to co-localise with tetraspanins, specifically CD63 and CD151 on the cell membrane. Interestingly, this work provided evidence of inhibited viral entry in their absence when knocked down by siRNA transfection or blocked by tetraspanin antibodies (Spoden et al., 2008). Once internalised, the virus interacts with the microtubule network, specifically dynein, and is trafficked to the trans-Golgi network in a furin cleavage-dependent manner (Day et al., 2013). HPV then awaits nuclear envelope breakdown (NEB), for nuclear entry, within a transport vesicle (Pyeon et al., 2009, DiGiuseppe et al., 2016). NEB is essential for viral entry to the nucleus; however, it seems that the virus itself does not promote mitotic activity and the concomitant NEB (Aydin et al., 2014). Instead, the use of the host cell's normal replication process to enter the nucleus is perhaps a viral adaptation acquired to stay undetected by the immune system (Aydin et al., 2014).

1.0.4. Viral replication

Upon entry to the nucleus there is a period of genome amplification which establishes the HPV infection within the host cell. During this stage, the virus exists as a low copy number episome (an extra-chromosomal element) of approximately 50-100 copies per cell within the cells of the basal lamina. The low copy number is established, likely to avoid immune detection, and forms a 'reservoir' of infection (Schneider, 1993, Doorbar, 2005, Bodily and Laimins, 2011, Egawa and Doorbar, 2017, Della Fera et al., 2021, Evande et al., 2023). The early proteins E1 and E2 are fundamental to this initial viral establishment, and work cooperatively at their target sequences to coordinate replication (Frattini and Laimins, 1994). E2 acts as an intermediate between viral DNA helicase, E1, and the viral genome, to allow E1 to unwind the viral dsDNA (Burley et al., 2020). The next phase is viral maintenance, and in spite of their early importance, there is suggestion that neither E1 nor E2 are required here (Angeletti et al., 2002, Egawa et al., 2012). However, others have described E2 as crucial in mediating even inheritance of HPV genomes into new host cells as E2 can interact with host proteins (for example, bromodomain containing protein - Brd4) to form an intermediate between host DNA and the HPV virus during mitosis to ensure segregation between daughter cells, and avoid genome loss to the cytoplasm (Burley et al., 2020).

As the infected keratinocyte differentiates and migrates through the layers of stratified squamous epithelium, the virus enters a period of amplification. Regulatory proteins, E6 and

E7, that are otherwise repressed by E2 at an early promoter, are upregulated (Bernard et al., 1989, Day et al., 2013). Through interaction with cell cycle regulators, p53 and pRb (Retinoblastoma protein), E6 and E7, respectively, promote re-entry of differentiated keratinocytes into mitosis, through S-phase re-entry and an extended G2-phase (Wang et al., 2009, Banerjee et al., 2011). The mechanism of E6 and E7 regulation of p53 and pRb is outlined at 1.1.1.3. The viral interactions that drive progression through the cell cycle are imperative to a productive infection, as HPV itself relies on its host cell's machinery for replication (Kajitani et al., 2012).

E4 and E5 proteins are also implicated in the viral life cycle, albeit indirectly. E4 protein levels rise, and cause an accumulation within the upper epithelium, as a result of its fibril formation capabilities (McIntosh et al., 2008, Doorbar, 2013). This upregulation may be important for post-genome amplification. However, the HR HPV-16 E4 protein can also promote viral amplification indirectly, though this activity is viral type-dependent (Egawa et al., 2017). Interestingly, both E5 and E6 appear to promote koilocytosis, a specific type of squamous cell that is a pathological indicator of a cervical HPV infection (Krawczyk et al., 2008). To conclude the cycle, E2 represses early gene expression from the LCR (Burley et al., 2020). The levels of capsid proteins L1 and L2 are subsequently elevated, and they coordinate genome packaging and virion release (Bodily and Laimins, 2011, Doorbar et al., 2012) in the cornified epithelial layers (Figure 2).

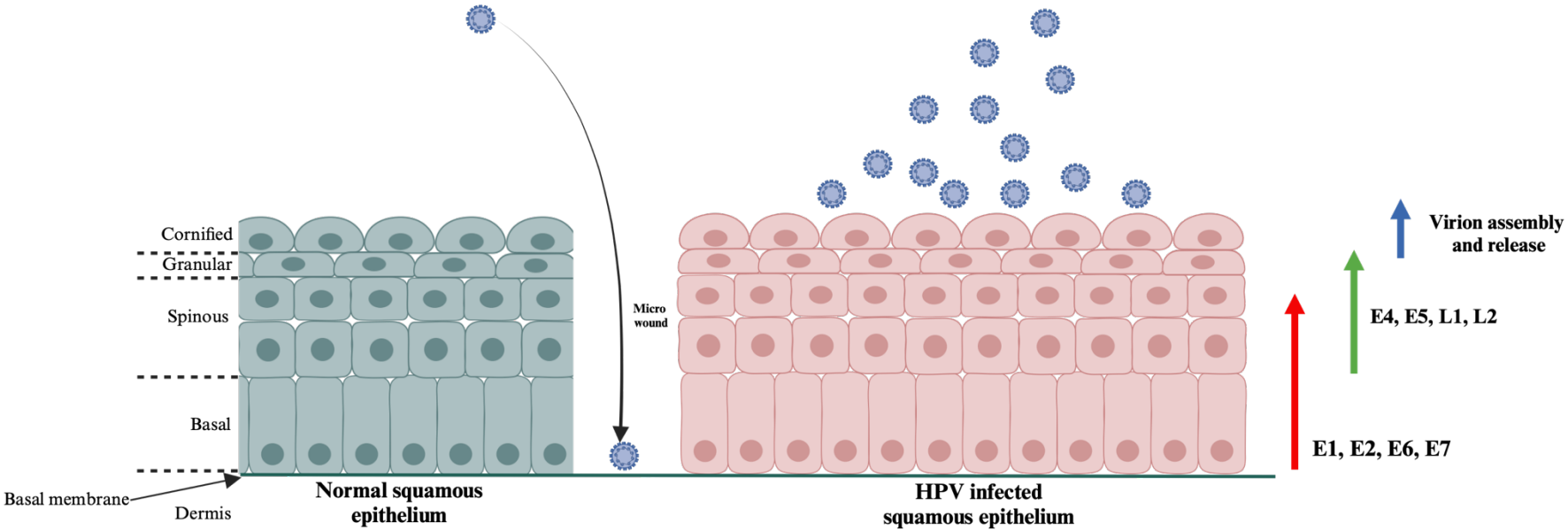


Figure 2. The lifecycle of HPV infection in the cells of the squamous stratified epithelium. The normal squamous epithelium is composed of the basal, spinous, granular, and cornified cell layers. HPV can gain access to the basal membrane through micro-abrasions or -wounds in the stratified epithelium, where it can infect its target cell. HPV genes are expressed in an intricate pattern to coordinate the viral life cycle. Early gene expression during infection establishment and maintenance include E1, E2, E6, and E7. As the infected cell migrates through the epidermal layers HPV enters a period of viral genome amplification (E6, E7, E4, E5 expression). Finally, late genes L1 and L2 are expressed to coordinate virion assembly and release. Image was created using BioRender.com.

Adapted from: (Moody, 2017)

1.0.5. Viral gene expression

To enable the correct pattern of viral gene expression required for successful completion of the viral life cycle, HPV transcription can be further regulated by direct genome modification, and post-transcriptional processing. For example, the methylation of CpG dinucleotides at E2 binding sites inhibits its binding *in vitro*, and results in the upregulation of otherwise repressed genes (Thain et al., 1996). Conversely, to promote E2 repression, HPV makes use of host cell factors including demethylase, SMCX, and Brd4. These have both been shown to work in conjunction with E2 at its binding sites (Wu et al., 2006, Smith et al., 2010). Interestingly, it appears that HPV can make use of the host protein Brd4, independently of E2, to repress downstream viral gene expression. For this to occur, tumour suppressor TIP60, a cellular histone deacetylase, is able to recruit Brd4 to the HPV genome. However, TIP60 itself is also a target for E6-mediated degradation in both LR- and HR-HPV types to then promote E6 expression (Jha et al., 2010).

Transcription from the viral genome does not produce monocistronic (i.e., those coding for a single protein) messenger RNA (mRNA) but rather polycistronic mRNA. Indeed, HPV mature mRNA transcripts are produced from pre-mRNAs containing more than one coding region, and so the post-transcriptional processing of pre-mRNAs is essential to enable the translation of the appropriate protein (Wang et al., 2024). HPV regulates this through polyadenylation at either the early or late poly(A) sites, as well as alternate RNA splicing to control the stability and content, respectively, of the mRNA transcript to be translated (Zheng and Baker, 2006). For example, E6 and E7 are encoded by the same pre-mRNA but production of the latter mRNA (E7) requires splicing; and whilst there have been various suggestions of the precise mechanism of this regulation, it involves interaction with host cellular proteins (for example, heterogeneous nuclear ribonucleoproteins (hnRNPs)) and the activity of some cell signalling pathways can also influence which mature mRNA transcript is produced (Wang et al., 2024).

1.1. HR-HPV – a virus with carcinogenic potential

HPV was identified as carcinogenic a number of decades ago by the work of Nobel Prize winner, the late Professor Harald zur Hausen, who was the first to describe HPVs as a causative agent of cervical cancer (Zur Hausen, 1977). The mechanisms of, and viral types involved in, the development of HPV-positive cervical cancer are now well-documented, of which ~99% of cases are attributable to the virus (World Health Organization, 2024a). The thorough understanding of the pathogenesis of HPV within the cervix, along with public health

interventions of screening and vaccination have had a huge impact on cervical disease across the world (Palmer et al., 2019, Falcaro et al., 2021, Lechner et al., 2022).

1.1.1. HPV-induced cervical carcinogenesis

HPV infections of the cervix, including those of HR viral types, are usually detected and cleared by the body's immune system. Within the cervix, carcinomas develop from lesions called cervical intra-epithelial neoplasias (CINs) (Snijders et al., 2006) that are graded I, II, and III, based on the severity of the cell transformation. HPV-driven carcinogenesis is not the natural nor definite endpoint of an infection. However, under certain circumstances, the viral infection persists (Crosbie et al., 2013). One example of a cause for viral persistence is a weakened immune system, caused by an existing infection with the human immunodeficiency virus (HIV), or more likely in the general population, due to immunosuppressive effects of alcohol consumption/smoking tobacco (Appleby et al., 2006, Adler et al., 2015, Oh et al., 2015), although the reasons for viral persistence differ. A persistent HPV infection is crucial for progression to malignancy (Koshiol et al., 2008, Castle et al., 2009). Viral oncoproteins, E6 and E7, drive this persistence, and although not as well understood, the E5 protein has also been implicated (Maufort et al., 2010). Another hallmark of HPV-positive cervical malignancies is the integration of E2 into the host genome at its hinge region (Vernon et al., 1997, Arias-Pulido et al., 2006, Cheung et al., 2006). The importance of the aforementioned proteins are discussed throughout this section.

1.1.1.2. Integrated or episomal viral forms

The classical model of HPV-induced cervical carcinogenesis is through the integration of the HPV genome into the host cell DNA, at the E2 viral gene. This is often the case within cervical malignancies, and the presence of integrated viral forms is usually linked to poor prognosis (Vernon et al., 1997, Kalantari et al., 1998). The likelihood of HPV integrating is largely influenced by the viral type, with HPV-16 and HPV-18 most frequently found integrated in pre-cancerous cervical cells and established cancers, reviewed in (Catalán-Castorena et al., 2024). Importantly, this event does not occur as part of the natural viral life cycle (Pett and Coleman, 2007). As a fundamental regulator of HPV transcription (Bernard et al., 1989), the disruption of E2 function results in the overexpression of both E6 and E7 (Jeon and Lambert, 1995), whose downstream oncogenic effects are discussed later within this section. In addition, the E2 protein is known to transcriptionally inactivate the human telomerase reverse transcriptase gene (hTERT) (Lee et al., 2002). Therefore, the loss of E2 function may also

promote oncogenic transformation through increased hTERT expression and the subsequent maintenance of telomere length, that allows the host cell to overcome the Hayflick limit of division (Rousseau and Autexier, 2015).

However, HPV integration is not considered the sole mechanism of oncogenic transformation within the cervix, with evidence of HPV-positive carcinomas harbouring both entirely episomal, and mixed (integrated and episomal) viral forms (Vernon et al., 1997, Cheung et al., 2006). Investigations into the presence of all forms of the viral genome have highlighted discrepancies between the viral type present within the tumour. One study identified episomal-only viral forms in HR HPV-16 positive cervical carcinomas (26% of all HPV-positive tumours assessed). However, no episomal-only genomes were detected in other HPV-associated tumours (including HPV-18, -25, and -31 positive) (Pirami et al., 1997). Contrastingly, others have determined that viral type HPV-16, as well as HR-HPV-18, and -45, are largely present as integrated into the host cell DNA (Vinokurova et al., 2008). It is, therefore, unclear whether the integration of HPV is dependent on viral type; however, it does show that integration cannot be the only pathway to malignancy.

These discrepancies in the viral forms detected may be due to the test method used and/or the material analysed (Kalantari et al., 2001, Gudleviciene et al., 2015). An example is the use of the polymerase chain reaction (PCR) for E1/E2 gene analysis (that looks for the loss of E2 expression), which struggles to identify the presence of mixed viral forms in a sample. As such, more accurate methods, including single-cell RNA-sequencing are being described to enable the detection of both episomal and integrated HPV (Kalantari et al., 2001).

The mechanisms involved in the oncogenic transformation of cells containing entirely episomal HPV is debated. The importance of E6 overexpression is clear, amid evidence that oncogenic E6 can be upregulated regardless of E2 disruption by integration (Hong et al., 2017). Furthermore, characterisation of cervical carcinomas containing either integrated or episomal HPV revealed similar levels of E6/E7 mRNA. This was speculated to be due to the methylation of E2 binding sites in episomal genomes, which presents an alternative mechanism to perturb protein function (Cheung et al., 2013). Recently, an *in vitro* cellular model of cervical carcinogenesis has been proposed for tumours containing episomal-only HPV-16. In brief, their data show that in cervical cancer cells with episomal HPV-16, activation of MAP-kinase (MAPK) signalling, along with mutations in the known key oncogenic proteins MYC and

phosphatidyl inositol-3 kinase (PIK3CA), can promote malignant transformation (Wongjampa et al., 2023). Indeed, a point mutation within the helical domain of exon 9 of PIK3CA is a common mutation within cervical cancers (McIntyre et al., 2013), and so, it is possible that episomal HPV-16 also uses this to drive carcinogenesis.

1.1.1.3. The roles of E6, E7, and E5 in cervical carcinogenesis

The mechanisms by which oncogenic E6 and E7 promote malignant transformation are well-defined. As previously discussed, overexpression of E6 and E7 is classically believed to be a result of the loss of their transcriptional regulation by E2, following viral integration. This integration event also causes the introduction of host DNA into the viral genome, and the subsequent loss of an instability motif, an A/U rich region in the 3' untranslated region (UTR) that increases the stability, and therefore the half-life, of the E6/E7 transcripts (Jeon and Lambert, 1995). The E6 and E7 proteins largely promote carcinogenesis through interaction with cell cycle regulators p53 and pRb, respectively (Crosbie et al., 2013), although both proteins have other carcinogenic activity.

Known as the 'guardian of the genome', p53 is a fundamental tumour suppressor protein, involved in the regulation of the normal cell cycle (Shaw, 1996). E6 is able to inhibit p53 through a ternary complex with the host ubiquitin E3 ligase, E6AP (E6 associated protein), that promotes p53 ubiquitination and proteasomal degradation (Scheffner et al., 1990, Scheffner et al., 1995). The subsequent loss of p53 enhances genomic instability, cell proliferation, and malignant transformation within HPV infected cells (Thomas et al., 1999). It is reported that E6-induced inhibition of p53 can occur in the absence of E6AP, with evidence of a phosphorylation-dependent feedback loop between the proteins (Vats et al., 2022). Moreover, E6 can indirectly target p53 function by binding to, and repressing, histone acetyltransferase p300, and CBP (cyclic AMP (cAMP) response element (CREB) binding protein) (Patel et al., 1999, Scarth et al., 2021), as well as indirect inactivation by promoting the degradation of p53 transcriptional co-activator, human alteration/deficiency in activation 3 (hADA3), through post-translational modifications (PTMs) including SUMOylation and ubiquitination (Chand et al., 2014).

Aside from the direct inactivation of p53, other pro-oncogenic effects of the E6 oncoprotein have been described. Firstly, contrastingly to E2's inhibition of hTERT, E6 possesses the ability to activate telomerase (Klingelhutz et al., 1996), and inactivate hADA3, as discussed

above, to further perturb p53 function and achieve pro-oncogenic effects to the infected cell. Furthermore, an important distinction between the HR- and LR-HPV viral types is the presence of a PDZ-binding domain at the C-terminus of HR E6 proteins (Spanos et al., 2008). Indeed, this domain appears to be crucial to the initiation of malignant transformation in HR-HPV infected cells (Yoshimatsu et al., 2017). There is evidence that the various cellular targets of E6 (i.e., p53 inactivation versus PDZ-domain containing protein targeting) are critical to different stages in cancer development (Simonson et al., 2005).

Overexpression of the E7 oncoprotein is also linked to carcinogenesis. HR-HPV E7 promotes degradation of pRb by ubiquitination, and this indirectly stabilises p53 (Boyer et al., 1996, Crosbie et al., 2013). However, this effect on p53 is opposed through its own interaction with E6 (Jones et al., 1997b, Eichten et al., 2002, Vande Pol and Klingelhutz, 2013). Moreover, E7 can perturb pRb function through preventing its interaction with transcription factor E2F (Wu et al., 1993), to promote uncontrolled cell cycle progression. The loss of pRb regulation of E2F promotes progression through the cell cycle, from G_1 to S-phase, enabling the virus to replicate, persist, and continue to promote oncogenic changes. In addition, E7 promotes DNA synthesis in differentiated keratinocytes (Jones et al., 1997a); by preventing the activity of cyclin-dependent kinase (Cdk) inhibitor, p21, E7 can interfere with the otherwise inverse relationship between differentiation and DNA synthesis (Jones et al., 1997a), resulting in increased cell proliferation.

Although, perhaps not as well understood (as the E6/E7 oncoproteins), E5 is also implicated in driving the viral persistence that can lead to carcinogenesis (Maufort et al., 2010). Several studies have described E5 as an activator of the epidermal growth factor receptor (EGFR), promoting the proliferation of the host cell (Leechanachai et al., 1992, Pim et al., 1992, Kim et al., 2006). Studies in mice have provided further evidence that E5 can function alone or in combination with a single other oncoprotein (E6 or E7) (Maufort et al., 2010). However, the role of E5 remains unclear.

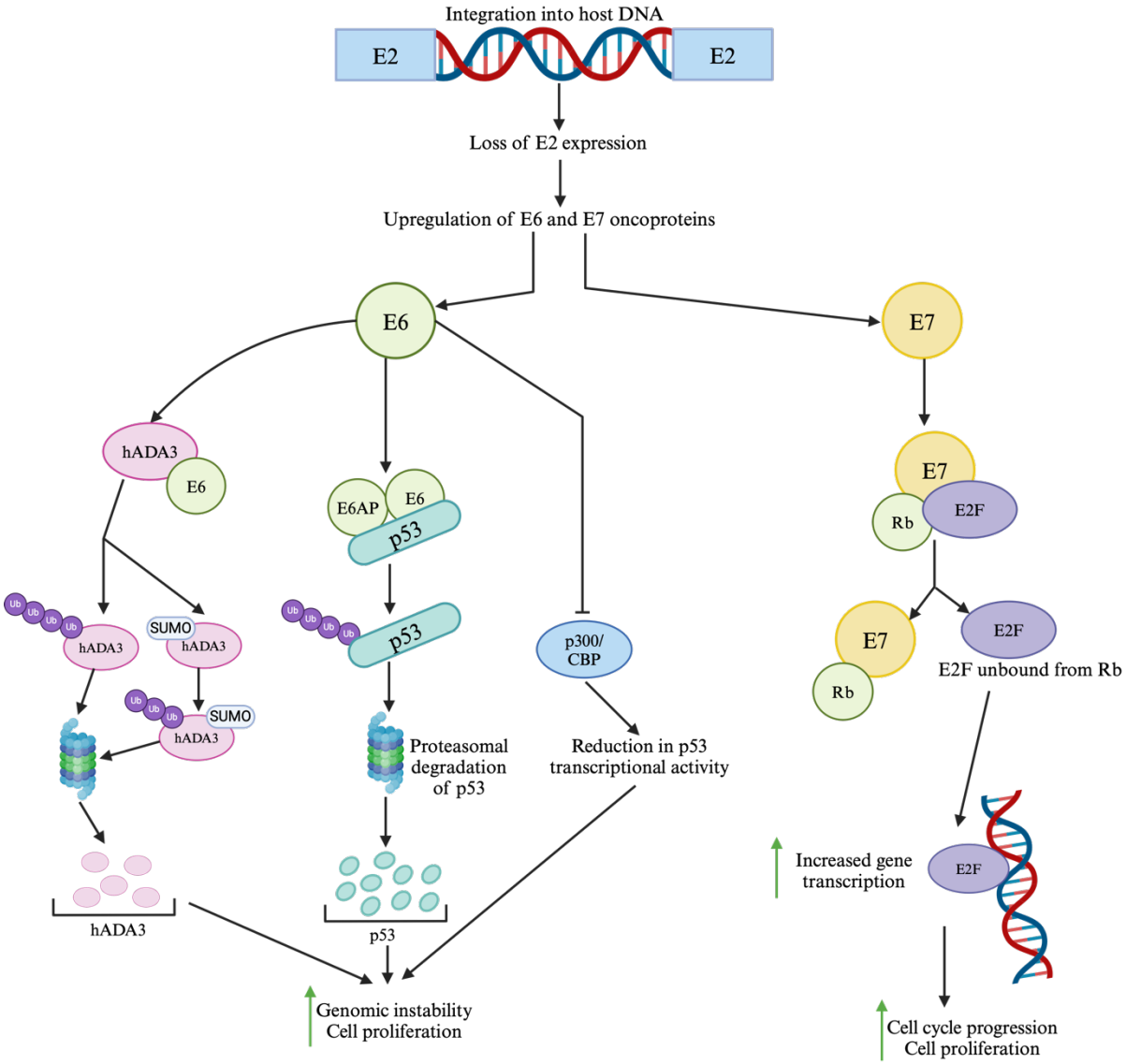


Figure 3. Interaction of HR-HPV oncoproteins, E6 and E7, with host cell proteins. Integration of HPV into the host cell DNA usually results in the loss of E2 expression, and the subsequent upregulation of E6 and E7 oncogenes. HR-HPV E6 binds p53 in a ternary complex with E6AP (a ubiquitin E3 ligase) to promote its proteasomal degradation. Indirect actions on p53 by E6 include ubiquitination and/or SUMOylation of hADA3 and repression of transcriptional co-activator p300/CBP. One common mechanism of HR-HPV E7 proteins is to prevent the interaction of Rb and E2F. Together these interactions lead to increased genomic instability, cell proliferation, and progression through the cell cycle. Figure created in BioRender.com

1.1.2. HPV-positive Oropharyngeal Cancer (HPV-OPC)

HPV-OPC is a HNC of the base of the tongue and tonsils. In contrast to cases of cervical cancer, HPV-OPC diagnoses are increasing at a global rate of ~4% per year (Tim Waterboer, personal communication), and the incidence of male OPC has now surpassed cervical cancer in both the UK and the United States (US) (Lechner et al., 2022) (Figure 4). Indeed, HPV-OPC currently comprise 30% of all OPC diagnoses, and this equates to >29,000 cases worldwide (de Martel et al., 2017). As with HPV-positive cervical malignancies, HR HPV-16 is the most common HR type in the head and neck, dominating HPV-OPC (Schache et al., 2016, Lechner et al., 2022) and HPV-16 is responsible for over 90% of all HPV-OPCs (Gillison et al., 2015).

The human oropharynx is composed of the soft palate, left and right palatine tonsils, the back of the throat, and the base of the tongue (Shiba and Chhetri, 2019) (Figure 5). The symptoms of OPC regardless of HPV-status as described by the NHS, are as follows: 'a neck lump, difficulty swallowing, or a persistent sore throat' (NHS UK, 2021).

HPV-OPCs are considered a distinct disease to HPV-negative OPCs. The two cancer types are linked to different risk factors; however, there are several crossovers. In brief, the traditional blueprint for HPV-negative HNCs is more frequent participation in lifestyle behaviours such as alcohol consumption and/or tobacco smoking (Hashibe et al., 2009), whilst HPV-OPCs are commonly linked to engagement in certain sexual behaviours, notably oral sex (Gougousis et al., 2020).

HPV driven OPCs are assumed to originate within cells of the tonsillar crypts (Kim et al., 2007). The crypts are invaginations of reticulated epithelium within the tonsil epithelium that have a significant immune presence (Faulkner et al., 2000). Promisingly, HPV-positivity is linked to better overall survival (OS) rates (Ang et al., 2010); this, along with their diverse molecular profiles, mean that HPV-OPCs are possibly suitable for different treatment regimes.

Yet despite the rising cases of HPV-OPC, we simply do not know the precise mechanism of HPV-driven carcinogenesis within the oropharynx. What is known of the pathogenesis of HPV within OPC, risk factors for oral HPV infection/HPV-OPCs, and biomarkers of, and treatments for, HPV-OPCs are now discussed in detail.

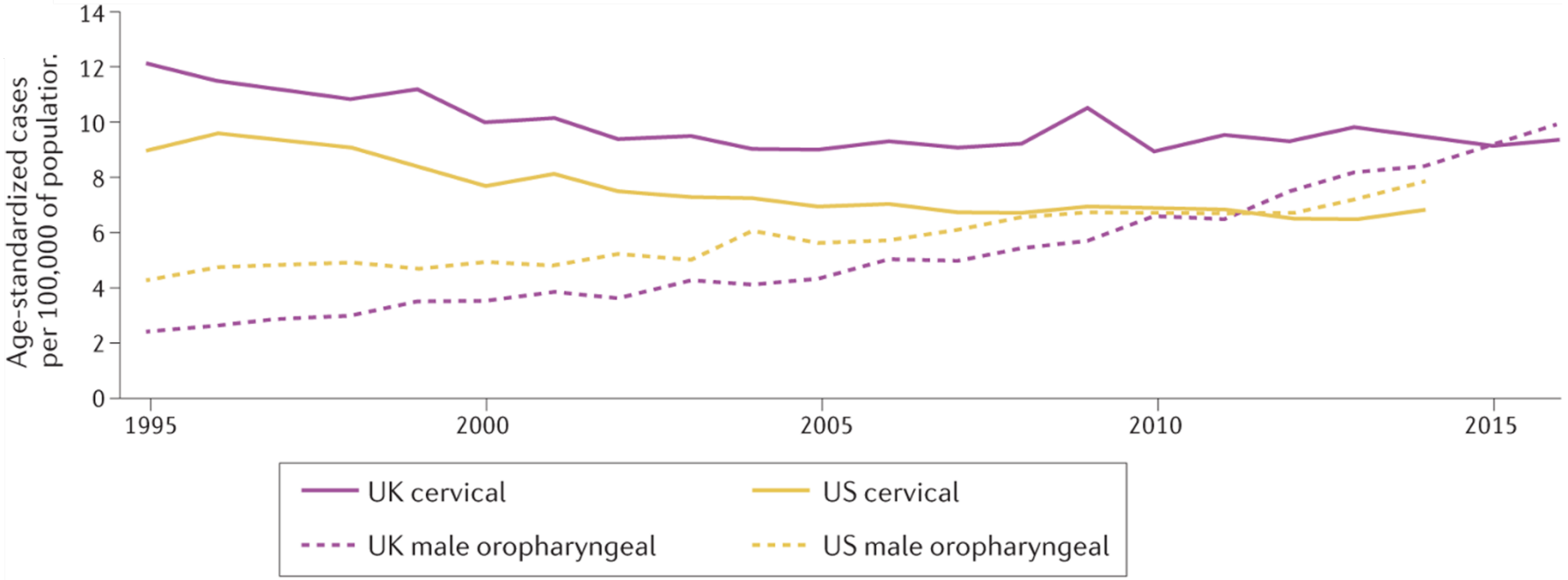


Figure 4. Age-standardised cases of cervical, and male oropharyngeal cancer (OPC). Over a 20-year period, between 1995 and 2015, the age-standardised cases of male OPC cases in both UK and US, have surpassed those of cervical cancer.

Figure from: (Lechner et al., 2022)

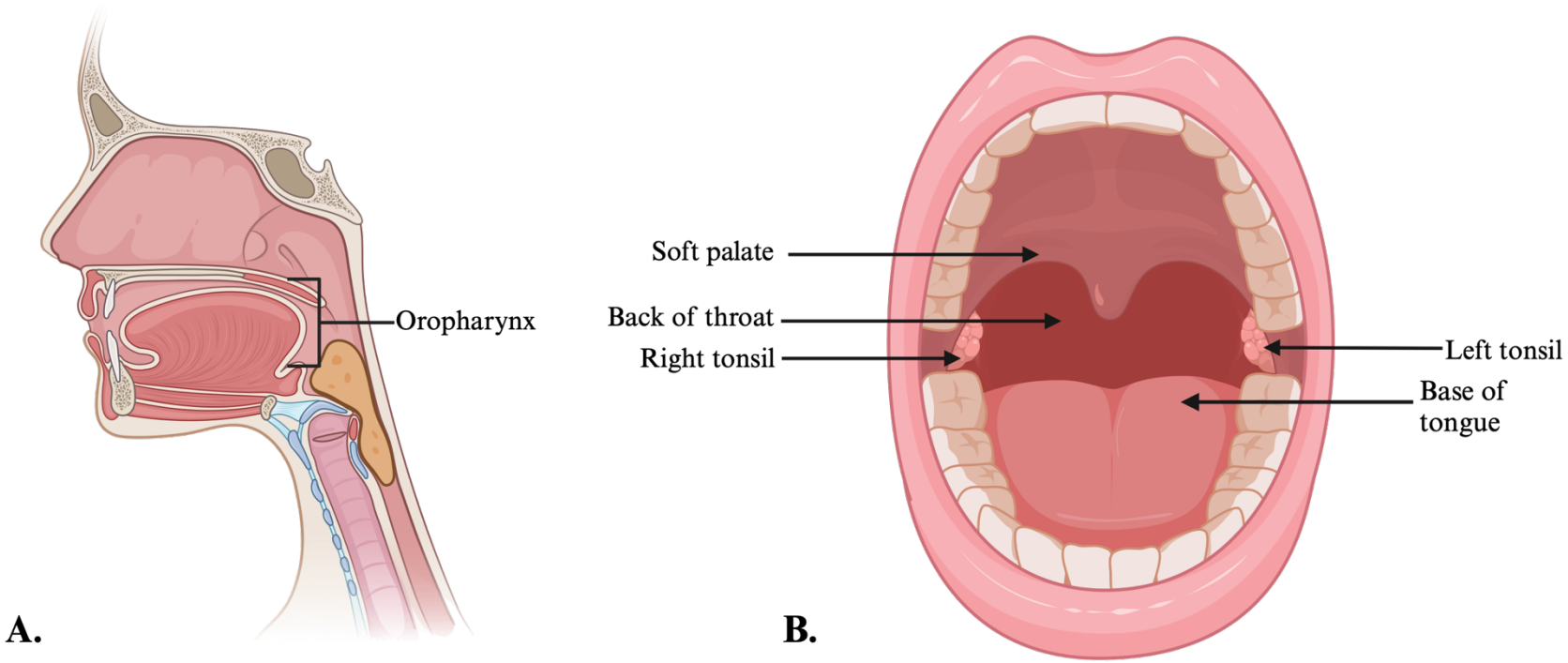


Figure 5. Anatomy of the human oropharynx. A side profile view of the location of the oropharynx (A). Individual components of the oropharynx are labelled within the mouth including: left and right palatine tonsils, base of tongue (lingual tonsil), back of the throat and the soft palate.

Image created using BioRender.com

1.2. The natural history of an oral HPV infection

As previously discussed, the natural history of an HPV infection of the cervix has been extensively studied. Conversely, the characterisation of the natural history of an oral HPV infection is sparse, and additional studies are sought to benefit the field. It has been suggested that cervical and HPV-positive HNC may possess similar molecular mechanisms of carcinogenesis, reviewed in (Gillison et al., 2014). However, the pathogenesis (viral strain, oncogene/oncoprotein expression, and viral integration) of an oral HPV infection, as well as potential biomarkers (host and/or viral) that may assist in the development of OPC, remains largely unexplored within pre-cancerous tissues.

1.2.1. The prevalence of oral HPV within the non-cancerous oral cavity/oropharynx

Within the UK specifically, estimates for oral HPV prevalence within the healthy population are varied, and possible explanations for this are outlined throughout this section. However, in most instances, oral HPV prevalence is described to be relatively low at 2-6% (Conway et al., 2016, Knight et al., 2016, Hearnden et al., 2018). Much of the data on oral HPV prevalence is from the US and is, unsurprisingly, more varied, given the much larger size of the country and the extensive diversity between individual states. However, often the reported rates remain similar to the UK, with estimates of ~1-10% (Smith et al., 2004, D'Souza et al., 2009, Turner et al., 2011, Edelstein et al., 2012, Gillison et al., 2012, Pickard et al., 2012, Cook et al., 2014, Fu et al., 2015, Raymond et al., 2018, Meites et al., 2020, Stankiewicz Karita et al., 2020, Giuliano et al., 2023, McIntosh et al., 2024). The US benefits from large health studies, such as the National Health and Nutritional Examination Survey (NHANES) study, which occurs every year. Here, oral prevalence is reported to be 7.5% in adults (Raymond et al., 2018). Together, these numbers are relatively low, which is somewhat surprising given that it is within these countries that recent data has shown the rapid rise in cases of OPC (Lechner et al., 2022). One explanation for this could be the HPV vaccination programmes that are available (and this is particularly accessible within the UK) that may have inferred a level of immunity within these populations, or that oral infections are much lower than cervical infections, and that these oral HPV infections are more likely to progress to disease.

In marked contrast, a recent study investigating the prevalence of oral HPV within native Australians reported a 35% prevalence rate (Jamieson et al., 2020); and analysis of their target population shows that they included a high number of current smokers and individuals that

reported an unvaccinated status (for HPV) (Jamieson et al., 2020), which may explain this high detection rate. Another report shows oral HPV detected within 10% of a different Australian population (Antonsson et al., 2021), and this sample was composed of a similar demographic of individuals (in terms of age/gender); however, unlike the aforementioned Australian study, this population was not limited to indigenous Australian individuals.

Indeed, it is clear that oral HPV prevalence rates are different dependent on the geographical location of the population analysed. For example, similar data that demonstrates high oral HPV prevalence (ranging from ~40-80%) have emerged from countries including Japan, India, Austria, US, North-Eastern Brazil, and Mexico (Terai et al., 1999, Pattanshetty et al., 2014, Dalla Torre et al., 2015, Lam et al., 2016, Shigeishi et al., 2018, Ribeiro et al., 2019, Morán-Torres et al., 2021) and this is largely surprising, as with the exception of the US and Austria (as part of Europe), HPV-OPCs are relatively low within these regions (Lu et al., 2022). Nonetheless, these oral HPV prevalence rates are considerably higher than those from other geographical locations (~0-6%) (Esquenazi et al., 2010, Hang et al., 2014a, Bumrungthai et al., 2019).

In some instances, there are discrepancies even within the same country/region, as is the case in many countries mentioned above (Kurose et al., 2004, Kreimer et al., 2011, Cho et al., 2020, Kofler et al., 2020, Kumari et al., 2021, Riddell et al., 2022), which implies that other factors may impact the prevalence estimate of an individual study. The broad range of oral HPV prevalence estimates across individual reports that describe the same geographical regions is highlighted in Table 2. Indeed, a study from our own group using a UK based university population found a similarly high rate to Jamieson *et al*., 2020, and one much higher than other reports from the UK, as we showed an ~23% oral HPV prevalence. We believe this observation is because our target sample was university students, who are more likely to report frequent participation in high-risk activities for oral HPV infection (Whitton et al., 2024). Here, we suggest that populations should be sampled in isolation, in order to identify at-risk groups that should be targeted by public health interventions (Whitton et al., 2024).

Furthermore, it is important to note that there is not currently a universal testing method for oral HPV, which means that variation is introduced based on sampling, and HPV detection methods used to describe oral prevalence. Some studies report on only HR-HPV prevalence (often HPV-16 and/or HPV-18) only, whereas others provide a much broader range of viral

type analysis, which is likely to effect the overall oral HPV prevalence reported. Additionally, Combes *et al.,* 2017 demonstrated poor concordance in HPV detection within oral rinse samples and tonsillar brushings from the same patient (Combes et al., 2017). Therefore, it is difficult to draw direct comparisons and conclusions on the prevalence of oral HPV infection across multiple studies.

Indeed, another study from the UK interestingly reported a 0% oral HPV prevalence within their sample. They tested for 20 mucosal viral types in tonsil tissues obtained from patients undergoing routine tonsillectomy surgery (N=3,888) (Palmer et al., 2014). These findings in non-cancerous tonsil tissue were corroborated in an American study which also detected no HPV-DNA within 226 number of tonsillar samples (Ernster et al., 2009). In the former study, one explanation for a 0% prevalence could be the inclusion of tonsil specimens from juvenile individuals, whom would be unlikely to participate in activities linked to HPV infection risk. Limitations to the latter study includes the analysis of viral types: HR HPV-16 and HPV-18 alone, and that their analysis included specimens collected between 1979-2001, making it an unlikely reflection of the current burden of oral HPV. Thus, despite it being believed that most HPV-OPCs arise within the cells of the tonsillar crypt (Kim et al., 2007), and are predominately HPV-16 positive (Lechner et al., 2022), this does not correlate with an oral HPV infection in a non-cancerous setting. Instead, evidence suggests that HPV prevalence specifically within tonsil tissue is low and other viral types to HPV-16 are commonly more prevalent within the mouth (Antonsson et al., 2014, Ribeiro et al., 2019).

Table 2. Oral HPV prevalence estimates summarised by region. Prevalence ranges taken from individual studies reporting on oral HPV prevalence within the adult population without HNC, in each region outlined are presented. These are representative of the lowest and highest oral HPV prevalence values in the literature for each region, presented by continent, except for the USA and UK, which are shown separately.

Geographical region	**Prevalence range reported**	**References**
Africa	3.5 – 5.7%	Marais et al., 2006 Davidson et al., 2014
Asia	0 – 75.8%	Ramesh et al., 2021 Shigeishi et al., 2018
Europe (not inc. UK)	0.8 – 42.1%	Napolitano et al., 2024 Dalla Torre et al., 2015
Oceania	2 – 35.3 %	Rumianek et al., 2023 Jamieson et al., 2020
USA	0 - 40.1%	Ernster et al., 2009 Lam et al., 2016
UK	2.2 – 23%	Hearnden et al., 2018 Whitton et al., 2024
Northern America (not inc. USA)	1.3 – 46.8%	González-Losa Mdel et al., 2008 Morán-Torres et al., 2021
Southern America	0 – 81.6%	Esquenazi et al., 2010 Ribeiro et al., 2019

Meta-analytical studies have attempted to overcome these sampling biases and observed differences reported in individual populations where the prevalence of oral HPV within people without diagnosis of any type of HNC has been investigated. These global estimates of oral HPV prevalence in healthy individuals range from 4.9 – 7.7%; however, the most recent of these reports includes data only up to 2017 (Tam et al., 2018), meaning that these figures are outdated given the rapid increase in incidence of HPV-OPC over the past 10 years. Nonetheless, these outdated figures still demonstrate the burden of oral HPV within the healthy population but an updated global prevalence estimate by meta-analysis would serve to provide a basis for future public health interventions.

Overall, it appears that oral HPV prevalence rates differ by region but there are also clear patterns in prevalence rates associated with other factors. A study from the USA highlighted the link between HPV infection and age, with a bimodal pattern of infection. Interestingly, in their cohort, there was a peak in prevalence at ages around 30-34, and again later at 60-64 years of age (Gillison et al., 2012), which could be due to more frequent participation in activities linked to oral HPV infection, and viral persistence within the younger and older peaks, respectively. Additionally, analysis of the NHANES data from 2011-2014 highlighted a discrepancy on the burden of infection between men and women, with prevalence rates of 11.5% and 3.2% detected, respectively (Sonawane et al., 2017), suggesting that gender is a risk factor for oral HPV infections. Conversely, others have reported a similar prevalence, irrespective of gender (Kreimer et al., 2010); risk factors for oral HPV infection, that impact prevalence data presented here, are discussed in detail within section 1.4.

1.2.2. The pathogenesis of HPV within the oropharynx

Most of the current knowledge of HPV within the oropharynx is learned from cancerous lesions, and thus provides a retrospective view of viral pathogenesis within this region. As such, the precise process from viral infection in the healthy oropharynx to malignancy remains unknown.

1.2.2.1. HPV pathogenesis within OPC

HPV-OPCs present with diverse molecular profiles to their HPV-negative counterparts, which suggests different pathways to malignancy. Indeed, it is well known that HPV-positivity is strongly correlated with improved prognosis in OPC (Ang et al., 2010, Goodman et al., 2015, Wang et al., 2015), and biomarkers of, and targeted treatments for, HPV-positive OPCs and

other HNCs are described at section 1.3. Perhaps the most significant difference between the two groups is the disparity in p53 status. As is the case in cervical malignancies, wild-type (WT)-p53 is the dominant form in HPV-HNCs, and therefore, these tumours present with a lower frequency of p53 mutation (Braakhuis et al., 2004, Westra et al., 2008). The E6*I isoform has been reported within HPV-OPC (Cerasuolo et al., 2017), which could explain the presence of WT-p53 as this protein prevents p53 degradation (Pim et al., 1997); however, the situation within the head and neck is unclear and still debated. Similarly to the situation with p53, regions prone to allelic loss in HPV-negative OPC are often not affected in HPV-positive carcinomas (Braakhuis et al., 2004) due to manipulation of the host cell by the virus itself.

The classical pathway to malignancy described in the cervix is based upon viral integration, and the subsequent overexpression of oncogenic proteins, E6 and E7. Integration appears to be rarer in the head and neck than within cervical malignancies, and the virus is often found in episomal forms in OPCs (Nulton et al., 2017, Anayannis et al., 2018), but it is not clear why. Furthermore, it appears that HPV can exist as an episome containing both viral and human DNA (a viral human-hybrid episome) (Nulton et al., 2017). The significance of this structure is yet to be determined, but such hybrid episomes are not limited to tumours of the head and neck, and have also been identified in some cervical cancers (Zhou et al., 2022). There is the possibility that these human-hybrid episomes may be a unique tumour subtype that were originally thought to contain the virus in a 'mixed' form (tumours containing some episomal and some integrated virus) (Nulton et al., 2017), and, thus, could present a subset of HPV-positive cancers with different biology. Could these tumours, therefore, be suitable for alternate targeted treatments? Interestingly, the frequent detection of episomal HPV in HPV-HNCs (Nulton et al., 2017) could imply that there may be other important pathways to malignancy used by HPV in this anatomical region, aside from the classical E6/E7 route present in most cervical cancers, as E2 expression is preserved.

Indeed, characterisation of several HPV-HNCs revealed that a subset of HPV-OPCs with non-integrated episomal virus showed a specific pattern of expression as follows; upregulation of E2, E4, and E5 but low levels of the E6/E7 transcripts (Parfenov et al., 2014). These differences in the viral expression of integrated versus non-integrated HPV subtypes demonstrate that an alternate route from HPV infection to malignancy may be more common within the head and neck. Ren *et al.,* 2020 characterised this potential novel oncogenic pathway in HPV-positive cancers that is thought to be dependent on episomal expression of the E2, E4, E5 proteins using

various models, not limited to those of the head and neck (i.e., *in vitro* cellular models and the use of transgenic mice). They provide evidence for a potential p53-dependent mechanism of inhibiting cell cycle arrest between G_1 to S-phase, and driving progression through the cell cycle by E2, E4, and E5 proteins. In addition, they demonstrate an increase in fibroblast growth factor receptor (FGFR) signalling in an E2/E4/E5 HPV-HNC pharyngeal subtype (Ren et al., 2020), which may provide one route of HPV-driven carcinogenesis, due to the downstream proliferative effects of FGFR signalling (Krook et al., 2021), and could be a potential therapeutic target in episomal HPV-positive cancers. Indeed, Ren and colleagues demonstrated that FGFR inhibition, alongside inhibition of other proteins, can halt proliferation in this subtype (Ren et al., 2020).

1.2.2.2. HPV infection of the tonsillar crypts

The cells of the tonsillar crypts are composed of stratified squamous epithelium and reticulated epithelium, and these are infiltrated with cells of the immune system (Perry, 1994) (Figure 6). As the tonsillar crypts are the assumed predominant site of HPV-OPC development, one would assume that it is within these cells that, under certain circumstances, an oral HPV infection can persist and promote carcinogenesis. Given that this area is at the direct interphase with the immune system, the establishment of a persistent viral infection within the tonsil crypts is somewhat surprising.

One explanation for this is that there are high levels of programmed cell death ligand (PD-L1) on the surface of the cells of the reticulated epithelium (Lyford-Pike et al., 2013) that under normal conditions functions to bind its receptor, programmed cell death protein-1 (PD-1) that is expressed on T-cells and this interaction helps to mark the cell as 'safe' to avoid the death of healthy cells in an immune response. However, cancer cells have hijacked this axis by expressing high levels of PD-L1 on their surface to activate this immune checkpoint, and evade detection, and promote tolerance (Iwai et al., 2002). Therefore, it could be speculated that HPV is able to exploit the high levels of PD-L1 within the keratinocyte cells of the tonsillar crypt, to manipulate their function to avoid immune detection and so persist. Indeed, high PD-L1 expression can be associated with HPV-positivity in OPCs (Hong et al., 2016b). The prognostic implications of PD-L1 expression is discussed in detail at 1.3.4.

There are limited appropriate models to investigate the pathogenesis of the virus in the healthy oropharynx. Roberts *et al.,* 2019 provide a comprehensive review of options to model the

biology of HPV in this region (Roberts et al., 2019). They discuss the use of transgenic mice, whom, despite their success in informing of HPV pathogenesis within other anatomical regions, are not appropriate for investigations within the tonsil, as mice do not have a tonsil/equivalent organ. In addition, they highlight the potential of techniques such as tonsil tissue explants, and the use of organotypic epithelial 'raft' (3D) culture to generate a population of primary keratinocytes to aid pathogenesis investigations (Roberts et al., 2019). Indeed, *in vitro* systems have been previously described and used by Iain Morgan and colleagues, where in one such study they used hTERT-immortalised normal oral keratinocytes (NOKs) to examine the viral life cycle specifically within oral keratinocytes expressing episomal HPV-16 (Evans et al., 2017). Other cell lines used for *in vitro* modelling of the HPV life cycle include immortalised human foreskin keratinocytes (HFKs) (Ferguson et al., 2021); although, oral cells are perhaps more useful to scrutinise oral HPV pathogenesis.

Improving knowledge within the field of HPV pathogenesis and the progression to HPV-induced malignancy within the oropharynx, would prove invaluable to our understanding of the virus and its influence on host cell biology, but also in the development of screening, treatment, and prevention interventions, like those available for HPV-induced cervical disease.

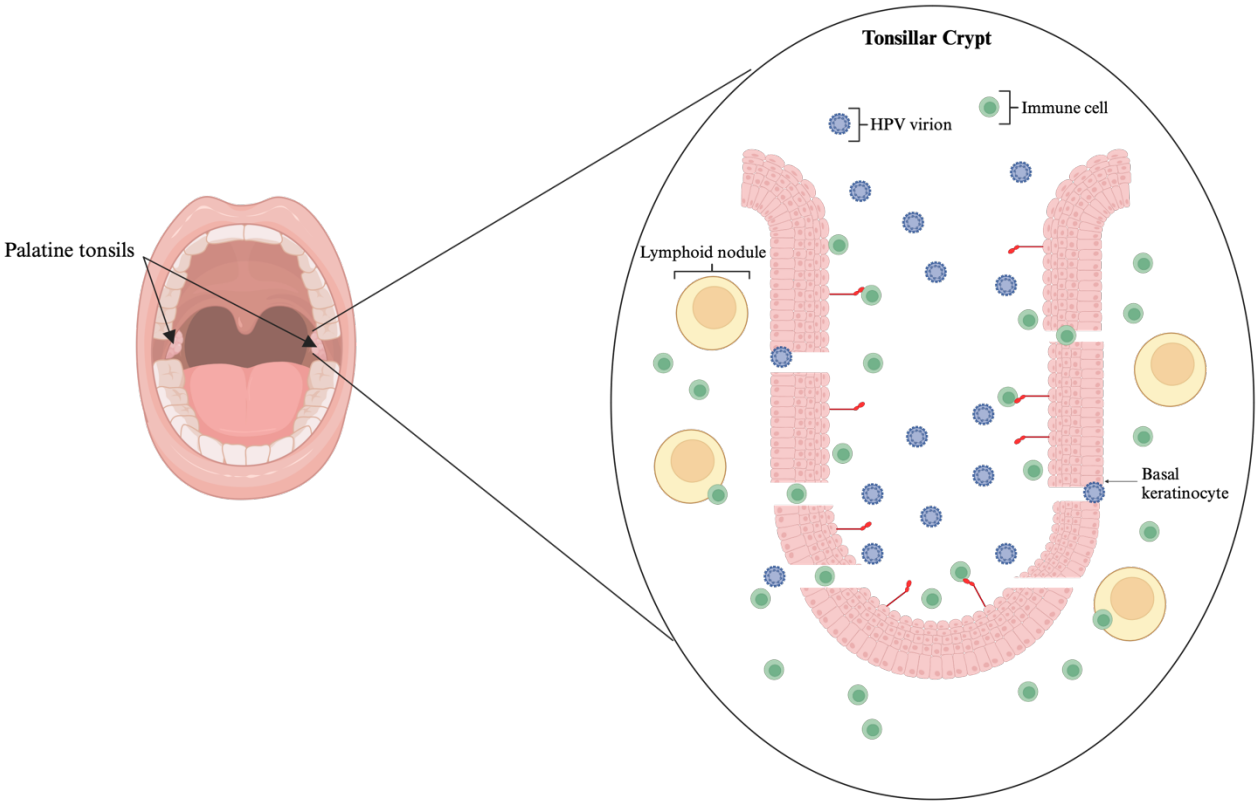


Figure 6. HPV infection of the tonsillar crypts within the human oropharynx. HPV virions are shown within the tonsillar crypt structure, where most HPV-OPCs are assumed to originate within. The tonsillar crypt structure is shown as invaginations in the tonsillar cell surface and the microenvironment of high immune infiltration within these structures is displayed. HPV gains access to its target cell, the differentiating basal keratinocyte, through intracellular spaces within the non-uniform stratified epithelium. HPV virions are shown in blue, cells of the immune system are shown in pale yellow or green, and PD-L1 receptors in red on the cell surface of the stratified squamous epithelium, as indicated. Figure produced using BioRender.com.

Adapted from: (Lim et al., 2023)

1.3. Biomarkers of, and treatments for, HPV-OPC

With limited knowledge of the pathogenesis of HPV within the healthy oropharynx there are a lack of biomarkers of HPV-positive non-cancerous oropharyngeal tissue that may be indicative of an infection that is at risk of malignant transformation. Biomarkers offer the potential to risk-stratify individuals for treatment, and/or group individuals that would benefit from monitoring for future disease progression. Furthermore, exploration of biomarkers associated with HPV-OPC, and perhaps other HPV influenced HNCs (themselves incredibly rare), provides an insight into the biology of HPV-positive malignancies that could potentially highlight targets of early HPV-induced cellular changes.

HPV-OPC does not appear to be a distinct disease, and there are conflicting reports on the levels of biomarkers associated with HPV-positivity in OPC and how these impact the survival outcomes of patients. Thus, HPV-positivity in OPC does not fully define a precise expression profile of biomarkers, but more likely that HPV-OPCs are composed of smaller groups that would respond differently and require alternate treatments, based on their biomarker expression blueprint.

1.3.1. HR-HPV and viral oncoprotein expression in HPV-OPC

With improved outcomes, HPV-status itself is considered a useful biomarker for favourable prognosis for oropharyngeal squamous cell carcinoma (OPSCC) patients. However, as stated at 1.3.2., expression levels of other cellular proteins (i.e., p16) contribute to the impact of HPV-status on prognosis (Weinberger et al., 2006). HPV-16 dominates cases of HPV-OPC, mediating over 90% of all HPV-positive OPC diagnoses (Gillison et al., 2015). As a result, from a prospective point of view, having a persistent HPV-16 infection in the oropharynx is considered a significant risk factor to cancer development, and therefore some have suggested that research of oral HPV in healthy individuals should be focussed on those with HPV-16 infection (Fakhry et al., 2020). However, given that the situation of oral HPV within the oropharynx remains unclear, inclusion of a broad range of HPV types in the non-cancerous oropharynx allows scrutiny of how HPV influences host cell biology, regardless of viral type, and may highlight differences between HPV-16 infections that are at risk of malignant transformation compared to HPV infection of other viral types.

1.3.1.1. p53 is not mutated in most HPV-positive oropharyngeal malignancies

WT-p53 is biomarker of HPV-positive cancer, including HPV-OPC (Gillison et al., 2000, Licitra et al., 2006, Hong et al., 2016a, Castellanos and Pan, 2016). Targeting of E6 to restore WT-p53 levels and function, offers a promising therapeutic avenue in HPV-positive disease and data shows that targeting of this viral protein has proved successful in reviving apoptotic activity in the *in vitro* setting (Rampias et al., 2009).

1.3.2. $p16^{INK4A}$ expression as a surrogate marker for HPV within HNC

Many studies that have investigated HPV in healthy or cancerous tissues from anogenital and head and neck regions have used p16 positivity to indicate HPV infection. Indeed, p16 is frequently overexpressed in HPV-positive malignancies as result of manipulation of host cell biology by HR-HPV oncoprotein E7 that indirectly increases p16 levels (Chung and Gillison, 2009). However, more recent data suggest that p16 is not an accurate biomarker for HPV-positivity in OPC and non-cancerous tissue, given oropharyngeal p16 can be overexpressed, or absent, regardless of HPV-status (Klingenberg et al., 2010, Arsa et al., 2021). The concordance of p16- and HPV-status is reported to differ dependent on regional incidence of HPV-OPCs; with p16 a good, or inappropriate, marker of HPV-positivity in countries with a higher or lower incidence of HPV-OPC, respectively (Mehanna et al., 2023).

Nonetheless, in conjunction with HPV-status, p16 is still a biomarker for clinical outcome. Interestingly, a subset of HPV-positive OPSCCs, that are negative for p16 expression have a worse prognosis than their p16-positive counterparts (Rasmussen et al., 2019, Mehanna et al., 2023); thus, this indicates that p16 expression has prognostic value in OPSCC, despite not being directly linked to HPV-status. In agreement, the HPV+/p16+ gene expression blueprint has previously been associated with better prognosis (Weinberger et al., 2006). This is unsurprising, given that p16 expression and HPV-status are both associated with improved prognosis, that the loss of either/or should indicate a worse prognostic outcome in patients.

The finding that there can be discordance between HPV-status and p16 expression in some instances of OPSCC is crucial to the field. Indeed, many studies of other biomarkers of HPV-associated HNC have used p16 as a surrogate for HPV expression, and this is a significant caveat to their data. Further research to describe the implication to treatments based on p16 status, with the knowledge that p16 expression is not necessarily correlated with HPV-status in HPV-OPC, would benefit HPV-OPC patients and the field.

1.3.3. microRNA (miRNA) expression profiles in HPV-OPC

The expression of sets of miRNA molecules have been identified as groups of biomarkers for HPV-OPC, and other HPV-positive HNCs (Krsek et al., 2024). Indeed, in the context of HNC, HPV-positive and HPV-negative malignancies appear to display different patterns of miRNA expression (Krsek et al., 2024). One study identified a set of five miRNAs linked to HPV-positivity and HPV pathogenesis when scrutinising HPV-OPC and HPV-positive cervical cancers, and they also found that HPV-OPC and HPV-negative OPC display distinct miRNA profiles (Lajer et al., 2012). However, other studies report different individual miRNAs that are dysregulated, and may have roles in HPV-OPCs (Hui et al., 2013, Vojtechova et al., 2016). miRNA expression is often dysregulated in disease, and the observed differences between miRNA expression profiles of HPV-HNC and HPV-negative HNCs render these a potential tool to aid distinction between HNCs of different HPV-status; but, thus far, there does not appear to be a universally reported cluster of miRNAs that are linked to HPV-OPC.

Prospective studies of changes in miRNA expression within the HPV-positive oropharynx in a non-cancerous setting are relatively unexplored; however, one study by Ekanayake Weeramange *et al*., 2023, compared miRNA expression within saliva of HPV-OPC individuals to HPV-positive healthy controls. They found a signature of 7 miRNAs that were dysregulated between these groups (Ekanayake Weeramange et al., 2023). The use of saliva is advantageous for diagnostic purposes, as it is a non-invasive and easily accessible sample type to collect. However, this study did not investigate whether this signature became informative as biomarkers for cellular changes within the HPV-positive control group, and whether any of these individuals developed HPV-OPC.

The dysregulation of miRNA in HPV-positive cancers is evident, and there have been distinct profiles described between HPV-positive and HPV-negative HNCs that support there being different molecular mechanisms within each cancer subtype. However, there is not yet a consensus panel of miRNAs that comprehensively represent HPV-positive HNC, or that can be used prospectively.

1.3.4. Other biomarkers for HPV-HNC

As a biomarker for HPV-OPC, expression of immune checkpoint protein PD-L1 is frequent in HPV-positive malignancies (Balermpas et al., 2017, Lechner et al., 2022), and this is associated with favourable prognosis in HPV-positive compared to HPV-negative OPCs (Hong et al.,

2019, Polesel et al., 2021). Indeed, there are a subset of HPV-positive HNSCCs that do not express PD-L1 (HPV+/PD-L1-) that demonstrates a worse prognosis than HPV+/PD-L1+ cancers, corroborating the potential prognostic implication of the HPV+/PD-L1+ genotype in disease (Zhu et al., 2023b). In contrast, however, one study found that within HNSCCs, PD-L1 overexpression was linked to better prognostic outcomes in HPV-negative but not in HPV-positive (specifically HPV-16) malignancies (Balermpas et al., 2017). Further evidence suggests that PD-L1 expression is associated with different a prognostic outlook dependent on the site of the tumour within the head and neck (Lenouvel et al., 2020, Wotman et al., 2020, Polesel et al., 2021).

The discrepancies in published data between the ideal combination of HPV-status and PD-L1 expression for favourable prognosis in HNSCCs may be because manipulation of the PD-1/PD-L1 axis is a well-established cancer mechanism that is not limited to HPV-positive tumours (Lin et al., 2024). Many types of cancer cells show the ability to hijack and activate the PD-1/PD-L1 immune checkpoint, expressing high levels of PD-L1 on their surface to become disguised as normal cells, evade detection by T cells, and ultimately, avoid tumour cell death (Iwai et al., 2002). Therefore, it is unsurprising that PD-L1 can be expressed regardless of HPV-status in OPC; however, it is reported that HPV-positive tumours express more PD-L1 proteins (Hong et al., 2019, Balermpas et al., 2017), possibly through viral integration into the gene encoding for PD-L1, causing dysregulation of its expression (Cao et al., 2019). Data from our group refute this, with PD-L1 not overexpressed in HPV-OPC (Mallender, 2024). Overall, although there remains conflict around the implication of PD-L1 expression in HNSCC based on HPV-status and cancer site, PD-L1+ tumours have been linked to an overall better response to PD-1/PD-L1 inhibitor targeted treatments within the head and neck (Paderno et al., 2024).

Epithelial cytoskeleton protein, cytokeratin-7 (CK7) has been linked to HPV-OPC (Woods et al., 2017); however, it has not been as extensively studied within the oropharynx. CK7 promotes oncogenesis in OPC as it has a SEQIKA amino acid sequence that can bind and stabilise oncoprotein E7 (Kanduc, 2002), and CK7 expression has been linked to HR-HPV activity in HPV-OPC (Morbini et al., 2015). Nonetheless, CK7 is not exclusively expressed within HPV-OPC, and it is present at similar levels within cells of the stratified epithelium of the normal tonsillar crypts (Morbini et al., 2015, Woods et al., 2017). CK7 is also a marker of the reticulated epithelium of the tonsillar crypts; here it is stabilisation of E7 mRNA may render the crypts vulnerable to malignant viral transformation (Woods et al., 2017). The expression of

CK7 here is one explanation for the tonsillar crypts as the assumed site of HPV-OPC development (Kim et al., 2007).

Finally, another interesting biomarker for HPV-positive OPC prognosis is Kiel-67 (Ki-67), a protein that is present during proliferation and indicates that the cell is in a state of active division (Oliveira and Ribeiro-Silva, 2011). This biomarker is commonly assessed by immunohistochemistry (IHC) in cancer diagnostics, as it is used to indicate the speed of tumour proliferation (Azémar et al., 2011). Unsurprisingly, Ki-67 expression is associated with poor prognosis in OPSCC (regardless of HPV-status); however, interestingly, Ki-67 has been described as more frequent in HPV-OPC, and when expressed in the presence of HPV infection, it yields a better prognosis (Liu et al., 2015). Given that Ki-67 is generally considered a marker of unfavourable prognosis, this link between Ki-67, HPV, and a better prognosis, is likely due to HPV-positivity positively impacting prognosis and treatment response.

1.3.5. HPV-HNC targeted treatments

As discussed throughout, HPV-HNCs are a distinct cancer type. Despite this, these cancers are currently presented with the same treatment options as HPV-negative OPCs, with individual treatments differing based on cancer stage, rather than HPV-status. It is clear that HPV-HNCs respond better to treatment than their HPV-negative counterparts (Mallen-St Clair et al., 2016), and this, along with the better prognosis associated with HPV-OPCs, highlighted a group of OPC patients suitable for de-escalation therapy trials.

Mehanna and colleagues closed their phase III clinical trial early. De-ESCALaTE HPV, used a treatment regime of EGFR-inhibitor cetuximab, administered in combination with radiotherapy (Mehanna et al., 2019c). Cetuximab was selected to replace the commonly used cisplatin drug. Unfortunately, these trials were unsuccessful and lead to reduced OS in the HPV-OPC patients (Mehanna et al., 2019c). At the same time, another trial investigating the use of cetuximab and radiotherapy for HPV-OPC patients (RTOG 1016) agreed with these findings, and described a reduced OS in their HPV-OPC patients compared to patients with HPV-OPC treated with cisplatin and radiotherapy (Gillison et al., 2019). The authors of these trials both made use of p16 IHC to either identify tumours suitable for additional HPV-DNA testing (Mehanna et al., 2019c), or as a surrogate marker for HPV-positivity (Gillison et al., 2019); and this perhaps resulted in some participants having an incorrect HPV-status, given that p16 is not always correlated with HPV-positivity (Klingenberg et al., 2010, Arsa et al.,

2021). Alternatively, the tragic findings could be due to a lack of understanding of HPV pathogenesis within the oropharynx, as EGFR is clearly not a suitable target for HPV-mediated disease in this locality.

The failures within these trials accentuated the lack of fundamental knowledge of the basic biology of HPV within the oropharynx; in particular, the impacts of the virus on the host cell that may influence response to treatments. Indeed, it supports the notion that there are molecularly distinct subsets of cancer types within the HPV-OPC group (i.e., HPV+/p16-), that would themselves respond differently to targeted treatments. Moreover, there are biomarkers that have been identified within HPV-OPC that may not be as specific to HPV-positivity as several of these are present/absent in either HPV-status OPC. These may be reasons for these failures to implemented de-escalation therapies; thus, highlighting the necessity for further research to address gaps in the knowledge of HPV pathogenesis in the oropharynx.

To conclude, there continues to be gaps in knowledge of fundamental biological differences in HPV-positive and HPV-negative malignancies that could translate to improved treatments.

1.4. The risk factors for oral HPV infection

There are several salient risk factors that show significant association with one's risk of oral HPV infection; these include both demographic factors and lifestyle characteristics (Figure 7).

Given the nature of many of these factors, a significant limitation to publications that study risk factors for oral HPV infection is that participation in most behavioural activities is self-reported. Therefore, the conclusions drawn from such studies are dependent on the accuracy of self-reported behaviours, and there is the possibility of misleading data if an untrue account of any behaviours are given. Moreover, another caveat of these studies is that the prevalence estimates used to determine statistical association of certain risk factors with HPV infection are collected using a variety of sampling methods as discussed, which inevitably introduces variation into reports, and may impact the conclusions in regards to HPV infection and its association with a potential risk factor.

1.4.1. Gender and age

HPV-OPC cases are more common amongst males than females (Guo et al., 2022), and so it is unsurprising that HPV-DNA is also often found to be more prevalent within healthy males

without OPC or other HNCs (Kreimer et al., 2004, Gillison et al., 2012, Antonsson et al., 2014, Sonawane et al., 2017, Tam et al., 2018, Cho et al., 2020). Indeed, the discrepancy in HPV-OPC cases by gender may be explained in part by HPV-OPC dominant viral type HR HPV-16 also being more commonly detected in the oropharynx of healthy males (Riddell et al., 2022, Giuliano et al., 2023). Here, HPV-16 was strikingly reported at a 5-fold higher prevalence in males than females in one study (Gillison et al., 2012). Interestingly, there is also evidence that a persistent oral HPV infection is more common in males (D'Souza et al., 2016, D'Souza et al., 2020). Men are more likely to report a greater number of sexual partners (both for vaginal and oral sexual activity) (D'Souza et al., 2016, Pauli et al., 2022) and this is a possible contributor to a greater incidence of male oral HPV infection, discussed later. However, these data of oral HPV and/or specifically HR HPV-16 dominance in males is not always corroborated (Pauli et al., 2022). Nonetheless, based on several reports that do describe an increased detection of HPV-16 within healthy males, one could infer that this is a significant contributing factor to the greater burden of HPV-OPC to men; however, this does not provide an explanation for the dominance and potential persistence of oral HPV in healthy males, as most infections are resolved and do not lead to cancer.

There have been several hypotheses, aside from behavioural differences, to explain the generally accepted disparity in oral HPV prevalence of males and females, with the former being at greater risk. Firstly, one explanation is that these gender differences may be due to an unexplored difference of the sex hormones (for example, androgens) on viral pathogenesis (Gillison et al., 2014). Secondly, there is argument for a general discrepancy in gender immune responses (Chaturvedi et al., 2015). Indeed, there is evidence for a greater rate of seroconversion within females, specifically antibodies for the quadrivalent vaccine types HPV-6, -11, -16, and -18, even before the introduction of the HPV vaccine (Thompson et al., 2004, Markowitz et al., 2009, Introcaso et al., 2014, Windon et al., 2019), which would provide women with greater protection against novel infection (Safaeian et al., 2010), and therefore, could account for a lower rate of oral HPV detection in women.

Undoubtably, advancements in the knowledge of how HPV influences host cell biology in the oropharynx and oral cavity of males compared to females or, vice versa, how fundamental differences in male/female biology, as well as behaviours, impact viral pathogenesis would be invaluable to fully decipher the rationale and mechanism of oral HPV dominance, in both infection and disease, in males.

Age is commonly associated with an increased risk of oral HPV infection; however, there are conflicted reports in the correlation coefficient between infection risk and age. Several publications have described older age as associated with an increased risk of both oral HPV infection (Kreimer et al., 2004, Pickard et al., 2012, Chaturvedi et al., 2014, Sonawane et al., 2017) and viral load (Chaturvedi et al., 2014), whereas there is also evidence of a positive relationship between younger age and oral HPV risk (Hang et al., 2014b, D'Souza et al., 2016). On the other hand, the importance of age as a risk factor at all has been questioned, with reports that age had no significant impact on oral HPV infection (Ribeiro et al., 2019).

A large study that analysed data from the NHANES between 2009-2010, highlighted an interesting 'bimodal pattern of infection' with two separate prevalence peaks, one at 30-34 and then 60-64 years old (Gillison et al., 2012). Indeed, this supports that both the younger age and older age groups may be concurrently at risk of oral HPV infection. It is possible that the differences in reported age association with oral HPV prevalence may be a result of incident versus persistent infection. For example, younger people are more likely to report frequent participation in activities well-established to be oral HPV infection risk factors (Mercer et al., 2013), that increase risk of acquiring an infection and could explain the high oral HPV prevalence within this group. On the other hand, the peak in prevalence at older ages could be a result of the virus; notably HR HPV-16, having a better ability to persist in these individuals (Pierce Campbell et al., 2015, D'Souza et al., 2020), perhaps due to a generic decrease in immune responses that accompany older age (Castelo-Branco and Soveral, 2014). However, it is likely that in future years, younger age groups will have a greater protection in response to vaccination interventions (1.5.2), that could result in a decline in circulating HPV-16 that would level out the bimodal pattern of infection.

1.4.2. Sexual behaviours

Increased numbers of sexual partners is a well-established risk factor of HPV-HNC (Gillison et al., 2014), and this is also associated with oral HPV prevalence in healthy individuals, as is the case with HPV infection of the cervix (Knight et al., 2016, Hearnden et al., 2018, Gonçalves et al., 2021, Bruno et al., 2023) due to a greater chance of exposure to the virus. Some studies have found that an increased number of vaginal intercourse partners is associated with oral HPV infection (D'Souza et al., 2009, Gillison et al., 2012, Cook et al., 2014), and there is indication that this is an particularly impactful risk factor, as one study has shown that the

statistical significance between males and females for oral HPV decreases when accounting for total vaginal intercourse partners (Pickard et al., 2012).

Given the nature of oral sex, it is not surprising that performance of this is also associated with an increased risk of oral HPV infection (Pickard et al., 2012, D'Souza et al., 2016, Shigeishi and Sugiyama, 2016), and transmission of oncogenic viral types such as HPV-16 and HPV-18 (Hearnden et al., 2018), as it provides a direct genital to oral route of viral transmission. Indeed, HPV commonly infects anogenital organs, and evidence suggests that the virus can be sexually transmitted from here to the oral cavity and oropharynx through performance of oral sex (Dahlstrom et al., 2014), of which it has been shown that performance of cunnilingus, yet not fellatio, increases risk of oral HPV, but only to males (D'Souza et al., 2016). Interestingly, one cohort study that included male university students, found a link between recent performance of oral sex and oral HPV infection; however, these infections were transient and did not persist until the next 4 month check-up (Edelstein et al., 2012); although this transience could be attributable to a better ability to clear viral infection within the age range of their cohort (18-24 years old). In marked contrast, within a population of men from North America, no difference in oral HPV prevalence rate was found dependent on whether an individual has or, has never partaken in oral sex (Kreimer et al., 2011). However, a younger age of debut for, and more frequent participation in, oral sex has been linked to HPV-associated malignancy in the oropharynx (Drake et al., 2021), supporting the generally accepted notion that this is an important factor to oral HPV.

Others have suggested that the nature of sexual activities performed are also a significant factor to risk of oral HPV, and they have shown that more risky, less safe, practices (for example, one-night stands and/or not using a condom, including during oro-genital sex) are linked to increased risk (Knight et al., 2016, Sethi et al., 2022). Indeed, a large scale analysis of NHANES data from 2009-2014 showed that individuals who reported the use of protection during oral sex had statistically lower HR HPV-16 and HPV-18 prevalence (Gupta et al., 2019). However, use of barriers (i.e., use of condom or dental dam) would not provide complete protection for HPV transmission during oral sex, and so, unsurprisingly, others have found no statistical association with use of protection during oral sex and oral HPV infection (Antonsson et al., 2014). Moreover, men who have sex with men (MSM) have been highlighted as a group at-risk for higher oral HPV prevalence (Shigeishi and Sugiyama, 2016), which is perhaps a result of more frequent participation in risky sexual behaviours (Hess et al., 2017); although, it

may also be in part a due to MSM being vulnerable to HIV infection (Centers for Disease and Control and Prevention, 2024) that itself can influence HPV pathogenesis (discussed at 1.4.5). Alternatively, it may be due to having not being protected from the originally female-only HPV vaccination programme, that until 2018 was not offered to MSM (GOV UK, 2018).

The widely accepted association of oral sex (both cunnilingus and fellatio) and increased risk of oral HPV infection provide evidence for oral-genital routes of HPV transmission; however, transmission of HPV to the oral cavity/oropharynx can also be transmitted through oral-oral routes. Open-mouthed kissing has been linked to oral HPV infection (Pickard et al., 2012, Fu et al., 2015, Gonçalves et al., 2021, Minhas et al., 2023), and interestingly, even within young male individuals that indicated no to any history of oral sex but stated they participate in open-mouthed kissing (D'Souza et al., 2009). Finally, non-sexual routes to oral infection of the mouth are possible; for instance as a result of self-inoculation (for example, through use of sex toys in both the mouth and vagina) (Cook et al., 2014). Moreover, some infections may be a result of vertical transmission from mother to baby (Khayargoli et al., 2023). Overall, it appears that transmission of HPV to the oral cavity is most common through sexual activity; however, it is likely there are alternative pathways of infection.

1.4.3. Lifestyle practices – alcohol consumption, smoking, and oral hygiene

Alcohol consumption and smoking are risk factors for OPC/HNCs that are HPV-negative malignancies (Gillison et al., 2014, Farsi et al., 2017, Thakral et al., 2024); however, these are both considered risk factors for HPV-HNC development (Farsi et al., 2017, Thakral et al., 2024), and for oral HPV infection within healthy individuals (Gillison et al., 2012). The relationship between alcohol consumption and oral HPV infection and its significance as a risk factor remain unclear, and there are conflicted reports to whether this is a genuine risk factor throughout the literature (Hearnden et al., 2018, Bumrungthai et al., 2019). However, D'Souza and colleagues, find there to be no association with alcohol and viral clearance (D'Souza et al., 2020), or increased likelihood of progression to HPV-associated disease (D'Souza et al., 2007). If assumed to be an oral HPV infection risk factor, with alcohol an immunosuppressive substance (Szabo, 1997, Szabo and Saha, 2015), it is likely that when consumed in excess, an individual could be more susceptible to HPV infection.

Smoking is a well-established risk factor for oral HPV infection (D'Souza et al., 2009, Kreimer et al., 2011, Gillison et al., 2012, Shigeishi and Sugiyama, 2016, Sonawane et al., 2017,

Hearnden et al., 2018), and participation in tobacco smoking has been reported to cause a higher viral load within HPV-positive individuals (Chaturvedi et al., 2014). Smoking habits may promote HPV infection through increased risk of periodontitis, and smoking-facilitated mucosal damage that enables virion access to target cells (Bagaitkar et al., 2008). Moreover, smoking causes immune disruption that might enable viral infection and influence HPV pathogenesis within the oral cavity to promote persistence (Sopori and Kozak, 1998, Knight et al., 2016). However, one study found smoking is not associated with increased risk of disease in the presence of HPV-16 infection (D'Souza et al., 2007).

In addition, marijuana smoking has been identified as a risk factor for oral HPV infection in some studies (Sonawane et al., 2017); however, others have reported no association between marijuana smoking and oral HPV infection (Ortiz et al., 2018b). Marijuana is a substance that also imposes immunosuppressive effects (Maggirwar and Khalsa, 2021, Milad et al., 2023), and mechanistically, it is therefore likely to increase oral HPV infection risk in a similar manner proposed for tobacco smoking. Additional research conducted within individuals without any form of HNC would help elucidate how lifestyle activities, such as smoking and heavy/any alcohol consumption, may impact the pathogenesis of HPV infection within the healthy oral cavity and oropharynx.

Finally, poor oral hygiene is believed to increase the risk of oral HPV infection (Bui et al., 2013, Bui et al., 2015, Dalla Torre et al., 2019); which is perhaps due to periodontitis, and/or inflammation and mucosal damage, caused by poor oral hygiene (Bui et al., 2013, Ortiz et al., 2018b, Dalla Torre et al., 2019), both of which would provide access for HPV to their target cells to establish an infection through epithelial damage.

1.4.4. Ethnicity

It is has been shown that HPV-OPC prevalence and outcomes are linked to ethnicity (Liederbach et al., 2017, Mazul et al., 2021, Sheth et al., 2021), but there is limited evidence examining the association of oral HPV prevalence rates with ethnicity. It has been reported that oral HPV prevalence is highest amongst Black individuals (Choi et al., 2020, Berenson et al., 2022); whereas Asian groups have been described to have lower oral HPV prevalence, with respect to White individuals (Choi et al., 2020).

The reasons for these patterns are not yet clear, but it is likely that other risk factors act in combination with ethnicity to influence an individual's likelihood of oral HPV infection. Indeed, the potential significance of ethnicity to oral HPV prevalence can be accounted for through sexual behaviour (Choi et al., 2020). Furthermore, NHANES data from the USA shows that Black male participants are most likely to have oral HPV; but only in the unvaccinated population (Berenson et al., 2022), which likely supports the importance of vaccination status to oral HPV prevalence status, irrespective of other potential risk factors.

Interestingly, although many HPV prevalence studies do attempt to include an ethnically diverse sample, it is important to note that there are limited studies that directly compare different ethnic groups for oral HPV prevalence. With this in mind, it is difficult to accurately determine how this factor impacts oral HPV risk in healthy individuals. Indeed, the nature of these data continue to highlight the unlikely linear relationship between oral HPV infection and a single risk factor. To decipher the influence of ethnicity and oral HPV prevalence, additional studies are sought to analyse potential associations, both biological and social.

1.4.5. HIV infection

HIV-positivity has been associated with an increased risk of oral HPV infection (Kreimer et al., 2004, Beachler et al., 2012, Antonsson et al., 2014, Oliver et al., 2018). Indeed, people living with HIV (PLWH) are more likely to develop HPV-OPC, as a result of decreased efficiency in HPV infection clearance (Beachler et al., 2014). The potential importance of HIV as a risk factor is supported with evidence that the prevalence of oral HPV between MSM and heterosexual men (men who have sex with women – MSW) is not significant when accounting for HIV status (Gonçalves et al., 2021). Thus, it suggests that MSM may be highlighted as a group at risk of oral HPV infection due to a higher HIV prevalence within this population.

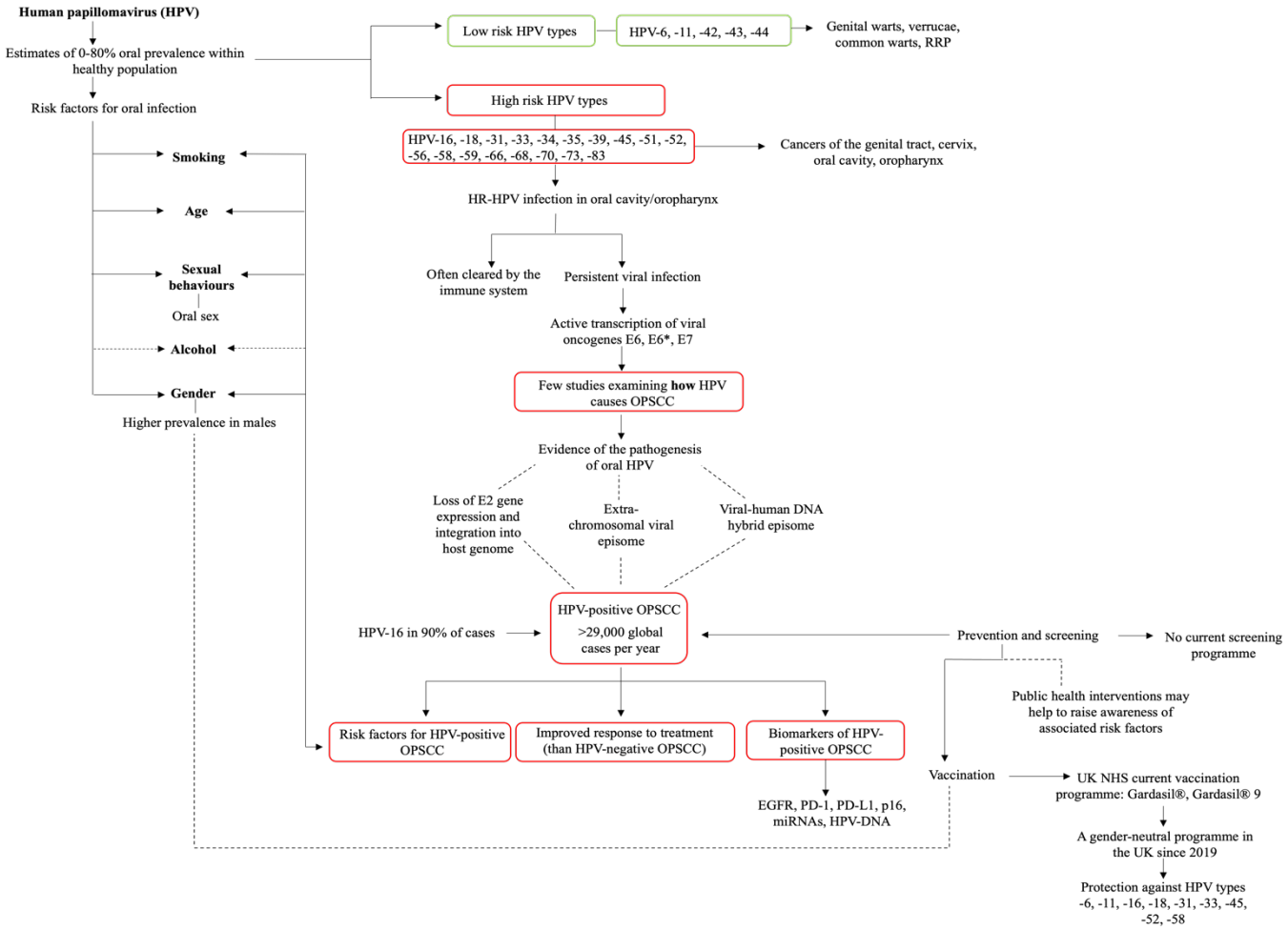


Figure 7. Oral HPV infection and its clinical relevance in HNC. The risk factors of oral HPV infection, LR- and HR-HPV viral types, current knowledge of oral HPV pathogenesis, salient points of HPV-OPC, and preventative measures against infection. Dotted lines indicate potential relationships; solid lines show greater confidence in associations.

Updated from: (Freckleton, 2023)

1.5. Public health interventions

The established link between HPV and cancer, most notably of the cervix, showed that cervical cancer could be a preventable disease. It set in motion public health interventions to protect eligible groups of people from prevalent viral types, both LR and HR. In the UK, women over the age of 25 years are invited for routine cervical screening. In addition, an efficacious HPV vaccination programme, that was initially offered to just girls of school year 8, is now available to both boys and girls aged 12-13. Both the progress, and impact of such public health interventions are discussed within this section.

1.5.1. The Cervical Screening Programme

Cervical screening was originally based upon the findings of Traut and Papanicolaou (Traut and Papanicolaou, 1943), which led to the development of the Papanicolaou test (Pap test) that aimed to detect morphological changes within the cells of the cervix (Papanicolaou and Traut, 1997). In the UK, the NHS Cervical Screening Programme (NHSCSP) commenced in the 1960s, and has been developed gradually over time to improve the accuracy and sensitivity of the results. For example, in 2004, liquid based cytology (LBC) tests were introduced. Then, in 2019, the cervical test changed again to molecularly detect the presence of HR-HPV viral types within the cervix. Here, HPV results form the basis of whether there is a need for further cytological analysis, with an HPV-positive result warranting such additional tests and monitoring of women (Public Health England, 2020b).

The selection of HR-HPV testing as the primary cervical screening method was considered controversial. In particular, there was concern for women who fall into the younger age bracket that the cervical screening programme targets. Individuals within this group are more likely to have a non-persistent HPV infection that is cleared by the immune system. Indeed, in these instances, such infections would be detected, cause concern for the individual, and result in women who are recalled to receive additional, invasive tests (i.e., a colposcopy) that could be unnecessary if the detected virus is a transient infection. Together, there is the potential for unnecessary psychological and physical harm for these women. However, in support of primary HPV testing, evidence suggests a greater ability to detect CIN III+ through HR-HPV than Pap testing within this age bracket (Feldstein et al., 2023). Nevertheless, it is important to remember that cervical cancer is not an entirely HPV-positive disease (Xing et al., 2020) and so, analysis for HR-HPV as the primary screening method could miss HPV-negative

individuals at risk of malignant changes; however, it is suitable for the majority of at-risk women given the overall burden of HPV infection in cervical cancers.

Cervical screening is currently offered for women, and individuals with a cervix, between the ages of 25 to 64 in both the UK and US. In Scotland, Wales, and the US, the test is available every five years (assuming a normal test result) (National Cancer Institute, 2020, Public Health Wales, NHS Inform - Scotland, 2023). However, in England, cervical screening is performed every three years up until the age of 50, then five-yearly after this (Public Health England, 2021b). The five-year gap between tests has been described as most cost-effective (Kim et al., 2017); however, a proposal to change this was met with concerns from some women over the effects on their health; i.e., greater anxiety and distress related to the impacts of increased wait times between screening on the likelihood of disease development (Rickford et al., 2023). As such, it is of importance to educate the public of changes to the programme and the reasons for this.

Although the cervical screening programme is offered routinely through the NHSCSP, not all eligible individuals will have the test. In the UK in 2018, a significant proportion of 1 in 4 women were recorded as not receiving their test (Public Health England, 2021a). As a direct result, the 'Cervical Screening Saves Lives' campaign was launched, which successfully corresponded with an increase in the number of women attending their test. Unfortunately, this improvement was only recorded during the period of the campaign (Public Health England, 2022a). This highlights the benefits of improved education on the risks of an HPV infection. In the UK, Jo's Cervical Cancer Trust have been instrumental in raising awareness of cervical cancer prevention and screening for 25 years, and have coordinated events and campaigns such as 'cervical cancer prevention week' (Hormone Health, 2024). Unfortunately, the charity closed this year (2024) and the impacts of this loss from an educational perspective will be seen over time. There are campaigns such as 'World HNC Day' and the 'Make Sense Campaign – HNC Awareness Week' for HNC (IFHNOS, 2016, EHNS, 2024); however, these are designed to raise awareness about the disease but are not related to disease prevention by screening. Currently, there is no similar screen or test for the presence of a high-risk oral HPV infection, and we do not have evidence of precancerous lesions in HPV-OPCs, as for the cervix.

1.5.2. The HPV vaccination programme

In September 2008, the HPV vaccination programme was introduced in the UK (Department of Health). Those who were eligible were girls in school year 8 (aged 12-13), and at this time, girls aged 14-17 were encouraged to take the vaccine as part of a 'catch-up' programme (Mehanna et al., 2019a). The bivalent vaccine (Cervarix, GSK) was the first to be offered, providing protection against HR-HPV types -16 and -18 (JCVI, 2008, Mehanna et al., 2019a). Subsequently, in 2012, the quadrivalent vaccine (Gardasil, Merck), was introduced to the programme. In addition to HPV-16 and -18, this vaccine also provides protection against the LR-HPV types -6 and -11, known to cause genital warts and RRP (Zheng and Baker, 2006). This quadrivalent vaccine is approved for use in both males and females of ages 9-26 (Markowitz et al., 2014, Petrosky et al., 2015, Mehanna et al., 2019a). At present, there are 3 HPV vaccines licensed for use in the UK. In addition to the bivalent and quadrivalent, there is a 9-valent vaccine (Gardasil® 9, licensed by Merck), that also offers protection against HR-HPV types -31, -33, -45, -51, and -58 (Petrosky et al., 2015, Mehanna et al., 2019a, MSD, 2024), which has been used in the UK vaccination programme since 2023 in a single dose schedule (GOV UK, 2023).

Analysis of the number of eligible individuals who receive the full vaccination course enables conclusions to be drawn on the effectiveness of the programme. The vaccine is highly efficacious; in the UK, recent estimates suggest an 87% reduction in cases of cervical cancer within girls who received their vaccine at the youngest eligible age (Falcaro et al., 2021), and specifically within Scotland, similar reports of an 89% reduction of cervical cancer, as well as CIN III lesions, within vaccinated women have emerged (Palmer et al., 2019). Furthermore, the prevalence of cervical HR HPV-16 and -18 are considerably lower within the vaccinated population, and this reduction extends to un-vaccinated women whose age group were eligible for the HPV vaccination at 12-13 years old (Kavanagh et al., 2017).

These data demonstrating the success of the HPV vaccination programme are coupled with generally widespread immunisation coverage in the region: a considerable number of year 8 girls received their first vaccines in the academic year of 2018-2019 (88%). However, as a result of the SARS-CoV-2 pandemic, and the closure of schools, the numbers of year 8 girls receiving their first vaccine fell to ~59% (2019-2020), though they rose again by ~17% in the following academic year (Public Health England, 2020a); however, these statistics are considerably lower than the WHO's global target of 90% vaccination coverage by 2030 (World

Health Organization, 2024e). Long-term effects of this reduction in HPV vaccination uptake should be monitored in coming years.

In the US, the vaccine was available in 2006, two years earlier than in the UK (Centers for Disease Control and Prevention, 2021). Data gathered by the Centres for Disease Control and Prevention (CDC) describe a coverage increase from only 26.7% in 2009 to 58.6% in 2020 (Centers for Disease and Control and Prevention, 2011, Pingali et al., 2021), and this still remains low. High vaccination coverage numbers, more akin to those of the UK, are evident in other developed European countries such as Italy, Spain, and Sweden (Sheikh et al., 2018); and in 2023, both Spain and Sweden reported a higher percentage of females completing the vaccination programme (>75%) than was recorded for the UK within the same year (~55%) (World Health Organization, 2023)

One explanation for the lower vaccination coverage in the US is that here the vaccination is not routinely offered through schools. It is, however, covered by most health insurance plans (Centers for Disease Control and Prevention, 2021). Some vaccinations are school requirements in the US, yet the HPV vaccination is currently only a requirement in five states (Calo et al., 2022). North and Niccolai., 2016 provide a useful review of the pros and cons of making the HPV vaccination mandatory for school entry. Broadly, they discuss a likely increased coverage of an effective and safe vaccine against the potential financial burden of the programme. Importantly, they point out successes in implementation of school entry requirements for other vaccinations in the US (i.e., Hepatitis B) to improve vaccination coverage between ethnic groups (North and Niccolai, 2016). Thus, these advancements could be translated to the HPV vaccination programme, although this would need to be met with educational and practical interventions to ensure participation.

Overall, the prevalence of HPV has fallen, in both men and women alike, since the first administration of the vaccine (Cameron et al., 2016, Oliver et al., 2017, Chaturvedi et al., 2019, Mehanna et al., 2019a). Indeed, it is evident that the prevalence of HPV has declined because of the immunisation programme with a strong correlation between prevalence and vaccination coverage (Markowitz et al., 2013, Oliver et al., 2017). This has been reported as high as a 34% decrease of HPV in cervicovaginal samples from pre- versus post-immunisation eras, within the US (Markowitz et al., 2016).

Given that the original purpose of the vaccination was to combat HPV-positive cervical disease in women, the statement made recently that the vaccine has 'successfully almost eliminated cervical cancer in women born since 1st September 1995' (Falcaro et al., 2021), is remarkable. Falcaro *et al.*, 2021 used statistical analysis to model risk of cervical cancer in vaccinated vs. non-vaccinated persons, and have clearly demonstrated the efficacy of the programme (Falcaro et al., 2021). As such, the elimination of cervical cancer is now a real possibility (World Health Organization, 2024b).

The extension of the immunisation programme to include boys (gender-neutral vaccination) is much debated across the world. In the UK in 2019, a decision was made by the Joint Committee on Vaccination and Immunisation (JCVI), and the government, to extend the HPV vaccination programme to year 8 schoolboys (Public Health England, 2019, UK Health Security Agency, 2019). The choice to do so was made despite evidence of a level of herd immunity within unvaccinated men (Canfell et al., 2012, Machalek et al., 2017, Chaturvedi et al., 2019).

In 2019-2020, the first year that the programme was accessible to boys, 54.4% of eligible boys received their first dose (Public Health England, 2019). These numbers are lower than those gathered for girls in previous years, but similar to the reported female uptake rate in 2020 (Public Health England, 2020a); therefore, one could assume these numbers were impacted by the global COVID-19 pandemic at the time. However, data from a recent academic year (2021-2022) demonstrated only a small increase of 8% in year 8 males who received their first dose (Public Health England, 2022b). As such, there is a clear need for better education amongst the public of the risks associated with HPV infection for boys, to promote vaccination uptake. This is discussed later at 1.5.3.

Within the head and neck, evidence suggests that the vaccination programme has impacted the prevalence of the virus here, as it has for the cervix. Oral HPV prevalence is lower in vaccinated versus non-vaccinated individuals (Hirth et al., 2017, Chaturvedi et al., 2018, Mehanna et al., 2019a) and the vaccine itself is described to have an >90% efficacy against oral HPV (Herrero et al., 2013). Contrastingly, others have concluded there a lack of concordance between vaccination status and oral HPV presence (Pickard et al., 2012), and the vaccine remains unlicensed for protection for OPC (MSD, 2024). Overall, the efficacy of the vaccination upon oropharyngeal infections, however, remains unclear.

1.5.3. Education on the risk of HPV infection within the public

Providing an education of the risks of HPV infection amongst the public is considered a useful tool to increase vaccination coverage and participation in HPV-related public health interventions. The reports on the public's understanding of the virus, and its risk to one's health have varied. Regarding the risks of HPV and cancer of the cervix, there have been conflicting reports of awareness, particularly within the vaccine target age group. One UK based study described a good understanding of the association of HPV and cervical cancer (Knight and Roberts, 2020), yet another found that one-third of their cohort were unaware of this link (Thanasas et al., 2020). In other reports specifically investigating individuals of at least cervical screening age, HPV awareness was linked to factors including having had a higher education or positive HPV test in the past (Kola-Palmer and Dhingra, 2020). Additional interventions would be useful to maintain and increase the reported awareness of HPV infection and cervical cancer, but specific interventions are required for target groups that are uninformed of such risks.

Prior to the UK HPV vaccination programme being extended to include boys, there was limited knowledge within the general population on the risks of the virus to these individuals. One study that investigated the understanding of the risk the virus presents to boys found that parents of teenage boys had less knowledge of the risks of HPV to boys than those to girls (Sherman and Nailer, 2018). Indeed, both the link between HPV and oral cancers, and HPV affecting men in general, is poorly understood (Knight and Roberts, 2020, Dodd et al., 2021). This lack of core knowledge of the virus, coupled with the historical association of the vaccination with women (i.e., being referred to as the 'cervical cancer' jab rather than HPV vaccination) may explain discrepancies in the uptake of the vaccination between males and females. It is probable that raising awareness of the risks of the virus to men would enhance the understanding of the importance of public health interventions designed to combat HPV-induced disease, and encourage everyone to partake in these.

1.6. Conclusion

To conclude, increased data on the prevalence within the general population would enable improved associations between risk factors, lifestyle choices, and oral HPV infection to be made, thereby informing public health interventions.

Several lifestyle and demographic factors have been linked to an increased risk of acquiring an incident oral HPV infection, and in some instances, these factors have also been linked to viral persistence. Currently, most of these risk factors are met with opposing reports that demonstrate that they have no association with oral HPV, and this can be partly explained by variation in population demographics (including lifestyle practices/behaviours) between individual studies. Moreover, the biology of the interactions these factors have with HPV to influence viral pathogenesis is largely unclear.

Further research to better describe how HPV influences the host cell biology and viral pathogenesis in the healthy oropharynx would enable improved identification of groups of individuals at risk of oral HPV and/or HPV-OPC, and identify subsets of individuals suitable for targeted therapies. Addressing the gaps in knowledge described here would be beneficial to the field, given the rapidly increasing cases of HPV-OPC worldwide.

1.7. Aims and objectives of this project

This project therefore aimed to address the following:

A. To investigate the prevalence of oropharyngeal HPV within the non-cancerous tonsil tissue of adults undergoing elective tonsillectomy surgery, and to examine viral pathogenesis in any HPV-positive tissues.

B. To provide an up-to-date estimate of the overall global burden of oral HPV within the healthy adult population.

C. To determine a 'protein expression blueprint' for HPV-DNA positivity within non-cancerous tonsil tissue to provide insight into how HPV influences host cell biology in the oropharynx.

To accomplish these aims, the following objectives were proposed:

1. To establish an NIHR-portfolio registered clinical study as part of a team of clinicians and scientists from several UK-based NHS hospitals to enable the recruitment of adult patients undergoing routine tonsillectomy surgery. The left and right palatine non-cancerous tonsil tissues, an oral swab, and a lifestyle questionnaire will be collected from all consenting individuals.

2. To develop a quantitative PCR (qPCR) screen that makes use of melt curve analysis with consensus primers to indicate the presence of the HPV L1 gene and to allow for the analysis of a broad range of HPV viral types in tonsillar genomic DNA (gDNA).

3. To define the final sample of non-cancerous tonsils collected from routine tonsillectomy surgeries as HPV-positive or HPV-negative by qPCR, and to perform additional pathogenesis studies for HPV-positive tissues (i.e., viral type determination).

4. The lifestyle characteristics and participant demographics of HPV-positive individuals will be analysed to indicate any associations between oral HPV prevalence and risk factors for infection.

5. To provide an updated global oral HPV prevalence estimate, a systematic review and meta-analysis of existing literature reporting on oral HPV prevalence within healthy adults will be conducted. Additional sub-analyses will be performed to examine HR-HPV prevalence and HR HPV-16 prevalence globally and then further analysed to highlight discrepancies in the burden of oral HPV between males versus females.

6. Protein isolated from HPV-positive and HPV-negative tonsil tissue will undergo proteomic analysis, using a non-biased technique – SWATH mass spectrometry. SWATH data will be subjected to differential protein expression analysis to highlight differences in protein expression between HPV-positive and HPV-negative tissues to describe a ‘protein expression blueprint’ for HPV presence in the healthy oropharynx.

We hypothesised that oral HPV would be prevalent in non-cancerous adult palatine tonsil tissues, and that the presence of HPV-DNA would influence host cell biology. With this in mind, we addressed the following research question: How prevalent is oral/oropharyngeal HPV in adults without corresponding malignancy, and what is the pathogenesis of the virus in this anatomical region?

CHAPTER 2: The development of a protocol to extract DNA from formalin-fixed tonsil tissue.

2.0. Introduction

To investigate the pathogenesis of oral HPV within non-cancerous tonsil tissues a clinical study was established on the NIHR-portfolio to recruit our prospective cohort of adult patients undergoing routine tonsillectomy surgery. Non-cancerous 'healthy' tonsil tissue, an oral swab, and a lifestyle questionnaire were collected from participants. Both the clinical study itself, and those eligible to participate are described in detail within the final research methodology (Chapter 3). When the study commenced at University Hospitals of Derby and Burton (UHDB) in 2016, tonsil tissue specimens were immediately frozen at -80°C at the point of removal to promote the preservation of nucleic acid for molecular analyses (Shabihkhani et al., 2014).

Sample collection for this project was ongoing at the time of the outbreak of the coronavirus pandemic in 2020. In line with government guidelines at the time, all tissue collection for the study ceased with the postponement of routine surgeries. When elective tonsillectomies were allowed to recommence in November 2021, changes to the study protocol were made to ensure the safety of individuals working on the project. It was agreed with the study hospitals and University that tonsil tissues would be immersed in 10% neutral buffered formalin (NBF) immediately after surgery; a change which allowed sample collection to continue throughout the pandemic. 10% NBF was selected to inactivate any potential SARS-CoV-2 present within tissues (Jureka et al., 2020, Kumar et al., 2021). In addition, 10% NBF was considered suitable as it is available within hospital pathology laboratories, making it easily accessible for use by participating hospitals.

Here, we discuss the changes to the sample collection protocol made because of the COVID-19 pandemic, and their lasting impacts on the project. An extensive optimisation process was required to develop a protocol to isolate DNA of sufficient quantity and quality from formalin-fixed tonsil tissue for use in subsequent molecular analyses. In this chapter, we therefore present our optimised method for DNA extraction from formalin-fixed tonsil tissue. We also outline a way in which to fix tonsil tissues to render them safe to researchers, whilst still preserving DNA integrity as much as possible. Unfortunately, as discussed at 2.7, we were unable to isolate RNA from tonsil tissue following fixation. Therefore, in samples collected from 2021 onwards we base our analysis on DNA alone.

2.1. The impact of SARS-CoV-2 on sample collection

The outbreak of the coronavirus pandemic in 2020 had a significant and lasting impact on this project for several reasons. Importantly, there was no sample collection for a period of 10 months. The project relied entirely on the tonsil tissues obtained from consenting individuals. Given that SARS-CoV-2 causes respiratory disease (World Health Organization, 2024c), our tissue of interest (palatine tonsils) were at risk of harbouring the virus. An amendment to the sample collection protocol (fixation in 10% NBF during surgery; IRAS 194857 amendment 3.1, July 2021) was necessary, as it meant that any potential SARS-CoV-2 present within specimens would be inactivated prior to their processing (Jureka et al., 2020, Kumar et al., 2021). This allowed the safe continuation of the study. However, formalin fixation of tonsil tissues meant that the DNA and RNA isolated from these tissues were not sufficient for use in this project using the extraction protocol we had developed for frozen tissue. Therefore, a significant amount of time was dedicated to re-optimising extraction methodology, in line with the protocol change to sample collection.

2.2. The isolation of nucleic acid and protein from tonsil tissues

The tonsil crypts are believed to be the primary site of most HPV-OPCs (Kim et al., 2007). As such, the palatine tonsils were selected as the most appropriate biological specimen to investigate the pathogenesis of HPV within the oropharynx. Given that HPV does not uniformly infect all cells of the cervix (Burd, 2003), we hypothesised that the virus would not be present across the whole tissue in an HPV-positive tonsil. With this in mind, it was imperative from the outset to design a protocol to sequentially isolate DNA, RNA, and protein from each tonsil specimen. If our hypothesis of HPV positivity varying across an infected tissue was correct, then if only one section was taken for DNA analysis, we could not be as confident of the HPV-DNA status across the whole tonsil. As such, we sought to develop a cost- and time-effective method to simultaneously extract nucleic acid and protein from tissues.

Human tonsil tissue is composed of lymphoid connective tissue and an outer layer of non-keratinising stratified squamous epithelium (Shiba and Chhetri, 2019). Formalin fixation unfortunately resulted in the tonsil specimen becoming rubbery in texture and pale in colour. Consequently, it was difficult to distinguish connective ‘fatty’ tissue when dissecting for extractions. If this fatty tissue was unintentionally used as the starting material, it resulted in a lower DNA concentration and/or quality, despite other replicates of the same conditions yielding good DNA. Therefore, at most stages of the methodology development presented

within this chapter, the various protocol modifications described were trialled with multiple replicates. The selected values displayed in the results tables are representative of biological replicates of the same set of conditions tested, the number of which is indicated in each table. There was wide variation in these results, and not all repeats of a given set of parameters were as good or bad as those displayed here.

2.2.1. TRIzol™ reagent for DNA extraction from formalin-fixed tonsil tissue

In the first instance, TRIzol™ reagent [Invitrogen™] was selected as a more cost-effective alternative to commercially available extraction kits, as tonsils are generally large tissues (~1.5 g (Michaels, 1987)). TRIzol™ reagent is designed to enable the sequential isolation of RNA, DNA, and then protein from animal tissue. The protocol recommends 1 mL reagent per 50-100 mg tissue (ThermoFisher Scientific, 2016). By contrast, one example of a kit designed for simultaneous DNA, RNA, and protein extraction from tissue, recommends no more than 10 mg tissue per spin column (Merck, 2024), which would add significant expense. Hence, TRIzol™ reagent was selected for the subsequent isolation of DNA, RNA, and protein following manufacturer's protocol. In brief, tissues were cut into 4-5 relatively even, sequential, sections and homogenised for 14, 16, or 20 minutes, in 1 mL TRIzol™ reagent using the TissueLyser LT instrument [Qiagen]. A 5 mm stainless steel bead was added to each sample for homogenisation. The TissueLyser LT [Qiagen] was selected due to its higher throughput (12 samples at any one time) in comparison to the TissueRuptor [Qiagen] (1 sample at any one time). Unfortunately, we discovered that DNA pellets extracted in this manner would not resuspend in TE buffer, regardless of homogenisation time (Table 3).

To address this issue, we next trialled a protocol designed by Shirley Zhu at Stanford University, California, to 'back extract' DNA from the organic phase generated during TRIzol™ phase separation. Tissue sections were prepared and homogenised as before. Then, to improve DNA isolation, 500 µL back extraction buffer (BEB) [4 M guanidine thiocyanate, 50 mM sodium citrate, 1 M Tris base] was added to the remaining interphase/organic phase. To attempt to increase the DNA yield, the BEB protocol was simultaneously trialled on two small sections of tissue. To generate these, as before, relatively even sequential sections of tissue were cut and each section was then further cut in half. These small portions of tissue were then processed, and extraction performed in 1 mL TRIzol™ reagent. After resuspension of DNA pellets in TE buffer, DNA from sections were pooled to give 1 DNA sample. This

modification is referred to as ‘A/B’ sections. Fortunately, with and without the use of ‘A/B’ sections, these DNA pellets did resuspend in elution buffer, allowing further DNA analysis (Table 3).

DNA concentration was subsequently measured using a Nanodrop™ 2000 spectrophotometer [ThermoFisher] – this method was routinely used for DNA/RNA concentration measurements throughout this chapter. The quantity of DNA measured as high as 224.1 ng/μL with a 16-minute homogenisation (Table 3). Unfortunately, however, the 260:280 ratios indicated poor quality or impure DNA (range of 1.44-1.49 at 14 minutes and 1.55-2.39 at 16 minutes). Indeed, when gDNA samples were analysed for the presence of housekeeping gene, beta-actin, by qPCR (see 3.8.1), it was not detectable. This was the case irrespective of whether the BEB protocol modification had been performed with ‘A/B’ tissue sections and pooled together, or not. Unfortunately, this confirmed that no viable DNA had been isolated from the formalin-fixed tissue using TRIzol™ together with ‘back extraction’ modification, despite a reasonable nanodrop reading.

Frozen tissue is the recommended starting material for use with TRIzol™ reagent. We believed that the issues we were experiencing were because the tissues were formalin-fixed. To confirm this we trialled the protocol with and without BEB and ‘A/B’ sections using the desired starting material (frozen tonsil tissue collected pre-pandemic). Indeed, DNA concentration, 260:280 ratio, and theoretical cell number per μL improved significantly, with beta-actin detected at PCR cycle ~22-27, regardless of protocol modification used (Table 3).

These data demonstrate that both the quantity and quality of DNA isolated from formalin-fixed tonsil tissue, using TRIzol™ reagent, was not sufficient for further analysis, despite protocol modifications. Although both RNA and protein were also isolated simultaneously from the samples, given that DNA was so poor, we were unable to complete the first experimental aim of defining the HPV-status of that tonsil tissue section. Later, both RNA and protein isolated from formalin-fixed tissue using TRIzol™ reagent were also found to be of poor quality. Consequently, the decision was made to start investigating alternative methods to obtain nucleic acid that could be used in this work.

Table 3. Summary of concentration and integrity of DNA extracted using TRIzol™ reagent. DNA was isolated using TRIzol™ reagent, with various homogenisation times and extraction modifications, as indicated. Where the number of replicates was greater than 1, the mean value and range of the replicates is shown. DNA concentration was measured using the Nanodrop™ 2000 spectrophotometer.

Extraction conditions/ modifications	Fresh-frozen or formalin-fixed	Homogenisation time (minutes)	Protocol modification	Representative (N=X)	Mean DNA concentration and range (ng/uL)	260:280	Mean beta-actin PCR Ct value and range	Mean theoretical number of cells / µL and range	Notes
1	Formalin-fixed	8	N/A	1	31.1	-	Undetected	-	N/A
2	Formalin-fixed	14	N/A	4	-	-	-	-	DNA pellets would not resuspend in buffer
3	Formalin-fixed	16	N/A	4	-	-	-	-	DNA pellets would not resuspend in buffer
4	Formalin-fixed	20	N/A	4	-	-	-	-	DNA pellets would not resuspend in buffer
5	Formalin-fixed	14	Back extraction buffer	2	35.5 (13.8-57.2)	1.47 (1.44-1.49)	Undetected	N/A	N/A
6	Formalin-fixed	16	Back extraction buffer	2	178.5 (132.8-224.1)	1.97 (1.55-2.39)	Undetected	N/A	N/A
7	Formalin-fixed	20	Back extraction buffer	2	100.2 (34.0-166.4)	1.54 (1.50-1.57)	Undetected	N/A	N/A
8	Formalin-fixed	14	BEB, A/B sections	1	104	1.49	Undetected	N/A	N/A
9	Formalin-fixed	16	BEB, A/B sections	1	147.7	1.49	Undetected	N/A	N/A
10	Formalin-fixed	20	BEB, A/B sections	1	105.4	1.47	Undetected	N/A	N/A
11	Fresh-frozen	6	N/A	3	1722.3 (805.4 – 2776.8)	1.88 (1.82-1.92)	22.5** (22.-23.0)	20128 (14079-26558)	N/A
12	Fresh-frozen	10	BEB, A/B sections	1	231	1.81	27.0**	1977	N/A
13	Fresh-frozen	14	BEB, A/B sections	1	514.3	1.98	26.7**	2443	N/A
14	Fresh-frozen	16	BEB, A/B sections	1	383.3	1.89	26.3**	3006	N/A

'-' indicates the sample was not sufficient for reading.

** = 25 ng gDNA added per reaction for beta-actin qPCR analysis.

2.3. The AllPrep® DNA/RNA FFPE kit [Qiagen] for the extraction of DNA from formalin-fixed tonsil tissue

We therefore sought to identify an alternative method for DNA extraction from formalin-fixed tonsil tissue. Human tissues are often formalin-fixed and paraffin embedded (FFPE) for subsequent analysis and research. As a result, commercial extraction kits have been designed for FFPE tissue. Following extensive discussions with technical experts, manufacturer Qiagen recommended adapting their AllPrep® DNA/RNA FFPE kit as potentially suitable for formalin-fixed (but not paraffin embedded tissue). We were willing to compromise simultaneous DNA, RNA, and protein extraction to improve the overall quality of each for further analysis.

Thus, we next outline the steps taken to optimise a protocol using the AllPrep® DNA/RNA FFPE kit to isolate DNA sufficient for molecular analysis from tonsil tissue fixed in 10% NBF for prolonged periods (months). Commercial extraction kits from other manufacturers were trialled concurrently (including the RNA/DNA/Protein Purification Plus Kit [Geneflow]) as we sought to improve our DNA yield and quality (data not shown), with promising results from the AllPrep® kit, which was therefore selected for protocol development.

2.3.1. Initial AllPrep® DNA/RNA FFPE kit trial

In the first instance, the manufacturer's protocol was followed with no modification to determine a baseline for the quality (purity and integrity) and quantity of DNA that could be obtained from prolonged formalin-fixed tonsil tissue. Qiagen advised us to start the protocol from step 5, after deparaffinisation steps (Qiagen, 2020). In brief, a ~40 mg section of tissue was weighed and DNA was isolated, following manufacturer's instructions. DNA was eluted in 40 µL Buffer ATE [Qiagen]. In this way, we were able to isolate DNA with a concentration and 260:280 ratio of 324.3 ng/µL, and 1.78, respectively (Table 4), reassuringly similar to the ~1.80 ratio used as an indication of pure DNA (ThermoFisher Scientific, 2012). Importantly, using the AllPrep® kit, we were also able to detect beta-actin by qPCR analysis. Despite this, Ct values were very late (~34.8 out of 40 cycles); given that genomic beta-actin is present at 2 copies per cell we suspected that, if we were unable to detect an abundant housekeeping gene until a late PCR cycle, then a low viral load in an HPV-positive individual would go undetected.

Nonetheless, as we were able to detect viable DNA, we decided to attempt to optimise this kit for the extraction of DNA from formalin-fixed tonsil tissues.

2.3.2. Optimisation of AllPrep® DNA/RNA FFPE kit extractions

The first consideration of our methodology development was to determine the mass of tissue to be added per spin column. As stated previously, we were keen to isolate DNA/RNA sequentially across most of each tonsil to give a good representation of its HPV-status but if only ~40 mg tissue was processed this would require ~25 columns per tonsil, which was financially prohibitive. With this in mind, we aimed to increase the mass of the tissue per column, without compromising DNA quality/quantity. However, when extractions were increased from 40 to 100 mg of tissue with no protocol amendments, we were unable to amplify any beta-actin in the sample by qPCR (Ct value = undetected) (Table 4). As a result, we sought to determine whether there were any additional protocol modifications that would allow for a greater mass of tissue per spin column to be used, whilst preserving DNA concentration and quality.

2.3.3. Tonsil tissue storage

We felt confident that the poor DNA integrity as indicated by beta-actin qPCR was a result of prolonged periods of fixation in 10% NBF (Vitošević et al., 2023). In other experimental application such as IHC techniques, a transfer to 70% EtOH following a relatively short fixation period is considered desirable for antibody recognition of antigens (Otali et al., 2013). In this project for our molecular analysis, we could not change the length of time that tonsils collected up to this point had been immersed in 10% NBF but on this basis, we speculated that a transfer to 70% EtOH before DNA extraction could potentially improve the process and prevent further DNA damage. Moreover, we still aimed to increase the mass of tissue per spin column. With this in mind, we attempted to extract DNA from formalin-fixed tissue following a 48-hour incubation in 70% EtOH with increasing masses of tissue.

To test this, a tonsil tissue that had been subjected to a long period (several months) of fixation was transferred to 70% EtOH for a 48-hour room-temperature (RT) incubation, before sectioning. After incubation, both a 40 mg and a 200 mg section of tissue were weighed, and DNA was isolated with the AllPrep® DNA/RNA FFPE kit following manufacturer's protocol. With the addition of an incubation in 70% EtOH, a 40 mg tissue section yielded DNA

concentrations and 260:280 ratios within the range of ~819-1195 ng/μL and 1.89-1.93, respectively. Importantly, the beta-actin Ct value decreased from 34.8 (with no 70% EtOH transfer) to 30.6 (mean Ct value calculated from replicates). The 48-hour 70% EtOH incubation also improved extracted DNA from greater tissue masses of ~150-200 mg (mean concentration: 525.4 ng/μL, mean 260:280: 1.70, mean Ct value: 33.7) (Table 4) permitting a more cost-effective extraction from sections of tissue that would represent a greater proportion of tissue from across the whole tonsil.

Together, these data indicated that a 70% EtOH transfer from formalin is necessary for improved DNA quality using the AllPrep® extraction kit, and that a smaller mass of tissue per spin column is still more effective, regardless of 70% EtOH transfer. As a result of these improved yields and quality of DNA we decided to transfer all the tissues already collected and stored in 10% NBF to 70% EtOH until processing. As such, unless otherwise stated, all extractions discussed from here on were performed on formalin-fixed tissue that was subsequently incubated in 70% EtOH.

2.3.4. Physical disruption and homogenisation of formalin-fixed tissue

We continued to increase the mass of tissue per AllPrep® spin column. Our next step was to explore homogenisation methods for physical tissue disruption, for larger masses of tissue. Once again, we selected the TissueLyser LT instrument [Qiagen] with one 5 mm stainless steel bead added for physical disruption.

Tonsil tissue was cut into sections with a mass range of 80-120 mg. Individual sections were then cut into fine pieces with a scalpel to promote lysis. All tissue homogenisation was performed in 150 μL Buffer PKD with 10 μL proteinase K solution (both from the AllPrep® extraction kit). Following manufacturer's instruction, the recommended homogenisation time for DNA extraction from animal tissue is <6 minutes at 20-30 Hz (Qiagen, 2009). Therefore, we started with 2-minutes for our formalin-fixed tissues and increased the time in increments of 1 minute. The homogenisation time was optimised to 10 minutes at 50 Hz as a compromise between how homogenous the tissue appeared visually whilst not wanting to introduce additional DNA shearing. DNA was extracted from the homogenised tissue with the AllPrep® kit. The difference in DNA concentration, 260:280 ratio and beta-actin Ct values did not differ much with the mass of tissue used for the extraction. Overall, replicates from 90-100 mg tissue

appeared to be the optimal tissue mass when comparing the mean values across all 3 measured parameters: 564.9 ng/μL, 260:280 – 1.85, beta-actin Ct value – 31.3 (range of 28.9-32.5; Table 4). This was noticeably better than that obtained without a 70% incubation or any homogenisation. Mean values of DNA concentration, 260:280, and Ct values for replicates of ~80 mg and ~120 mg extracted tissue are also shown in Table 4.

We therefore selected to use ~100 mg sections of tissue per spin column for AllPrep® kit DNA extractions. The approach was seen as a good compromise to keep the cost of extraction reasonable, whilst still achieving good quality nucleic acid, and enabling the scrutiny of HPV-status across the entire tonsil tissue. As such, in all further optimisation of the extraction kit, ~100 mg tissue was used alongside a 10-minute homogenisation step (50 Hz) before extraction.

2.3.5. Final protocol modification for prolonged formalin-fixed tonsil tissue

After discussing our progress with adapting this kit for our tissues with the manufacturer's technical support team, they recommended washing tissue in 1X phosphate buffered saline (PBS) prior to extracting. This would remove any formalin/70% EtOH traces which should improve the performance of the spin columns. Hence, 1X PBS washes were introduced to the protocol. 100 mg tissue samples were subject to 2x 5-minute 1X PBS washes prior to undertaking homogenisation and nucleic acid extraction as described above.

In comparison to DNA isolated from formalin-fixed tissue incubated in 70% EtOH, 1X PBS washes in tissues that had had no 70% EtOH transfer did not improve DNA purity, and the mean DNA concentration was similar (Table 4), but this varied hugely between replicates of samples isolated using the same extraction parameters. However, we were again able to detect beta-actin from these samples (Table 4) suggesting that the 1X PBS washes may be a good additional step to improve the DNA yield from formalin-fixed tissues. Although these data in Table 4 suggest that the average Ct value of beta-actin detection did not differ between 1X PBS washed (Ct = 31.2) and 70% EtOH incubated (Ct = 31.3) tissues, please note that 75 ng total gDNA was added to the PCR reaction mix of the PBS-washed tissues, whereas only 25 ng gDNA was added to reactions of the 70% EtOH washed tissues; therefore, the theoretical cell number per μL was higher in the samples following an 70% EtOH transfer (this arose because the PCR protocol – Chapter 3 – was simultaneously being adapted for formalin-fixed tissues, with different DNA concentrations tested).

We tested a combination of the two modifications (a 70% EtOH transfer and 1X PBS washes) but the mean beta-actin detection of replicates subjected to these protocol amendments (Ct = 30.6 with 70% EtOH and 1X PBS, Ct = 31.2 with 1X PBS no 70% EtOH) suggested that introducing a 70% EtOH incubation before 1X PBS washes made little difference to the quality of extracted DNA. Given that the tissues had been fixed in 10% NBF for several months, we suspected DNA was already very fragmented. However, we did find that the ranges of DNA purity measured were marginally improved (260:280 ratio was calculated between 1.79-1.93, Table 4), consequently, on this basis and the advice of the manufacturer, we decided to use both a 70% EtOH transfer and 1X PBS washes prior to extraction.

In summary, as we were, unfortunately, confident that DNA fragmentation within the formalin-fixed tonsils was already high, we aimed to extract the best quantity and quality possible from these tissues. Our focus was to isolate DNA of sufficient integrity for downstream molecular analyses (such as qPCR for HPV-DNA detection). Taken together, we concluded the following tonsil storage and AllPrep® protocol modifications (Table 4) produced the best quality DNA with these formalin-fixed tonsil tissues:

- A transfer to 70% EtOH from 10% NBF for at least 48-hours prior to DNA extraction
- 2x 5-minute 1X PBS washes
- ~100 mg tissue fragments per AllPrep® spin column
- 10-minute homogenisation using TissueLyser LT instrument, at 50 Hz

This final method is outlined in Figure 8 and detailed in Chapter 3, section 3.5.2.

Table 4. Summary of measured concentration and integrity of tonsil DNA extracted using the AllPrep® kit, with indicated protocol modifications. DNA was isolated using the AllPrep® DNA/RNA FFPE Kit [Qiagen] from tissue fixed in 10% NBF for prolonged periods. Various protocol modifications were trialled to improve DNA quality: tissue mass, homogenisation time, and additional protocol modifications are shown. The TissueLyser LT instrument [Qiagen] with 1X 5 mm stainless steel bead [Qiagen] was used for any homogenisation. Where the number of replicates is greater than 1, the mean value and range of the replicates is shown. The number of replicates per set of conditions tested is indicated (N=X). DNA concentration was measured using a Nanodrop™ 2000 spectrophotometer [ThermoFisher].

Extraction conditions/ modifications	Tissue mass (mg)	Homogenisation time (minutes)	Protocol modification	Representative (N=X)	Mean DNA concentration and range (ng/µL)	260:280	Mean beta-actin PCR Ct value and range	Mean theoretical number of cells / µLand range	Notes
1	40	N/A	N/A	1	324.3	1.78	34.8**	10	No EtOH transfer
2	100	N/A	N/A	1	154.1	1.58	Undetected	N/A	N/A
3	40	N/A	70% EtOH transfer	2	1007.1 (819.2-1195.0)	1.91 (1.89-1.93)	30.6** (30.5-30.7)	172 (161-182)	N/A
4	150-200	N/A	70% EtOH transfer	2	525.4 (118.1-932.6)	1.7 (1.70-1.70)	33.7** (33.2-34.2)	22 (15-29)	No 1x PBS washes
5	70-85	10	70% EtOH transfer	6	539.2 (368.2-711.0)	1.87 (1.84-1.89)	31.2** (30.6-32.1)	127 (62-174)	No 1x PBS washes
6	90-100	10	70% EtOH transfer	6	564.9 (173.3-1180.6)	1.85 (1.78-1.88)	31.3** (28.9-32.5)	164 (47-555)	No 1x PBS washes
7	120-150	10	70% EtOH transfer	1	381.5	1.87	30.7**	163	No 1x PBS washes
8	100	10	1x PBS washes	6	547.3 (308.3-678.1)	1.86 (1.70-1.91)	31.2 (29.9*-35.5**)	69 (6-102)	No EtOH transfer
9	100	10	70% EtOH transfer, 1x PBS	26	703.4 (279.2-1037.9)	1.86 (1.79-1.93)	30.6 (28.6*-33.6**)	111 (28-229)	N/A

* = 75 ng, ** = 25 ng, *** = 50 ng, total gDNA added to each qPCR reaction analysing the presence of beta-actin.

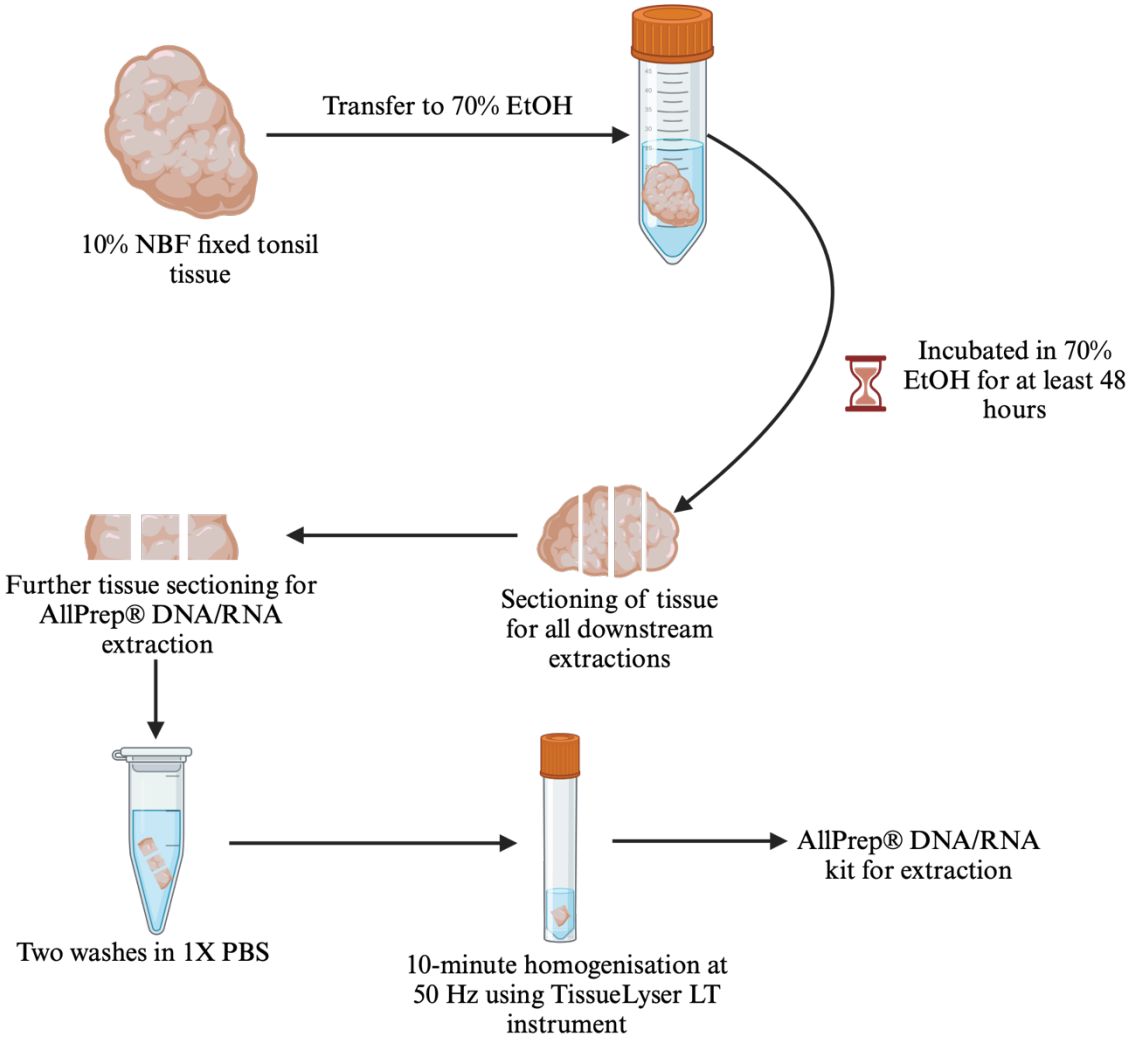


Figure 8. Final tissue homogenisation protocol for the AllPrep® DNA/RNA FFPE kit. Fixed tonsil tissue was transferred to 70% EtOH at least 48 hours before sectioning for an AllPrep® DNA/RNA extraction. Other modifications to yield optimal DNA concentration and quality included 1X PBS washes and a 10-minute homogenisation using the TissueLyser LT instrument [Qiagen]. DNA was extracted from prepared tissues in accordance with the manufacturer's instructions commencing at step 5, having omitted deparaffinisation (AllPrep® DNA/RNA FFPE kit [Qiagen]).

Image was created using BioRender.com

2.4. 24-hour formalin fixation period of tonsil tissues

The methodology development process thus far strongly suggested that prolonged periods in 10% NBF was having a negative impact on the integrity of DNA. We discuss here further amendments made to the study sample collection protocol to improve DNA quality. Any changes needed to be practical for participating hospitals and still render potential SARS-CoV-2 inactive for the safety of the researchers. We anticipated that decreasing the fixation period down to only 24 hours may better preserve DNA integrity for downstream PCR analysis. It would not be practically possible to immediately process tissues after this period, but we hypothesised that by introducing a transfer to 70% EtOH step at the hospitals, tissues and their DNA would be better preserved until processing at the University. 70% EtOH can also inactivate SARS-CoV-2 (Kumar et al., 2021, Nomura et al., 2021), and so the fixation period would continue without continued DNA damage inflicted by 10% NBF.

To test our hypothesis, fresh-frozen tonsil tissue was immersed in 10% formalin for 24 hours at RT. The tissue was transferred to 70% EtOH for ~48 hours before beginning extraction. 100 mg sections of tissue were then homogenised, and DNA extracted using the AllPrep® kit with all protocol amendments summarised above. There was a striking improvement in DNA concentration and integrity following only a 24-hour formalin fixation period. The mean and range for replicates of these parameters was: 4019.1 ng/μL (939-6760.6 ng/μL), 1.88 260:280 ratio (1.84-1.91) and, 26.0 beta-actin Ct value (25.3-26.7; Table 5) with this earlier detection of beta-actin by qPCR suggesting less DNA fragmentation within the sample. We concluded that for future tissue collection, the period of formalin fixation needed to be as short as possible.

2.5. Amendment to the HPV clinical study protocol

In March 2023, an amendment to the HPV clinical study protocol was made (IRAS amendment 5.0). From this point all tonsils collected in line with the clinical study were still fixed in 10% NBF at the point of removal. However, tonsil tissues were now to be transferred to 70% EtOH 24-96 hours post fixation. Whilst 24 hours would have been the most ideal fixation period, the timeframe was extended to 96 hours to keep the process practical for the pathology laboratories at the hospitals, particularly given the frequency of tonsillectomies scheduled on Fridays and Saturdays.

2.6. Bead type

Finally, we considered use of the different beads available for use with the TissueLyser LT instrument [Qiagen] for the homogenisation step. As discussed at 2.3.4., we observed that the fixed tissue did not give a homogenous solution following physical disruption. Two or three 3 mm tungsten carbide beads [Qiagen] were added per sample for a 10-minute homogenisation period (50 Hz) to determine if smaller beads would yield improved physical disruption. This did not improve the DNA quality and quantity in comparison to homogenisation with the 5 mm stainless steel beads (Table 5); in fact, DNA yield notably dropped, but as this was tested using tonsil tissues from the hospital, the fixation time could have been up to 96 hours. Nonetheless, we decided to continue to add one 5 mm stainless steel bead per sample for optimal tissue homogenisation; based on the rationale that 3 mm tungsten carbide beads did not improve physical disruption of tissue sections and that the larger (5 mm) beads were easier to work with.

Table 5. Summary of concentration and integrity of DNA isolated from 24-96 hour formalin-fixed tonsil tissue, using the AllPrep® DNA/RNA FFPE kit. Data are representative of replicates for the same set of conditions tested. The TissueLyser LT instrument was used for all homogenisation, with stainless steel or tungsten carbide beads, as indicated. All tissue was transferred to 70% EtOH 24-96 hours after fixation in 10% NBF. Tissues were washed in 1X PBS prior to homogenisation. DNA was measured using a Nanodrop™ 2000 spectrophotometer [ThermoFisher].

Extraction conditions/ modifications	Homogenisation beads	Protocol modification	Representative (N=X)	Mean DNA concentration and range (ng/µL)	Mean 260:280 and range	Mean beta-actin PCR Ct Value and range	Mean theoretical number of cells / µL and range
1	Stainless Steel	70% EtOH transfer, 1x PBS	6	4019.1 (939 – 6760.6)	1.88 (1.84 – 1.91)	26.0** (25.3 – 26.7)	3643 (2420 – 4670)
2	Tungsten Carbide	70% EtOH transfer, 1 x PBS	7	409.9 (333.2 – 490.4)	1.91 (1.90 – 1.92)	24.4*** (24.0 – 24.6)	2317 (2042 – 2935)

** = 25 ng, *** = 50 ng, total gDNA added to each qPCR reaction analysing the presence of beta-actin.

2.7. The extraction of RNA from formalin-fixed tonsil tissue

The AllPrep® kit is designed for a concurrent DNA and RNA extraction meaning that in all protocol amendments discussed above, RNA and DNA were isolated sequentially, as per manufacturer's instructions. When optimising the kit for formalin-fixed tissue, we focussed on the development of a protocol to extract DNA to enable determination of HPV-DNA status by qPCR. Here, we describe the quantity and integrity of RNA that was also obtained with each of the protocol modifications discussed above.

To best represent the various protocol modifications introduced during DNA extraction optimisation, we selected several RNA samples. RNA concentration and quality were assessed with both the Nanodrop™ 2000 spectrophotometer [ThermoFisher] and the TapeStation 4200 analyser [Agilent] using their High Sensitivity RNA ScreenTape® kit [Agilent] (Table 6). Comparatively, the RNA concentration measured using the Nanodrop™ was consistently higher than that with the TapeStation analyser. Regardless of concentration, the 260:280 ratios measured by the Nanodrop™ were too low to be pure RNA (Table 6); pure RNA has an 260:280 ratio of ~2.0 (ThermoFisher Scientific, 2012).

In addition to RNA concentration the TapeStation Analyser 4200 [Agilent] provides an RNA Integrity Number equivalent (RINe) value (Agilent, 2021). The RINe value is calculated as a measure of the ratio of the 28S and 18S ribosomal fragments. The presence of these fragments intact suggests good quality RNA (Walker et al., 2023). For all RNA isolated from formalin-fixed tonsil tissue, the RIN values were <2.0. Therefore, in all cases, the 18S ribosomal fragment was absent, suggestive of RNA degradation.

Finally, we assessed RNA quality by quantitative reverse transcriptase polymerase chain reaction (qRT-PCR) for beta-actin. In brief, RNA was treated with the TURBO™ DNA-*free* Kit [Ambion] to remove any contaminating DNA. Conversion to complementary DNA (cDNA) was performed using the Precision nanoScript™ 2 Reverse Transcription kit [Primerdesign]. Manufacturer's protocol was followed in both cases. The presence of beta-actin was assessed by qPCR (described in full at 3.8). Most samples displayed no amplification of beta-actin, or their Ct value measured outside of the range of the standard curve (Figure 9). Beta-actin was detected at cycle 28/40 in one formalin-fixed sample (extraction condition 4; Table 6); we suspect this to be due to contaminating gDNA, rather than cDNA, given that no concentration/RIN value was measured for the same sample by the TapeStation analyser. In

marked contrast to the majority of formalin-fixed RNA samples, cDNA from a fresh-frozen sample (extraction condition 1) extracted by TRIzol™ and converted to cDNA as above, showed detectable beta-actin amplification during cycle 19/40 (Table 6) despite still having a low RIN value.

Having used multiple methods to assess RNA integrity, we concluded that any RNA isolated from formalin-fixed tissue using the AllPrep® kit would not be sufficient for RNA analysis by sequencing or qRT-PCR for both host and viral markers. RNA from formalin-fixed tissues was not included in this work. However, a section of tissue from these all the included specimens has been stored at -80°C in 70% EtOH for potential future work.

Table 6. Summary of the quantity and integrity of RNA isolated from formalin-fixed tonsil tissue. RNA was extracted using the AllPrep® DNA/RNA FFPE kit [Qiagen] from tonsil tissue fixed in formalin for various times. For frozen tonsil tissue, RNA was extracted using TRIzol™ reagent [Invitrogen™]. Tissue was homogenised using either stainless steel or tungsten carbide beads. RNA concentration and 260:280 values were measured using the Nanodrop™ 2000 spectrophotometer [ThermoFisher]. RNA concentration and RINe values were analysed with the TapeStation 4200 analyser [Agilent]. cDNA beta-actin qPCR Ct values are displayed. Data presented are those that gave the highest RIN value from any replicates of the same set of conditions analysed.

	Tonsil storage and RNA extraction				Nanodrop™ 2000 spectrophotometer		Agilent Tapestation Analysis			cDNA analysis
Extraction conditions /modifications	**Frozen or formalin fixed**	**Beads**	**Extraction method**	**Representative (N=X)**	**RNA concentration (ng/µL)**	**260:280**	**RNA concentration (µg/ml)**	**RIN value**	**Comment**	**Beta-actin qPCR Ct value**
1	Frozen	Stainless Steel	TRIzol™	1	9571.4	2.03	22.3	2.8	Lower ribosomal fragment is missing	19.6
2	Prolonged formalin fixation → 70% EtOH	N/A	AllPrep®	1	39.8	1.72	1.11	1.6	Lower ribosomal fragment is missing	36.2
3	Formalin → 70% IMS	Tungsten Carbide	AllPrep®	8	339.1	1.85	1.26	1.4	Lower ribosomal fragment is missing	35.2
4	Formalin → 70% IMS	Stainless Steel	AllPrep®	4	1116	1.96	-	-	Lower ribosomal fragment is missing	28.0
5	Formalin fixed → 70% EtOH	Tungsten Carbide	AllPrep®	4	7.2	1.89	0.768	1.8	Lower ribosomal fragment is missing	Undetermined
6	Formalin fixed → 70% EtOH	Stainless Steel	AllPrep®	4	28.6	1.82	1.20	1.5	Lower ribosomal fragment is missing	Undetermined
7	Prolonged formalin fixation → 70% EtOH	Stainless Steel	AllPrep®	4	41	1.78	5.67	1.5	Lower ribosomal fragment is missing	35.8

'-' indicates the sample was gave no reading.

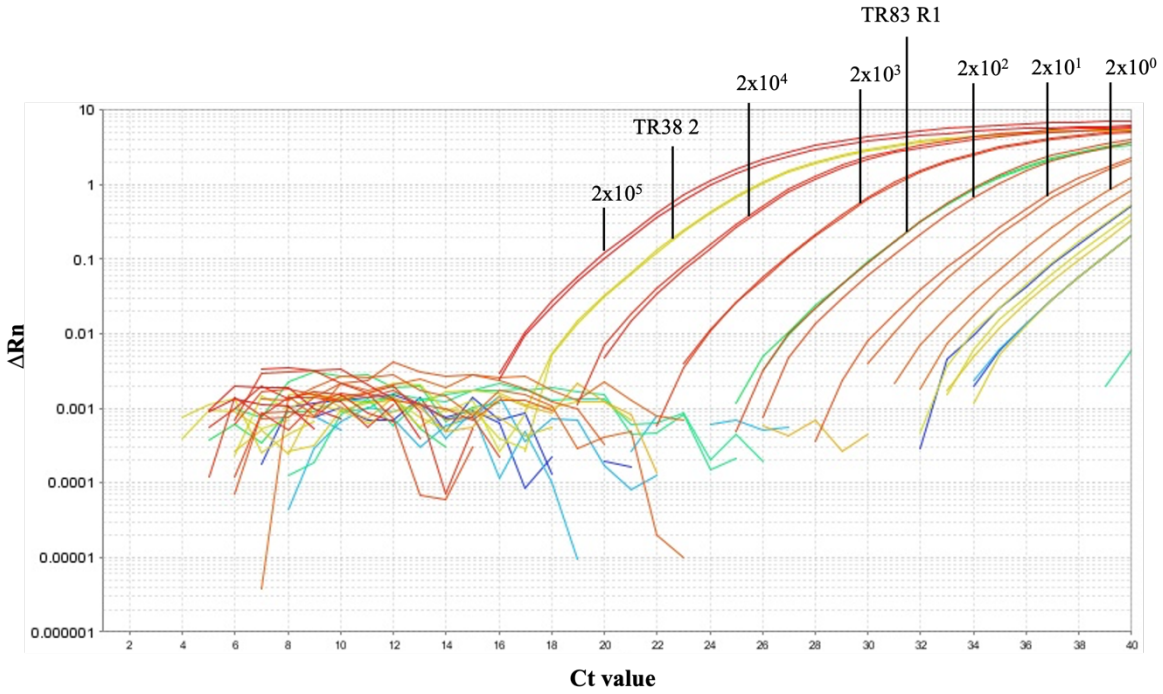


Figure 9. Beta-actin amplification in cDNA samples generated from formalin-fixed tonsil RNA. cDNA samples generated from tonsil RNA were assessed for the presence of beta-actin against a standard curve of known copy number, ranging from 200,000 to 2 beta-actin copies. The signal intensity (ΔRn) is plotted against the cycle number for each sample. cDNA samples (from TR38 2 and TR83 R1) amplified within the boundaries of the standard curve. All other cDNA samples displayed no amplification of beta-actin, or amplification was outside the lower boundary of the standard curve. There was no amplification in the H_2O control.

2.8. Discussion

10% formalin, most commonly NBF, is routinely used to fix tissues for histological analysis. Its frequent use has led to well-established protocols to ensure the appropriate use of formalin for this purpose. For histology, tissues are formalin-fixed and then embedded in paraffin wax in preparation for various staining. In this project, formalin was selected based on its ability to render potential SARS-CoV-2 inactive (Jureka et al., 2020, Kumar et al., 2021) during the coronavirus pandemic of 2019-onwards, to ensure that tonsil tissues collected from study participants were safe for use in this research. These specimens were not embedded in paraffin after fixation.

Prolonged periods of tissue fixation in NBF are known to cause significant DNA degradation (Ferrer et al., 2007). Indeed, because of formalin-induced DNA degradation, we were unable to detect housekeeping gene beta-actin in DNA isolated from tissues subjected to long periods of fixation. This was unsurprising given the reported positive correlation between the time spent in formalin and the extent of DNA fragmentation (Vitošević et al., 2018, Vitošević et al., 2023). Moreover, we found that RNA isolated from formalin-fixed tissue was severely degraded, likely due to nucleic acid fragmentation by formalin (Chung et al., 2008) and so, RNA was not used in this project.

When reducing the formalin fixation time to 24-hours, we improved DNA integrity immensely. However, formalin itself is a relatively slow acting fixative agent. It can penetrate tissue at a rate of approximately 1 mm per hour (Medawar, 1941). Healthy tonsil tissues are large tissues of an average mass of 1.5 g (Michaels, 1987) and so would not be entirely penetrated in a reduced formalin incubation. 70% EtOH also acts as an anti-viral agent and can deactivate SARS-CoV-2 (Kumar et al., 2021, Nomura et al., 2021). We found that when reducing the period of formalin fixation and transferring to 70% EtOH for long term storage, we were able to isolate DNA of better quality whilst still inactivating potential coronavirus.

2.8.1. Conclusion

Taken together, this work unfortunately suggests that prolonged periods of fixation in formalin is not compatible with nucleic acid being used for molecular analyses. We found that a formalin-fixation period as low as 24 hours, followed by a transfer to 70% EtOH, yields the best quantity and quality of extracted DNA whilst ensuring inactivation of SARS-CoV-2, and

describe a new protocol optimised to isolate DNA from formalin-fixed tonsil tissue, which is detailed in Chapter 3 section 3.5.2.

Finally, in March 2023, we were able to resume the collection of fresh-frozen tissue, following the change in the status of the ongoing SARS-CoV-2 pandemic, and subject to significant revision to our risk assessment (IRAS amendment 5.0). At this point, hospitals were able to continue to collect fixed tissues, although we had stated our preference was for fresh-frozen, but many of the study sites preferred the convenience of fixed tissues. Consequently, in April 2024, we removed the option to collect fixed tissue from the study protocol and stated that only fresh-frozen tissue could be collected for the study (IRAS amendment 5.2).

CHAPTER 3: Final Research Methodology

3.1. Meta-analysis of oral HPV prevalence in the healthy adult population

A meta-analysis was performed to determine oral HPV prevalence in the healthy adult population across reports in the literature. Our meta-analysis was registered to PROSPERO with ID: CRD42024534499 (https://www.crd.york.ac.uk/prospero/).

3.1.1. Literature searching and selection of papers for analysis

To investigate the reports of oral HPV prevalence within the healthy adult population, we searched Embase, Ovid MEDLINE, and PubMed databases with the following search terms for each: Embase and Ovid MEDLINE: ((HPV OR Human Papillomavirus OR papillomaviridae OR papillomavirus OR alphapapillomavirus OR alphapapillomavirus OR papilloma) AND (oral OR mouth) AND (prevalence OR epidemiology)), and PubMed: (((HPV) OR (Human Papillomavirus)) OR (papillomaviridae)) OR (papillomavirus)) OR (alphapapillomavirus)) OR (papillloma)) OR (HPV[Other Term])) OR (Human papillomavirus[MeSH Terms]) OR (papillomaviridae[MeSH Terms]) OR (papillomavirus[MeSH Terms]) OR (alphapapillomavirus[MeSH Terms]) OR (papilloma[MeSH Terms])) AND ((oral) OR (mouth)) AND ((prevalence) OR (epidemiology) OR (epidemiology[MeSH Terms]))). All hits generated through database searching were subjected to an initial title review by two independent researchers Mary-Anne Freckleton (MAF) and Elizabeth Marsh (EM) based on the inclusion and exclusion criteria of the meta-analysis.

The inclusion criteria were primary research papers describing the frequency or prevalence of any of the oral human papillomavirus subtypes, specifically:

- Individuals over 18 years old.
- Individuals must have no oral HPV-related pathology.
- Individuals without head and neck cancer.
- For HPV detection, must have used either PCR based methods, hybrid capture 2, or sequencing techniques including NGS.
- Samples must be collected from the oral cavity/oropharynx by: oral rinse/gargle, oral swab, oral/tonsillar brush, saliva samples, and/or tonsillectomy.
- Have over 50 individuals in the study sample.

- Primary literature.
- Samples must be collected in 1995 or later.
- Provided detailed methodological description of oral HPV sampling techniques.

The exclusion criteria for this meta-analysis were as follows:

- Studies where individuals under 18 cannot be removed from the cohort.

So as not to over- or under-estimate oral prevalence, studies conducted **exclusively** in the following populations were also excluded:

- Individuals under 18 years old.
- People who have been vaccinated against HPV.
- People not vaccinated against HPV.
- Men who have sex with men (MSM)
- HIV-positive individuals.
- Individuals with, and partners of individuals with, HPV related pathology.
- HPV-positive individuals (cervical infection) and their partners.
- Drug users.
- Sex workers.
- Studies looking at beta and gamma HPV prevalence alone.
- Individuals with clinically relevant risk factors (including oral lichen planus, Fanconi anaemia, oral lesions).

3.1.1.1. Study abstract and full-text screening

All reports still included following title screening were uploaded to Rayyan – an online software designed to assist with meta-analysis/systematic reviewing of papers (Mourad Ouzzani, 2016). Again, two independent researchers (MAF and EM) reviewed all abstracts and categorised them as 'include', 'maybe', or 'exclude'. The screening was done with 'blind mode' switched on to anonymise each reviewer's response to the other. Any conflicts between the two reviewers were discussed with a third individual Laurice Fretwell (LF) or Thomas Illingworth (TI) until a unanimous decision was reached.

All reports within the 'maybe' list were subjected to a full-text review and included/excluded accordingly; where there was ambiguity, these were discussed within the team.

3.1.2. Data extraction

Data were extracted from all included studies by two researchers (MAF and EM) independently. The data extracted from each paper were the country of origin, study period, age (mean, median, range), sex, study population or setting, sample collection method/s, HPV detection method, viral types included in analyses, and selection of study population. For HPV prevalence, data extracted included: total sample size, total HPV-positive, total HPV-negative, overall HPV prevalence, prevalence of HR-HPV types, overall HPV-16 prevalence, total females, overall female HPV prevalence, female HPV-16 prevalence, total males, overall male HPV prevalence, and male HPV-16 prevalence. In longitudinal studies that reported HPV prevalence at multiple time points, we extracted the baseline data in all cases.

If the study population was described in more than one publication, the paper that reported the greatest sample size was included (or the original study if this was the same). Where there was more than one sampling method used (e.g., swab and rinse), if this was not already combined, we included the method that gave the highest prevalence. Data were extracted from alpha, beta, and gamma HPV types for the overall HPV prevalence; beta and gamma types were only excluded if the paper was exclusively investigating the prevalence of these. For data extraction of HR-HPV prevalence, the viral types included differed dependent on the study. If a study only reported on particular viral types (for example, HPV-16 and -18), these data were included for the study overall HR-HPV prevalence. When a study provided a pre-calculated HR-HPV prevalence, this value was extracted despite no knowledge of the viral types included. If a paper did not give a HR-HPV prevalence, this was manually calculated from data on the prevalence of HPV-16, -18, -31, -33, -34, -35, -39, -45. -51, -52, -56, -58, -59, -66, -68, and -70 (Burd, 2003). During data extraction, studies that then appeared unsuitable based on our criteria were excluded. All included studies were assessed for their risk of bias for prevalence studies using tool guidelines as described by Hoy *et al*., 2012 (Hoy et al., 2012) by MAF, EM, and LF.

Tables for each statistical analysis were prepared in Microsoft Excel for the following: Overall HPV prevalence, overall HR-HPV prevalence, overall HPV-16 prevalence, overall male HPV prevalence, overall male HPV-16 prevalence, overall women HPV prevalence, and overall women HPV-16 prevalence.

3.1.3. Statistical analysis of oral HPV prevalence

All statistical analyses were carried out by James Hind (JH) and NTU as follows; for statistical analysis of oral HPV prevalence in each of the comparative groups, data was weighted by

inverse variance for analysis; to do this, data was initially transformed using the Freeman Tukey double arcsine formula to stabilise the variance based on a high number of zeros reported by individual studies. Data was then back transformed to percentages; but use of the harmonic mean in the back transform was difficult (Schwarzer et al., 2019). Finally, a combined inverse variance weighted analysis of the prevalence for each group (prevalence of total HPV, total HR-HPV, total HPV-16, total male, total female, male HPV-16, and female HPV-16) was performed using Microsoft Excel and Minitab Statistical Software (version 21.4). The pooled, transformed variance was used as a substitute for the harmonic mean (Barendregt et al., 2013).

3.2. NIHR-portfolio HPV clinical study

This project is part of an NIHR-portfolio registered clinical study (REC reference: 16/SW/0079, IRAS project ID: 194857). Patient recruitment was carried out by GCP-trained clinicians and researchers working on this project at the following hospitals:

- University Hospitals of Derby and Burton (UHDB)
- University Hospital Leicester (UHL)
- University Hospital Nottingham (UHN)
- Princess Royal Hospital, Telford (PRH)
- Calderdale Royal Hospital (CRH)

3.2.1. Recruitment to the study of routine tonsillectomy patients

Patients undergoing routine tonsillectomy surgery at participating hospitals were invited to take part in this clinical study. The inclusion criteria were all adult individuals (over 18 years of age) undergoing routine tonsillectomy surgery. Potential candidates were to be excluded if their tissues were cancerous, this was analysed by a consultant histopathologist at the hospital of collection who confirmed tissues to be non-cancerous before use in this research. Eligible individuals were approached in advance of their surgery and were provided with a letter of invitation and Patient Information Sheet (PIS) (Appendix A; Figures 33-34); then following invitation to participate, potential candidates were given time to make an informed decision to consent and were provided with relevant contact details for any questions regarding the study. All persons were made aware of how to withdraw from the study, and informed that they would not be provided with any results from laboratory analyses. As yet, participants have been given no official debriefing, but subjects were notified of the ClinicalTrials.gov identifier number.

Upon consenting to the study, a unique identifier number (UIN) was generated by participants to enable to use of their tissues and questionnaire data pseudo-anonymously throughout this project. Samples were anonymised by the clinical teams.

3.2.2. Sample collection and specimen storage/fixation requirements

Consenting individuals to the study were required to provide both their left and right non-cancerous palatine tonsil tissues, a tonsillar swab, and a lifestyle questionnaire. The lifestyle questionnaire was completed by participants before their surgery either online, or a paper version was provided.

Tissues were stored differently dependent on their date of collection; before 2020 tissues were frozen at -80°C immediately after surgery. Whereas tonsil tissues collected following the recommencement of routine surgeries (during the Coronavirus pandemic) were fixed in 10% NBF at the point of removal, following an amendment to the sample collection protocol (IRAS 194657 amendment 3.1, July 2021) to render any potential SARS-CoV-2 inactive. In March 2023 IRAS amendment 5.0 was confirmed which meant that tissues were fixed in 10% NBF at the point of removal, then transferred to 70% EtOH 24-96 hours post fixation. Tonsils collected after April 2024 were frozen immediately after surgery (IRAS amendment 5.2). Brushings of the tonsil tissue surface were taken under general anaesthesia by the operating clinician from all participants using a new sterile swab for each individual; swabs were also fixed in 10% NBF during surgery after July 2021, and frozen at -80°C after collection in April 2024. The tonsil swabs are stored for future research but did not contribute to this project.

Regardless of storage requirements determined by Coronavirus guidelines, all non-cancerous tonsil tissues and lifestyle questionnaires used in this project were collected between 2016 to 2024.

3.2.3. Demographics and lifestyle characteristics of participants

The lifestyle questionnaire was designed to provide information on the cohort demographics (including their age, gender, ethnicity, and relationship status). There were also questions to provide information on each individual's participation/frequency of participation in activities known to impact the risk of oral HPV infection (Appendix A; Figure 36). Data from questionnaires was uploaded to Microsoft Excel and any charts to represent the distribution of

demographic characteristics of the study sample were produced in this software. There has been no statistical analysis of these data.

3.3. University of Birmingham Human Biomaterials Resource Centre (HBRC) tissues

Twenty non-cancerous palatine fresh-frozen tonsil tissues (from 10 people) were purchased from the University of Birmingham's HBRC in 2021 (Appendix C; Figure 39) to facilitate protocol optimisation in the interim months when no routine surgeries were scheduled. Tissues were stored at -80°C upon arrival at the University. We did not know the tissue orientation (left or right) of each pair, or receive any information of the person's demographics.

3.4. Sectioning of non-cancerous palatine tonsil tissue

Tonsil tissues were already sectioned when received at the University; 2-4 sections had been taken at the collecting hospital for histological analysis to confirm tissues as non-malignant.

Twenty-three participants were recruited from UHDB to the clinical study, and samples collected, prior to the start of this project. These were sectioned for a separate DNA, RNA, and protein extraction by other researchers within the team upon receipt to the University laboratories.

3.4.1. 10% NBF fixed palatine tonsils

10% NBF fixed tissues were sectioned for a simultaneous DNA/RNA extraction, protein isolation, histological analysis, and potential future RNA extractions to give the best representation of the entire tonsil for each analysis. Each portion of tissue (as left by the hospital) was divided into 4 relatively even pieces using a petri dish, scalpel and forceps. These were all sterile and replaced in-between tonsils (including the left and right of the same donor) to prevent cross-contamination. The 4 pieces of tissue were divided evenly for each downstream analysis (Figure 10). If the number of tissue sections was greater than the 4 shown in Figure 10, the pattern for distributing the tissue was repeated from number 1 onwards until all tissue was sectioned. Tissue segments were immersed in 70% EtOH (70% absolute EtOH [Honeywell International Inc., North Carolina, USA], 30% nuclease free water [National Diagnostics, Georgia, USA]), and stored at RT, until further processing. Tonsil tissue saved for RNA work in future research was stored at -20°C in ~500 µL 70% EtOH.

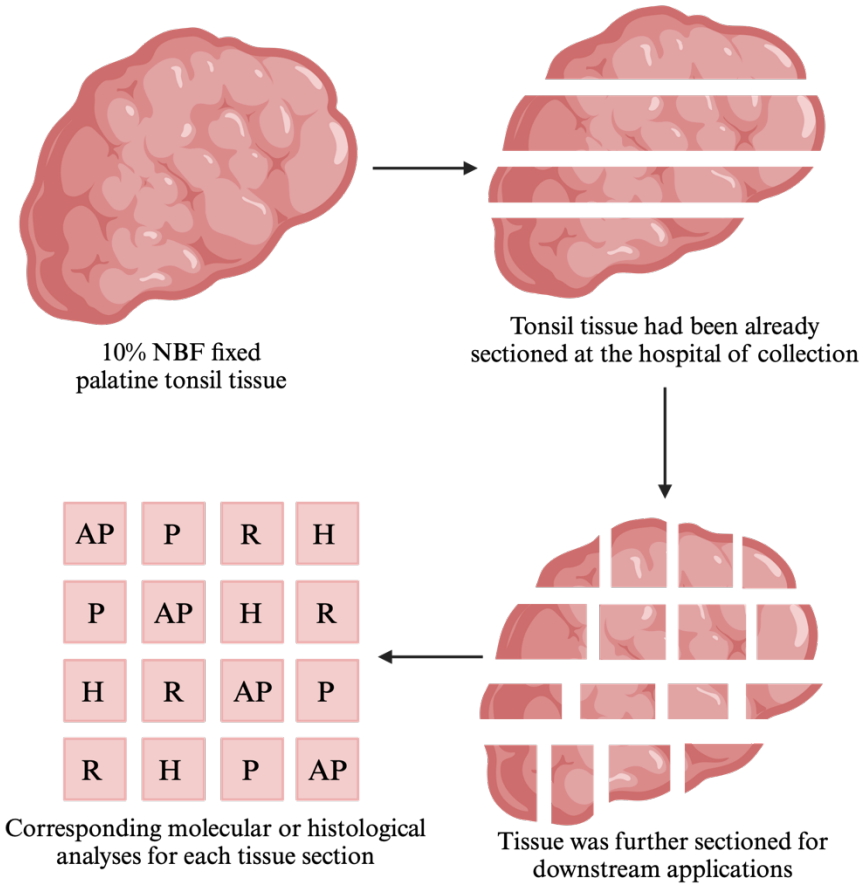


Figure 10. Sectioning of 10% NBF fixed palatine tonsil tissues for DNA, RNA, and protein extraction, and histological analyses. The tonsil tissues arrived already sectioned following histological analysis at the hospital of collection (left of arrow). The number of sections received depended on the size of the tissue specimen. A diagrammatic example of how a tonsil tissue of 4 sections was further divided for the following: an AllPrep® DNA/RNA FFPE kit extraction (AP), 8 M urea protein isolation (P), future RNA work (R), and histological analyses (H), to give the most accurate representation across the whole tonsil for all downstream analyses.

Image was created using BioRender.com.

3.4.2. Fresh-frozen tonsil tissue

Tonsil tissue specimens frozen at the point of removal were transferred from the hospital of collection to the University laboratory on dry ice and stored at -80°C until processing (different tissue storage conditions (i.e., 10% NBF fixed or fresh-frozen) are discussed at Chapter 2). ~50 mg sections of tonsil tissue were taken in a petri dish using a sterile scalpel and sterile forceps within a fume hood to isolate the user from potential harmful aerosols. All cutting apparatus was changed between each tonsil tissue to prevent cross-contamination. After ~50 mg tissue sections were taken (with total sections ranging from 2-4 per tissue), the remaining tissue was immediately returned to -80°C to prevent degradation of nucleic acid. ~50 mg tissue sections were then weighed to within ±5 mg of the ideal mass. Each section was further cut into finer fragments of tissue to aid homogenisation and immediately immersed in buffer for homogenisation (described at 3.6.1). Tissues from the HBRC were sectioned following this same sterile technique; however, as these were generally used for protocol optimisation they were sectioned at various masses as and when required, not following this standardised protocol.

3.5. DNA and RNA extraction from formalin-fixed tonsil tissues

RNA and DNA extractions from 10% NBF fixed tonsil tissue were performed using the AllPrep® DNA/RNA FFPE kit [Qiagen, Venlo, Netherlands]. This kit is designed for nucleic acid extraction from FFPE sections. The deparaffinisation steps outlined in the manufacturer's protocol were unnecessary for our sample type (formalin-fixed, not paraffin embedded). Following advice from Qiagen, the deparaffinisation steps were missed and the protocol was followed from step 5 (Qiagen, 2020) starting with tissue homogenisation steps in the lysis buffer provided in the kit.

3.5.1. Homogenisation and lysis of 10% NBF fixed tonsil tissues

Tonsil tissue sections for the AllPrep® DNA/RNA FFPE kit [Qiagen, Venlo, Netherlands] extraction were weighed to a total of ~100 mg (range ~80-110 mg) tissue per spin column. The number of spin columns used for each tonsil tissue varied dependent on the overall mass of the entire tonsil. Each individual ~100 mg portion of tissue was cut into fine pieces using a scalpel and was immersed in 500 µL 1X PBS [Sigma-Aldrich, Massachusetts, USA] for a 5-minute RT incubation before all 1X PBS was aspirated. This was repeated for a total of two washes. 150 µL Buffer PKD and 10 µL Proteinase K [both AllPrep® kit, Qiagen, Venlo, Netherlands]

was added to each sample and vortexed. Tissue was then homogenised in the suspension using the TissueLyser LT instrument [Qiagen, Venlo, Netherlands] for 10 minutes at 50 Hz with 1x 5 mm stainless steel bead [Qiagen, Venlo, Netherlands].

3.5.2. AllPrep® DNA/RNA FFPE kit [Qiagen] for extraction of DNA from 10% NBF fixed tonsil tissue

RNA and subsequent DNA extraction from lysed tonsil tissue was performed using the AllPrep® DNA/RNA FFPE Kit [Qiagen, Venlo, Netherlands] following manufacturer's protocol, with minor modifications (Qiagen, 2020). All buffers described were included in the kit and were prepared as per the protocol instructions.

In brief, the protocol was as follows: homogenised tissue was incubated at 56°C for 20 minutes before 3 minutes on wet ice, and a then a 15 minute centrifuge (17,200 x g) to pellet large debris, as formalin-fixed tissue does not visibly homogenise, and tissue was still present following lysis. Supernatant for RNA isolation was removed away from the remaining tissue. Care was taken to not aspirate any tissue/fatty material. Described as a 'pellet' in the manufacturer's protocol, these tonsil tissue remnants were stored either overnight at 4°C, or at -20°C for any longer periods in preparation for later DNA extraction.

RNA isolation was then carried out and the supernatant containing RNA was incubated at 80°C for 15 minutes to reverse cross-linking inflicted by formalin fixation. The protocol was then followed exactly: 320 µL Buffer RLT was added to supernatant and vortexed to mix. The entire RNA sample was then bound to the spin column membrane, with 700 µL of the sample passed through the column at a time by centrifugation. The speed and time used throughout all spin steps varied dependent on the viscosity of the RNA sample. 8,000 x g for 15 seconds was used as a starting point (unless max. speed indicated), but often increased to up to 15,000 x g for 1 minute to promote the viscous liquid moving through the column. 350 µL Buffer FRN was then passed through the column by centrifugation. For DNase treatment of RNA, 10 µL DNase I and 70 µL Buffer RDD was added directly to the spin column membrane. Samples were subjected to a 15 minute RT incubation, before addition of 500 µL Buffer FRN, and immediately spun down to elute. At this stage, eluate (containing RNA and small RNAs) was not discarded, and was added directly back onto the membrane. Eluate was then centrifuged through the column and flow-through discarded. Two sequential washes of the membrane in

500 µL Buffer RPE were performed, with eluate discarded after each spin. An additional 5 minute spin followed to dry the spin column membrane (max. speed). 20 µL RNase-free water was added to each column and incubated at RT for 5 minutes. RNA was eluted (max. speed spin, 3 minutes), whereupon eluate was immediately frozen at -20°C.

For DNA extraction, 180 µL Buffer ATL and 40 µL Proteinase K was added to the tissue /DNA 'pellet' saved before RNA extraction and vortexed. The tissue was incubated at 56°C for 1 hour and then 90°C for 2 hours to remove cross-linking/modification to DNA resulting from formalin fixation. After incubation, samples were spun down briefly and 200 µL Buffer AL, then 200 µL absolute EtOH [Honeywell International Inc., North Carolina, USA] was added and vortexed after each addition. As connective tonsil tissue remained in most cases, care was taken to not aspirate tissue when removing supernatant to avoid blocking the spin column. Three sequential washes were performed by addition and centrifugation through the spin column of 700 µL Buffer AW1, then 700 µL Buffer AW2, and finally 700 µL absolute EtOH. The flow-through was discarded after each wash. As with RNA isolation, the centrifugation speeds used were dependent on sample viscosity. 8,000 x g was used as a starting point (unless max. speed stated) and increased until all solution had passed through the column. An additional 5 minute spin followed to dry the spin column membrane (max. speed). 75 µL Buffer ATE was added to each spin column and incubated at RT for 5 minutes, then centrifuged at max. speed for 3 minutes to elute DNA. This elution step was repeated to give a total elution volume of 150 µL in Buffer ATE. DNA was immediately stored at -20°C until further processing.

All centrifugation steps in the protocol were performed at RT using a Prism™ R refrigerated MicroCentrifuge [Labnet International, New Jersey, USA]. As this equipment has a maximum relative centrifugal force (rcf) of 17,200 x g, any centrifugation steps to be performed at 20,000 x g in the manufacturer's protocol were done at the machine's own maximum spin speed (17,200 x g).

3.6. DNA isolation from fresh-frozen tonsil tissue

DNA extractions from the 23 tonsil pairs collected prior to this project were either performed using the DNeasy® Blood & Tissue Kit [Qiagen, Venlo, Netherlands] as described at 3.6.2., or using a phenol:chloroform method, by previous researchers in the lab. All fresh-frozen

tissues collected during this project were subjected to extraction using the DNeasy® kit. All DNA was stored at -80°C until use.

3.6.1. Homogenisation and lysis of fresh-frozen tonsil tissue

Minced tissue sections were placed in a pre-cooled 2 mL screw top tube containing 1x 5 mm stainless steel bead [Qiagen, Venlo, Netherlands]. 180 µL Buffer ATL [DNeasy® Blood & Tissue Kit, Qiagen, Venlo, Netherlands] was added and tissue was subjected to homogenisation (2 minutes, 30 Hz) using the TissueLyser LT instrument [Qiagen, Venlo, Netherlands]. After 2 minutes, the 12 hole homogeniser insert was spun by 90° to alter the orientation of the samples. Samples were further homogenised for 2 minutes, 50 Hz, and then placed on ice.

3.6.2. DNeasy® Blood & Tissue Kit [Qiagen] for DNA extraction

For DNA isolation from fresh-frozen tonsil tissue, the DNeasy® Blood & Tissue Kit [Qiagen, Venlo, Netherlands] was used. Unless otherwise stated, all buffers used in this protocol were provided as part of the extraction kit. Homogenised samples were centrifuged (1 minute, 17,200 x g, RT) and 40 µL Proteinase K was added immediately after. Tubes were incubated at 56°C overnight to promote tissue lysis.

DNA isolation was then performed generally following manufacturer's protocol for total DNA isolation from animal tissues (Qiagen, 2023), with minor modifications. In short, samples were vigorously vortexed for 15 seconds. Then, 200 µL Buffer AL and 200 µL absolute ethanol [Honeywell International Inc., North Carolina, USA] were added and immediately vortexed. The mixture was then added to a DNeasy® spin column; however, the liquid was often extremely viscous and so was vortexed during transfer to the spin column to aid pipetting. Care was taken to not transfer any tissue remnants/connective tissue not lysed onto the spin column membrane. DNA was then bound to the column by centrifugation (1 minute, 17,200 x g, RT) and subjected to 2 washes by centrifugation of 500 µL Buffer AW1, then 500 µL Buffer AW2 through the spin column (1 minute, 15,000 x g, RT). Eluate was discarded after each wash. An additional spin followed to dry the spin column membrane (3 minutes, 17,200 x g, RT). 200 µL Buffer AE was added to each spin column and incubated at RT for 5 minutes, then centrifuged (3 minutes, 17, 200 x g, RT) to elute DNA. All 200 µL eluate was pipetted back onto the membrane for another 5 minute RT incubation before a final centrifugation (3 minutes, 17, 200 x g, RT). Eluted DNA was immediately stored at -20°C until use.

3.7. DNA quantity and quality analysis and dilution for PCR assays

DNA concentration, 260:280, and 260:230 ratios were analysed with the NanoDrop™ 2000 Spectrophotometer [ThermoFisher Scientific, Massachusetts, USA], using the corresponding software (NanoDrop™ 2000 Software Version 1.6.198). All DNA samples with a concentration >400 ng/μL were diluted 1:10 in either Buffer ATE for DNA from formalin-fixed tissues [AllPrep® kit, Qiagen, Venlo, Netherlands] or Buffer AE [DNeasy Blood & Tissue Kit, Qiagen, Venlo, Netherlands] for DNA of fresh-frozen tissue. All DNA samples were produced to a final concentration of 25 ng/μL for beta-actin and HPV-DNA qPCR analysis. Stocks were produced in Buffer ATE [Qiagen, Venlo, Netherlands] for 10% NBF fixed tissue and in Buffer AE [Qiagen, Venlo, Netherlands] for fresh-frozen tissue.

3.8. Beta-actin housekeeping gene PCR for DNA integrity analysis

To confirm viable DNA, tonsillar gDNA samples were all screened for the presence of housekeeping gene, beta-actin, using either an endpoint standard PCR or a qPCR approach.

3.8.1. Beta-actin TaqMan Real-Time PCR screening

All reactions for beta-actin (ACTB) qPCR screening were set up as follows: 1X Precision®FAST master mix (with ROX) [Primerdesign, Hampshire, UK], ACTB primer/probe mix (300 nM/100 nM) [Primerdesign, Hampshire, UK], 25, 50, or 75 ng total gDNA, and sterile DEPC-treated water [National Diagnostics, Georgia, USA] to a total volume of 20 μL. The amount of gDNA added to each reaction was dependent on the tonsil tissue of origin: 25 ng gDNA for fresh-frozen tonsil tissue, 50 ng gDNA for 10% NBF fixed tissue transferred to 70% EtOH 24-96 hours post-surgery, and 75 ng gDNA for tissues fixed in 10% NBF for prolonged periods (months). Tonsil gDNA samples were assayed against a standard curve of known beta-actin copy number, produced following manufacturer's protocol (Primer Design, 2017). A top standard from the kit of $2x10^5$ beta-actin copies was diluted 1:10 in series, with the lowest copy number $2x10^0$. All reactions were thermocycled for 40 cycles of an enzyme activation step for 2 minutes at 95°C, DNA denaturation for 5 seconds at 95°C and a 20 second extension step at 60°C.

All tonsil gDNA samples were ran in duplicate and DNA that showed amplification of beta-actin within the boundaries of the standard curve of known beta-actin copy number in both replicates were considered to have viable gDNA. These gDNA samples were considered to

have passed the quality control screen and were included in the cohort for subsequent HPV-DNA qPCR consensus screening. Any gDNA samples where only 1 out of 2 replicates showed PCR amplification of beta-actin were repeated. When optimising DNA extractions (Chapter 2) the standard curve of known copy number was used to calculate an estimated cell number per sample. An example of a standard plate set-up for qPCR beta-actin screening is shown in Figure 11.

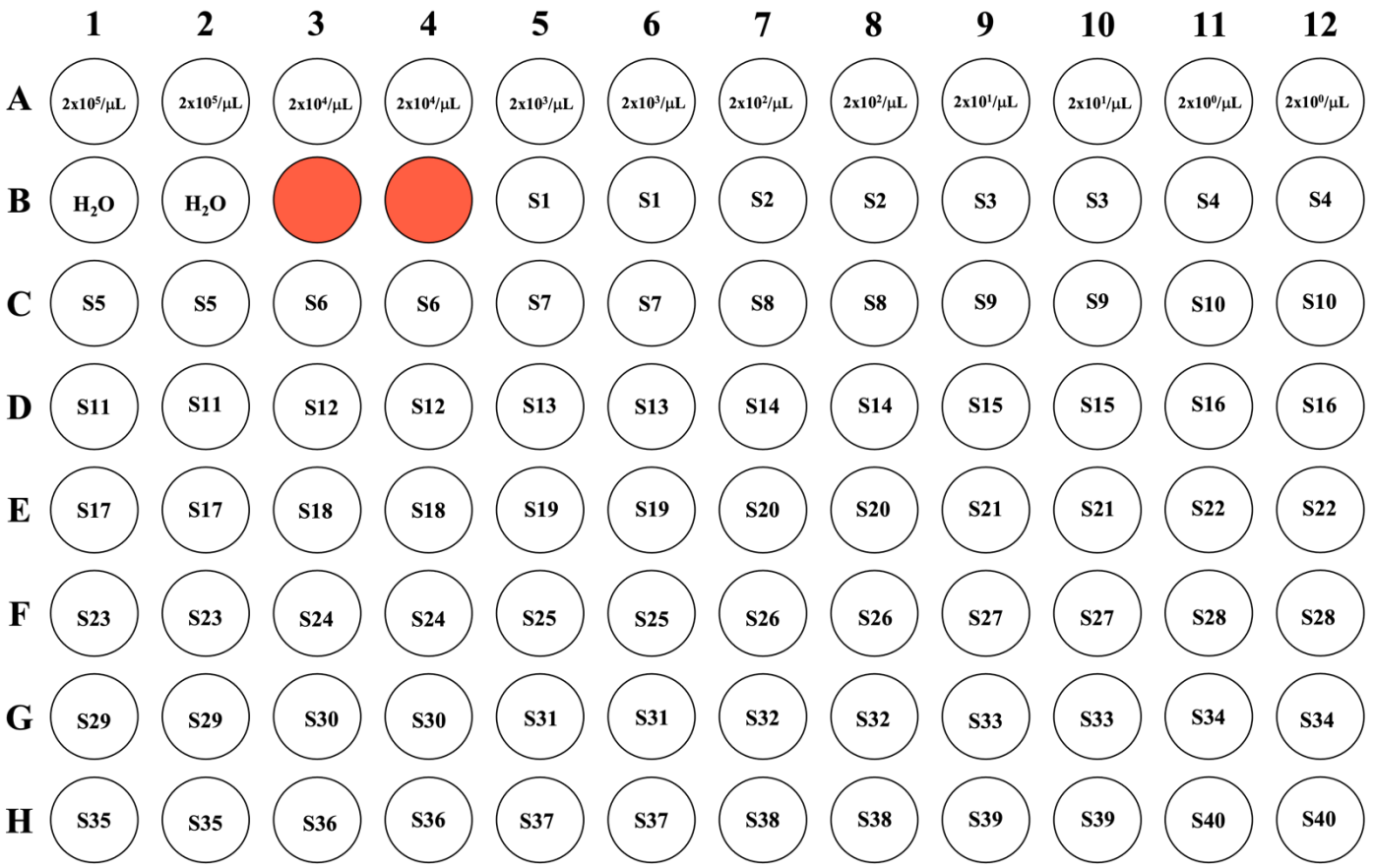


Figure 11. Standard layout of a 96-well plate for beta-actin qPCR amplification analysis of non-cancerous tonsil gDNA. Row A was used for the standard curve of known beta-actin copies, that ranged from $2x10^5$ copies per µL to $2x10^0$ copies per µL. All tonsil gDNA samples were run in duplicate, with the final gDNA concentration dependent on the storage conditions of the tissue that DNA had originated from (i.e., fresh-frozen, formalin-fixed transferred to 70% EtOH 24-96 hours post-surgery, or tissue fixed in 10% NBF for prolonged periods (months)).

3.8.2. Beta-actin amplification analysis by standard endpoint PCR

Due to cost limitations, several tonsillar gDNA were screened for viable DNA using a standard PCR assay for beta-actin detection. In brief, all PCR reactions were produced to a total volume of 20 µL, consisting of 1X Platinum™ II Hot-Start Green PCR Master Mix [ThermoFisher Scientific, Massachusetts, USA], 200 nM ACTB forward primer, 200 nM ACTB reverse primer, and a final concentration of gDNA added dependent on tonsil tissue of origin (outlined at 3.8.1). The total volume was made up with sterile DEPC-treated water [National Diagnostics, Georgia, USA]. All reactions were thermocycled for an initial denaturation of 2 minutes at 94°C, then 35 cycles of 15 seconds at 94°C, 15 seconds at 60°C, and 15 seconds at 68°C. This was followed by a hold stage at 4°C. Beta-actin primer sequences used are shown in Table 7.

3.9. Gel electrophoresis of PCR amplimers

Any PCR products analysed by gel electrophoresis were separated on an agarose gel (agarose [Sigma-Aldrich, Massachusetts, USA] diluted in 1X TAE buffer [National Diagnostics, Georgia, USA] with 1X GelRed® Nucleic Acid Stain in DMSO [Biotium, California, USA]), of an appropriate percentage for the expected amplicon size.

PCR products were prepared in 1X Orange DNA Loading Dye [ThermoFisher Scientific, Massachusetts, USA]. Samples were electrophoresed at 100 V for ~1 hour (or until complete separation). 5 µL HyperLadder™ 25 bp DNA ladder [Bioline, Meridian Bioscience Inc., Ohio, USA] was run on all gels to determine amplicon size determination. For beta-actin PCR, successful amplification was confirmed by presence of a DNA band at ~166 bp. For PCR products analysed after HPV-DNA qPCR consensus screening (described at 3.10.), amplification of HPV L1 was confirmed by the presence of a DNA band at ~150 bp.

Following electrophoresis, agarose gels were imaged with either a Chemi-Doc MP Imaging System [Bio-Rad, California, USA] or a Geneflash Bio Imaging Gel Doc System [Syngene International Ltd, Bangalore, India], as indicated in each figure.

3.10. HPV-DNA qPCR consensus screening

Tonsil gDNA samples that had passed quality control checks were subjected to HPV-DNA analysis using our modified HPV-DNA qPCR consensus assay (Appendix D). GP5+/GP6+

consensus primers that amplify a 150 bp region of the HPV L1 gene were used in this assay (de Roda Husman et al., 1995); primer sequences are shown in Table 7. HPV-positivity was assessed through melt curve analysis of PCR products, based on the negative derivative reporter value (-Rn). All gDNA samples were screened in triplicate and were defined as HPV-positive if at least 2/3 replicates had peaks in their melt curve analysis suggesting HPV-DNA presence. If 1/3 replicates had a peak of HPV L1 amplification in their melt curve plot, that sample was repeated. DNA samples with no replicates or 1/3 replicates (on repeat) having an L1 peak were considered HPV-negative. All definite HPV-positive samples and representative samples of those with inconsistent peaks in their melt curve analysis were taken forward to HPV genotyping (3.11) or further analysis by gel electrophoresis. An example of a standard plate set up for HPV-DNA qPCR consensus screening is shown in Figure 12.

All PCR reactions were set up to a total volume of 20 µL, consisting of 1X PowerTrack™ SYBR Green master mix [ThermoFisher Scientific, Massachusetts, USA], 200 nM GP5+ forward primer, 200 nM GP6+ reverse primer, 50, 100, or 150 ng gDNA, and sterile DEPC-treated water [National Diagnostics, Georgia, USA] to the total volume. The amount of gDNA added to the reaction was based on the tissue of origin, and was optimised to be double that required for beta-actin amplification (fresh-frozen tissue = 50 ng, 10% NBF transferred to 70% EtOH 24-96 hours post-surgery = 100 ng, 10% NBF fixed tissue for prolonged periods (several months) = 150 ng). 50 ng total gDNA was added for HPV controls (HPV-18 pGEMII-WT plasmid (kind gift from Dr Sally Roberts, University of Birmingham) and, HPV-16 positive SiHa DNA (section 3.14). All reactions were thermocycled for an initial denaturation of 3 minutes at 95°C, then 40 cycles of 20 seconds at 95°C, 30 seconds at 52°C and then an extension step at 72°C for 10 seconds. A melt curve stage followed: 95°C - 15 seconds, 60°C – 1 minute increasing up to 95°C, with data collected every 0.3°C.

3.10.1. Statistical analysis of qPCR data

Statistical comparisons of HPV-DNA qPCR prevalence estimates based on gender were performed using GraphPad Prism Software [version 10, Massachusetts, USA]. A two-sided Fisher's exact test was selected using Koopman asymptomatic score for confidence interval calculations.

Table 7. Oligonucleotides for PCR assays. The sequences and expected amplicon size of primers used in beta-actin screening by standard PCR (ACTB F and R) and consensus primers used in our modified HPV-DNA qPCR assay (GP5+ F and GP6+ R).

Primer name	Sequence (5' → 3')	Expected amplicon size (bp)
ACTB (F)	GCC CGG CTA TTC TCG CAG	166
ACTB (R)	ACC CAG CCA GCT CCC CTA	166
GP5+ (F)	TTT GTT ACT GTG GTA GAT ACT AC	150
GP6+ (R)	GAA AAA TAA ACT GTA AAT CAT ATT C	150

(F) = forward primer, (R) = reverse primer

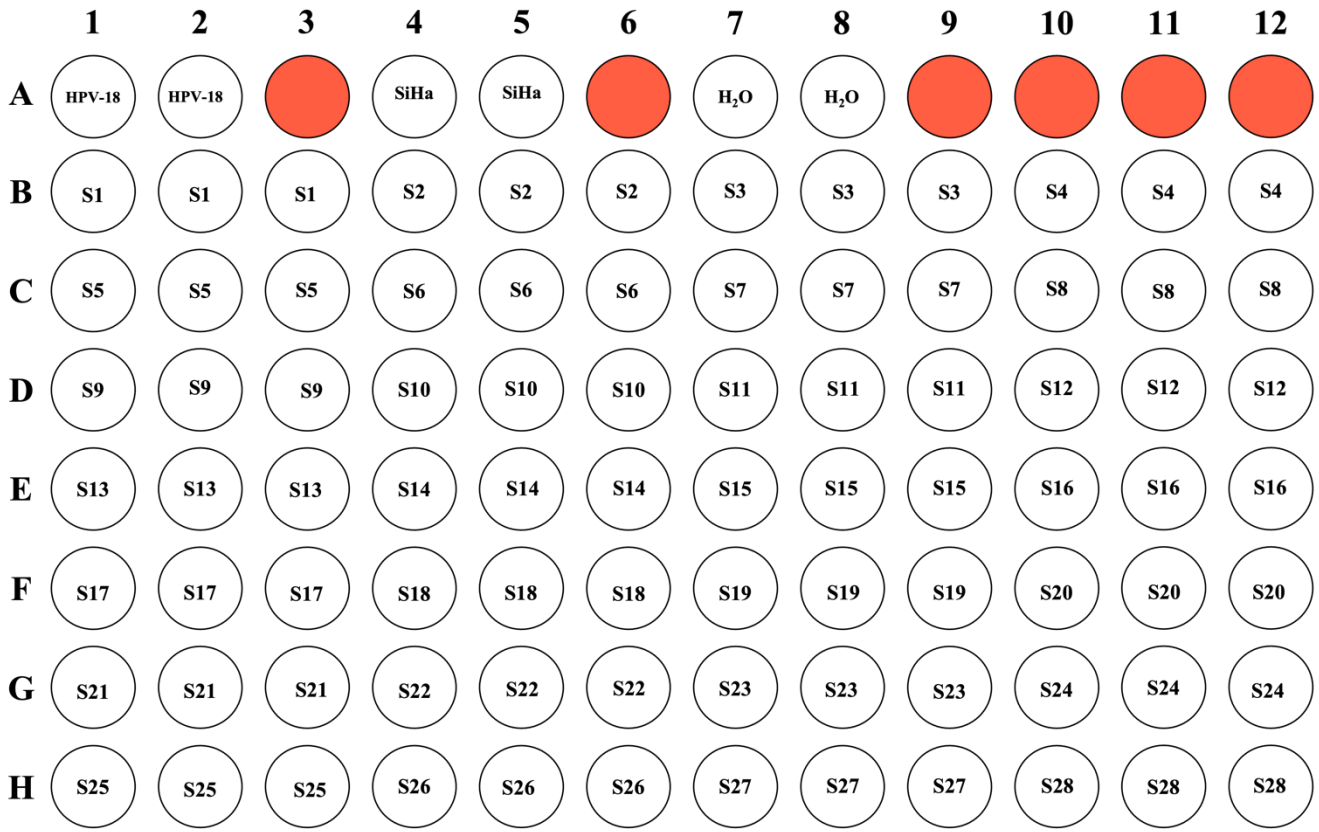


Figure 12. Standard layout of a 96-well plate for HPV-DNA qPCR consensus screening. Row A was used for HPV and negative controls (HPV-18 = HPV-18 pGEMII-WT plasmid, SiHa = HPV-16 positive SiHa DNA, H_2O for negative control). All tonsil gDNA samples were run in triplicate, with the final gDNA concentration dependent on the storage conditions of the tissue that DNA had originated from (i.e., fresh-frozen, formalin-fixed transferred to 70% EtOH 24-96 hours post-surgery, or tissue fixed in 10% NBF for prolonged periods (months)). 28 samples were run in each 96 well plate.

3.11. HPV genotyping by Sanger sequencing

Sanger sequencing was used to sequence PCR products during HPV-DNA qPCR assay optimisation (Appendix D). Samples were electrophoresed on 2% agarose gels (see 3.9). DNA bands were visualised with Chromato-Vue TL-33 UV Transilluminator [UVP, California, USA] and excised using a new sterile scalpel for each band. Gel fragments were stored at -20°C in preparation for DNA extraction from the gel.

3.11.1. DNA purification from agarose gels

DNA was purified from gel fragments using the QIAquick® Gel Extraction Kit [Qiagen, Venlo, Netherlands]. Manufacturer's protocol was followed, with all buffers used being part of the kit. In brief, 100 μL Buffer QG was added to dissolve the gel fragment at a ratio of 3:1 to each 100 mg of gel, and heated for 50°C for 10 minutes with regular vortexing. An equal volume of isopropanol to gel mass (i.e., 1 mg = 1 μL) was added and vortexed. The DNA mixture was added directly onto the membrane of a spin column and centrifuged to bind to the membrane. An additional step designed to remove all agarose was incorporated (500 μL Buffer QG added to the membrane and centrifuged), then, membranes were washed once in 750 μL Buffer PE (1 minute centrifuge). An additional 1 minute spin followed to dry the spin column membrane. 30 μL Buffer EB was added to each spin column and incubated at RT for 4 minutes, then centrifuged for 1 minute to elute DNA; this longer incubation time in elution buffer (Buffer EB) and smaller volume was selected for increased concentration of eluted DNA. All centrifugation steps were performed in a tabletop microcentrifuge (Sigma 1-14 [Fisher Scientific, Massachusetts, USA]) at 13,000 rpm. Extracted purified DNA was quantified using the NanoDrop™ Lite Spectrophotometer [ThermoFisher Scientific, Massachusetts, USA] and DNA stored at -20°C until further required.

3.11.2. Sample preparation for Sanger sequencing

Purified DNA samples were prepared for the TubeSeq service provided by Eurofins Genomics for Sanger sequencing, following their sample preparation guidelines. Samples were prepared at a concentration of 1 ng/μL in a total volume of 15 μL. If necessary, the volume was made up with Water, DEPC Treated, Sterile [National Diagnostics, Georgia, USA]. For each purified DNA sample, 2 samples were prepared with either GP5+ or GP6+ primer, added at 2 μL of a 10 μM stock.

3.11.3. Data analysis of FASTA sequence data (NCBI BLAST)

FASTA sequences were searched for their alignment using the NCBI Basic Local Alignment Search Tool (BLAST) (https://blast.ncbi.nlm.nih.gov) in the blastn suite, with the following parameters selected for all searches: standard databases (nr etc.) = core nucleotide database (core_nt), program selection optimise for: highly similar sequences (megablast).

3.12. Genotyping of HPV-positive tonsil gDNA

DNA samples considered HPV-positive, or potentially positive, by melt curve analysis as part of routine HPV-DNA qPCR consensus screening were subjected to HPV genotyping using the INNO-LiPA® Genotyping Extra II kit [Fujirebio, Tokyo, Japan] that can detect up to 32 HPV viral types (including 13 HR viral types) and the presence of multiple viral types in a sample.

3.12.1. PCR amplification

The manufacturer's protocol for PCR amplification and subsequent hybridisation was followed with minor modification (Fujirebio, 2016). In brief, 50 µL PCR reactions were set up as follows: 40 µL SPF10 supplied master mix [Fujirebio, Tokyo, Japan] and 9 µL nuclease-free H_2O + 1 µL tonsil gDNA for each sample, or 10 µL positive control [Fujirebio, Tokyo, Japan]. For all tonsil gDNA samples, 1µL of tonsil gDNA was added to each reaction regardless of concentration. All samples were subjected to two initial stages of decontamination at 37°C for 10 minutes then denaturation at 94°C for 9 minutes. This was followed by 40 cycles of 94°C – 30 seconds, 52°C – 45 seconds, 72°C – 45 seconds and a final hold stage at 72°C. PCR cycling was performed using a StepOnePlus™ Real-Time PCR system [Applied Biosystems, ThermoFisher Scientific, Massachusetts, USA]. PCR products were stored at -20°C for hybridisation and viral typing.

3.12.2. Hybridisation of PCR products

The next stages of hybridisation and visualisation of viral types were performed in a separate space to PCR set-up, to avoid contamination with PCR amplicons. All INNO-LiPA® HPV Genotyping Extra II strips were handled with sterile forceps. As with PCR amplification, hybridisation of PCR products was performed following manufacturer's protocol. In short, 10 uL of PCR products from each reaction were denatured for 5 minutes with 10 uL denaturation solution [Fujirebio, Tokyo, Japan] in an individual trough for each tonsil PCR product, at 20-25°C. 2 mL of hybridization solution [Fujirebio, Tokyo, Japan] was added along with an

INNO-LiPA® HPV Genotyping Extra II strip for binding of PCR products. Hybridisation was carried out for 1 hour at 49°C, in a shaking water bath (35 rpm).

All liquid was then aspirated from each trough. Strips were washed twice in 2 mL pre-warmed Stringent Wash solution (preheated to 37°C) [Fujirebio, Tokyo, Japan]. Troughs were gently shaken in between washes to encourage the liquid to wash the surface of each strip. After the second wash, a final 2mL of Stringent wash solution was added, and strips were incubated at 49°C, 30 minutes, in a shaking water bath (45 rpm).

The strips were washed twice in 2 mL Rinse Solution (1 minute incubation, all liquid was removed between washes); all incubations during this colour development stage were performed at RT on an orbital shaker at 135 rpm. The conjugate solution was then added into each trough (2 mL) for a 30 minute incubation. The strips were then subjected to 2 washes in 2 mL rinse solution (liquid removed after each wash) followed by a wash in 2 mL substrate buffer. After substrate buffer was removed, visualisation of any hybridisation was initiated by addition of 2 mL substrate solution for a 30 minute incubation, and aspirated. 2 mL distilled water was then added to wash the strips (2 x 3 minute washes). After development, strips were dried, and then compared to the template strip provided to determine success of the assay, and viral types present.

3.13. Tissue Sections and Staining

FFPE blocks were sectioned at 4 µm from representative locations across the tonsil tissue, and stained with haematoxylin and eosin (H&E) and mouse-anti-$p16^{INK4a}$ [clone IHC116-1, GeneAb, GenomeMe, Richmond, Canada] at UHDB using the Dako Omnis IHC platform [Agilent, Santa Clara, USA], as per the IRAS study protocol. H&E and p16 staining was scored by eye, and reported by Dr Amandeep Mann, Consultant Histopathologist, UHDB. Further sections were de-paraffinised and stained with mouse-anti-CK7 [clone RN7, Novocastra, Leica Microsystems Ltd, Milton Keynes, UK] and then incubated in Mixed DAB Refine using the BOND™ RX Fully Automated Research Stainer [Leica Microsystems Ltd, Milton Keynes, UK] at HistologiX Ltd [Nottingham, UK] by Dr Rebecca Mallender, Research Scientist, University of Derby. All slides were imaged with a NanoZoomer S360 Digital Slide Scanner and NZAcquire Software to scan slides and viewing in NDP.view2 software [Hamamatsu Photonics, Hamamatsu, Japan].

3.14. Cell culture

Adherent cell lines SiHa ATCC HTB-35 (human uterine epithelial squamous cell carcinoma, 1-2 integrated copies HPV-16 per cell), VU-SCC-147T (human oesophageal epithelial squamous cell carcinoma, 1-2 integrated copies HPV-16 per cell), and VU-SCC-040T (human oesophageal epithelial squamous cell carcinoma, HPV-negative), latterly both gifts from Professor H Joenje, VU Medical Centre, Amsterdam via Dr S Roberts, University of Birmingham, were grown in Dulbecco's modified eagle medium (DMEM) containing 584 mg/L L-glutamine [Scientific Laboratory Supplies Ltd, Nottingham, UK], supplemented with 10% foetal bovine serum (FBS), 1% penicillin-streptomycin (10,000 U/mL), and 1% non-essential amino acids [all GIBCO, ThermoFisher Scientific, Massachusetts, USA], in a humidified incubator at 37°C with 5% CO_2. Media was replaced every 2-3 days and cells were passaged at 80% confluency using 0.05% Trypsin-EDTA [GIBCO, ThermoFisher Scientific, Massachusetts, USA]. Cells were also harvested by trypsinisation, washed in 1 x PBS, and collected as a cell pellet through centrifugation (160 x g, 5 minutes), whereupon they were immediately snap frozen at -80°C.

3.15. Tonsil tissue lysis and 8 M urea protein extraction

There was an extensive optimisation process to determine the optimal protocol for the isolation of protein from 10% NBF fixed tonsil tissues to be analysed using SWATH mass spectrometry (MS) – as with DNA and RNA extraction, this was a challenging process from fixed tissues compared to fresh-frozen. Here, we outline the final protocol that gave the highest protein concentration as well as protein identities following data searching of SWATH mass spectrometry data downstream.

203 tonsil tissues were selected based on their HPV-DNA status, and donor gender, for MS analysis. 2 additional protein samples from cell lines (see 3.14) were processed following 3.15.2 onwards). Some of the tissues had been collected and sectioned by other researchers in the laboratory prior to the start of this project. These had been subjected to 8 M urea protein extraction previously, and so, these protein lysates were excluded from the methodology presented here until protein quantification (section 3.15.3); but were still subjected to sonication as described at 3.15.2.

3.15.1. Preparation of tonsil tissue for lysis

Tonsil tissues had been sectioned upon arrival in the University laboratory and sections of tissue had been stored in 70% EtOH, RT, in preparation for protein extraction. From these tonsil sections, ~200 mg tissue was cut using a sterile scalpel in a petri dish. Cutting apparatus was changed between every tonsil tissue, including between left and right tissues of the same donor. Tissue was weighed to 200 mg ± 10 mg per sample. Dependent on total tissue mass available, 1 to 2 ~200 mg sections were taken for each left and right tonsil (total of 2-4 sections for extraction per participant). Each tissue section was then cut further into fine slices to aid homogenisation. All samples were then subjected to 2x 5 minute washes in 500 µL 1X PBS [Sigma-Aldrich, Massachusetts, USA] at RT, with all 1X PBS aspirated after each wash. A 5 mm stainless steel bead [Qiagen, Venlo, Netherlands] was then added to each tube ready for physical disruption.

3.15.2. Tissue homogenisation and sonication

400 µL protein extraction buffer was added (8 M urea [ThermoFisher Scientific, Massachusetts, USA], 25 mM Tris [Sigma-Aldrich, Massachusetts, USA], pH 7.5], 1% Triton X-100 [Sigma-Aldrich, Massachusetts, USA] with protease inhibitor cocktail [Sigma-Aldrich, Massachusetts, USA]; 2.5 µL per 50 mg tissue). Samples were homogenised in protein extraction buffer using the TissueLyser LT instrument [Qiagen, Venlo, Netherlands] for 10 minutes at 50 Hz. Homogenised tissue was incubated on wet ice for 20 minutes before centrifugation (300 x g, 1 minute, 4°C). Stainless steel beads were then removed using forceps cleaned in household bleach, UltraPure H_2O, acetonitrile (ACN) [Fisher Scientific, Massachusetts, USA], and rinsed again in UltraPure H_2O, between samples. All centrifugation steps described throughout section 3.15. were performed in a tabletop Prism™ R Refrigerated MicroCentrifuge [Labnet International, New Jersey, USA].

To lyse tonsil tissue further, samples were next sonicated on ice with a Fisherbrand™ Model 120 Sonic Dismembrator [Fisher Scientific, Massachusetts, USA] for 1 minute (sonication programme: 5 seconds on, 5 seconds off, 100% amplitude). The sonicating probe was immersed and cleaned in UltraPure H_2O then ACN, and then again in UltraPure H_2O, after every sample to prevent cross-contamination. After sonicating, samples were centrifuged at 10,000 x g, 10 minutes, 4°C. Supernatant was transferred to a new 1.5 mL tube, away from remaining tissue debris. This centrifugation step, and removal of supernatant, was repeated to remove any remaining contaminating cell debris/tonsil tissue. A total of 205 protein samples

were stored at -80°C, before transfer on wet ice to Nottingham Trent University (NTU) for further processing.

3.15.3. Protein quantification

Tonsil protein was quantified by the Qubit™ 4 Fluorometer [ThermoFisher Scientific, Massachusetts, USA] using their Qubit™ protein assay kit [ThermoFisher Scientific, Massachusetts, USA], following manufacturer's protocol (ThermoFisher Scientific, 2022b). The kit is designed to quantify protein within the range of 12.5 ng/μL – 5000 μg/mL. In brief, a working solution was produced containing Qubit™ Protein Reagent diluted 1:200 in Qubit™ Protein Buffer. Standards 1, 2, and 3, were diluted 1:20 in working solution (final concentrations: 0 ng/μL, 200 ng/μL, and 400 ng/μL, respectively). All protein samples were diluted 1:200 in working solution (total volume per sample 200 μL). Standards and diluted protein were vortexed thoroughly (2-3 seconds) and incubated at RT for 15 minutes.

Protein concentration (μg/μL) was measured using the Qubit™ 4 fluorometer. Any samples with a concentration outside of the upper range of the detection limit (≥5 μg/μL) were diluted 1:2 (with further dilutions carried out if necessary) and read again. Calculated values were used to determine the volume required for a total protein mass of 75 μg. The appropriate volume of each sample to give 75 μg total protein was aliquot and samples stored at -80°C until required for digestion.

3.16. Tryptic protein digestion for SWATH-MS

To digest 75 μg total tonsil tissue protein, the S-Trap™ Rapid Universal MS Prep [Protifi, New York, USA] protocol was selected along with the S-Trap™ micro columns [Protifi, New York, USA] that are designed to digest ≤100 μg protein. As recommended, all buffers (Table 8) were produced either the day before, or on the day of use. All aliquots containing 75 μg total protein were dried down using the Concentrator Plus [Eppendorf, Hamburg, Germany] with the MZ 2C NT vacuum pump [Vacuubrand] (programme; 60°C, V-AQ mode), until all liquid had evaporated.

All protein samples (N=205) were resuspended in 25 μL lysis buffer [5% sodium dodecyl sulphate (SDS) [Sigma-Aldrich, Massachusetts, USA], 50 mM Triethylammonium bicarbonate buffer (TEAB) [Sigma-Aldrich, Massachusetts, USA], pH 7.55]. Resuspended samples were

subjected to a 10 minute sonication in a sonicating water bath and then centrifuged (4-minutes, 13,000 x g, RT). 1 µL reducing agent, 0.5 M dithiothreitol (DTT) [Sigma-Aldrich, Massachusetts, USA] was added before a 20 minute incubation on a thermomixer (56°C, 400 rpm) with shaking to disrupt protein disulphide bonds. Then, 2 µL 0.5 M iodoacetamide (IDA) [Sigma-Aldrich, Massachusetts, USA] was added and each solution was incubated for 15 minutes at RT in the dark. This was to alkylate any reduced cysteine residues.

12% phosphoric acid solution [Sigma-Aldrich, Massachusetts, USA] was added to the reduced/alkylated protein solution, to a final concentration of 1.2%. Samples were vortexed thoroughly to ensure phosphoric acid had fully dissolved. 185 µL S-Trap buffer [100 mM TEAB [Sigma-Aldrich, Massachusetts, USA] in 90% MeOH [Merck, Darmstadt, Germany]] was then added, followed by a vigorous vortex. 200 µL of the protein mixture was added directly onto the membrane of an S-Trap™ micro column [Protifi, New York, USA] and centrifuged at 4,000 x g, 30-seconds, RT. Samples were spun until all lysate had passed through the column, increasing rcf 100 x g each time for those that would not pass through.

To wash protein bound to the S-Trap™ micro column, 150 µL S-Trap wash buffer was added to the column and centrifuged (4,000 x g, 30-seconds, RT). A total of 3 additional washes were performed. Then, to digest protein, 25 µL digestion buffer [trypsin/Lys-C mix [Promega, Wisconsin, USA], 50 mM TEAB pH 7.5] was added so that the final trypsin mass was at a 1:10 ratio to the total protein (e.g., 5 µg trypsin for 50 µg protein) and samples were incubated at 47 °C for 1 hour 30 minutes.

Peptide elution was then performed into Protein LoBind tubes [Eppendorf, Hamburg, Germany]. 40 µL of 3 elution buffers were added sequentially, each followed by a centrifugation step (4,000 x g, 30-seconds, RT). Elution buffer 1; 50 mM TEAB pH 8.5, elution buffer 2; 0.2% formic acid (FA) [Merck, Darmstadt, Germany], elution buffer 3; 50% ACN [Merck, Darmstadt, Germany] containing 0.2% FA. Eluate, containing peptides, was saved, and pooled together following each centrifuge step. Protein pellets were dried down in the Concentrator Plus [Eppendorf, Hamburg, Germany] with the MZ 2C NT vacuum pump [Vacuubrand] (programme; 60°C, V-AQ mode) until all liquid had evaporated. Protein pellets were stored at -80°C, until SWATH mass spectrometry analysis.

Table 8. Solutions for S-Trap™ Rapid Universal MS Sample Prep [Protifi] protein digestion. The composition of each solution and the pH value if adjusted is shown.

Solution	Composition	pH
Lysis buffer	5% SDS, 50 mM TEAB	7.55
Reductant	0.5 M DTT	Unadjusted
Alkylator	0.5 M IDA	Unadjusted
1 M TEAB stock	1 M TEAB	7.1 or 2.55
Wash buffer	100 mM TEAB in 90% MeOH	7.1 or 7.55
Digestion buffer	Trypsin/Lys-C mix, 50 mM TEAB	7.5
Elution buffer 1	50 mM TEAB in H_2O	Unadjusted, pH 8.5
Elution buffer 2	0.2% FA	Unadjusted
Elution buffer 3	50% ACN containing 0.2% FA	Unadjusted

3.17. Proteomic analysis of non-cancerous palatine tonsils

206 digested protein samples (including 1 digestion repeat) were next prepared for SWATH mass spectrometry. Peptides were resuspended in 100 µL 10% ACN [Merck, Darmstadt, Germany], 0.1% FA [Merck, Darmstadt, Germany] before vortexing. Samples were incubated on a shaker at 185 rpm, 4°C for 90 minutes. A higher concentration of ACN was used here to encourage any peptides into solution that had adhered to the tube during drying down following tryptic digestion. After the incubation, 100 µL 0.1% FA was added to give a final concentration of 5% ACN, 0.1% FA in a total volume of 200 µL. Resuspended peptides were vigorously vortexed and then centrifuged at 12,900 rpm, 1 minute. The total volume (~200 µL) was transferred to a 300 µL clear PP vial 12x33 mm crimp or snap tube [Chromatography Direct, Northamptonshire, UK], with care taken to not introduce air into the liquid. 11 mm SNAP CAP 10mil PFTE lined caps [Chromatography Direct, Northamptonshire, UK] were added to each tube. All peptide samples were categorised based on the gender of each participant that tonsil protein was isolated from and randomised into 3 batches for mass spectrometry analysis, based on a roughly even split of gender by batch.

3.17.1. HPLC and SWATH-MS

2 µL resuspended peptides (approx. 750 ng material) were injected by the autosampler of the ACQUITY UPLC (Ultra Performance Liquid Chromatography) M-Class System [Waters Corporation, Massachusetts, USA] onto a Kinetex 2.6 µM XB-C18 100 Å, LC Column of 150 x 0.3 mm [Phenomenex, California, USA] at 30°C. Reverse phase HPLC was carried out and peptides were eluted from the column into a ZenoTOF 7600 mass spectrometer [Sciex, Massachusetts, USA] at a gradient elution of flow rate 10 µL/minute. The total run time of HPLC was 16.5 minutes. The starting ratio of mobile phase A [0.1% FA, Merck, Darmstadt, Germany] to mobile phase B [ACN, Merck, Darmstadt, Germany with 0.1% FA] was 97% : 3%. The concentration of mobile phase B was increased to 35% in a 12-minute period, then again to 80% at 13 minutes for 2 minutes. At 15.5 minutes, the column was re-equilibrated to initial conditions [3% mobile phase B] over 1 minute (12 µL/minute flow rate).

Eluted peptides were sprayed into the MS and ionised using a OptiFlow Turbo V ion source [Sciex, Massachusetts, USA] with a Microflow probe, flow range 1-50 µL/minute, using zeno pulsing at 4700 V. Before SWATH, MS1 data for all intact peptides was generated. Then, peptides were fragmented in the collision cell and all peptides were analysed across a

mass/charge (m/z) range of 400-700 m/z split into 65 windows of varied m/z range. Each window was scanned for 12 milliseconds (ms), giving a total of 780 ms for the total m/z range of all windows and fragmentation data for all peptides in each window was collected (MS2).

3.17.2. Protein data searching – DIA-NN

Data generated from SWATH mass spectrometry was processed in DIA-NN software (Demichev et al., 2020), version 1.9, with library (provided by Dr Nick Morrice, Sciex, generated from multiple high/low pH fractions from HeLa and K562 protein digest containing 11268 protein groups) provided as the spectral library. This spectral library was generated using K562 and HeLa cell lines. For data searching in DIA-NN, the following parameters were selected for analysis: All default settings except 210 files loaded, Threads 48: 0.02 FDR filtering, MBR off. The command line used was as follows: --lib HeLA_K562_SciexLibrary.speclib --threads 48 --verbose 2 --out Outputreport.tsv --qvalue 0.02 --matrices --unimod4 --rt-profiling. After processing, samples were examined for their total precursor number and protein identities to ensure samples with similar protein identities (IDs) were grouped for differential expression analysis (DEA).

3.17.3. DEA - Amica

DEA of protein intensities was performed using the online software, Amica [Version 3.0.1 - https://bioapps.maxperutzlabs.ac.at/app/amica]. This online software makes use of an R-based package, LIMMA, to allow calculation of p-value, adjusted (adj.) p-value, and log fold change (FC) of imputed intensities between comparative groupings of samples (Didusch et al., 2022). A total of 2 analyses were carried out that allowed the statistical analysis of protein expression in tonsil protein samples of different groups. These were: HPV-positive versus HPV-negative (fresh-frozen tonsil protein) and HPV-positive male versus HPV-positive female (fresh-frozen tonsil protein).

For each comparison a contrast matrix and an experimental design was produced, as following the requirements of Amica software. The pg_matrix file generated from SWATH data searching in DIA-NN was edited to contain relevant samples for each DEA. The advanced parameters were left as the default settings; except for ‘filter on valid values per group’ whereby the minimum valid values in each or one group was altered dependent on each individual comparative analysis (4_in_one_group – HPV-positive versus HPV-negative,

5_in_one_group – HPV-positive males versus HPV-positive females). Normal imputation method was used for all analyses.

Differentially expressed protein hits identified by Amica were visualised by volcano plots of each comparison. Data (including logFC, and adj.p-value/p-value) were uploaded to VolcaNoseR2. To generate tables of top significantly increased and decreased protein hits for each comparison, data were filtered on logFC and p-value of various stringency, dependent on the data set.

3.17.3.1. Ontological analysis

A list of UniProtIDs of the top significantly changed protein hits for each analysis (e.g., HPV-positive versus HPV-negative) were uploaded to Panther (version 19.0 - https://www.pantherdb.org). Protein hits were searched with organism – *homo sapiens* and select analysis – functional classification viewed in graphic charts as bar chart, selected as the search parameters. Hits were categorised based on their protein classification, biological process, molecular function, and cellular component.

3.17.3.2. Pathway enrichment analysis

Pathway enrichment analysis of top protein hits, filtered based on various p-value/logFC stringencies, from each comparison was performed in reactome.org. The appropriate list of top hits was downloaded from VolcaNoseR2 and these differentially expressed protein hits were searched in Analysis Tools without 'project to human' or 'include interactors' selected for database searching. The analysis report from the software was downloaded to determine pathway enrichment for each comparison.

CHAPTER 4: Meta-analyses of global oral HPV prevalence estimates amongst the adult population.

4.0. Introduction

Rates of HPV detection within the mouths of healthy individuals are inconsistent between studies, ranging from several reports of a <1% prevalence (Kurose et al., 2004, Ernster et al., 2009, Esquenazi et al., 2010, Stankiewicz Karita et al., 2020, Kumari et al., 2021, Güran et al., 2021, Napolitano et al., 2024) to ~35% and even as high as ~80% (Pattanshetty et al., 2014, Dalla Torre et al., 2015, Lam et al., 2016, Shigeishi et al., 2018, Ribeiro et al., 2019, Jamieson et al., 2020, Morán-Torres et al., 2021). These discrepancies could be attributed to specific characteristics of the target population of the study; for instance the age group, country of origin, or gender, amongst other demographic factors; but there is also considerable variation in sample collection (e.g., oral rinse versus swab/brushing) and HPV detection methods, including specific viral types analysed between publications. Indeed, there is currently no standard protocol for HPV testing within the oral cavity and oropharynx. Taken together, this means that through scrutinising studies on an individual basis it is difficult to ascertain an accurate estimate of oral HPV prevalence across the global population.

To be able to estimate the prevalence of oral HPV on a global scale, we selected to perform a systematic review and meta-analysis of the literature as a suitable approach to combine reports investigating different populations and regions, whilst weighting them for variance in their sample size, to determine an overall global oral HPV prevalence estimate. Previous meta-analyses of oral HPV prevalence have been conducted (Tam et al., 2018, Mena et al., 2019); however, given the significant increases in rates of OPC in recent years (Lechner et al., 2022), together with the large proportion of these attributable to HPV (de Martel et al., 2017), we sought to provide an updated estimate of oral HPV prevalence within 'healthy' individuals to provide an insight into HPV prevalence in studies published from any region globally between 1995-present (2024) within persons without any type of HNC.

We carried out advanced database searches to identify appropriate studies based on an intricate inclusion/exclusion criteria (described in full at Chapter 3.1.1) that, in brief, were studies that reported on adults without HNC. We were interested to see how prevalence estimates within these studies would compare to our own findings of oropharyngeal HPV prevalence within

non-cancerous tonsil tissue; thus, these meta-analyses were designed to include a similar participant demographic to be broadly comparable. We performed meta-analyses of HPV infection within the healthy adult population, for any HPV type, HR-HPVs, and HPV-16 which is the most common viral type found in HPV-OPC (Chaturvedi et al., 2013, Schache et al., 2016, Lechner et al., 2022). From the literature, we hypothesised that HPV would be prevalent within the oral cavity and oropharynx of the healthy population globally, and that HR HPV-16 would be more frequently detected in males than females (Gillison et al., 2012).

In this chapter, we describe an oral HPV prevalence estimate based on individual reports from healthy adult populations across the globe, and present additional sub meta-analyses to summarise the contribution of HR-HPV types, and specifically HPV-16.

As outlined in the final research methodology, this meta-analysis was kindly supported by Dr Elizabeth Marsh, Dr Tom Illingworth, Dr Laurice Fretwell, and Dr James Hind.

4.1. Selection of studies for meta-analysis

Embase, Ovid MEDLINE®, and PubMed Advanced Search Builder were searched using terms that had been specifically designed for each database. A total of 8086 records were identified, with all hits found within Embase (N=5380) and/or PubMed (N=2706) databases. During a rigorous process of title, abstract and then full-text screening, any studies that did not meet the inclusion criteria were removed at each stage of study selection; a total of 7997 studies were excluded. Where necessary, authors were contacted for further information (by email or other means i.e., ResearchGate); in most cases, this was to confirm an absolute age-range of participants of their study. Those studies whereby the authors did not respond to any queries were excluded from the study.

Overall, we identified 89 studies that reported on a total of 86,372 adult individuals for inclusion in this meta-analysis of global oral HPV prevalence within the adult population (Figure 13). For each study included in any of the individual analyses, data were extracted where available for the following categories: Study period, study population/setting, selection of study population, absolute age range, sample collection methods, HPV detection methods, sample size (including total number of males and females), overall HPV prevalence, overall HR-HPV prevalence, overall HPV-16 prevalence, female HPV prevalence (overall and HPV-16), and male HPV prevalence (overall and HPV-16) (Table 9). When data extraction was

performed, data from papers that reported on the prevalence of oral HPV within adults and children were split and only data from individuals over 18 years old were extracted. In addition, in longitudinal studies where participants were monitored over time for new and/or persistent HPV infection, we extracted baseline prevalence data only for use in these meta-analyses, and if more than one sampling method had been used within a study and reported on as separate prevalences, we extracted data from the sampling method with the highest prevalence.

4.1.1. Study classification and demographics of their cohorts for the final list of reports included in these meta-analyses

The final included studies were only analysed based on their reports of oral HPV prevalence (overall and HR/HPV-16 estimates), as well as oral HPV prevalence based on biological sex. Statistical comparisons of oral HPV prevalence were not performed for any other parameters. Nonetheless, we reviewed the study classification and certain demographics of participants included in each study to provide an overall summary of the studies that contribute to the final oral HPV prevalence estimates presented here. In brief, the majority of included studies were classified as cross-sectional studies (N=61), with cohort studies the second most common study classification (N=9).

Studies were then grouped based on geographical location; we found representation of oral HPV within populations across the globe within the final sample. Most studies provided data on individuals based in North America (N=28) or Europe (N=25); with the remainder from Asia (N=15), South America (N=12), Oceania (N=6), and Africa (N=3). One of the inclusion criteria for study period was data collection from 1995 onwards; however, where this information was available, the majority of data collection and participant recruitment of these studies occurred between 2010-2019 (N=52), followed by 2000-2009 (N=18), 2020-2024 (N=7), and 1995-1999 (N=3). If data were collected within more than one period, here it is counted within each group where its data collection falls into. Finally, age ranges of individual studies were assessed to determine the breadth of the age representation of the healthy adult population across the reports. The overall age range was 18-97, and we found 45/89 studies that included individuals as young as 18 years old, our lowest age requirement for inclusion, with 35/89 studies sampling participants over 60 years old.

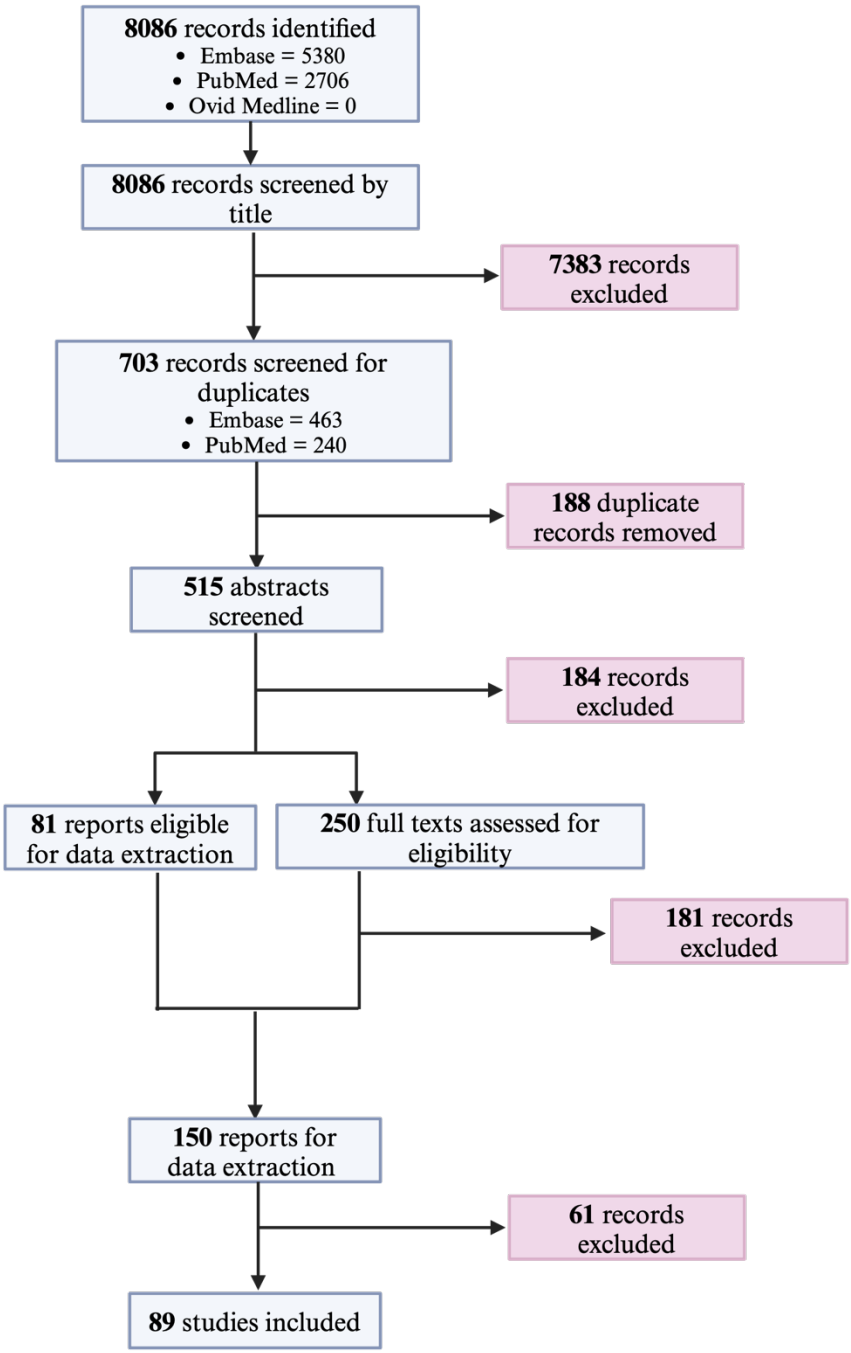


Figure 13. PRISMA flowchart of report selection for meta-analysis of oral HPV prevalence estimates. 8086 records were identified from the search terms between Embase, PubMed, and Ovid MEDLINE® databases. All records were subjected to an initial title screen, and duplicates removed. 515 records were uploaded to Rayyan [Rayyan Systems Inc. Cambridge, MA] for abstract screening and were assigned 'include', 'exclude' or 'maybe', in regard to inclusion by two independent researchers. A full text review was carried out for reports within the maybe category. Data extraction was then performed, and additional reports were excluded if unclear or unsuitable against the inclusion criteria. 89 studies were included for final data-analyses. At all stages, records considered not suitable for inclusion based on the criteria in the analyses were excluded.

Figure created with: BioRender.com

Table 9. Studies included in meta-analysis of oral HPV prevalence in the healthy adult population. 89 studies were included in one or more of the meta-analyses comparisons (overall HPV, overall HR-HPV, overall HPV-16, overall female, female HPV-16, overall male and/or male HPV-16; with inclusion in any comparisons indicated) and are shown alongside extracted data from each report. Overall rating for risk of bias is also indicated.

										HPV prevalence estimates								Inclusion in analyses						
Study	**Study period**	**Study design**	**Study population/ setting**	**Selection of study population**	**Age range**	**Sample collection methods**	**HPV detection methods**	**Risk score**	**Sample size**	**Overall HPV prevalence**	**HR-HPV prevalence**	**HPV-16 prevalence**	**Female (n)**	**Female HPV prevalence**	**Female HPV-16 prevalence**	**Male (n)**	**Male HPV prevalence**	**Male HPV-16 prevalence**	**Overall HPV**	**Overall HR-HPV**	**Overall HPV-16**	**Female**	**Female HPV-16**	**Male**	**Male HPV-16**
Giuliano et al., 2023	Nov. 2021 - Mar. 2022	Cross-sectional	Primary care (dentist)	Convenience	18-60	Oral gargle specimen	SPF10/DEIA/LiPA25 system	Low	3196	6.6%	2.0%	0.7%	1773	4.6%	0.4%	1423	9.1%	1.1%	Yes	Yes	Yes	Yes	Yes	Yes	Yes
de Souza et al., 2023	Oct. 2020 - Nov. 2021	Cross-sectional	Community (Oral Diversity Study)	Population-based	18-70	Saliva samples	MY/GP+ PCR system	Low	911	7.2%	3.8%	1.1%	678	6.3%	NR	227	9.7%	NR	Yes	Yes	Yes	Yes	No	Yes	No
Gheit et al., 2023	Jun. 2015 - Oct. 2016	Sub-study (cross-sectional)	Community (IARC cohort)	Population-based	18-25	Oral gargle specimen	E7-MPG assay (PCR based)	Low	997	6.8%	2.6%	1.9%	997	6.8%	1.9%	N/A	NR	NR	Yes	Yes	Yes	Yes	Yes	No	No
Riddell et al., 2022	2015 - 2017	Cohort study	University population	Convenience	18-50+	Saliva and oral rinse samples	Multiplex competitive PCR amplification, followed by probe specific single base extension and MALDI-ToF-MS	Low	443	0.2%	13.0%	2.7%	201	9.5%	0.0%	238	23.1%	3.8%	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Kolesnik et al., 2022	NR.	Convenience sample	Primary care (dentist)	Convenience	20-60+	Cytobrush swab of cheeks by dentist	INNO-LiPA HPV genotyping assay	Low	246	13.8%	4.5%	1.2%	122	4.9%	NR.	124	22.6%	NR.	Yes	Yes	Yes	Yes	No	Yes	No
Mosna-Torres et al., 2021	Jan. 2018 - Jun. 2018	Prospective study	University population	Convenience	18-55	Cytobrush of mouth and oropharynx	MPCR kit for Human Papillomavirus Set 2	Moderate	94	46.8%	NR.	11.7%	53	41.5%	NR.	41	53.7%	NR.	Yes	No	Yes	Yes	No	Yes	No
Jamieson et al., 2020	Feb. 2018 - Jan. 2019	Cross-sectional	Community (ACCHO)	Convenience	18+	Saliva samples	Nested PCR (MY09/11, GP5+/GP6+)	Low	910	35.3%	7.5%	1.5%	595	36.1%	1.7%	315	33.7%	1.3%	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sastre-Cantón et al., 2019	Nov. 2012	Cross-sectional	University population	Convenience	18-25	Oral gargle specimen	SPF10/DEIA/LiPA25 system	Low	543	7.2%	3.5%	0.4%	288	6.6%	0.4%	255	7.8%	0.4%	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Zhu et al., 2023	2021	Prospective cohort study	Community (DLCC)	Population-based	20-80	Oral cavity swabs	Nested PCR (MY09/11, GP5+/GP6+) and sequencing	Moderate	9867	3.0%	1.3%	1.1%	6819	2.7%	NR.	3048	3.7%	NR.	Yes	Yes	Yes	Yes	No	Yes	No
Rajendra Santosh et al., 2019	Sept. 2016 - Feb. 2017	Cross-sectional	Primary care (dentist and HIV services centre)	Convenience	18-64	Saliva samples	Linear Array HPV genotyping kit	Low	104	8.7%	1.0%	0.0%	55	9.1%	0.0%	49	8.2%	0.0%	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Fu et al., 2015	Mar. 2011 - Jan. 2012	Longitudinal cohort study	Affiliated with University	Convenience	30-50	Oral sweb and rinse samples	MY09/11 PCR and Roche Linear Array	Moderate	409	2.4%	2.2%	0.7%	409	2.4%	0.7%	N/A	NR.	NR.	Yes	Yes	Yes	Yes	Yes	No	No
Wong et al., 2018	NR.	Population based study	Community	Population-based	18-64	Oral rinse samples	PCR NGS based assays	Low	1426	15.5%	0.8%	0.4%	745	1.6%	NR.	680	3.5%	NR.	Yes	Yes	Yes	Yes	No	Yes	No
Cook et al., 2014	Apr. 2011 - Sept. 2011	Cohort study	University population	Convenience	18-54	Oral rinse samples	Roche Linear Array	Moderate	1010	1.9%	0.7%	0.2%	1010	1.9%	0.2%	N/A	NR.	NR.	Yes	Yes	Yes	Yes	Yes	No	No
Antonsson et al., 2021	Jun. 2014 - Oct. 2016	Cohort study	Workplaces, medical clinics, University, Sexual health clinics	Convenience	20-70	Saliva samples	Nested PCR (MY09/11, GP5+/GP6+)	Low	636	10.7%	6.4%	1.7%	346	9.5%	NR.	284	12.3%	NR.	Yes	Yes	Yes	Yes	No	Yes	No
Cavenaghi et al., 2013	2011	Cross-sectional	Community	Population-based	18-89	Oral rinse/ mouthwash	Roche Linear Array	Low	124	2.4%	0.8%	0.0%	77	NR.	0.0%	47	NR.	0.0%	Yes	Yes	Yes	No	Yes	No	Yes
Turner et al., 2011	Jun. 2010 - Oct. 2010	Cross-sectional	Primary care (dentist)	Convenience	18-65+	Saliva samples	HPV-16/-18 type-specific PCR	Low	151	2.6%	2.6%	2.6%	79	5.0%	5.1%	72	0.0%	0.0%	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Esquenazi et al., 2010	May 2006 - Apr. 2007	Cross-sectional	University population	Convenience	20-31	Oral brushing	PCR (GP5+/GP6+, MY09/MY11)	Moderate	100	0.0%	0.0%	0.0%	60	0.0%	0.0%	40	0.0%	0.0%	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Araújo et al., 2014	2013	Cross-sectional	Primary care (dentist)	Convenience	18-79	Oral mucosa scraping	PCR (MY09/MY11)	Low	166	24.1%	3.6%	0.0%	103	NR.	0.0%	63	0.0%	NR.	Yes	Yes	Yes	No	Yes	Yes	No
de la Garza-Ramos et al., 2020	Apr. 2017	Cross-sectional	Primary care (dental brigade)	Convenience	18+	Cytobrush of mouth and oropharynx	Nested PCR (MY09/11, GP5+/GP6+)	Moderate	198	6.6%	1.0%	1.0%	137	8.8%	NR.	61	19.7%	NR.	Yes	Yes	Yes	Yes	No	Yes	No
de Oliveira et al., 2023	Apr. 2018 - Aug. 2019	Cross-sectional	Community	Population-based	22+	Oral brushing of gums and cheeks	INNO-LiPA HPV genotyping assay	Moderate	200	11.7%	NR.	NR.	NR.	NR.	NR.	NR.	NR.	NR.	Yes	No	No	No	No	No	No
Tsikis et al., 2018	Jul. 2015 - Oct. 2015	Cross-sectional	Primary care (hospital - STI/HIV clinics	Convenience	18-55	Oral rinse samples	PCR MY09/11 and Illumina sequencing	Low	294	3.7%	1.0%	NR.	N/A	NR.	NR.	294	3.7%	NR.	Yes	Yes	No	No	No	Yes	No
Mosmann et al., 2021	NR.	Cross-sectional	Primary care (Hospital)	Convenience	18-67	Scrapes of oral cavity	PCR (MY09/MY11) then RFLP analysis for viral type determination	Low	100	14.0%	10.0%	10.0%	100	14.0%	10.0%	N/A	NR.	NR.	Yes	Yes	Yes	Yes	Yes	No	No
Lucas-Roxburgh et al., 2015	Jun. 2013 - Nov. 2013	Cross-sectional	Primary care (SexualS tudent health centres)	Convenience	18-25	Oral brushing of left and right buccal surfaces	PCR (PGMY) then Roche Cobas 4800 for viral type determination	Low	216	3.2%	0.9%	0.0%	216	3.2%	0.0%	N/A	NR.	NR.	Yes	Yes	Yes	Yes	Yes	No	No
Edelstein et al., 2012	Sept. 2008 - Apr. 2010	Cohort study	University population	Convenience	18-25	Gargle/rinse and 2 self-collected oropharyngeal swabs	MY09/11/HME01 PCR and genotyping by liquid bead microarray assay	Moderate	212	7.5%	13.7%	2.8%	N/A	NR.	NR.	212	7.5%	2.8%	Yes	Yes	Yes	No	No	Yes	Yes
Lang Kuhs et al., 2018	Jun. 2008 - Dec. 2009	Randomised control trial	Community (Costa Rican Vaccine Trial)	Population-based	22-29	30 second gargle rinse	SPF10/DEIA/LiPA25 system	Low	5838	1.7%	1.0%	0.2%	5838	1.7%	0.2%	N/A	NR.	NR.	Yes	Yes	Yes	Yes	Yes	No	No
Rumianek et al., 2023	NR.	Cross-sectional	Primary care (dentist)	Convenience	18+	Oral rinse and gargle samples	PGMY09/11 PCR then INNO-LiPA or Abbott RealTime High-Risk HPV test	Low	298	2.0%	2.0%	0.3%	167	0.0%	0.0%	131	4.6%	0.8%	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sias et al., 2019	NR.	Retrospective cross-sectional	Samples stored in hospital laboratory	Convenience	NR.	Swab of right/left buccal mucosa, palatine tonsils, upper and lower pharynx area	MY09/11 and FAP59/64 PCR, then Genomica genotyping or sequencing	Low	124	20.9%	0.8%	0.8%	NR.	NR.	NR.	NR.	NR.	0.0%	Yes	Yes	Yes	No	No	No	No
Auguste et al., 2017	NR.	Population based case-control study	Primary care (Hospital)	Convenience	18-75	Saliva samples	INNO-LiPA HPV genotyping assay	Moderate	308	26.0%	9.7%	0.6%	82	14.6%	NR.	226	30.1%	NR.	Yes	Yes	Yes	Yes	No	Yes	No

										HPV prevalence estimates									Inclusion in analyses						
Study	**Study period**	**Study design**	**Study population/ setting**	**Selection of study population**	**Age range**	**Sample collection methods**	**HPV detection methods**	**Risk score**	**Sample size**	**Overall HPV prevalence**	**HR-HPV prevalence**	**HPV-16 prevalence**	**Female (n)**	**Female HPV prevalence**	**Female HPV-16 prevalence**	**Male (n)**	**Male HPV prevalence**	**Male HPV-16 prevalence**	**Overall HPV**	**Overall HR-HPV**	**Overall HPV-16**	**Female**	**Female HPV-16**	**Male**	**Male HPV-16**
Hang et al., 2014	2009 - 2011	Population based prospective cohort study	Unclear	Population-based	22-65	Swab of upper/lower lips, left/right sides of hard palate, buccal mucosa, top/bottom of tongue, surface of gingiva	SPF1/GP6+ mediated PCR and sequencing	Moderate	5351	6.1%	0.5%	0.4%	2816	5.1%	0.4%	2535	5.8%	0.4%	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Kofbed et al., 2015	2010 - 2011	Cross-sectional	Primary care (Hospital)	Convenience	22-68	Oral rinse samples	Luminex-based HPV genotyping	Low	271	1.1%	0.4%	0.4%	223	NR.	NR.	64	NR.	NR.	Yes	Yes	Yes	No	No	No	No
Hearnden et al., 2018	Apr. 2013 - Aug. 2014	Observational case-control study	Primary care (Hospital)	Convenience	18-60	Oral rinse and gargle samples	Roche Linear Array	Low	680	2.2%	2.2%	0.6%	270	0.7%	NR.	410	3.2%	NR.	Yes	Yes	Yes	Yes	No	Yes	No
McIntosh et al., 2024	2020 - 2021	Cross-sectional	Community	Convenience	21-59	Oral rinse samples	Cobas HR HPV test (Roche)	Moderate	91	8.8%	8.8%	1.1%	N/A	NR.	NR.	N/A	NR.	NR.	Yes	Yes	Yes	No	No	No	No
González-Losa Mdel et al., 2008	NR.	Cross-sectional	Primary care (dentist)	Convenience	19-64	Oral brushing	Hybrid capture 2	Low	77	1.3%	1.4%	NR.	39	0.0%	0.0%	38	2.6%	NR.	Yes	Yes	No	Yes	Yes	Yes	No
Kumari et al., 2021	NR.	Cross-sectional	University population	Convenience	18-25	Saliva samples	HPV-16/-18 type-specific PCR.	Moderate	100	0.0%	0.0%	0.0%	NR.	0.0%	0.0%	NR.	0.0%	0.0%	Yes	Yes	Yes	No	No	No	No
Knight et al., 2016	NR.	Cross-sectional	University population	Convenience	18-25	Buccal cell mouth swab	PCR - L1 detection	Low	124	4.0%	NR.	NR.	57	3.5%	NR.	67	4.5%	NR.	Yes	No	No	Yes	No	Yes	No
Meites et al., 2020	Feb. 2016 - Sept. 2018	Cross-sectional	Community	Convenience	18-26	Oral rinse specimen	Roche Linear Array	Low	1767	7.5%	1.4%	0.5%	N/A	NR.	NR.	N/A	NR.	NR.	Yes	Yes	Yes	No	No	No	No
Visalli et al., 2021	NR.	Cross-sectional	Primary care (Hospital)	Convenience	19-59	Saliva samples	MY09/11 PCR then type-specific PCR.	Low	92	26.1%	12.0%	5.4%	24	NR.	NR.	68	NR.	NR.	Yes	Yes	Yes	No	No	No	No
Shigeishi et al., 2016	2014 - 2015	Cross-sectional	Primary care (Hospital)	Convenience	20-89	Oral gargle specimen	HPV-16 type-specific PCR.	Low	91	28.6%	28.6%	28.6%	NR.	NR.	NR.	NR.	NR.	NR.	Yes	Yes	Yes	No	No	No	No
Nascimento et al., 2019	Oct. 2014 - Oct. 2016	Cross-sectional	Primary care (dentist)	Convenience	30-80	Oral brushing	Nested PCR (MY09/11, GP5+/GP6+) then RFLP and type-specific primers	Low	83	27.7%	4.8%	0.0%	57	NR.	NR.	33	NR.	NR.	Yes	Yes	Yes	No	No	No	No
Suehiro et al., 2020	Sept. 2017 - May 2018	Cross-sectional	Primary care (Hospital)	Convenience	19-66	Full mouth oral scraping (cheeks, tongue, tonsils and oropharynx) and gargle sample	Kit Multiplex XGEN MULTI HPV CHIP HS12 (Mobius Life Sciences)	Low	254	11.8%	4.7%	0.0%	254	11.8%	0.0%	N/A	NR.	NR.	Yes	Yes	Yes	Yes	Yes	No	No
Sánchez-Siles et al., 2020	May 2015 - Feb. 2018	Cross-sectional	Primary care (Hospital)	Convenience	18+	Saliva samples	PGMY09/11 PCR and HPV probe hybridisation for type determination	Low	100	13.0%	12.0%	2.0%	100	13.0%	2.0%	N/A	NR.	NR.	Yes	Yes	Yes	Yes	Yes	No	No
Lupato et al., 2017	Jul. 2014 - Jan. 2015	Cross-sectional	University population	Convenience	19-35	Oral rinse specimen	BSGP5+/6+ PCR MPG assay	Low	500	4.0%	2.2%	1.2%	253	4.0%	NR.	247	4.0%	NR.	Yes	Yes	Yes	Yes	No	Yes	No
Stankiewicz Karita et al., 2020	Nov. 2016 - Nov. 2018	Cross-sectional	Primary care (dentist)	Convenience	18-96	Oral rinse specimen	Type-specific qPCR.	Low	15308	1.0%	1.0%	0.8%	8351	0.6%	0.4%	6957	1.6%	1.3%	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Machado et al., 2014	NR.	Cross-sectional	Unclear (population men)	Convenience	18-68	Oral brushing	MY09/11 PCR then type-specific PCR and RFLP	Moderate	514	1.3%	0.4%	NR.	N/A	NR.	NR.	514	1.3%	NR.	Yes	Yes	No	No	No	Yes	No
Gichki et al., 2012	NR.	Cross-sectional	Primary care (Hospital)	Convenience	18-60	Oral brushing	HPV 1003 and 1004 qPCR primers then HPV-16/-18 type-specific PCR.	Low	192	24.5%	7.3%	2.1%	72	37.9%	NR.	120	27.5%	NR.	Yes	Yes	Yes	Yes	No	Yes	No
Colon-López et al., 2014	NR.	Cross-sectional	STI clinic	Convenience	25+	Oral rinse specimen	INNO-LiPA HPV genotyping assay	Low	157	22.3%	NR.	NR.	N/A	NR.	NR.	157	22.3%	NR.	Yes	No	No	No	No	Yes	No
Kunose et al., 2004	Oct. 2000 - Dec. 2002	Cross-sectional	Primary care (dentist)	Convenience	20-85	Oral scraping (cytobrush)	MY09/11 PCR then sequencing	Low	572	0.7%	0.2%	0.2%	Unsure - hacked data to remove children	NR.	NR.	Unsure - hacked data to remove children	NR.	NR.	Yes	Yes	Yes	No	No	No	No
Pattanshetty et al., 2014	NR.	Cross-sectional	Primary care (dentist)	Convenience	20-45	Oral brushing	Type-specific qPCR.	Moderate	60	65.0%	65.0%	46.6%	27	NR.	NR.	33	NR.	NR.	Yes	Yes	Yes	No	No	No	No
Ragin et al., 2007	Jul. 2004 - Sept. 2004	Cross-sectional	Primary care (Hospital - Cancer clinic)	Convenience	18-65	Oral rinse specimen	MY09/11 and GP5+/6+ nested, then typed using Roche linear array or sequencing (if didn't hybridise)	Moderate	212	6.6%	1.4%	0.9%	212	6.6%	0.9%	N/A	NR.	NR.	Yes	Yes	Yes	Yes	Yes	No	No
Conway et al., 2016	Nov. 2013 - Mar. 2015	Prospective study	Primary care (dentist)	Convenience	26-69	Oral gargle/rinse	Luminex-based HPV genotyping	Moderate	341	5.9%	NR.	NR.	Can't interpret - hacked data for adults	NR.	NR.	Can't interpret - hacked data for adults	NR.	NR.	Yes	No	No	No	No	No	No
Kreimer et al., 2004	2001 - 2002	Cross-sectional	Community - multi-centre	Convenience	18-85	Oral brushing and rinse	MY09/11 PCR.	Moderate	597	13.6%	7.0%	2.5%	NR.	NR.	NR.	NR.	NR.	NR.	Yes	Yes	Yes	No	No	No	No
Visalli et al., 2016	Jul. 2014 - Oct. 2015	Cross-sectional	Primary care (Hospital)	Convenience	20-52	Saliva samples	MY09/11 PCR then type-specific PCR. (7 viral types)	Moderate	125	20.8%	NR.	NR.	125	20.8%	NR.	N/A	NR.	NR.	Yes	No	No	Yes	No	No	No
Lam et al., 2016	2010 - 2014	Longitudinal cohort study	Community	Convenience	18+	Oral rinse samples	PGMY primers and genotyping by reverse line-blot hybridisation	Low	766	40.1%	NR.	NR.	364	39.6%	NR.	402	40.5%	NR.	Yes	No	No	Yes	No	Yes	No
Rosen et al., 2016	Jan. 2010 - Jun. 2011	Cross-sectional	Community	Population-based	19-85	Oral rinse samples	Roche Linear Array	Moderate	845	7.7%	2.1%	NR.	N/A	NR.	NR.	N/A	NR.	NR.	Yes	Yes	No	No	No	No	No
Bui et al., 2015	Aug. 2013 - Oct. 2013	Cross-sectional	Primary care (Hospital)	Convenience	18-45	Oral rinse samples	MY09/11 PCR and genotyping by hybridisation onto ELISA plates	Low	126	24.6%	16.6%	2.4%	126	24.6%	2.4%	N/A	NR.	NR.	Yes	Yes	Yes	Yes	Yes	No	No

										HPV prevalence estimates									Inclusion in analyses						
Study	**Study period**	**Study design**	**Study population/ setting**	**Selection of study population**	**Age range**	**Sample collection methods**	**HPV detection methods**	**Risk score**	**Sample size**	**Overall HPV prevalence**	**HR-HPV prevalence**	**HPV-16 prevalence**	**Female (n)**	**Female HPV prevalence**	**Female HPV-16 prevalence**	**Male (n)**	**Male HPV prevalence**	**Male HPV-16 prevalence**	**Overall HPV**	**Overall HR-HPV**	**Overall HPV-16**	**Female**	**Female HPV-16**	**Male**	**Male HPV-16**
Skoczyński et al., 2015	Mar. 2009 - Jun. 2011	Cross-sectional	Primary care (Hospital)	Convenience	NR	Buccal mucosa smear	MY09/11 PCR then type-specific PCR	Moderate	152	12.5%	NR	NR	152	12.5%	NR	N/A	NR	NR	Yes	No	No	Yes	No	No	No
Wimardhani et al., 2015	Nov. 2013 - Jan. 2014	Pilot study	University population	Convenience	18+	Saliva samples	HPV-16/-18 type-specific PCR (HPV-16 & -18), then sequencing	Moderate	200	6.0%	6.0%	4.0%	200	6.0%	4.0%	N/A	NR	NR	Yes	Yes	Yes	Yes	Yes	No	No
Ernster et al., 2009	Jan. 1997 - Dec. 2001	Cross-sectional	Primary care (Hospital)	Convenience	21+	Non-cancerous palatine tonsils	Type-specific qPCR	Low	107	0.0%	0.0%	0.0%	59	0.0%	0.0%	48	0.0%	0.0%	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sindvadi Ramesh et al., 2021	Mar. 2019 - Dec. 2019	Cross-sectional	University population	Convenience	18+	Saliva samples/ oral rinse	MY09/11 PCR	High	200	0.0%	0.0%	0.0%	NR	NR	NR	NR	NR	NR	Yes	Yes	Yes	No	No	No	No
Guan et al., 2021	NR	Cross-sectional	University population	Convenience	20-25	Oral rinse samples	PGMY09/11 consensus screening and genotyping using the Molgentix f-HPV genotyping kit	Moderate	452	0.7%	0.0%	0.0%	123	0.8%	0.0%	329	0.6%	0.0%	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Combes et al., 2017	Feb. 2012 - Jun. 2016	Cross-sectional	Primary care (Hospital)	Convenience	25+	Tonsil brushings (tonsillectomies)	Type specific E7 PCR bead-based multiplex genotyping assay	Low	384	3.4%	NR	NR	Can't interpret - backed data for adults	NR	NR	Can't interpret - backed data for adults	NR	NR	Yes	No	No	No	No	No	No
Ribeiro et al., 2019	Jan. 2013 - Mar. 2014	Cross-sectional	Primary care (clinic based)	Convenience	18+	Oral brushing	Nested PCR (MY09/11, GP5+/GP6+) and HPV-6, -11, -16, -18 type-specific PCR	High	98	81.6%	32.7%	4.1%	Can't interpret - backed data for adults	NR	NR	Can't interpret - backed data for adults	NR	NR	Yes	Yes	Yes	No	No	No	No
Rob et al., 2022	Jan. 2018 - Feb. 2020	Cross-sectional	Primary care (Hospital - dermatology)	Convenience	18-64	Oral rinse and gargle samples	GP5+/6+ PCR and reverse line blot hybridisation or sequenced	Moderate	157	14.0%	NR	NR	Can't interpret	NR	NR	Can't interpret	NR	NR	Yes	No	No	No	No	No	No
Cho et al., 2020	Dec. 2016 - Mar. 2017	Cross-sectional	Primary care (Hospital)	Convenience	20-69	Oropharyngeal 'washing'	Nested PCR (MY09/11, GP5+/GP6+), sequencing, GENOSEARCH HPV-31 typing	Moderate	435	5.7%	4.4%	1.4%	193	2.6%	NR	242	8.3%	NR	Yes	Yes	Yes	Yes	No	Yes	No
Dalla Torre et al., 2015	NR	Cross-sectional	Primary care (Hospital - Craniomaxillofacial and Oral Surgery)	Convenience	19-50	Oral brush smear	HC2-HPV-DNA-testing (QIAGEN, NL) using Hybrid Capture 2	Moderate	214	42.1%	24.8%	NR	Can't interpret - backed data for non-cancerous	NR	NR	Can't interpret - backed data for non-cancerous	NR	NR	Yes	Yes	No	No	No	No	No
Davidson et al., 2014	2012	Cross-sectional	Community	Convenience	40+	Oral rinse samples	Roche Linear Array	Moderate	87	5.7%	2.3%	1.1%	N/A	NR	NR	214	5.7%	1.2%	Yes	Yes	Yes	No	No	Yes	Yes
Malerova et al., 2020	Sept. 2015 - Dec. 2017	Cross-sectional	University population (and participants of other studies)	Convenience	18+	Oral rinse samples	Biotin-labelled GP5+/6+ then reverse line blot hybridisation	Low	294	2.0%	1.0%	0.3%	210	2.9%	0.5%	84	0.0%	0.0%	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Enerly et al., 2019	Sept. 2016 - Feb. 2017	Cross-sectional	Community	Population-based	18-20	Oral swab	Multiplex HPV genotyping (based on GP5+/6+ primers followed by hybridisation of type-specific oligonucleotide probes)	Low	312	1.3%	0.3%	0.3%	312	1.3%	0.3%	N/A	NR	NR	Yes	Yes	Yes	Yes	Yes	No	No
Herrmn et al., 2022	Jan. 2016 - Dec. 2018	Cross-sectional	Primary care (Hospital - dermatology and STI clinics)	Convenience	18+	Oral brush (cytobrush)	PCR (primers not recorded)	Low	116	31.0%	15.5%	NR	116	31.0%	NR	N/A	NR	NR	Yes	Yes	No	Yes	No	No	No
Zargami and Azizi Jalilian, 2020	NR	Cross-sectional	Primary care (dentist)	Convenience	19-65	Oral gargle/rinse	PCR (primers not recorded) then INNO-LiPA HPV genotyping assay	Low	300	12.0%	0.3%	0.0%	143	10.5%	0.0%	157	13.4%	0.0%	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Napolitano et al., 2024	Nov. 2022 - Sept. 2023	Cross-sectional	University population	Cluster	18-30	Saliva samples	Anyplex™II HPV 28 Detection system	Low	1002	0.8%	0.7%	NR	690	0.9%	NR	312	0.6%	NR	Yes	Yes	No	Yes	No	Yes	No
Kofler et al., 2020	Oct. 2019 - Dec. 2019	Prospective case-control study	Primary care (Hospital - maxillofacial outpatients)	Convenience	45+	Oropharyngeal brushing	Single-step PCR (5 µL of purified nucleic acid), genotyping followed using allele-specific reverse line-blo hybridisation of the PCR products	Moderate	112	1.8%	0.0%	0.0%	19	NR	0.0%	93	NR	0.0%	Yes	Yes	Yes	No	Yes	No	Yes
Wood et al., 2020	NR	Cross-sectional	Primary care (Hospital - dental and HIV clinics	Convenience	20-74	Oral gargle/rinse	Abbott m2000™ RealTime system, Riatol qPCR HPV genotyping assay; AML lab-developed assay	Moderate	221	3.6%	2.2%	0.9%	116	1.7%	0.9%	101	5.9%	1.0%	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Mehanna et al., 2019	2013 - 2015	Cross-sectional	Primary care (Hospital - tonsillectomies)	Convenience	18-56	Oral rinse, oral brushings, tonsillectomy	10-primer HPV short PCR fragment (SPF10) PCR DNA enzyme immunoassay (DEIA)–line probe assay (LiPA25) version 1, then E6-based multiplex type-specific system (MPTS123)	Low	168	17.3%	8.3%	4.8%	84	15.5%	2.4%	84	19.1%	7.1%	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Dalla Torre et al., 2019	Jan. 2008 - Dec. 2012	Cross-sectional	Primary care (dentist - hospital based)	Convenience	23-46	Oral brush smear	Hybrid capture 2	Low	187	20.9%	14.4%	NR	91	20.9%	NR	96	20.8%	NR	Yes	Yes	No	Yes	No	Yes	No

										HPV prevalence estimates									Inclusion in analyses						
Study	**Study period**	**Study design**	**Study population/ setting**	**Selection of study population**	**Age range**	**Sample collection methods**	**HPV detection methods**	**Risk score**	**Sample size**	**Overall HPV prevalence**	**HR-HPV prevalence**	**HPV-16 prevalence**	**Female (n)**	**Female HPV prevalence**	**Female HPV-16 prevalence**	**Male (n)**	**Male HPV prevalence**	**Male HPV-16 prevalence**	**Overall HPV**	**Overall HR-HPV**	**Overall HPV-16**	**Female**	**Female HPV-16**	**Male**	**Male HPV-16**
Ortiz et al., 2018	Sept. 2014 - May 2016	Cross-sectional	Community (SOALS)	Established cohort (SOALS)	40-70	Oral rinse samples	MY09/11 PCR then dot-blot hybridisation of PCR products	Moderate	735	5.7%	NR.	NR.	531	3.9%	NR.	204	10.3%	NR.	Yes	No	No	Yes	No	Yes	No
de Souza et al., 2018	NR.	Cross-sectional	Primary care	Convenience	20-70	Oral rinse samples	Nested PCR (MY09/11, GP5+/GP6+) and sequencing	Low	96	11.5%	NR.	NR.	49	NR.	NR.	47	NR.	NR.	Yes	No	No	No	No	No	No
Shigeishi et al., 2018	Nov. 2016 - Aug. 2017	Cross-sectional	Primary care (Hospital)	Convenience	20-97	Oral rinse samples	HPV-16 E6 qPCR	Low	124	75.8%	75.8%	75.8%	76	73.7%	73.7%	48	79.2%	79.2%	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Laprise et al., 2017	Jan. 2005 - Nov. 2013	Case-control study	Primary care (Hospital)	Convenience	18+	Oral rinse samples	Roche Linear Array	Moderate	422	14.5%	9.5%	2.4%	132	NR.	NR.	297	NR.	NR.	Yes	Yes	Yes	No	No	No	No
Raymond et al., 2018	2009 - 2014	Cross-sectional	Community (NHANES)	Population-based	18-69	Oral gargle/rinse	Roche Linear Array	Low	13981	7.5%	4.2%	0.9%	7102	NR.	NR.	6879	NR.	NR.	Yes	Yes	Yes	No	No	No	No
Marais et al., 2006	2003	Cross-sectional	Primary care (dentist)	Convenience	20-61	Mucosal brushing	Nested PCR (degenerate primers)	Low	116	3.5%	0.9%	0.9%	69	2.9%	1.5%	47	4.3%	0.0%	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Smith et al., 2004	Sept. 1997 - Nov. 2000	Cross-sectional	Primary care (Hospital - obstetrics)	Convenience	NR.	Oral rinse samples	MY09/11 PCR and a primer to improve HPV51 detection (PCR), then sequencing for typing	Moderate	645	2.8%	1.9%	0.5%	577	2.4%	0.5%	68	5.9%	0.0%	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Brouwer et al., 2022	Apr. 2015 - Dec. 2017	Prospective longitudinal study	University population	Convenience	18+	Saliva samples	PCR MassArray	Low	338	10.1%	6.5%	NR.	228	10.5%	NR.	110	9.0%	NR.	Yes	Yes	No	Yes	No	Yes	No
Bettampadi et al., 2020	Mar. 2005 - Dec. 2009	Prospective longitudinal study	Community (HIM study)	Convenience	18-73	Oral gargle specimen	SPF10/DEIA/LiPA25 system	Moderate	3098	8.8%	5.9%	1.7%	N/A	NR.	NR.	3098	8.8%	1.7%	Yes	Yes	Yes	No	No	Yes	Yes
Kero et al., 2012	1998 - 2002	Prospective longitudinal study	Primary care (Hospital)	Convenience	19-46	Oral scrape	Nested PCR (MY09/11, GP5+/GP6+) then hybridisation for viral type determination	Low	455	17.4%	12.3%	9.2%	N/A	NR.	NR.	131	18.3%	6.1%	Yes	Yes	Yes	No	No	Yes	Yes
Rautava et al., 2012	1998 - 2002	Prospective longitudinal study	Primary care (Hospital)	Convenience	18-46	Oral scrape	Nested PCR (MY09/11, GP5+/GP6+) then hybridisation for viral type determination	Low					324	17.0%	10.5%	N/A	NR.	NR.	Yes			Yes	Yes	No	No
D'Souza et al., 2023	Apr. 2017 - Jan. 2022	Multi-centre prospective study	Community (MOUTH study)	Convenience	24-86	Oral gargle/rinse	SPF10 PCR, DEIA, Cobas, Aptima for genotyping	Moderate	1105	N/A	7.3%	1.8%	115	NR.	0.0%	993	NR.	2.0%	No	Yes	Yes	No	No	No	No
D'Souza et al., 2009	2000 - 2007	Case-control cross-sectional	University population and Primary care (hospital case-control)	Convenience	18-65+	30 second oral rinse	Multiplex PCR by PGYMD9/11 followed by Roche line blot hybridisation	Low	542	4.1%	2.8%	0.4%	81	1.2%	NR.	461	4.6%	NR.	Yes	Yes	Yes	Yes	No	Yes	No
Pickard et al., 2012	Sept. 2009	Longitudinal cohort study	University population	Convenience	18-30	Oral rinse samples	Roche Linear Array	Low	1000	2.4%	1.2%	0.2%	537	1.7%	NR.	463	3.2%	NR.	Yes	Yes	Yes	Yes	No	No	No

4.2. Meta-analysis of oral HPV prevalence estimates amongst the adult population globally

The following comparisons were selected to generate prevalence estimates: overall HPV prevalence, overall HR-HPV prevalence, and overall HPV-16 prevalence. In all analyses, data were weighted by inverse variance using the Freeman Tukey double arcsine formula, which was selected to enable weighting of percentage data, and back transformed to percentages before an combined inverse weighted analysis was performed to generate the prevalence values for each group presented.

To determine the overall global oral HPV prevalence amongst the healthy population, we identified 88/89 studies that reported on oral HPV prevalence from 85,267 individuals, for inclusion in this analysis; one study was excluded as the authors only reported on HR-HPV prevalence, despite using sampling methods that also allowed for detection of multiple LR-HPV types (D'Souza et al., 2023). We found the prevalence of oral HPV within these studies to be 5.044% (95% CI 4.897-5.193%) (Figure 14).

Next, we scrutinised the prevalence of HR-HPV viral types. To calculate overall HR-HPV prevalence, 77 studies that reported on the HR-HPV status from 83,041 participants were included in the analysis, and we found the prevalence of HR viral types within these studies to be 2.133% (95% CI 2.033-2.235%). However, the definition of HR-HPV types varied dependent on the individual report of data extraction. In some instances, studies had reported on just HPV-16, or HPV-16 and HPV-18 prevalence for example, whilst other studies gave a value for HR-HPV within their cohort but did not provide the exact viral types included. Moreover, where studies did not provide an HR-HPV prevalence, the data within the paper were split to determine HR-HPV prevalence based on the numbers reported for following viral types: HPV-16, -18, -31, -33, -35, -39, -45, -51, -52, -56, -58, -59, -66, -68, and -70 (Burd, 2003). As a result of these potential discrepancies in viral types considered HR-HPV between the reports, the individual reports pooled to generate an overall HR-HPV prevalence do not provide data on the same number of HR viral types.

Finally we investigated the most dominant viral type in HPV-OPC, HPV-16 within the healthy population. The prevalence of HR HPV-16 was found to be 0.724% (95% CI 0.661-0.790%), with 69 studies that reported on 79,548 participants included in this calculation.

4.2.1. Oral HPV infection in males is almost twice as prevalent than within females

To examine the prevalence of HPV (regardless of viral type), and HR HPV-16 within the oropharynx and oral cavity of males and females, the final 89 studies that fitted the inclusion criteria were filtered to identify studies that specifically reported oral prevalence rates for male HPV, female HPV, male HPV-16 and female HPV-16. Please note, that for all four prevalence comparisons here, males and females were included based on biological sex.

For the meta-analysis of overall oral HPV prevalence in males, 48 studies that reported on oral HPV prevalence in 25,400 males were included and the estimate prevalence was 5.224% (95% CI 4.949-5.506%). We found that oral HPV-16 prevalence was approximately 0.966% (95% CI 0.809-1.135%) within males, based on reports from 25 studies that reported on oral HPV-16 within 16,649 male individuals. Interestingly, these data were both significantly higher than the prevalence estimates calculated for any HPV and HR HPV-16 within females (Figure 14).

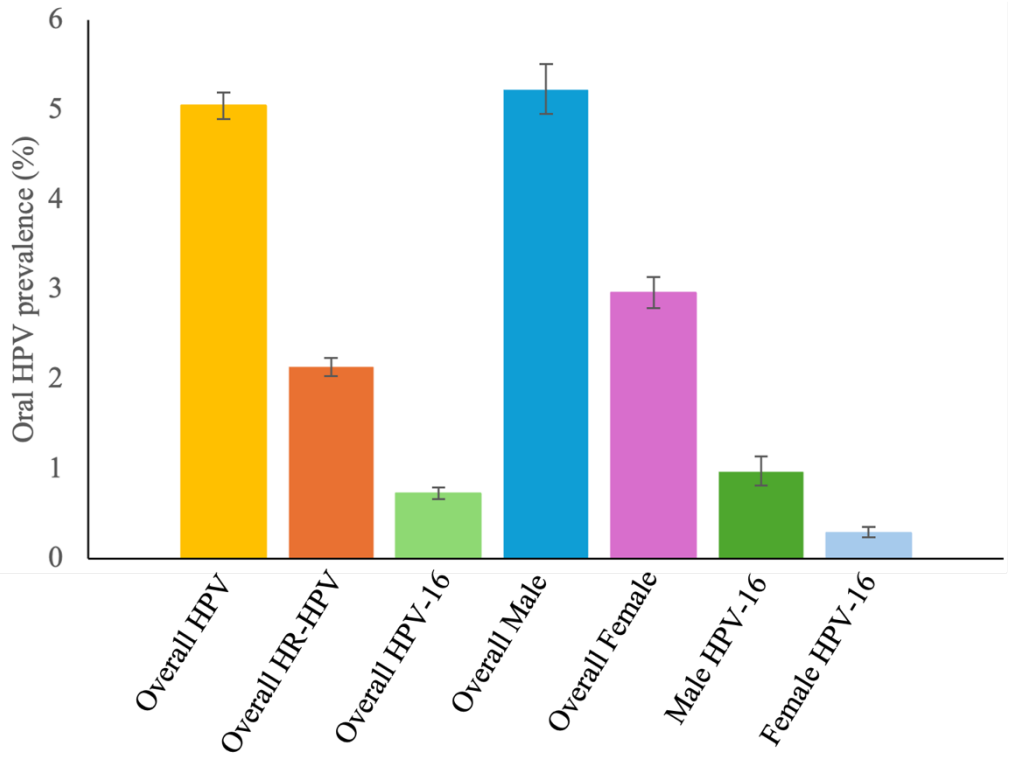


Figure 14. Oral HPV infection is almost twice as prevalent in males than females, with males over three times more likely to have an oral HR HPV-16 infection. Global oral HPV prevalence estimates calculated from studies selected in line with the meta-analyses criteria are shown for overall HPV, overall HR-HPV, overall HPV-16, and then the overall oral HPV prevalence for males, females and the prevalence of HR HPV-16 within males and females. Estimate of oral HPV prevalence within males was calculated at the highest overall percentage, almost twice as prevalent than within females. Prevalence estimates were calculated in Microsoft Excel and Minitab (v 21.4) using an combined inverse variance weighted analysis. Data are presented as a percentage prevalence with 95% CI intervals.

Taken together, these data indicate that based on 89 reports of oral HPV prevalence within the healthy adult population, the overall prevalence of oral HPV is approximately 5%. We found that prevalence estimates between all individual comparative groups analysed were significantly different from each other, with the exception of the calculated overall prevalence and overall oral HPV prevalence within males, which are indistinguishable. Moreover, our data show that men are nearly twice as likely to have an oral HPV infection than females (odds ratio: 1.805). Interestingly, the HPV-OPC dominating viral type HR HPV-16 is more than twice as likely to be detected in the oropharynx/oral cavity of males than females (odds ratio: 3.347).

4.3. Discussion

Here, we performed a systematic review and meta-analysis of reports of oral HPV prevalence within healthy individuals and describe an overall prevalence of 5.044% for any HPV viral type (95% CI: 4.897-5.193%). Our estimate supports those of several studies that too describe an oral HPV prevalence of ~5-6%, when analysing for a wide variety of HPV viral types (Ragin et al., 2007, Davidson et al., 2014, Hang et al., 2014b, Conway et al., 2016, Ortiz et al., 2018a, Cho et al., 2020, de la Garza-Ramos et al., 2020, Gheit et al., 2023, Giuliano et al., 2023). However, it is of note that whilst our data of overall oral HPV prevalence is similar to these individual studies, reports from others are extremely varied, and analyses at a population-level remains imperative to understanding nuances between studies. Indeed, we appreciate that such discrepancies in individual reports, which may be a result of a number of factors (i.e., variation in research methodology or geographical location investigated), would influence the overall prevalence estimates described in this chapter.

Other meta-analyses of oral HPV prevalence have been published that included data up to 2017, but to our knowledge, this is the most up-to-date estimate and our systematic review identified a total of 17 studies that reported on novel data from 2018 onwards for inclusion in one or more meta-analyses comparisons. Our finding for global oral HPV prevalence of any HPV viral type is strongly supported by Mena *et al.*, 2019, who described a strikingly similar global prevalence of 4.9% within healthy teenagers and adults (Mena et al., 2019), this suggesting a minimal increase from the final year of its study inclusion criteria (2015) to 2024, as included in our analysis. Another study that also included only data on prevalence within individuals over 18 (collected between 1995-2017) found 7.7% prevalence of oral HPV (Tam

et al., 2018); however, this does not necessarily indicate a decrease in prevalence with respect to our oral HPV prevalence estimate, as there is likely variation introduced due to the studies included; however, their inclusion criteria are similar, so this may be an emerging trend, perhaps a result of HPV vaccination programmes, and this would need to be monitored over time.

Investigations of HPV-16 within the healthy population are salient given its domination within HPV-positive cancers of the oropharynx, as well as being accepted as a causative agent of HPV-OPC development. Oral HPV-16 prevalence has previously been reported between 1.0-1.4% (Tam et al., 2018, Mena et al., 2019); however, we found HR HPV-16 prevalence to be less than 1% (0.724%), and although this a marginal decrease, it is possible that this may represent increased herd immunity within several populations following the HPV vaccination programme, of which increasing numbers of young people vaccinated against certain viral types have now reached adulthood.

Regardless of viral type, men are statistically more likely to have an oral HPV infection (Gillison et al., 2012, Tam et al., 2018), due to a variety of factors including possible discrepancies in immune responses (Chaturvedi et al., 2015), or behaviours such as reports of more total sexual partners (D'Souza et al., 2016, Pauli et al., 2022). Indeed, in this work we corroborate this finding by unsurprisingly reporting males to be twice as likely to have an oral HPV infection than females. Consistently with other reports of HPV-16 specifically being more prevalent in healthy males (Gillison et al., 2012, Riddell et al., 2022, Giuliano et al., 2023), we found males to be three times more likely to have an oral HPV-16 infection than females, which supports the burden of HPV-associated disease to males (Guo et al., 2022). In future work, it would be interesting to perform these analyses (i.e., oral HPV prevalence in males) based on geographic and demographic factors.

4.3.1. Conclusion

To conclude, these data indicate that HPV is present within the oral cavity/oropharynx of healthy individuals globally at a reasonable level prevalence (5.044%). Importantly, that as with HPV-OPC, both the likelihood of having any type of HPV as well as an HR HPV-16 infection is significantly more within males than females, which has significant implications for public health strategies to prevent HPV-mediated disease.

CHAPTER 5: HPV prevalence within the non-cancerous tonsil tissue of routine tonsillectomy patients.

5.0. Introduction

In recent years, the cases of HPV-positive OPCs have been increasing (Mourad et al., 2017), and male OPC cases have now overtaken the incidence of cervical cancer in women (Lechner et al., 2022). In addition to the significance of HPV in OPCs, it is apparent that the virus is often present within the non-cancerous oropharynx and oral cavity of healthy individuals. The overall oral HPV prevalence varies between studies, being dependent on several experimental and demographic factors; however, it has quite recently been shown to be as high as 35% within saliva samples (Jamieson et al., 2020) and from our own meta-analysis, we believe the overall global oral prevalence of HPV to be approximately 5% (Chapter 4).

The tonsillar crypts are the assumed site of HPV-OPC development (Kim et al., 2007). Given this, it is surprising that HPV has not been found within the non-cancerous tonsil in some reports, being described retrospectively at a prevalence of 0% (Ernster et al., 2009, Palmer et al., 2014). However, when HPV has been detected in these tissues, it is within the relatively low region of 1.9 - 4.9% (Combes et al., 2017, Rieth et al., 2018, Mehanna et al., 2019b). Thus, despite the rising number of HPV-OPCs, there is little knowledge of the pathogenesis of the virus within the non-cancerous oropharynx.

To address this knowledge gap, we selected the palatine tonsils as the basis for our prevalence investigation. These tissues were collected as part of our established clinical study (section 5.1). To begin to elucidate the pathogenesis of an HPV infection within the oropharynx, we firstly defined the HPV-status of the non-cancerous tonsil tissue of this prospective sample. We used our modified HPV-DNA qPCR assay (optimisation shown in Appendix D; details of methodology and primers sequences shown in Chapter 3) to assess the presence of HPV within tonsil tissue and categorise our sample (of adults undergoing routine tonsillectomy surgery) as HPV-positive or HPV-negative. We hypothesised that HPV would likely be detectable in some of the non-cancerous tonsil tissue of our adult population, and that we could use HPV-positive biospecimens to subsequently investigate viral pathogenesis in a non-cancerous setting. We hoped to improve knowledge of the natural history of the virus in the oropharynx, prior to potential malignant changes.

In this chapter, we describe an overall prevalence of oropharyngeal HPV by qPCR with consensus primers (GP5+/GP6+) within the non-cancerous tonsil tissue of our sample. We outline the differences between the detection of HPV in tonsil tissue stored as fresh-frozen or formalin-fixed. Furthermore, we begin to explore HPV pathogenesis in the oropharynx and describe the viral types present in HPV-positive tissues, as well as histologically characterising the tissues, and the localisation of important biomarkers.

5.1. Final study sample

The final sample included in this project consisted of adults undergoing routine tonsillectomy surgeries, whose tonsil tissue was required to be confirmed as non-dysplastic by a pathologist at one of the following hospitals:

- University Hospitals of Derby and Burton (UHDB)
- University Hospital Leicester (UHL)
- University Hospital Nottingham (UHN)
- Princess Royal Hospital, Telford (PRH)
- Calderdale Royal Hospital (CRH)

The convenience sample was generated by all eligible individuals (those over 18 years old, undergoing routine tonsillectomy surgery) being invited to participate in the study; to reduce the introduction of participant bias, there were no targeted sampling methods used. Left and right palatine tonsils, an oral swab, and a self-reported lifestyle questionnaire were collected from all consenting adults. Tissues were collected between 2016-2024, in line with the clinical study (REC reference: 16/SW/0079, IRAS project ID: 194857; Appendix A). Recruitment to the clinical study is ongoing; at present, the total number of participants recruited at any of the above hospitals is 156. Of these, 110 participants are included in this analysis. These participants had DNA extracted from multiple sections of their left and right tonsils, and gDNA from each section was subjected to beta-actin quality control screening. There was a significant gap in tissue collection between the beginning of 2020 to the end of 2021 following government-imposed Covid-19 guidelines at the time. As part of our contingency measures, we purchased tonsil tissues from the University of Birmingham's HBRC during this time, and these tissues are included within the final sample. Thus DNA was also extracted from an

additional 20 tissues (10 participants), purchased through the HBRC. Therefore the total participant number, prior to DNA quality assessment was 120.

5.1.1. Beta-actin qPCR for DNA integrity analysis

To determine tonsil DNA samples with viable material for HPV-DNA qPCR consensus screening, the presence of housekeeping gene beta-actin was assessed as a quality control. All gDNA isolated from sections across tonsil tissues was screened for beta-actin gene using either a quantitative or a standard PCR approach (individual samples screened: N=914). 474 tonsil DNA samples passed beta-actin screening by qPCR, as indicated by amplification within the upper and lower boundaries of the standard curve of known copy number (Figure 15). A further 428 samples passed beta-actin quality control checks by standard endpoint PCR, indicated by a band at ~166 bp by gel electrophoresis (Figure 16). 12 DNA samples failed beta-actin screening, and so were excluded from further analysis due to lack of sufficient viable DNA; together these correspond to the exclusion of 2 participants from the study.

As expected, we found that the detection of beta-actin by qPCR was much later (indicated by the PCR cycle in which gene amplification reached threshold) in gDNA isolated from tissues fixed in 10% NBF for long periods in comparison to that of those subjected to 24-hour 10% NBF fixation (Figure 17). As a result, depending on the method of tissue storage, different DNA concentrations were added per reaction for beta-actin PCR analysis as part of our extensive optimisation of DNA extraction and HPV-DNA qPCR screen (detailed within Chapter 2 and Appendix D, respectively).

The total number of samples with detectable beta-actin and therefore taken forward to be screened for HPV-DNA by qPCR was N=902. The number of DNA samples from each tonsil tissue varied, with the number of final participants being 118 individuals (comprising 233 tonsil tissues).

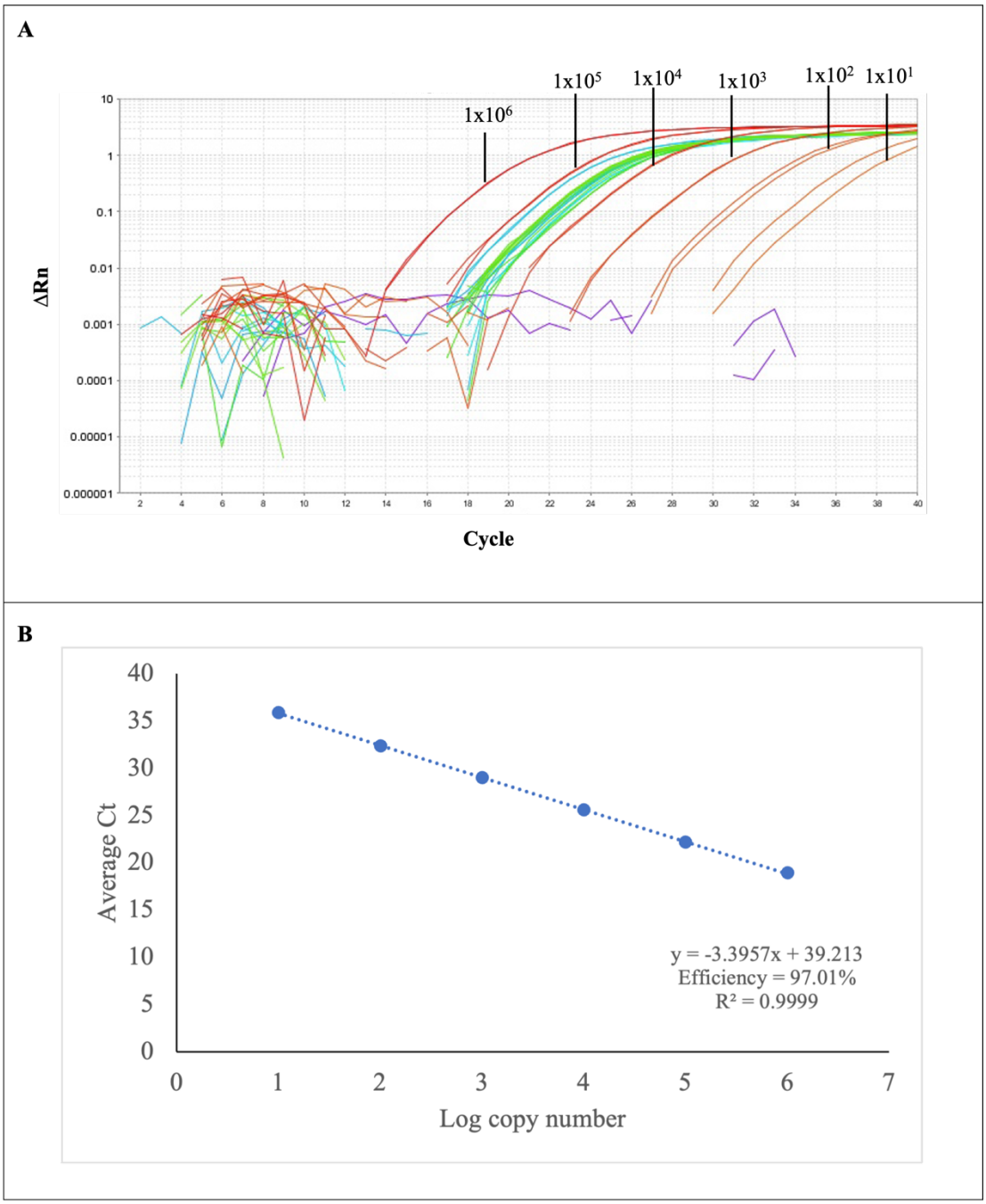


Figure 15. Amplification plot of the beta-actin gene in tonsillar gDNA and standard curve for copy number calculation. (A) Tonsil gDNA samples were assessed for beta-actin amplification by qPCR. Amplification plots were generated using the AB® StepOnePlus™ Real-Time PCR system, and data analysed in their software (v 2.2.2). The normalised fluorescence signal (ΔRn) is plotted against the cycle number. Tonsil gDNA samples were analysed against a standard curve of known beta-actin copy number ($1x10^6$, $1x10^5$, $1x10^4$, $1x10^3$, $1x10^2$, $1x10^1$). In this example, all tonsil gDNA samples [green and blue curves] amplified within the boundaries of the standard curve [red curves]. **(B)** A standard curve of the log value of known beta-actin copy number against average Ct value was generated using Microsoft Excel. The equation of the line, reaction efficiency, and R^2 value is displayed.

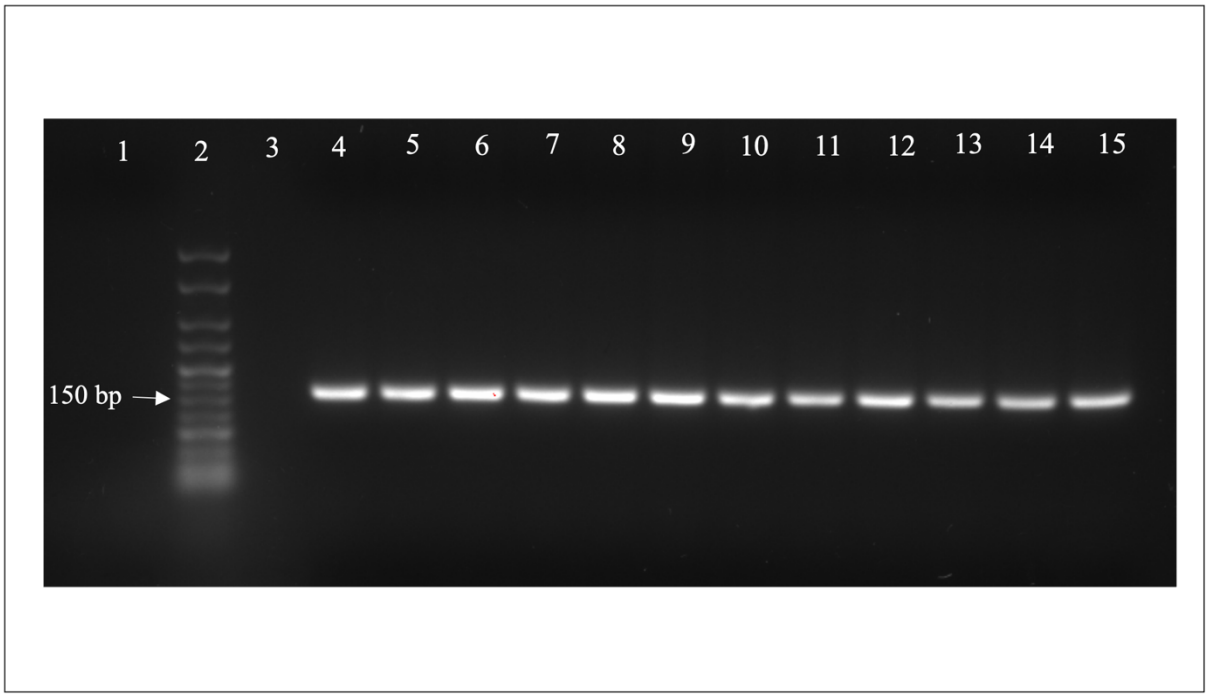


Figure 16. Analysis of standard PCR products for beta-actin amplification by gel electrophoresis. Amplicons produced with ACTB primers and Invitrogen™ Platinum™ II Hot-Start Green PCR master mix [ThermoFisher] using the AB® StepOnePlus™ Real-Time PCR system as a thermocycler were electrophoresed on a 2% agarose gel. A band at approximately 166 bp was used to indicate successful amplification of beta-actin in tonsillar gDNA. Gel images were taken using a ChemiDoc™ MP Imaging System [Bio-Rad].

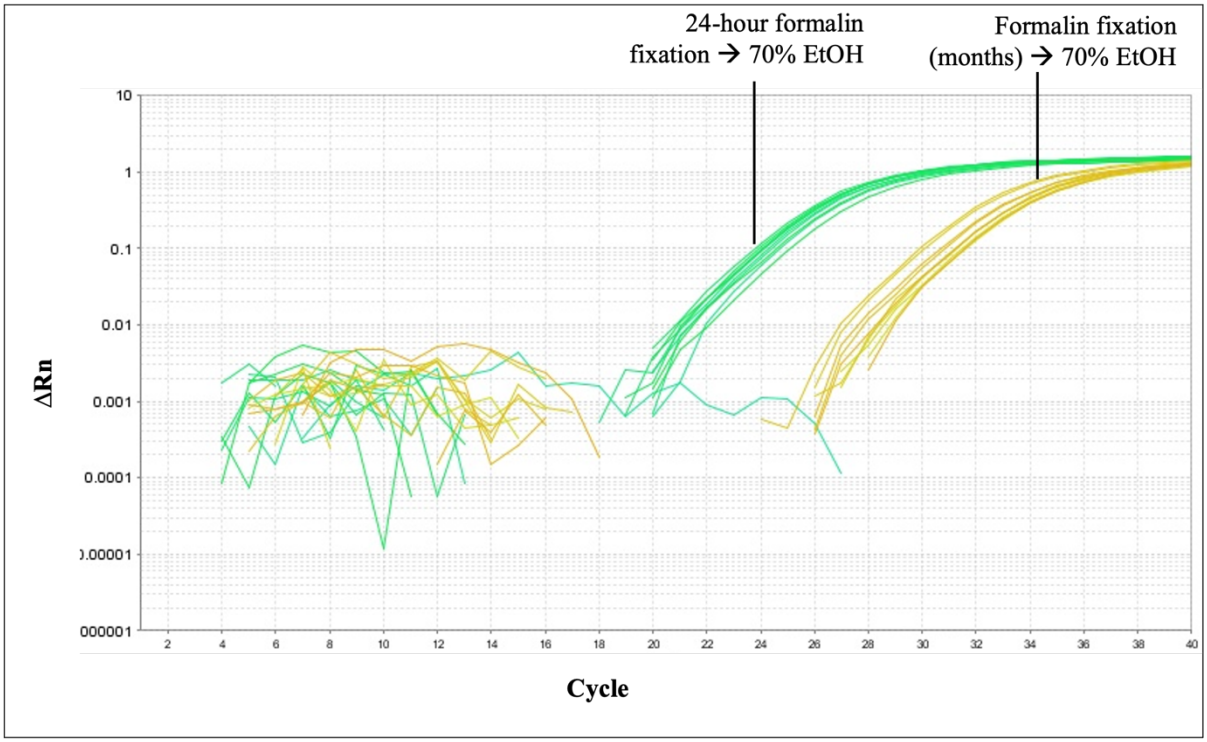


Figure 17. Amplification plot of housekeeping gene, beta-actin, within gDNA isolated from formalin-fixed tonsil tissues. Tonsil gDNA samples were assessed for the presence of beta-actin by qPCR. Amplification plots were generated using the AB® StepOnePlus™ Real-Time PCR system, and data analysed in their software (v 2.2.2). The normalised fluorescence signal (ΔRn) is displayed against the cycle number. DNA from tonsil tissue that had been fixed in 10% NBF for long periods (months) before transfer to 70% EtOH showed a much later cycle for beta-actin amplification than DNA isolated from tissue fixed for 24-hours before 70% EtOH transfer, suggesting poorer DNA quality.

5.1.2. Demographic information of the final sample

Demographic information of all consenting individuals was collected in a self-reported lifestyle questionnaire as part of the clinical study. It is of note that we did not have access to the same information for tonsil tissues obtained from the HBRC, and so these are recorded as 'unknown' for each characteristic. The following demographic information refers only to individuals whose DNA passed integrity checks and were therefore used for the overall oral HPV prevalence analysis. Of the final participants 25.4% were male (N=30/118) and 66.1% female (N=78/118), gender identity was unknown for 8.5% participants (N=10/118) (Figure 18A). The age distribution of the final sample was: 18-25 = 48.3% (N=57/118), 26-35 = 26.3% (N=31/118), 36-45 = 13.6% (N=16/118), 46-55 = 3.4 (N=4/118) and unknown = 8.5% (N=10/118) (Figure 18B). The gender distribution within each age group is represented in Figure 18C. The ethnic grouping of the sample was 66.1% White British (N=78/118), 1.7% Black British (N=2/118), 8.5% White Other (N=10/118), 2.5% British Other (N=3/118), 5.9% British Asian (N=7/118), 1.7% Asian (N=2/118), 3.4% Other (N=4/118), and 10.2% provided no information on their ethnicity (N=12/118, unknown) (Figure 18D). Although a mixture of genders, ages, and ethnicities were represented within the sample, unexpectedly, our sample was dominated by White British females aged 18-35 = 56.8% (N=67/118), as a result, our overall prevalence calculation was based on a high proportion of this specific participant demographic, and the final sample was therefore not an even representation of the target population (all adults over 18).

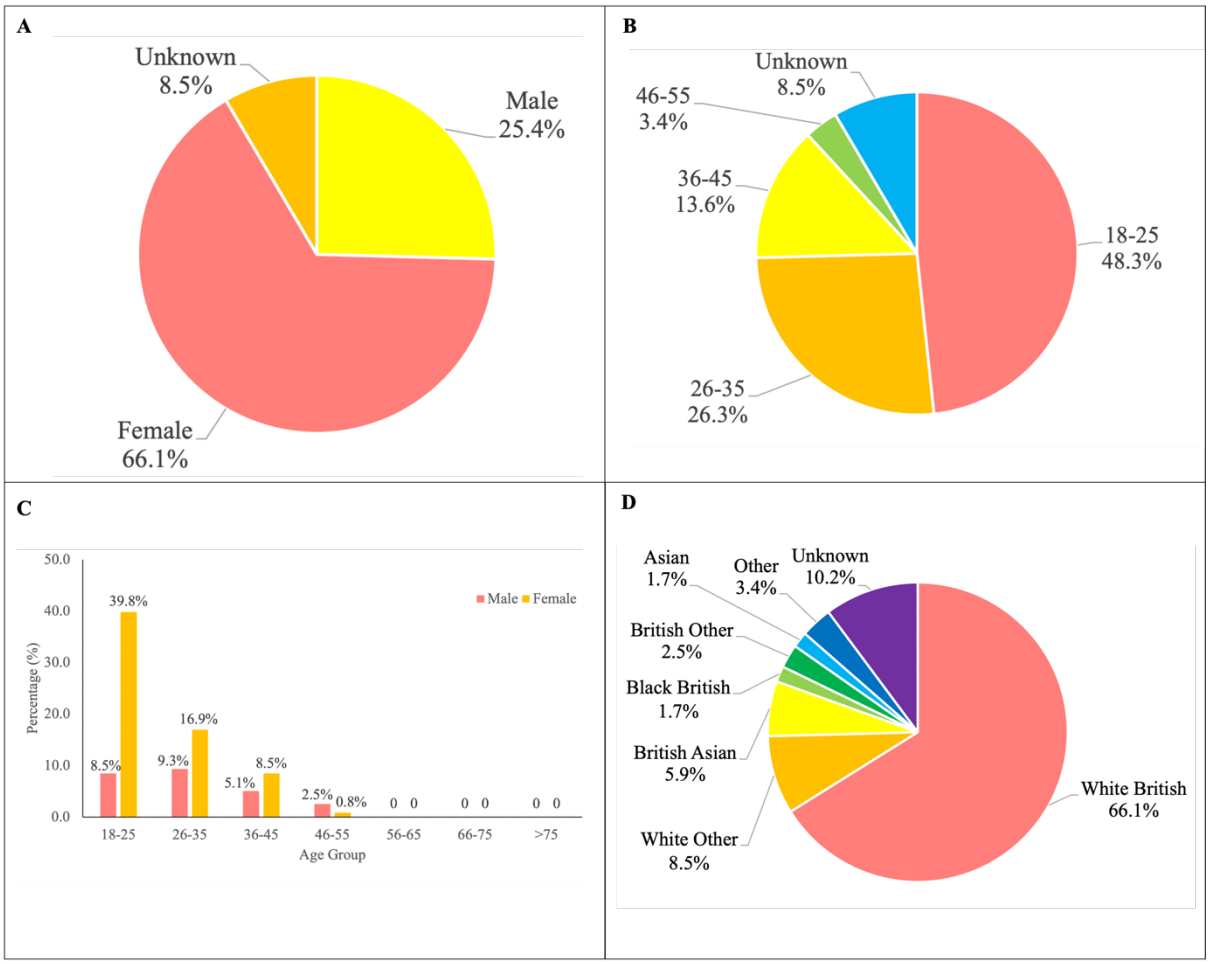


Figure 18. Demographic information of the final sample included in our prospective HPV prevalence analysis. All charts were generated using data provided by study participants in their self-reported lifestyle questionnaire. Percentage values were calculated of the total number of participants, including tissues obtained from the Birmingham Biobank. Tissues from the biobank are considered in the 'unknown' category for all characteristics outlined. **(A)** Participant gender distribution. **(B)** Age distribution of participants; grouped into the following categories: 18-25, 26-35, 36-45, 46-55. **(C)** The percentage distribution of males and females in each age group. **(D)** The ethnic distribution of all final participants.

5.2. Overall HPV prevalence in non-cancerous tonsil tissue by qPCR consensus screening

Next, we investigated the presence of HPV within the non-cancerous tonsil tissue of our sample. Using our modified HPV-DNA qPCR screen, we were able to simultaneously screen samples for multiple viral types.

Samples were considered positive based on a peak in their melt curve analysis that aligned with HPV-18 plasmid control, with peaks consistently present in 2/3 replicates. Although not quantitative, an example of different levels of HPV-positivity, based on amount of PCR product generated as indicated by -Rn value, and HPV-negative gDNA samples (all from fresh-frozen tissue are shown in Figure 19. If 1/3 replicates were positive, these were considered HPV-negative but repeated to confirm HPV-status. SiHa cells were selected as an additional positive control as they are HPV-16 positive, and HPV-16 is the most dominant viral type in HPV-OPC (Gillison et al., 2015). However, given our optimisation work (Appendix D), if the melt curve of PCR products aligned with HPV-16 the sample was initially deemed HPV-negative and assumed to be displaying non-specific primer binding; but, where possible, these samples were still taken forward to be genotyped by INNO-LiPA® assay in case of a genuine oropharyngeal HPV infection.

By consensus screening, we therefore found the overall HPV prevalence to be 10.2% (N=12/118), comprising 4 males, 7 females, and 1 unknown individual (Table 10).

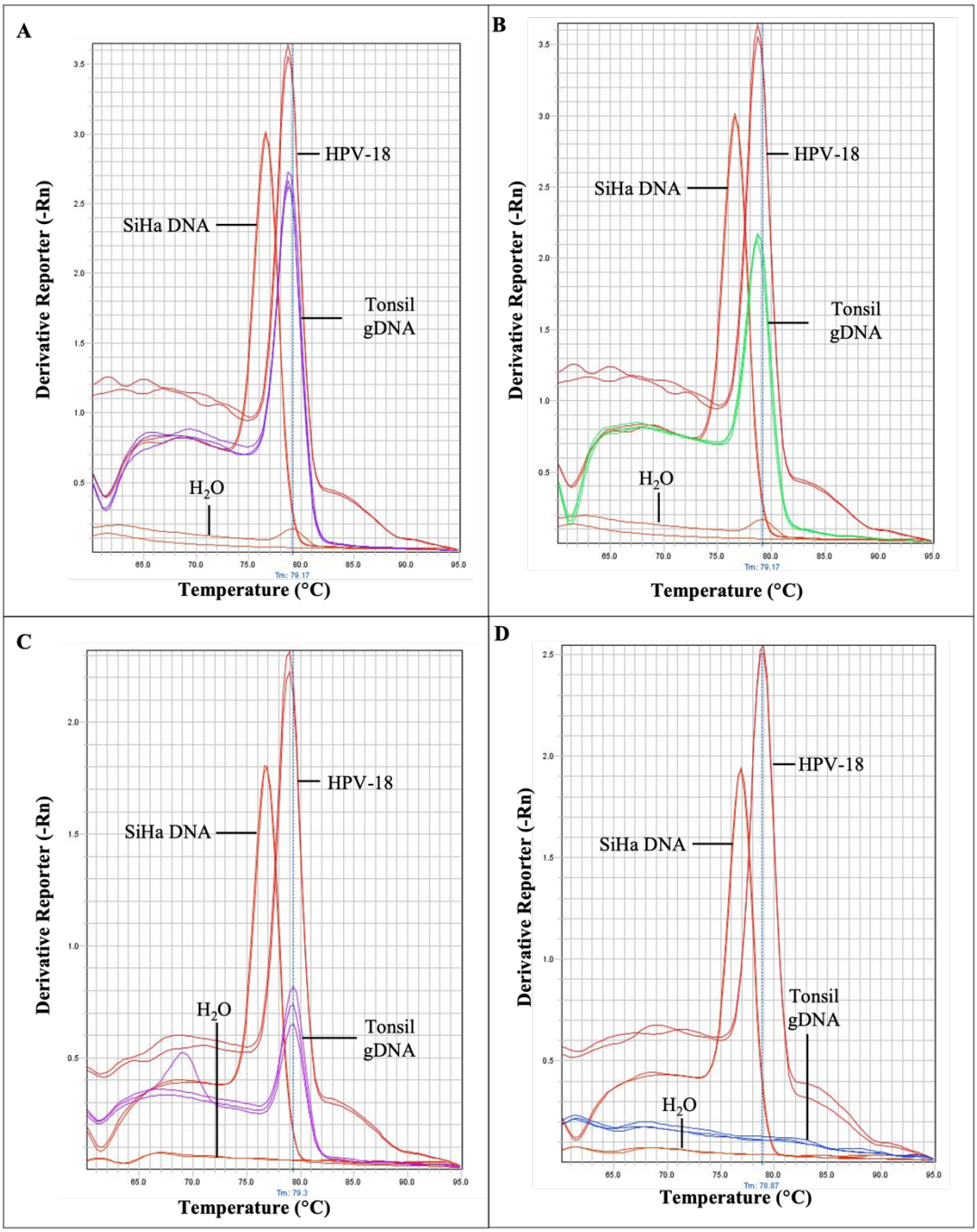


Figure 19. Detection of oropharyngeal HPV infection in DNA from non-cancerous tonsil tissues by qPCR consensus screening. Melt curve analysis of products generated using our modified HPV-DNA qPCR screen. PCR was performed using the AB® StepOnePlus™ Real-Time PCR system. Data was analysed using their software, version 2.2.2. Temperature (°C) is displayed against derivative reporter values (-Rn). In each plot, peaks from HPV controls (HPV-18 plasmid and SiHa DNA), and negative control (H_2O) are annotated. Melt curve analysis of representative HPV-positive tonsil gDNA samples with various amount of PCR product generated, as defined by derivative reporter value **(A)** strongly positive, **(B)** positive, **(C)** weak/possible positive. **(D)** A tonsil gDNA sample with no L1 amplification defined as certainly HPV-negative. All tonsil gDNA samples displayed were isolated from fresh-frozen tonsil tissue.

5.2.1. Demographics and behaviours of individuals with HPV in their palatine tonsils

We explored the characteristics of this HPV-positive group through the data provided in their self-reported questionnaire. We focussed our summary on characteristics known to impact one's risk of oral HPV and/or HPV-OPC. These include gender (men at greater risk), age (bimodal pattern of infection), alcohol consumption, smoking, and certain sexual behaviours (Gillison et al., 2012, Whitton et al., 2024). This revealed that HPV-DNA positive individuals were 4 males (comprising two 26-35, one 36-45, and one 46–55-year-old/s), 7 females (comprising five 18-25 and two 26–35-year-olds), and 1 unknown individual. Most were married or in a long-term relationship (N=8/12); 2 identified as single and 1 in a short-term relationship. 6 individuals identified as heterosexual, 3 as bisexual, 2 as homosexual, and sexuality was unknown for 1 HPV-positive individual. Finally, we found that of those whose ethnicity was known to us, 11 were White British and 1 British Asian individual (Table 10).

Within HPV-positive individuals, we observed a relatively even split between those who currently smoked, and those who did not smoke. Nearly all individuals (N=10/12) indicated that they do currently drink alcohol, and none indicated previously heavier alcohol drinking behaviours. All HPV-positive individuals engaged in open-mouth kissing, with most (N=10/12) stating they have had sexual intercourse in the past year, with the most common lifetime number of sexual partners being 1-5 (N=5/12). Finally, the data collected on HPV revealed that only 1 individual (female) has had a previous HPV infection diagnosed – cervical, CIN unknown – and only 1 individual (female) indicated they had received the HPV vaccination, but true vaccination status of the HPV-positive persons was unclear given that 4/12 were unsure of their vaccination status. Due to the limited size of the HPV-positive group, lifestyle behaviours cannot be statistically analysed against HPV-status; a descriptive summary of these lifestyle characteristics split by gender is instead shown in Table 10.

We found that males were more likely to be HPV-positive than females; with a prevalence of 13.3% compared to 9.0% of the total males compared to of the total number of females, respectively, but this finding was not statistically significant (p=0.49).

5.3. The pathogenesis of HPV within non-cancerous tonsil tissue

We next focussed on understanding the viral pathogenesis (e.g., viral type, concordance in each tonsil, and levels of positivity across a tissue) of the HPV-positive samples.

We assessed the concordance of HPV-positivity in the left and right tonsils of the 12 HPV-positive individuals. We found that 83.3% (N=10) of the HPV-positive cases had detectable HPV in both their left and right tonsils by HPV-DNA qPCR consensus screening. In those cases where HPV-status was non-concordant (N=2), HPV was detected in the right tonsil tissue or this information was unknown. Representative melt curve plots of DNA from left and right tonsil tissues that were either concordant or non-concordant in their HPV-positive status are shown in Figure 20.

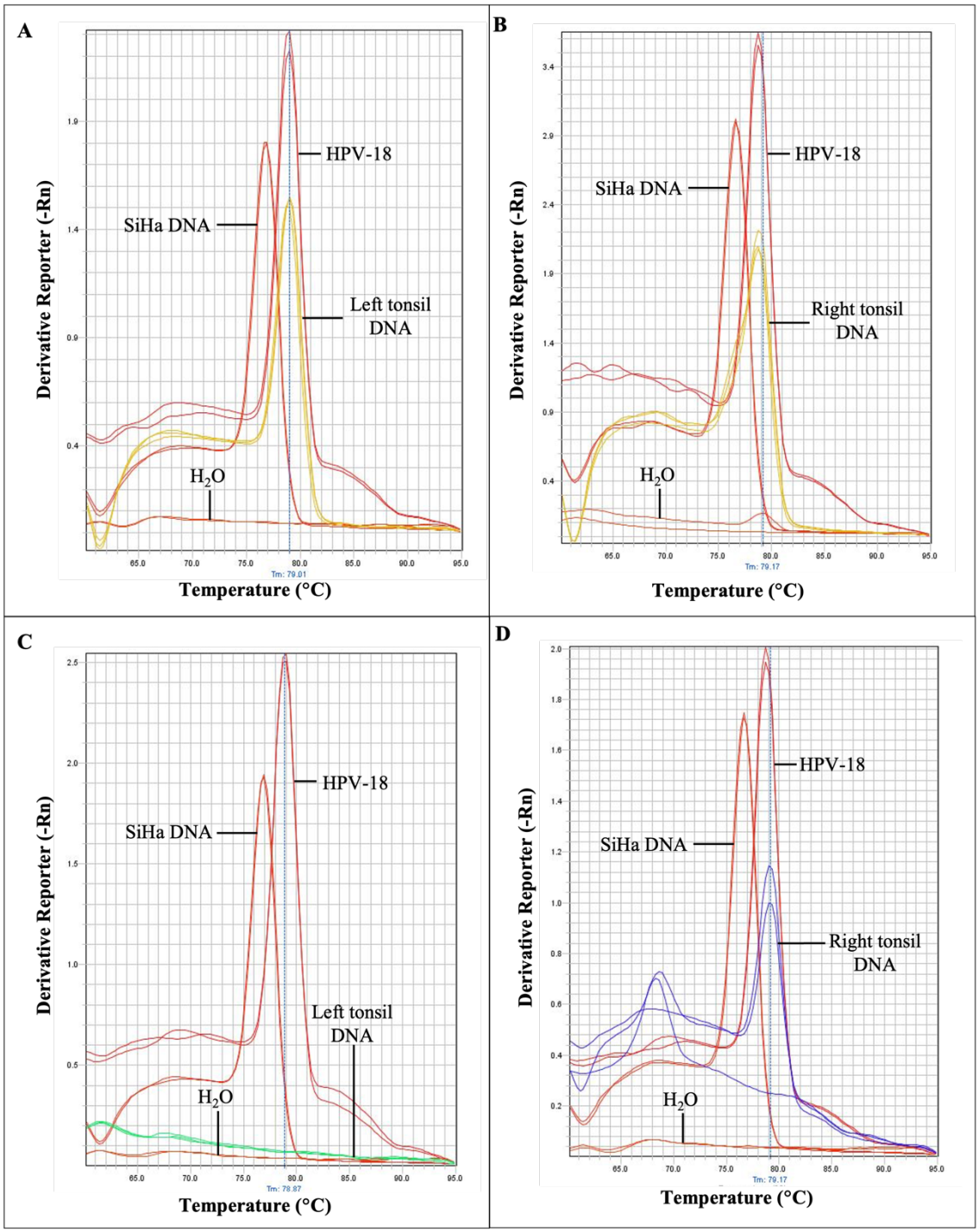


Figure 20. HPV-positivity is mostly concordant in the left and right non-cancerous palatine tonsil tissues. Melt curve analysis of products generated using our modified HPV-DNA qPCR screen. PCR was performed using the AB® StepOnePlus™ Real-Time PCR system. Data was analysed using their software, version 2.2.2. Temperature (°C) is displayed against derivative reporter values (-Rn). Each panel displays melt curve plots from HPV controls (HPV-18 and SiHa DNA) and negative controls (H_2O). Melt curve analysis of PCR products generated from an HPV-positive left tonsil tissue **(A)**, and from the corresponding HPV-positive right tissue of the same participant **(B)**. Melt curve analysis of PCR products generated from an HPV-negative left tonsil tissue **(C)**, and from the corresponding HPV-positive right tissue of the same participant **(D)**.

HPV does not uniformly infect cells of the cervix if an infection is present (Burd, 2003). With this in mind, we considered HPV positivity across a whole tonsil tissue. We hypothesised that HPV would not be present across a tonsil in a similar manner to a viral infection of the cervix. We used the negative derivative reporter value (-Rn) of PCR products generated by qPCR (SYBR green) from different sections of the same tonsil tissue (N=4). We observed that the amount of PCR product generated from viral gene L1 amplification differed between sections of the same positive tissue in nearly all cases, although we were unable to quantify viral load specifically. One HPV-positive tonsil was markedly positive in gDNA taken from all 8 sections. A representative example of varied HPV detection is displayed in Figure 21.

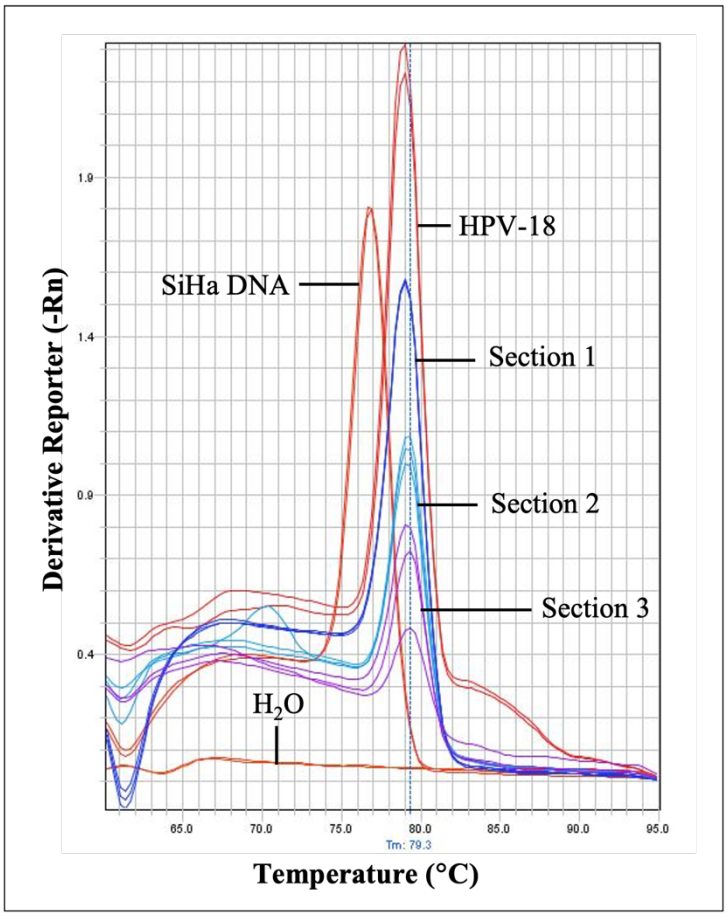


Figure 21. HPV-positivity varies across different sections of the same tonsil tissue. Melt curve analysis of products generated using our modified HPV-DNA qPCR screen. PCR was performed using the AB® StepOnePlus™ Real-Time PCR system. Data was analysed using the corresponding software, v 2.2.2. Temperature (°C) is displayed against derivative reporter values (-Rn). Melt curve plots from HPV controls (HPV-18 plasmid and SiHa DNA), and H_2O are displayed. PCR products generated from gDNA isolated from 3 different sections of the same tonsil tissue are also shown.

5.3.1. Viral type HPV-18 is most prevalent within non-cancerous tonsil tissue

DNA from tonsil tissues of the 12 participants considered HPV-positive by HPV-DNA qPCR consensus screening were examined for their viral type using the INNO-LiPA® HPV Genotyping Extra II kit [Fujirebio]. The INNO-LiPA® assay makes use of a PCR amplification stage with biotinylated primers (SPF10) followed by hybridisation of PCR products to detect multiple viral types (13 HR, 6 probable HR-HPV (pHR-HPV), 9 LR, and 4 unknown risk types) (Fujirebio, 2016).

In some instances there was only 1 DNA sample from each HPV-positive tonsil tissue; therefore, these were all analysed. However, when multiple sections across the tissue showed HPV L1 amplification, the most positive section (as indicated by melt curve analysis derivative reporter value) or a range of sections from the tissue were taken forward to genotyping.

The presence of HPV-DNA was confirmed in most of the tonsil tissues considered HPV-positive by qPCR; however, DNA from 2 tonsil tissues showed no hybridisation to any viral types or the HPV controls on the strip and were considered un-typable by INNO-LiPA® genotyping. One DNA sample (isolated from a female right tonsil tissue) showed hybridisation to an HPV control, but did not hybridise to one of the 32 specific HPV types on the strip and again, was considered un-typable.

Interestingly, all participants with HPV types identified by hybridisation contained HR-HPV types (Table 10). Indeed, the viral type distribution showed that only 3 HPV types were detected within our HPV-positive participants: 2 HR – HPV-16 and HPV-18, and 1 pHR – HPV-73. We found that HPV-18 was most prevalent in HPV-positive tonsil tissue from our sample and was detected in 83.3% (N=10) HPV-positive participants. 6 participants showed a HPV-16 infection and 3 participants showed a pHR HPV-73 infection; the prevalence of HPV-16 and HPV-73 were 50% and 25% of the HPV-positive participants, respectively.

Of those participants where HPV was concordant in the left and right tonsil tissue by consensus qPCR screening (10 participants), we found that viral types were not always concordant. 4 participants had the same viral types in the left and right tissues, and 6 participants differed in the viral types detected. It is of note that whilst the combination of viral types detected varied between the left and right tissue of the same individual, HPV-18 was detected in all cases where HPV was typable by INNO-LiPA®. Interestingly, HPV-16 was always present as a dual

infection with HPV-18 in the 6 participants with detectable HPV-16; similarly, where HPV-73 was detected, this was always in a dual infection with HPV-18. The HPV types detected for each of the 12 HPV-positive participants are shown alongside a summary of each participant demographics in Table 10.

Table 10. Summary of demographics, lifestyle characteristics, and HPV-types detected for each HPV-positive individual. For all HPV-positive males (N=4) and HPV-positive females (N=7), and one HPV-positive unknown gender participant (N=1), a summary of their demographics and responses to selected questions posed in the questionnaire are displayed. For each individual their HPV-status in each tonsil by HPV-DNA consensus screening and the viral types detected by INNO-LiPA® genotyping in each positive tonsil is shown.

	Gender	Age	Ethnicity	Relationship Status	Sexuality	Sexual intercourse in past year	Smoking	Alcohol Use	HPV vaccination	HPV tonsil positivity	HPV-type (right tonsil)	HPV-type (left tonsil)
1	Male	26-35	White British	Single	Heterosexual	Yes	Current smoker	Current (1-2 times/week)	Unsure	Both	16, 18	16, 18
2	Male	26-35	White British	Married/Long term	Homosexual	Yes	Never	Current (1-2 times/week)	No	Both	18, 73	18, 73
3	Male	46-55	White British	Married/Long term	Heterosexual	Yes	Previous smoker	Current (Daily)	No	Both	18	16, 18
4	Female	18-25	White British	Single	Heterosexual	Unknown	Current smoker	Current (few times/month)	Unsure	Both	18	18
5	Female	18-25	White British	Married/Long term	Homosexual	Yes	Never	Current (once a month)	Unsure	Both	Unknown	16, 18
6	Female	18-25	White British	Married/Long term	Bisexual	Yes	Current smoker	Current (once a month)	Yes	Both	16, 18	18
7	Male	36-45	White British	Married/Long term	Heterosexual	Yes	Never	Current (Daily)	No	Both	16, 18	18
8	Female	26-35	White British	Married/Long term	Bisexual	Yes	Never	Current (few times/month)	No	Both	18	18, 73
9	Female	26-35	British Asian	Married/Long term	Heterosexual	Yes	Never	Never	No	Both	16, 18	16, 18
10	Female	18-25	White British	Married/Long term	Heterosexual	Yes	Never	Current (few times/year)	No	Both	18	18, 73
11	Female	18-25	White British	Short term	Bisexual	Yes	Never	Current (few times/month)	Unsure	Right	Unknown	Negative
12	Unknown	Unknown	Unknown	Unknown	Unknown	Unknown	Unknown	Unknown	Unknown	1 – unknown which	Unknown	Unknown

5.3.2. Histological analysis of HPV-positive and HPV-negative tonsil tissue

We next examined the tissue morphology of, and select biomarkers present within, a sample of HPV-positive and HPV-negative tonsils using H&E staining and IHC. The H&E stains shown, p16 IHC, and histology reports referred to were performed at RDH by Dr Amandeep Mann (2021-present) and Dr Robinson (2016-2020). Sectioning of FFPE blocks and staining for CK7 using IHC, alongside all imaging, was performed by Dr Rebecca Mallender at HistologiX (Nottingham, UK).

To be included in this clinical study, tonsils were confirmed to be non-dysplastic and with normal tissue morphology by a pathologist at each respective hospital. We examined the H&E slides of a representative sample of molecularly HPV-positive and HPV-negative tissues to locate morphological features of the palatine tonsils. All tissues show a normal tonsil morphology, with tonsillar crypts and surrounding stratified epithelium, and the lymphoid follicles, which are composed of mantle zone with a germinal centre (shown in Figure 22 and 23, along the top panel). Histology reports of all five samples selected within this analysis were described as showing 'reactive lymphoid hyperplasia' as reported by Dr Amandeep Mann or Dr Robinson.

5.3.2.1. $p16^{INK4A}$ – a true biomarker for HPV?

$p16^{INK4A}$ overexpression is often considered a surrogate marker of HR-HPV positivity (Benevolo et al., 2006) although its accuracy in correctly correlating with HPV infection in non-cancerous tissues is disputed (Klingenberg et al., 2010). In accordance with the study protocol, all tonsil tissues were subjected to a p16 stain, as per standard hospital protocols. The scoring system for the specimens discussed and shown here was as follows: absent, focal, discontinuous, diffuse and either absent, weak, moderate, or intense, and was reported by each pathologist. A score was given for the surface epithelium, crypt epithelium, and germinal centres of tonsil tissue sections.

We were interested in whether the p16 scores showed any correlation or pattern with the HPV-status given to each sample by HPV-DNA qPCR screening. The p16 scores at the surface epithelium, within the tonsillar crypts and geminal centres given to those samples defined as HPV-positive by qPCR are displayed in Table 11. Nearly all HPV-positive samples showed no p16 staining at the surface epithelium. HPV-positive participant 11 scored 1 at the surface epithelium in their left tonsil (weak discontinuous or moderate focal immunoreactivity), which

was shown to be HPV-negative by qPCR; their right tonsil was HPV-positive but p16 negative. Within the geminal centres, p16 staining for most of our HPV-positive sample was given low scores of 0 (absent) or 1 (focal and moderate or discontinuous and weak) in 9/11 right tonsils, and 6/10 left HPV-positive tonsils. Again, the scores for p16 staining within the crypt epithelium of HPV-positive by qPCR were quite varied although they were generally higher (indicating stronger staining) than observed within the surface epithelium and germinal centres (right tonsil: two scored 1, five scored 2, and four scored 3; left tonsil: one scored 1, five scored 2, and four scored 3). Surprisingly, the two most HPV-positive tissues by molecular DNA analysis were characterised as having low p16 staining in their crypt epithelium (HPV-positive participant 1 scored 1 in each tonsil; HPV-positive participant 2 scored 1 in right and 2 in left). A representative sample of p16 IHC of sections from HPV-positive (participants 1, 2, 4, and 10) and an HPV-negative (participant 13), by qPCR, were imaged (Figure 22 and 23, middle panel).

Comparatively, p16 staining is 'focal and weak' within the germinal centres of the representative HPV-negative sample shown here (both left and right tonsil sections, Figure 22 and 23 middle panel). However, the p16 staining was graded highest within the tonsillar crypts in this sample; the 'discontinuous and intense' staining can be seen along the stratified epithelium of the tonsillar crypt. To corroborate these observations, we examined the p16 scores of fresh-frozen HPV-negative tonsil tissues (defined by qPCR), within their surface epithelium, crypt epithelium and germinal centres, the qualitative scores are displayed in Table 12. Interestingly, p16 immunoreactivity within the tonsillar crypts was scored 3 in almost half of the HPV-negative specimens, demonstrating stronger staining and suggesting either moderate/intense diffuse or intense discontinuous expression of p16 within their crypts.

Overall, p16 scores by IHC do not appear to correlate or display any clear and obvious patterns between an HPV-positive status by qPCR screening. We find that the p16 scores for each of the structures analysed are varied, but more intense expression of p16 by IHC is not correlated at all to HPV-status (whether HPV-positive or HPV-negative). The crypt epithelium was graded the strongest p16 staining regardless of whether HPV-DNA was present in the specimen by qPCR.

Table 11. p16 scores of the right and left palatine tonsils of tissues considered HPV-positive by HPV-DNA qPCR consensus screening. The intensity of p16 staining was scored from 0-3 based on their scoring chart and recorded by a histopathologist at RDH. p16 scores are for the surface epithelium, crypt epithelium, and germinal centres for the left and right tonsils.

		Right tonsil			Left tonsil		
Sample	**HPV-status by qPCR**	**Surface epithelium**	**Crypt epithelium**	**Germinal centres**	**Surface epithelium**	**Crypt epithelium**	**Germinal centres**
1	Positive	N/A	2	1	N/A	2	1
2	Positive	0	1	2	0	1	2
3	Positive	0	1	1	0	2	1
4	Positive	0	2	1	0	2	1
5	Positive	0	3	1	0	3	0
6	Positive	0	2	0	0	3	0
7	Positive	0	3	1	0	3	2
8	Positive	0	2	1	0	2	1
9	Positive	0	3	1	0	3	2
10	Positive	0	2	2	0	2	3
11	Positive in right, Negative in left	0	3	1	1	2	1
13	Negative	0	3	0	0	3	0

Table 12. p16 scores of the right and left palatine tonsils of fresh-frozen tissues considered HPV-negative by HPV-DNA qPCR consensus screening. The intensity of p16 staining was scored from 0-3 based on their scoring chart, and recorded by a histopathologist at RDH. p16 scores are for the surface epithelium, crypt epithelium, and germinal centres for the left and right tonsils.

	Surface Epithelium		Germinal Centres		Tonsillar Crypts	
Score	**Right (N=13)**	**Left (N=14)**	**Right (N=13)**	**Left (N=14)**	**Right (N=13)**	**Left (N=14)**
0	9	9	2	1	0	0
1	4	5	9	10	2	1
2	0	0	2	3	5	7
3	0	0	0	0	6	6

5.3.2.2. CK7 – a marker of the tonsillar crypt epithelium in non-cancerous tonsil tissue

We then examined the expression of CK7 by IHC in the same sections of tissue as discussed for p16 (Figure 22 and 23, bottom panel). CK7 was selected as a marker specific to cells surrounding the tonsillar crypts in non-cancerous tonsil tissues (Morbini et al., 2015, Woods et al., 2017); but it is a possible biomarker for an active HR-HPV infection in OPC (Morbini et al., 2015) as well as having a general association with HPV-positive OPC (Woods et al., 2017). We show positive staining for CK7 along the cells of stratified epithelium that line the crypts of the normal palatine tonsil, as expected. There is expression of CK7 that correlates with the site of immunostaining detected for p16, regardless of HPV-status by qPCR, and the intensity of staining for CK7 is not noticeably different between the HPV-positive and HPV-negative tissues.

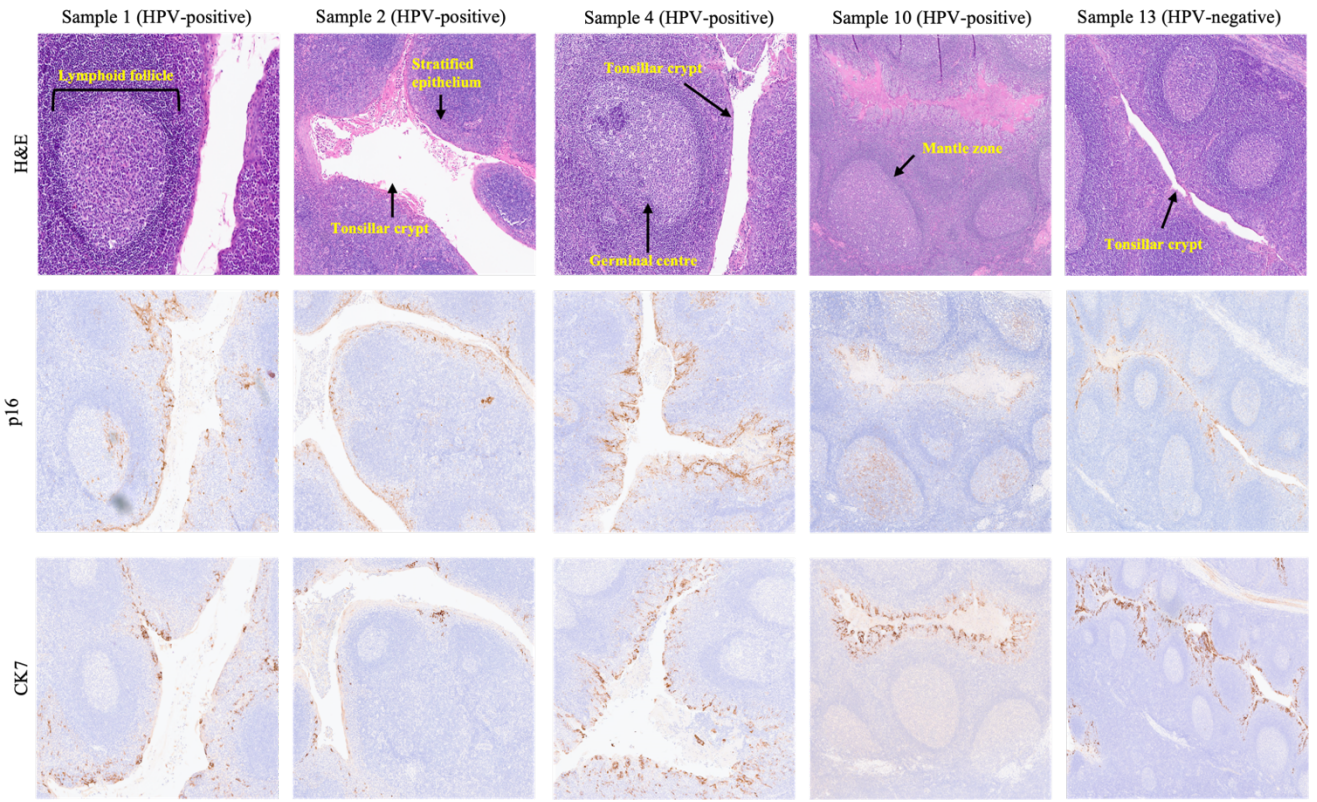


Figure 22. H&E, p16, and CK7 staining of sections taken from the right tissue of representative participant tonsils. Tonsil tissues had been defined as HPV-positive or HPV-negative based on our molecular analysis, and were fixed and sectioned. H&E staining is shown along the top row of the panel; several normal histological features of the palatine tonsils are indicated. The tonsillar crypt is shown, surrounded by stratified epithelium, itself a lighter stain. The lymphoid follicle is composed of the germinal centre (lighter stain) and surrounded by the mantle zone (darker region). p16 staining of tonsillar crypts and germinal centres is displayed along the middle row of the panel, with brown positive staining. CK7 staining of tonsillar crypts is displayed along the bottom row, with positive staining shown in brown. All slides were imaged using the NanoZoomer S360 Digital Slide Scanner with their corresponding software (NZAcquire Software). Images were taken when viewing scans in NDP.view2 software.

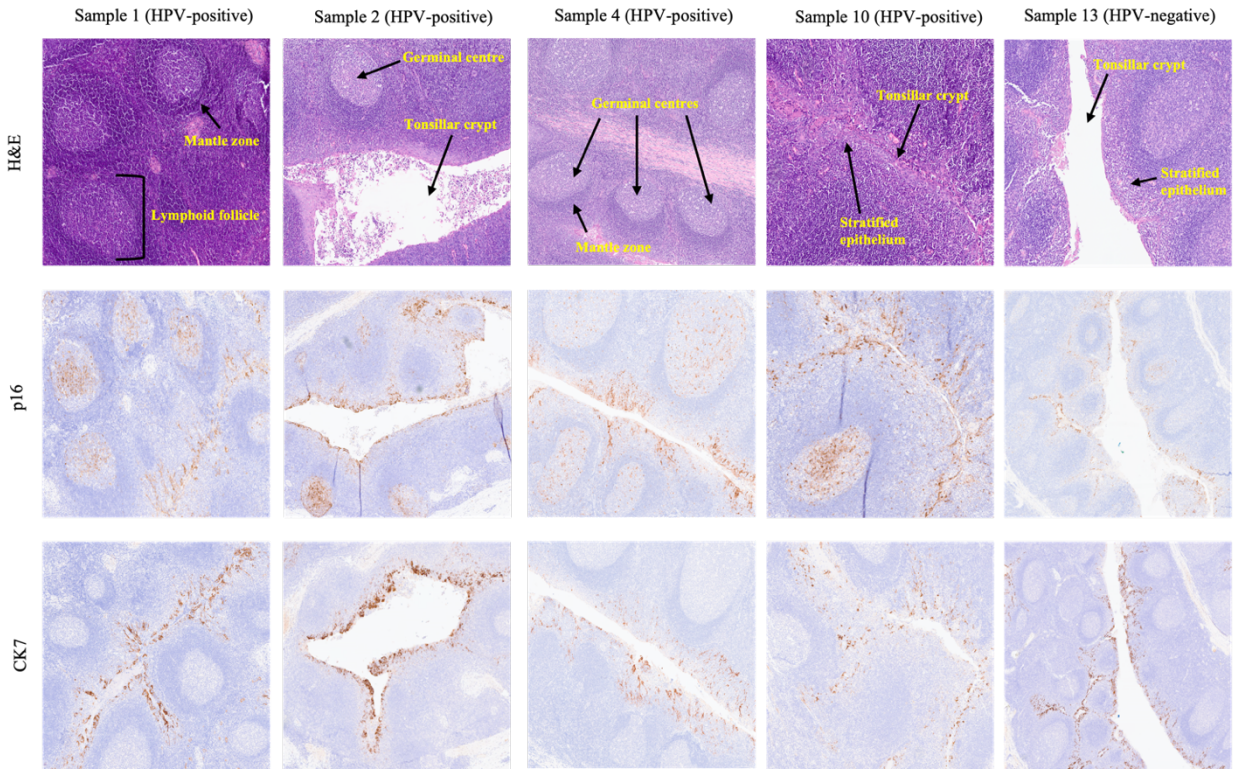


Figure 23. H&E, p16, and CK7 staining of sections taken from the left tissue of representative participant tonsils. Tonsil tissues had been defined as HPV-positive or HPV-negative based on our molecular analysis, and were fixed and sections. H&E staining is shown along the top row of the panel; several normal histological features of the palatine tonsils are indicated. The tonsillar crypt is shown, surrounded by stratified epithelium, itself a lighter stain. The lymphoid follicle is composed of the germinal centre (lighter stain) and surrounded by the mantle zone (darker region). p16 staining of tonsillar crypts and germinal centres is displayed along the middle row of the panel, with brown positive staining. CK7 staining of tonsillar crypts is displayed along the bottom row, with positive staining shown in brown. All slides were imaged using the NanoZoomer S360 Digital Slide Scanner with their corresponding software (NZAcquire Software). Images were taken when viewing scans in NDP.view2 software.

5.4. Formalin fixation of tonsil tissues – HPV-negative or undetected HPV-DNA

Interestingly, all the tonsil tissues from the 12 HPV-positive participants discussed above were collected and stored as fresh-frozen specimens. In most of the formalin-fixed samples, melt curve analysis was negative for HPV L1 amplification; however, in some instances, DNA from formalin-fixed samples did show what could be interpreted as a peak (with alignment to either HPV-18 plasmid or SiHa DNA controls) in one or more of the technical repeats within a plate. As per protocol, these samples were repeated by HPV-DNA consensus screening, but often still presented with inconsistent data that indicated the screen was unreliable in relation to these samples. For the formalin-fixed tonsil tissues, beta-actin housekeeping gene was always detected late and, consequently, we were unsurprised that the HPV-DNA L1 consensus screen was unable to consistently generate PCR products, given that HPV would be expected to be present at a lower copy number than beta-actin.

Due to cost limitations, we selected a representative selection of these samples from formalin-fixed tonsil tissue, that were inconsistent for HPV via L1 consensus screening, for further scrutiny with an alternative system (INNO-LiPA® HPV Genotyping Extra II kit [Fujirebio]). In 8 of these samples the use of SPF10 primers and subsequent hybridisation step showed no HPV positivity, which confirmed our interpretation of the inconsistent melt curve analysis by consensus screening that these samples were indeed HPV-negative; and therefore, have not been included in the overall HPV-positive sample.

One sample, however, that had shown only 1 peak in just 1 replicate that aligned with HPV-18 plasmid (of 3 technical repeats) in the L1 consensus screen was surprisingly positive for HPV-16 using the INNO-LiPA® genotyping kit. Indeed, this finding supports our hypothesis that in the formalin-fixed samples we were unable to detect HPV reliably, but that some these samples would be HPV-positive, and that a higher method of sensitivity would be required in order to detect HPV in these samples. Unfortunately, due to cost limitations we had to leave our investigations with the INNO-LiPA® genotyping kit, and we instead sought to explore alternative methods of PCR detection for the formalin-fixed tissues (Appendix D).

5.5. Discussion

HPV-positive OPC cases are increasing (Mourad et al., 2017), yet the knowledge of HPV infection within the healthy oropharynx remains sparse. In this project, we used our modified qPCR consensus assay to describe an oral HPV prevalence of 10.2% within the non-cancerous tonsil tissues of adults undergoing routine tonsillectomy surgery in several NHS hospitals in the UK. This being double the estimate of 5% for global oral HPV prevalence calculated in our own meta-analysis (Chapter 4), yet profoundly lower than other reports of an oral prevalence of 35% within healthy individuals (Jamieson et al., 2020).

These data show that HPV-status is mainly concordant between the left and right tonsil tissues of HPV-positive individuals; it is possible that this is a consequence of productive viral infection in one tonsil tissue that results in the spread of infection to the other tissue. Another explanation could be that HPV uniformly infects the oropharynx; however, within individual cells of the cervix, the nature of an HPV infection is often non-uniform (Burd, 2003). Indeed, we too found that the amount of PCR product generated from L1 amplification in DNA samples across the same HPV-positive tonsil tissue differed, and we speculated this to be due to a non-uniform viral infection across the tissue. Nonetheless, this discrepancy does not discount the suggestion of HPV uniformly infecting both tonsil tissues, as differences in viral detection across the tissue could be due to a tropism of HPV to certain structures within the tonsil, such as the crypts.

In this project, we found viral type HPV-18 to be most prevalent within non-cancerous HPV-positive tonsil tissues (detected in 83.3% of HPV-positive participants); and despite HPV-16 being the dominant viral type in HPV-OPC (Schache et al., 2016, Lechner et al., 2022), it was detected in only 50% of these HPV-positive individuals, suggesting that HPV-16 is not as prominent in infections as it is in disease.

p16 has frequently been used as a surrogate marker for HR-HPV infection (Benevolo et al., 2006); yet in non-cancerous tissues, its usefulness in predicting the presence of HPV has been disputed (Klingenberg et al., 2010, Ahn et al., 2022). Indeed, we found little difference in the scoring of p16 staining in HPV-positive and HPV-negative tissues (as defined by qPCR). A caveat of the scoring system is that it is a subjective process and it is possible that the scores given may differ slightly dependent on the pathologist scoring them; moreover, there may be discrepancies introduced based on reagents and imaging platforms used, that may impact

interpretation of p16 staining intensities. However, these data support the suggestion that p16 is not an appropriate indicator of HPV infection in non-cancerous tonsil tissues.

Formalin fixation of tissues causes severe fragmentation of DNA that renders it progressively more difficult to detect by PCR over time (Ferrer et al., 2007, Vitošević et al., 2018, Vitošević et al., 2023). We found that the melt curve analyses of DNA from formalin-fixed tissues generated through our modified HPV-DNA qPCR screen with GP5+/GP6+ primers were inconsistent and difficult to interpret. In marked contrast, but as expected, melt curve analysis of DNA from fresh-frozen tonsil tissue was distinctly HPV-positive or HPV-negative, in part a result of significantly better preservation of DNA (Shabihkhani et al., 2014). Due to this clear discrepancy in the viability of DNA from formalin-fixed versus fresh-frozen tonsil tissue, and the lack of/inconsistency in HPV detection in the formalin-fixed tissues, we hypothesised that there may be a difference in the ability to detect HPV due to the damaging fragmentation effects of formalin (Ferrer et al., 2007), and we believe that some of the formalin-fixed 'HPV-negative' by qPCR samples may be positive and that HPV is actually undetected in these samples. Indeed, our findings that using an alternative, more expensive, technique of genotyping with the INNO-LiPA® kit resulted in 1 formalin-fixed DNA sample being positive for HPV-16 that had only one peak in one replicate by consensus screening, supported this theory.

The use of alternative PCR enzymes and approaches that were designed to work with low copy number targets and complex PCRs provide further evidence that it is a most likely due to severe DNA fragmentation as there was either no HPV detected, or there was no stringency in detection of HPV with formalin-fixed samples (Appendix D).

Taken together, these data indicate that we cannot deduce whether the formalin-fixed tissues are genuinely negative or whether the virus is undetected, and it supports our hypothesis that formalin fixation of tissues is inappropriate for molecular analysis of viruses' that, if present, are most probably low copy number given the non-cancerous nature of the tissues scrutinised. From this work, we strongly suggest that human tissue is not fixed with formalin for subsequent molecular analyses; indeed, recent literature suggests 70% EtOH a suitable fixative to render samples safe of any potential class III pathogens, including SARS-CoV-2 (Kumar et al., 2021, Nomura et al., 2021). Given that 70% EtOH has been shown to have little effect on DNA degradation over a period of months (Bradbury et al., 2023), we speculate that HPV-DNA may

have been detectable in a higher proportion of this sample of non-cancerous tonsil tissue if 70% EtOH had been as the primary fixative for these tissues.

5.5.1. Conclusion

To conclude these data outlined in Chapter 5; here, we defined our sample of non-cancerous palatine tonsil tissues of adult routine tonsillectomy patients as HPV-positive or HPV-negative. We confirm the hypothesis that HPV was present in some of these tissues, and we show an overall prevalence of 10.2% (N=12/118). In pathogenesis studies of the HPV-positive cohort, we conclude that HR HPV-18 was the most prevalent viral type, followed by HR HPV-16 and pHR HPV-73, and that regardless of viral type present, HPV-status was largely concordant in both tonsil tissues. In addition, it is demonstrated in the samples reviewed in this project that the biomarkers p16 and CK7 used for characterising HPV-OPC are not good indicators of HPV-positivity in non-cancerous tissues.

Finally, these data further support our suggestion throughout this thesis that 10% NBF is not an appropriate fixative for tissues that are required for molecular analyses.

CHAPTER 6: Differential protein expression in non-cancerous palatine tonsils with HPV-DNA infection.

6.0. Introduction

HPV-positive OPCs respond better to treatment than HPV-negative OPCs (Mallen-St Clair et al., 2016); however, recent years have seen failures to successfully implement de-escalation therapies and alternate therapies based on HPV-positivity (Gillison et al., 2019, Mehanna et al., 2019c), and this is likely due to a lack of fundamental knowledge of the biology of HPV-OPCs (i.e., varying p16 status (Mehanna et al., 2023), amongst other things). Whilst several biomarkers have been linked to HPV in OPC; including p16, CK7, and PD-L1 (Weinberger et al., 2006, Balermpas et al., 2017, Woods et al., 2017), the impact of HPV infection on the proteome of the non-cancerous oropharynx remains largely unexplored. Moreover, it is evident that both oral HPV infection and HPV-OPC risk is greater within men than women but the mechanism behind this discrepancy remains unclear.

To analyse the proteome of non-cancerous tonsil tissue, we selected a non-biased approach, Sequential Window Acquisition of all Theoretical Mass Spectra (SWATH-MS) that is a form of data independent acquisition (DIA). SWATH-MS is a technique that enables analysis of all proteins present in a sample across a m/z range (Ludwig et al., 2018). SWATH-MS was considered the most appropriate approach for these analyses, given that we wanted to examine the entire proteome for changes in expression based on HPV-status, and we hoped this would give an overview of whether HPV has broad biological changes on protein expression.

To further understand how HPV influences the host cell biology in the non-cancerous oropharynx, we used a selection of palatine tonsil tissues collected as part of our established clinical study that had been defined as HPV-positive or HPV-negative (Chapter 5). We hypothesised that the host cell proteome of HPV-positive tonsil tissue would be altered by an HPV infection. Moreover, we speculated there may be a difference in protein expression impacting on normal protein function between males and females with HPV-positive tonsil tissue. In this project, we aimed to determine a ‘protein expression blueprint’ associated with oropharyngeal HPV infection, and to understand how changes in the proteome profile as a result of HPV infection influence host cell biology.

In this chapter, we describe proteins that are significantly changed in their expression within individuals with an oropharyngeal HPV-DNA infection in their non-cancerous tonsil tissues.

6.1. Final participant sample for protein extraction and SWATH-MS

The final sample of tonsil tissues selected for protein isolation and subsequent mass spectrometry was as follows: 96 formalin-fixed tissues (N=177/203 samples), and 16 fresh-frozen tissues (N=26/203 samples). These tissues were selected based on their HPV-status, and/or gender, in anticipation of downstream differential protein expression analyses. HPV-DNA qPCR screening revealed that all HPV-positive tissues were of fresh-frozen origin, and so we included most of these tissues in this sample. Other tissues were selected based on participant gender; given that oral HPV infection in healthy individuals and HPV-associated oral disease is higher within men than women (Gillison et al., 2012, Guo et al., 2022, Giuliano et al., 2023) we anticipated that this would be most informative to elucidate differences between host cell biology that may influence HPV pathogenesis. Moreover, all the formalin-fixed tissues had been defined as HPV-negative but with uncertainty of their true HPV-status, due to assumed DNA fragmentation by prolonged formalin fixation periods (Chapter 5), so it was considered inappropriate to analyse these tissues as HPV-negative. 23 participants were male, and 33 were female. The total number of tissue sections from which protein was extracted from varied dependent on the mass of tissue available for protein extraction; however, we aimed to have 2 sections from different locations of each left and right participant tonsils. In total, 203 protein samples were prepared for quantification and later analysis.

Given the small number of HPV-positive tonsil tissues available for SWATH-MS, analysis of differential protein expression by HPV-status was established as a sub-analyses of the overall sample (N=26/203). This is because it would not be statistically accurate to compare the proteome of only 6 HPV-positive individuals to the entire rest of the sample (all HPV-negative or HPV-status undecided in the case of formalin-fixed tissues), and so a small number of 2 HPV-negative individuals were selected for inclusion in this comparison. As a result of the different storage methods of the tissues (i.e., 10% NBF or frozen); it was decided that these HPV-negative tissues must also be from fresh-frozen origin, due to the clear impact 10% NBF had on both DNA and RNA (Chapter 2), we did not want to introduce another variable to the comparison aside from HPV-status. This subset of the total 203 protein samples was composed of 26 samples, taken from a total of 16 tonsil tissues from 8 participants (including - 4 males; 3 HPV-positive, 1 HPV-negative and 4 females; 3 HPV-positive, 1 HPV-negative).

6.2. Quantification of protein from formalin-fixed versus fresh-frozen tonsil tissue

All 203 protein samples were quantified using the Qubit™ 4 fluorimeter and their corresponding protein assay kit. We found there to generally be little difference in protein concentration of samples isolated from formalin-fixed tissue and fresh-frozen tonsil tissue, with formalin-fixed tissue generally having a lower concentration by their Qubit™ reading than fresh-frozen but not in all instances. Nonetheless, protein concentrations were more than sufficient for the small amount of total protein required for subsequent analysis (75 μg).

6.3. SWATH-MS analysis of the proteome of non-cancerous tonsil tissue

203 tryptic-digested protein samples were subjected to HPLC-SWATH-MS. Given that we were interested in the impact of biological gender on the normal tonsil proteome of HPV-positive and HPV-negative tissues in downstream differential protein expression analysis, batches for MS analysis were divided roughly based on gender of tissue of origin to reduce likelihood of bias between runs having impact on proteins quantified.

To assess the intensities of mass peaks across all m/z windows analysed, the total ion current (TIC) chromatograms were reviewed as a quality control. Despite extensive optimisation of protein isolation for formalin-fixed tonsil tissue, in this final experiment we found a striking difference between the shapes of the curves of TICs generated from protein samples of fresh-frozen compared to formalin-fixed tissue, with the latter displaying a flat trace in which the intensity of peaks detected over time were consistently low and returned to baseline at each time point (Figure 24A). In marked contrast, the traces of protein of fresh-frozen tonsil tissue displayed curves with a shape that suggested an abundance of protein detected across all m/z windows analysed (Figure 24A). We found there to be no obvious difference in the TIC traces of most of the fresh-frozen tonsillar protein samples based on their HPV-status, indicating that the proteins detected in fresh-frozen tissue were not impacted by HPV-status (Figure 24B).

The issues experienced isolating nucleic acid from formalin-fixed tonsil tissues meant that we speculated again these TICs were likely also a result of 10% NBF. To rule out the possibility that an incorrect protein quantification (by Qubit™ 4 fluorimeter assay) had caused an disproportionate ratio of trypsin:protein to be added during tryptic digestion, we measured the concentration of a selection of protein samples using an alternate method – Pierce 660 nm Protein Assay Kit [ThermoFisher]. Despite the concentration of these samples measuring lower

using this assay than with the Qubit™ kit (data not shown) and the subsequent alteration of the trypsin:protein concentration accordingly, the outcome following digestion and SWATH-MS was unchanged. Indeed, as expected samples isolated from formalin-fixed tissues presented with extremely low protein identities (IDs) when searched in DIA-NN.

These data suggest that formalin fixation of tonsil tissue resulted in protein lysates unsuitable for SWATH-MS. We speculated that 10% NBF had impacted the ability to successfully lyse cells and isolate protein. To conclude here, these data support the recurring theme that formalin fixation of tissue is not a suitable method of storage if downstream molecular analysis is required. Due to time limitations, it was not possible to re-extract proteins from 203 tonsil tissues, and so we continued with a sample of fresh-frozen protein isolates for the final analysis. Protein samples from formalin-fixed tonsil tissues were not used for differential protein expression analysis, due to insufficient protein ID that were considered too sparse for even a separate analysis within these tissues. Therefore, these data presented in this chapter are based on tonsil protein isolated from fresh-frozen tissues only.

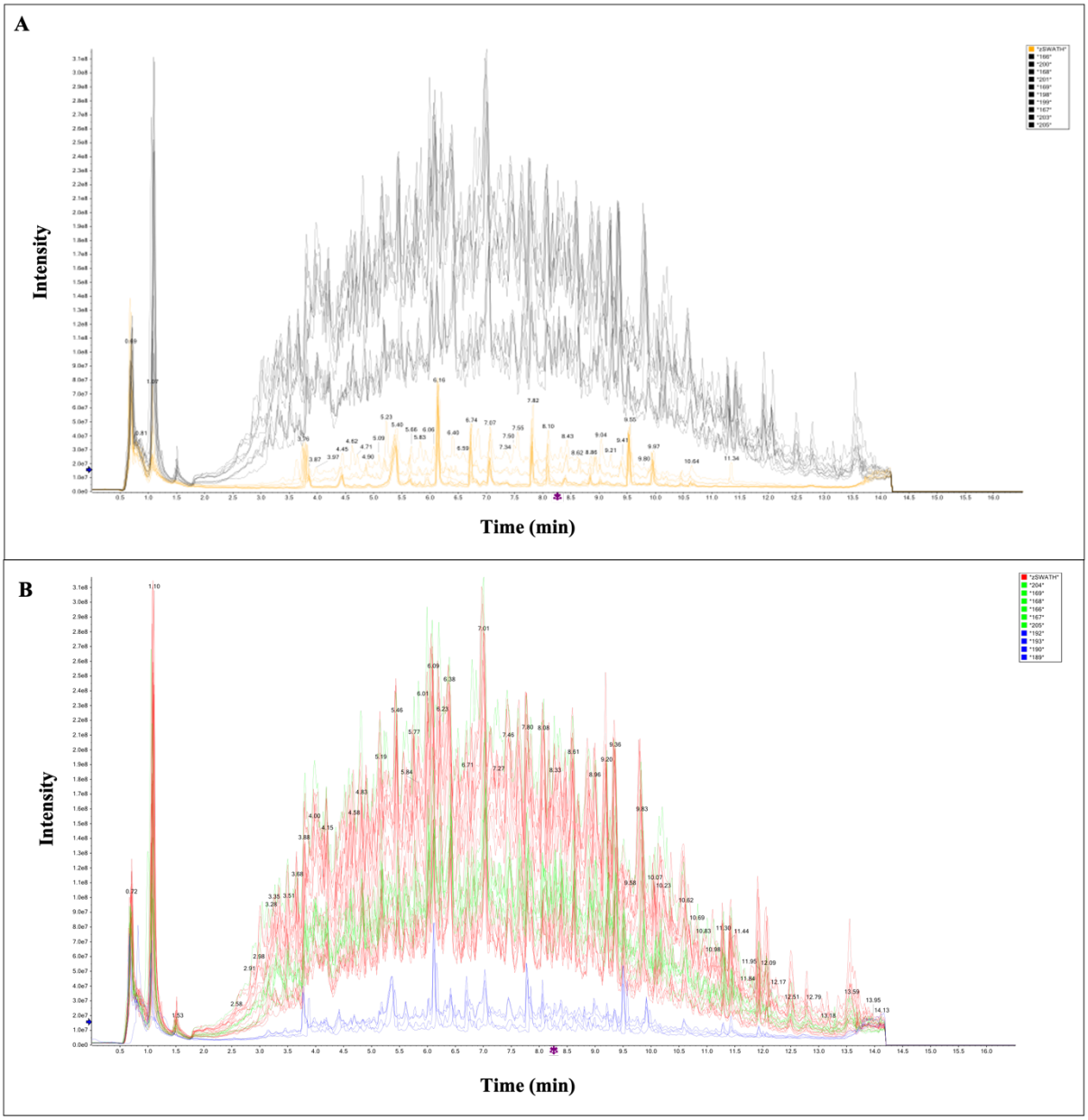


Figure 24. TIC chromatograms of tonsil protein samples following SWATH-MS. TIC chromatograms of the total charged peptides detected at each time point for selected samples. TIC chromatograms were produced. **(A)** TICs of a representative sample of protein isolated from fresh-frozen tonsil tissue (black) versus protein isolated from formalin-fixed tonsil tissue (orange). A greater abundance of peptides were detected at each time point in fresh-frozen samples compared to formalin-fixed samples. **(B)** TICs of representative protein isolates from HPV-positive (red) and HPV-negative (green) tonsil tissues demonstrate no clear difference between peptide abundance based on HPV-status. Traces from anomalous fresh-frozen samples that showed poor peptide abundance at each time point are shown in blue. All traces in panel B are representative of protein isolates from fresh-frozen tonsil tissue.

6.3.1. Proteins identified by SWATH data searching in DIA-NN

SWATH data was searched in DIA-NN software (v 9.1), and we found the range of proteins identified in samples isolated from fresh-frozen tonsil tissue to be 1385-5302. This presents a broad range, and so in individual statistical comparisons samples were selected from the total to have similar numbers suitable for statistical analysis.

6.4. Differential expression of HPV-positive and HPV-negative non-cancerous palatine tonsil protein

To compare the proteome of HPV-positive and HPV-negative tonsil tissue, a differential protein expression analysis was performed. We hypothesised that protein expression would be altered due to HPV infection. Protein from 8 HPV-positive and 6 HPV-negative non-cancerous tonsil tissue samples were included in the comparison. HPV-positive protein samples were selected to have representative protein from each HPV-DNA positive participant available for analyses, with HPV-negative protein samples with comparable protein ID numbers selected as controls; the number of proteins identified by DIA-NN for each individual sample is displayed in Table 13.

Table 13. Total raw precursor numbers and proteins identified by DIA-NN searching of samples included in HPV-positive versus HPV-negative DEA. The sample identifier (SWATH_number) and the HPV-DNA status by qPCR of the tonsil tissue that protein was isolated from is shown. The raw precursor numbers from SWATH-MS are shown against the total number of proteins identified through SWATH data searching in DIA-NN.

SWATH sample	HPV-DNA status by qPCR	Total precursors	Total protein IDs
SWATH_199	Positive	31146	5302
SWATH_201	Positive	29674	5203
SWATH_203	Positive	24780	4662
SWATH_197	Positive	24585	4497
SWATH_191	Positive	22900	4170
SWATH_184	Positive	18933	3900
SWATH_187	Positive	19508	3847
SWATH_186	Positive	15569	3346
SWATH_169	Negative	29264	5222
SWATH_204	Negative	28802	5207
SWATH_168	Negative	29466	5189
SWATH_166	Negative	28087	5035
SWATH_167	Negative	26217	4797
SWATH_205	Negative	24674	4617

Following filtering protein hits based on the selected imputation parameters, 1199 proteins were removed from the statistical analysis and 5103 proteins were quantified.

The quality control tests performed by the online software were assessed to analyse the imputed values. The principal component analysis (PCA) scatterplot revealed some clustering of HPV-negative protein samples. HPV-positive samples were more dispersed with some clustering closer to HPV-negative samples. The PC1 and PC2 percentages equate to over 50% of the variance of the imputed intensity values suggesting that some clustering is based on the variable analysed (HPV-status) (Figure 25A). We found that the median intensity and interquartile ranges (IQR) were consistent within the HPV-negative groups and that the IQRs of the HPV-positive samples were slightly greater. In both comparative groups, there were several imputed values outside of the maximum score, yet only a few outliers below the minimum score (Figure 25B). The coefficient of variation (CV) plot indicated a high level of variability of data in each group (both HPV-positive and -negative); however, this may have been skewed by the large number of outliers (Figure 25C). Finally, we assessed the Pearson correlation plot of imputed intensities to determine how replicates of the same group correlate. It showed strong correlation between samples of the HPV-negative group (Figure 25D).

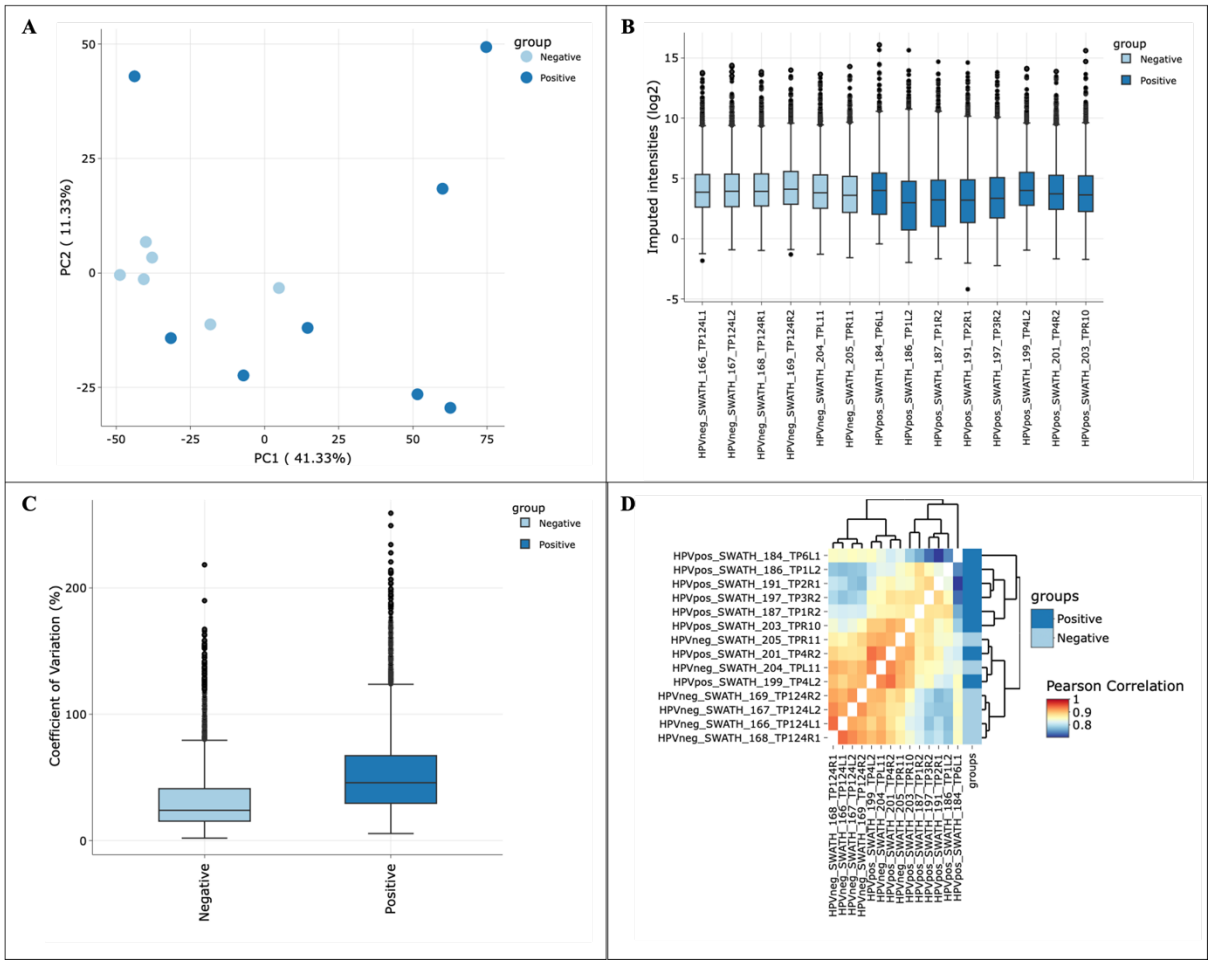


Figure 25. Quality control data generated by Amica software of differential abundance analysis of HPV-positive versus HPV-negative non-cancerous tonsillar protein. Protein expression analysis was performed in Amica software (v 3.0.1) using the limma package for statistical analysis. 5103 proteins were available for quantification. QC data was generated for the analysis of the imputed intensities based on the imputation method and parameters manually selected. **(A)** PCA plot of HPV-positive versus HPV-negative imputed intensities. **(B)** Box plot of imputed intensities of each sample. **(C)** Coefficient of Variation plot of imputed intensities between comparative groups. **(D)** Pearson correlation plot to compare replicates of HPV-positive and HPV-negative samples. In all panels, HPV-positive protein samples are represented in dark blue and HPV-negative in light blue.

6.4.1. A panel of 222 proteins with significantly changed expression in HPV-positive versus HPV-negative tonsil tissues

The differential abundance quantities that had been generated in Amica were then analysed. We selected to scrutinise the statistically significant hits of HPV-positive versus HPV-negative tissues, rather than HPV-negative versus HPV-positive, to determine a blueprint of protein expression in non-cancerous tonsil tissue based on the presence of HPV-DNA compared to the proteome in the absence of virus. Protein hits were filtered in Microsoft Excel initially for adjusted p-value, p-value, and logFC. We found that p-value <0.05 gave rise to 1107 protein hits, and a p-value <0.01, 234 hits. When filtering for a logFC of ±0.6 at each p-value there were 939 hits ($p<0.05$, logFC ± 0.6) and 220 hits ($p<0.01$, logFC ± 0.6). We found just two protein hits (BH3-interacting domain agonist (BID), and mitochondrial inner membrane m-AAA protease component paraplegin) when filtering for an adjusted p-value<0.05, logFC ± 0.6; however, given that we aimed to understand more about the fundamental biology of HPV and how it influences the host cell, more protein hits would be informative for this investigation, and so a less stringent approach and was used in further analysis.

Data were uploaded to VolcaNoseR2 to filter top hits of the HPV-positive versus HPV-negative tonsil proteins based on various statistical parameters. Using a $-\log_{10}$ p-value of 2 (equivalent to $p<0.01$) and a logFC threshold ± 0.6 (equivalent to absolute FC 1.5) a panel of 222 protein hits were generated ranked on their Manhattan distance and then on their p-value (Appendix E). We found within the top 100 proteins ranked on their Manhattan distance, only eight proteins were significantly higher in HPV-positive tonsil tissue. This trend can be seen in Figure 26 where most proteins show a negative log FC, meaning protein expression was significantly lower in the HPV-positive tonsil protein group than HPV-negative. Data points [blue and red] represent the 222 changed and significant protein hits that have a $-\log_{10}$ p-value > 2 and a logFC threshold ±0.6. The top 10 increased and top 10 decreased protein hits, as filtered on these parameters, are annotated with their UniProt ID to demonstrate the most changed proteins in HPV-positive tonsil tissue in this comparison (Figure 26).

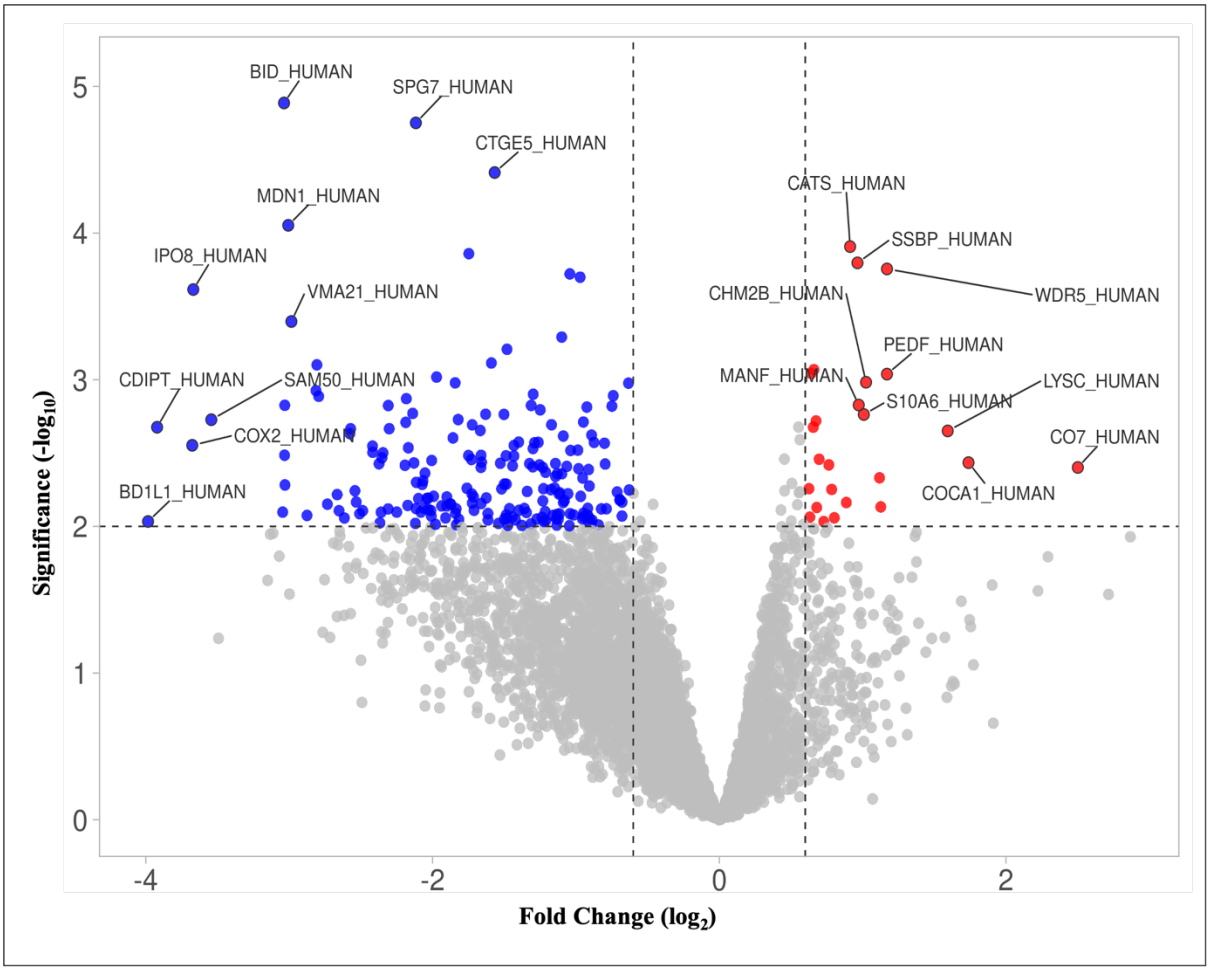


Figure 26. Volcano plot of the 5103 proteins quantified in Amica filtered on their log_2 FC and p-value ($-log_{10}$). Log_2FC and $-log_{10}$ p-values were generated in a differential abundance analysis by online software Amica (v 3.0.1). The values quantified for HPV-positive versus HPV-negative comparison were uploaded to VolcaNoseR2 software. All data points are representative of the 5103 proteins within the dataset available for quantification. The proteins were filtered based on a $-log_{10}$ p-value of 2 (equivalent to $p<0.01$) and a log_2FC ±0.6 (absolute FC 1.5). Blue data points indicate a negative FC, with expression significantly lower in HPV-positive samples compared to HPV-negative. Red data points indicate a positive FC, with expression significantly higher in HPV-positive samples compared to HPV-negative. The top 10 increased (and significant) and top 10 decreased (and significant) hits are labelled with their UniProt IDs.

6.4.2. Protein function of significantly changed protein hits in HPV-positive compared to HPV-negative non-cancerous tonsil tissue

The top 10 proteins hits, as determined by Manhattan distance, of each of the significantly higher/lower expression in HPV-positive tonsil tissues are annotated in Figure 26 and described in Table 14. Protein hits were determined based on their Manhattan distance, then re-ranked based on their p-value (highest to lowest).

To elucidate how the virus may be influencing the host cell function in the head and neck, we provide insight into the canonical protein function of these 20 hits. Interestingly, the most common protein category of the significantly increased expressed proteins in HPV-positive tonsil tissue was an immune role, which seems plausible considering the context of an viral infection. The most significant hit (increased) was Cathepsin S which has a role in the major histocompatibility complex (MHC) class II antigen presentation process (Beers et al., 2005). Other immune-related proteins overexpressed in the presence of oropharyngeal HPV were Lysozyme C, and also the Complement Component C7 (C7) of the membrane attack complex (MAC) which contribute to the adaptive and/or innate immune response (Ragland and Criss, 2017, Xie et al., 2020). Two of the top 10 overexpressed hits in HPV-positive tonsil protein were neurotrophic proteins, Pigment derived-epithelium factor (PEDF) (Becerra et al., 2012) and Mesencephalic astrocyte-derived neurotrophic factor (Petrova et al., 2003). Aside from its neurotrophic function, PEDF has a plethora of other roles including in antiangiogenic, anti-inflammatory, and in antimetastatic/tumourigenic pathways (Becerra et al., 2012). The remaining five upregulated protein hits are of varied function including involvement in calcium signalling (Protein S100-A6) (Yamaguchi et al., 2012), structural roles (Collagen alpha-1 (XII) chain), mitochondrial replication (single-stranded DNA-binding protein, mitochondrial) (Jiang et al., 2021) and protein trafficking (charged multivesicular body protein 2b) (Bodon et al., 2011).

The same data search was performed to examine the top 10 significantly decreased protein hits in HPV-positive tonsil protein compared to control HPV-negative tissues. We found that the top hit, BID, promotes apoptosis (Wang et al., 1996). Amongst the top 10 hits we too describe decreased expression of biorientation of chromosomes in cell division protein 1-like 1 which is able to promote genome stability (Higgs et al., 2015), perhaps eluding to reduced genomic stability in the presence of HPV-DNA in tonsil tissue. A few proteins were linked to mitochondrial function or structure including the mitochondrial inner membrane m-AAA

protease component paraplegin (Gerdes et al., 2012), cytochrome c oxidase subunit 2 (Capaldi et al., 1983), and sorting and assembly machinery component 50 homolog (Ott et al., 2012). Two proteins identified were ATPases/component of an ATPase, specifically midasin (an AAA ATPase) and vacuolar ATPase assembly integral membrane protein (VMA21, an (V)-ATPase component). Midasin is specifically involved in the maturation of small ribosomal subunit, 60S (Raman et al., 2016). Whereas the VMA21 protein, as part of the V-ATPases play a role in normal cellular processes such as protein degradation and transport across the membrane (Forgac, 2007). Other proteins with negative differential expression in HPV-positive tissues appeared to have roles in cell membrane structure, through the production of phospholipids (CDP-diacylglycerol-inositol 3-phosphatidyltransferase). Another protein, melanoma inhibitory activity protein 2 appears to act in cooperation with other proteins to traffic pre-chylomicrons and lipoproteins from the endoplasmic reticulum (ER) (Santos et al., 2016). Finally, we describe the statistically significant decrease in the expression of Importin-8 in HPV-positive tissues, whose role is to import proteins with a classical nuclear localisation signal (cNLS) from the cytoplasm to the nucleus (Miyamoto et al., 2016).

These data indicate that the top protein hits identified in the HPV-positive versus HPV-negative tonsil protein comparison harbour a wide variety of canonical functions, many of which fit the context of a viral infection.

Table 14. Top 10 significantly increased and decreased protein hits identified in HPV-positive tonsil tissue in comparison to HPV-negative tonsil tissue. The UniProt IDs and corresponding protein name for each hit is shown against the direction of expression change (e.g., increased = a positive FC in HPV-positive tissues compared to HPV-negative tonsil). The significance shown is equal to a $-\log_{10}$ (p-value). These hits were selected based on their Manhattan distance, with the greatest distance considered most significant and then re-ordered as most significantly changed (by p-value) to least changed.

UniProtID	Protein Name	Change	Fold change (log2)	Significance	Manhattan distance
CATS_HUMAN	Cathepsin S	Increased	0.912799731	3.907503927	4.820303658
SSBP_HUMAN	Single-stranded DNA-binding protein, mitochondrial	Increased	0.964095314	3.796662452	4.760757766
WDR5_HUMAN	WD repeat-containing protein 5	Increased	1.169944393	3.755294982	4.925239375
PEDF_HUMAN	Pigment epithelium-derived factor	Increased	1.169837139	3.038137242	4.207974381
CHM2B_HUMAN	Charged multivesicular body protein 2b	Increased	1.023752456	2.982736212	4.006488667999999
MANF_HUMAN	Mesencephalic astrocyte-derived neurotrophic factor	Increased	0.972832129	2.827530763	3.800362892
S10A6_HUMAN	Protein S100-A6	Increased	1.00845533	2.76294235	3.77139768
LYSC_HUMAN	Lysozyme C	Increased	1.59371374	2.650889421	4.244603161000001
COCA1_HUMAN	Collagen alpha-1(XII) chain	Increased	1.7387781	2.434712759	4.173490859
CO7_HUMAN	Complement component C7	Increased	2.501335641	2.401682442	4.903018082999999
BID_HUMAN	BH3-interacting domain death agonist	Decreased	-3.03590199	4.886791136	7.922693126
SPG7_HUMAN	Mitochondrial inner membrane m-AAA protease component paraplegin	Decreased	-2.116411311	4.751194734	6.867606045
CTGE5_HUMAN*	Melanoma inhibitory activity protein 2	Decreased	-1.566592541	4.412491737	5.979084278
MDN1_HUMAN	Midasin	Decreased	-3.00646278	4.052495722	7.058958501999999
IPO8_HUMAN	Importin-8	Decreased	-3.668944359	3.614778149	7.283722508
VMA21_HUMAN	Vacuolar ATPase assembly integral membrane protein VMA21	Decreased	-2.984638222	3.396752752	6.381390974
SAM50_HUMAN	Sorting and assembly machinery component 50 homolog	Decreased	-3.543452731	2.72736508	6.270817811000001
CDIPT_HUMAN	CDP-diacylglycerol--inositol 3-phosphatidyltransferase	Decreased	-3.921494999	2.675946896	6.597441895
COX2_HUMAN	Cytochrome c oxidase subunit 2	Decreased	-3.676255793	2.552754188	6.229009981000001
BD1L1_HUMAN	Biorientation of chromosomes in cell division protein 1-like 1	Decreased	-3.984533126	2.033316512	6.017849638

6.4.3. Classification of differentially expressed proteins in HPV-positive versus HPV-negative tonsil tissue

All proteins identified as significantly changed between HPV-positive and HPV-negative tonsil tissues (N=222), based on p-value <0.01, logFC ± 0.6 (Appendix E, Table 24) were subjected to gene ontology enrichment analysis using the Protein Analysis THrough Evolutionary Relationships (PANTHER) database (pantherdb.org). UniProt IDs were searched against the *homo sapiens* database and mapped to their corresponding gene ID. 24 IDs were not found in the database and were excluded from the classification. The remaining 198 gene hits were categorised based on their protein classification, biological process, molecular function, and cellular component (Figure 27). In all analyses of ontological classification when describing total number of gene hits for any category the values reported are as follows: X/total genes searched, % of total number of genes searched, % total hits for each classification.

When categorised for protein classification, 55/198 genes were unclassified by PANTHER and these removed from the graphical summary of protein classification distribution. The remaining 143 gene hits revealed most of the SWATH protein hits were within the metabolic interconversion enzyme family (46/143, 23.2%, 32.2%). Other notably enriched protein families included: protein modifying enzymes (23/143, 11.6%, 16.1%), transporters (17/143, 8.6%, 11.9%), and protein-binding activity modulators (16/143, 8.1%, 11.2%). All other gene hits for protein classification comprised <10% of the total protein classes identified (Figure 27A). Interestingly, despite proteins with immune functions (either direct or indirectly) being prominent in the top 10 increased and top 10 decreased protein hits, we found only two proteins of the overall proteins significantly changed between HPV-positive versus HPV-negative had their function primarily classed within the defence/immune protein family.

Next, 198 mapped genes of significantly changed proteins between HPV-positive versus HPV-negative tonsil tissue were searched for their biological process/es, to enable an overview of the broad process that the proteins contribute to directly (e.g., within a specific pathway/molecular process), or outwardly contribute to (Thomas, 2017). In this instance, 52 genes were unable to be classified by biological process in PANTHER database; again, these were excluded from the bar chart summary of enriched biological processes amongst the hits. The remaining 146 genes searched were mapped to have roles in 312 biological processes (Figure 27B). Most biological processes identified were classified as cellular processes (124/312, 62.6%, 39.7%), with many of the genes found to have involvement in metabolic,

localisation and biological regulation. There were also nine further categories of biological process identified within the list of differentially regulated proteins searched; however, each had less of the total genes within them. Similarly to protein classification, of all the significantly changed proteins (between HPV-positive versus HPV-negative), only two genes (1% total genes searched, 0.6% all hits) were identified as having a role in an immunological process.

To further scrutinise the identified protein hits differentially expressed between HPV-positive and HPV-negative non-cancerous tonsil tissues, 198 mapped genes were searched for their fundamental molecular roles to provide an insight into their activities. 82 genes were unable to be classified based on PANTHER database for molecular function, leaving 116 genes for analysis that generated 162 hits of molecular function. The vast majority of the genes searched had/or are involved in catalytic activity (N=64/116, 32.3%, 39.5%) and/or binding (N=56/116, 28.3%, 34.6%) (Figure 27C). The genes searched returned a smaller number of hits within 9 other classification of molecular function including: ATP-dependent, antioxidant, cytoskeletal motor, electron, transfer, molecular adaptor, molecular function regulator, structural molecule, transcription regulator, translation regulator, and transporter activity (Figure 27C).

Finally, the data of HPV-positive versus HPV-negative protein hits were searched for their cellular component domain, to show localisation of these proteins (i.e., whether they are present in the cytosol, within an organelle, or a larger protein complex (Mi and Thomas, 2009)). 43/198 genes were unclassified by the PANTHER database for cellular compartment. Most of the proteins exist as cellular anatomical entities (N=154/155, 77.8%, 73.7%) (Figure 27D); these were further categorised into a range of hits with the majority of genes classified as intracellular anatomical structures (N=137/154), cytoplasmic proteins (N=114/154), and localised to an organelle (N=113/154). Some of the 155 gene hits also suggest a role of some proteins in protein containing complexes (N=55/155, 27.8%, 26.3%) (Figure 27D). Further scrutiny revealed most of these genes are relevant to proteins that could be within catalytic complexes (N=30/55), or membrane protein complexes (N=23/55).

These data show that the proteins that were identified as significantly changed in HPV-positive versus HPV-negative non-cancerous tonsil tissues ($p<0.01$, logFC ± 0.6) belong to a range of protein families and are not localised to one particular family. Interestingly, 46/143 proteins classifiable by PANTHER were metabolic interconversion enzymes. Similarly, these proteins have roles within several different biological processes and possess a range of molecular

functions, with contribution to cellular processes and catalytic function the most common hits, respectively. Therefore, these proteins show involvement in a range of potential processes and further analysis of where these protein hits are collectively impacting biological pathways would be informative of possible impacts of HPV on host cell biology, as discussed below.

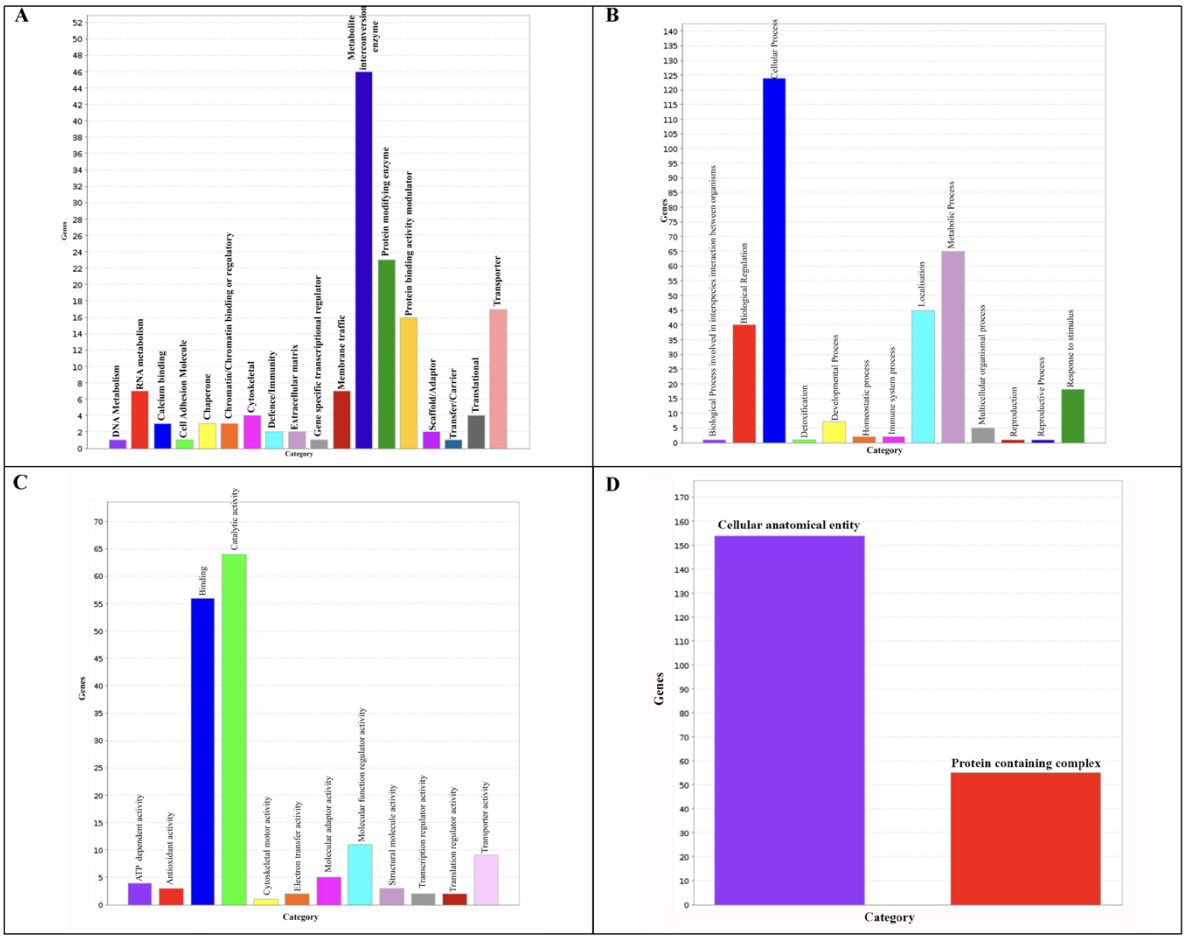


Figure 27. Ontological classification of proteins differentially expressed in HPV-positive compared to HPV-negative tonsil tissue. UniProt IDs of 222 protein hits identified as significantly differentially expressed in HPV-positive versus HPV-negative tonsil tissue were uploaded to PANTHER online software (version 19) and mapped to gene identities available in the database. 24 protein IDs in the search query were not found. The remaining 198 genes identified were examined based on their protein classification (A), biological process (B), molecular function (C), and cellular component (D). In all panels, the number of gene hits in each category is shown. In all analyses, genes that were considered unclassified in the Panther database in each category were removed; this equates to 55 genes in protein classification, 52 in biological process, 82 genes in molecular function, and 43 in cellular component.

6.4.4. Pathway enrichment analysis of differentially expressed proteins in HPV-positive tonsil tissue

To further scrutinise the impact of HPV within the non-cancerous oropharynx on the host cell biology, we performed pathway enrichment analysis in reactome.org software. Using the total statistically significant protein hits (N=222) identified by differential abundance analysis of HPV-positive versus HPV-negative tonsillar protein (Appendix E, Table 24) we searched the cell signalling pathways in which these proteins are involved to determine overrepresentation of these proteins in specific pathways/networks. UniProt IDs of 222 protein hits identified as differentially expressed between HPV-positive versus HPV-negative non-cancerous tonsil tissue were uploaded to reactome.org and searched in their database. 65 of these protein hits were not found in this analysis.

We first examined the overall functional networks that these protein hits appear within to determine whether differentially expressed proteins within HPV-positive tonsil tissues showed an association with any broad functional networks that would suggest general impacts on host cell biology (i.e., to metabolism through just one or few proteins appearing across several pathways in this network). We found that at least one of the search proteins were present within most of the networks searched in this analysis, with the exception of 'digestion and absorption' and 'drug absorption, distribution, metabolism, and excretion (ADME)', in which none of the searched proteins were mapped to a pathway in this network. Perhaps, as expected, with a small number of proteins analysed, none of the overall functional networks were significantly overrepresented by these proteins based on p-value adjusted for FDR. Although no significant association was found with a particular network, the network with the most proteins found within (number of entities found/total entities, FDR) was autophagy. These data indicate that the differentially expressed proteins identified between HPV-positive and HPV-negative tonsil tissue are not significantly localised to a particular network of biological pathways.

The next step was to determine individual pathways that were enriched within these protein hits of HPV-positive tonsil tissue to provide an insight into specific biological functions that may be impacted by HPV. The top 25 overrepresented pathways ranked on their significance are outlined in Table 15. These were sorted into their highest hierarchical overall network to determine whether these hits belonged to, which showed that these pathways were involved in: metabolism of protein pathways (N=5/25), four were involved in vesicle-mediated transport, three in metabolism, three in autophagy, two in disease pathways, two in protein localisation,

and one pathway was enriched in each of the following: immune system, cell cycle, programmed cell death, metabolism of RNA, signal transduction, and organelle biogenesis/maintenance.

Table 15. Top 25 significantly enriched pathways of 222 protein hits identified by DEA of HPV-positive versus HPV-negative tonsil tissue. The 25 most significant biological pathways of which the set of proteins searched are overrepresented in are ranked in order of their p-value. The number of the searched proteins (entities) of the total proteins present within that pathway and the entities ratio (of the number of proteins within this pathway of the total reactome database) is shown alongside the calculated p-value and p-value adjusted for false discovery (FDR). For reactions within the pathway, the number of reactions in which the found protein entities contribute to of the total reactions is shown alongside the reactions ratio (calculated as per the entities ratio). Pathways significant after adjusting for false discovery are highlighted in yellow.

Pathway name	Entities				Reactions	
	Found	Ratio	p-value	FDR*	Found	Ratio
Respiratory electron transport	16/158	0.013	8.19e-08	8.53e-05	27/47	0.003
Complex I biogenesis	10/68	0.006	9.15e-07	4.76e-04	10/14	9.23e-04
Asparagine N-linked glycosylation	19/307	0.026	7.70e-06	0.003	35/145	0.01
Aerobic respiration and respiratory electron transport	17/262	0.022	1.30e-05	0.003	30/117	0.008
Intra-Golgi Traffic	7/44	0.004	2.47e-05	0.005	6/10	6.59e-04
Mitochondrial protein degradation	9/98	0.008	1.23e-04	0.021	13/20	0.001
RAB geranylgeranylation	7/65	0.005	2.72e-04	0.04	2/5	3.30e-04
Activation and oligomerisation of BAK protein	2/2	1.69e-04	6.95e-04	0.085	3/3	1.98e-04
COP1-mediated anterograde transport	8/102	0.009	8.14e-04	0.085	11/12	7.91e-04
Selective autophagy	7/90	0.008	0.002	0.168	27/56	0.004
Intra-Golgi and retrograde Golgi-to-ER traffic	11/206	0.017	0.002	0.181	19/48	0.003
Metabolism of proteins	56/2073	0.175	0.003	0.219	123/887	0.058
ER to Golgi Anterograde Transport	9/156	0.013	0.003	0.219	25/39	0.003
Neutrophil degranulation	18/478	0.04	0.005	0.289	8/10	6.59e-04
Protein localisation	9/166	0.014	0.005	0.289	20/53	0.003
Maturation of spike protein	4/38	0.003	0.006	0.351	1/2	7.91e-04
Macroautophagy	8/144	0.012	0.007	0.351	36/95	0.006
Initiation of Nuclear Envelope (NE) Reformation	3/20	0.002	0.007	0.351	2/7	4.62e-04
DAG and IP3 signalling	4/41	0.003	0.008	0.353	13/28	0.002
Mitochondrial protein import	5/65	0.005	0.008	0.353	10/14	9.23e-04
Transport to the Golgi and subsequent modification	9/187	0.016	0.01	0.353	25/60	0.004
HuR (ELAVL1) binds and stabilises mRNA	2/8	6.76e-04	0.01	0.353	3/7	4.62e-04
Mitochondrial biogenesis	6/98	0.008	0.011	0.353	4/36	0.002
Autophagy	8/159	0.013	0.011	0.353	47/116	0.008
Maturation of hRSV A proteins	3/25	0.002	0.012	0.353	4/40	0.003

To increase the stringency of this analysis, we further scrutinised the enriched pathways with a $p<0.05$ when adjusted for false discovery (FDR value in Table 15). Seven out of the 25 most overrepresented pathways remained statistically significant when using the FDR value, these were as follows, ordered from most to least significant: respiratory electron transport, complex I biogenesis, asparagine N-linked glycosylation, aerobic respiration and respiratory electron transport, intra-Golgi traffic, mitochondrial protein degradation, and RAB geranylgeranylation. To note, the respiratory electron transport, complex I biogenesis, and aerobic respiration and respiratory transport pathways, and then asparagine N-linked glycosylation, mitochondrial protein degradation, and RAB geranylgeranylation all contribute to the same overall function network of metabolism and metabolism of proteins, respectively, strengthening the evidence for HPV-mediated change here.

To determine whether these pathways were over- or under-expressed in HPV-positive tonsil tissue with respect to HPV-negative, the protein hits found as entities within each pathway were analysed and their FC direction (positive or negative) was determined. We found that with the exception of just one individual protein hit in just one of the enriched pathways identified (SSBP_HUMAN, single-stranded DNA-binding protein, mitochondrial) (Table 16), all the proteins searched that were mapped to these pathways had demonstrated decreased expression in HPV-positive tonsil tissue with respect to HPV-negative tonsil tissue. We suspect that these pathways are likely to have their canonical function opposed rather than enhanced, and given that analysis of the protein roles within the top 7 pathways (those significant by FDR) showed they were involved in positive regulation or association, their decreased expression would suggest these pathways are more likely to be downregulated or repressed than upregulated in HPV-positive individuals.

These data show the respiratory electron transport pathway to be the most enriched within the protein hits identified as differentially expressed in HPV-positive (with respect to HPV-negative) tonsil tissues ($p=8.19e-08$, FDR=8.53e-05). 16 of a total 158 proteins within the pathway were found within the dataset, with these found in three of four complexes of the chain (complex I, III, and IV) and also involved in the synthesis of such compounds (Figure 28). In addition, complex I biosynthesis was found to be an enriched pathway individually contributing to the overall process.

Overall, this preliminary data suggests that broadly aerobic respiration and protein metabolism are most effected by downregulation of proteins in HPV-positive tonsil tissue.

Table 16. Significantly enriched pathways within HPV-positive versus HPV-negative protein hits by FDR and their mapped proteins for each pathway. Seven pathways remained significantly enriched when p-value was adjusted for false discovery (FDR<0.05). Significantly changed proteins (between HPV-positive versus HPV-negative tonsils) mapped to each pathway are shown, and are highlighted in blue or red to demonstrate whether each protein was identified as decreased or increased expression by differential abundance analysis, respectively. * indicates that the enriched pathways are all of the wider metabolism network and ** the protein metabolism network.

Pathway	Protein hits found				Overall change
Respiratory electron transport*	COX15_HUMAN	QCR10_HUMAN	NDUB6_HUMAN	NDUA9_HUMAN	Decreased
	NDUA3_HUMAN	TIDC1_HUMAN	NU5M_HUMAN	NDUBB_HUMAN	
	NDUB5_HUMAN	COX2_HUMAN	SURF1_HUMAN	OXA1L_HUMAN	
	NU2M_HUMAN	NDUA4_HUMAN	CY1_HUMAN	T126A_HUMAN	
Complex I biogenesis*	NDUA3_HUMAN	TIDC1_HUMAN	OXA1L_HUMAN	NU2M_HUMAN	Decreased
	NDUB6_HUMAN	NDUA9_HUMAN	NDUB5_HUMAN	T126A_HUMAN	
	NU5M_HUMAN	NDUBB_HUMAN			
Asparagine N-linked glycosylation**	ALG12_HUMAN	RPN2_HUMAN	GOSR1_HUMAN	GFPT1_HUMAN	Decreased
	COG4_HUMAN	YKT6_HUMAN	OST48_HUMAN	GOSR2_HUMAN	
	DAD1_HUMAN	ALG5_HUMAN	SAR1B_HUMAN	RPN1_HUMAN	
	GLCNE_HUMAN	COG5_HUMAN	ARF4_HUMAN	SYVN1_HUMAN	
	MPU1_HUMAN	DYHC1_HUMAN	COG6_HUMAN		
Aerobic respiration and respiratory electron transport*	ATP5S_HUMAN	T126A_HUMAN	QCR10_HUMAN	SURF1_HUMAN	Decreased
	CY1_HUMAN	COX15_HUMAN	TIDC1_HUMAN	NDUB6_HUMAN	
	NDUA9_HUMAN	NDUA3_HUMAN	COX2_HUMAN		
	NDUBB_HUMAN	NDUB5_HUMAN	NDUA4_HUMAN		
	OXA1L_HUMAN	NU2M_HUMAN	NU5M_HUMAN		
Intra-Golgi Traffic	COG4_HUMAN	YKT6_HUMAN	GOSR2_HUMAN	RAB30_HUMAN	Decreased
	GOSR1_HUMAN	COG5_HUMAN	COG6_HUMAN		
Mitochondrial protein degradation**	ADT3_HUMAN	COX2_HUMAN	SSBP_HUMAN		Decreased
	NDUB6_HUMAN	NU2M_HUMAN	KAPCA_HUMAN		
	SPG7_HUMAN	NU5M_HUMAN	YMEL1_HUMAN		
RAB geranylation**	RAB18_HUMAN	RAB4B_HUMAN	RAB3D_HUMAN	RAB4A_HUMAN	Decreased
	RAB34_HUMAN	RAB2A_HUMAN	RAB30_HUMAN		

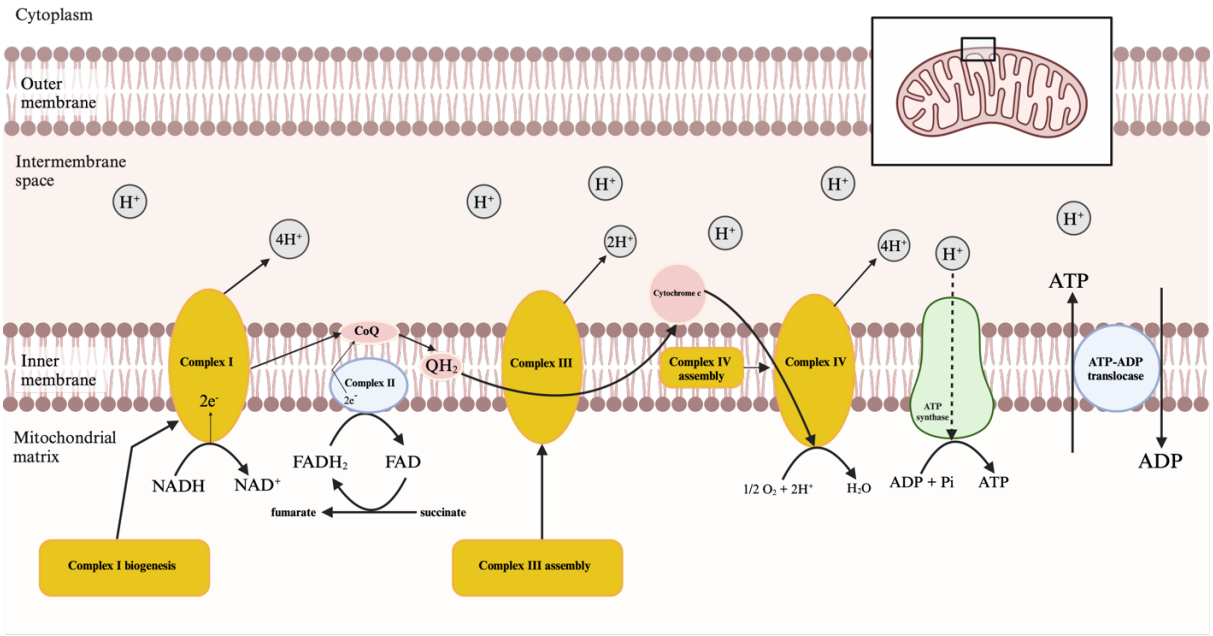


Figure 28. Respiratory electron transport pathway within the mitochondria is most significantly overrepresented within the statistically differentially expressed proteins of HPV-positive versus HPV-negative tonsil tissue. 222 significantly changed protein hits, identified by DEA (using Amica software) were analysed in Reactome.org for pathway enrichment analysis. The respiratory electron pathway was identified as most significantly overrepresented pathway within these proteins. 16 entities from the protein hits of a total 158 proteins within this pathway were found (FDR: 8.53e-05). A diagram of the pathway to represent the location of the proteins hits was produced in BioRender software; protein complexes or specific processes that these protein hits contribute to are coloured in yellow. All entities from the top hits search that were identified within this pathway were found to have decreased expression in HPV-positive tonsils in comparison to HPV-negative tissues.

Figure produced in BioRender.com

6.5. Analysis of HPV-positive males and HPV-positive females

There is a greater burden of oral HPV infection in healthy men than healthy women and cases of HPV-OPC are more frequent in men (Gillison et al., 2012, Guo et al., 2022). Whilst our primary aim was to determine protein expression changes arising from an oropharyngeal HPV infection within the tonsils, we hypothesised there may be further differences in HPV-positive individuals based on their biological gender. We aimed to determine the impact of tonsillar HPV in men on host cell protein expression, and how it may differ to that of women.

All protein samples selected for this comparison had been extracted from fresh-frozen tonsil tissue and they were chosen to have similar numbers of protein IDs. A seven HPV-positive female versus seven HPV-positive male comparison was performed to include protein samples from as many HPV-positive individuals as possible, with representation of protein from their left and right tonsil tissues. The number of protein IDs of HPV-positive males was 3346-4497 and of females, 3714-5302 (Table 17). Please note, the HPV-positive samples in this sub-comparison were not all used in the HPV-positive versus HPV-negative DEA (section 6.4; above), this was because there were fewer HPV-negative protein samples available and so a smaller number of HPV-positive samples were used to be statistically comparable.

Table 17. Total raw precursor numbers and proteins identified by DIA-NN searching of samples included in HPV-positive male versus HPV-positive female DEA. The sample identifier (SWATH_number) and the gender of the participant whose tonsil protein was isolated from is shown. The raw precursor numbers from SWATH-MS are shown against the total number of proteins identified through SWATH data searching in DIA-NN.

SWATH sample	Gender	Total precursors	Total protein IDs
SWATH_197	Male	30422	4497
SWATH_195	Male	23053	4483
SWATH_194	Male	24977	4291
SWATH_196	Male	24585	4288
SWATH_191	Male	10740	4170
SWATH_187	Male	19473	3847
SWATH_186	Male	19508	3346
SWATH_199	Female	28767	5302
SWATH_198	Female	31146	5286
SWATH_201	Female	18297	5203
SWATH_203	Female	28802	4662
SWATH_184	Female	17494	3900
SWATH_185	Female	15569	3721
SWATH_202	Female	24780	3714

As before, differential protein expression was performed using Amica online software (v 3.0.1). The parameters for data analysis were as the software default; however, normal imputation was selected at a stringency of a minimum valid values 'in_one_group' of five for imputation. The number of proteins in the pg_matrix was 6302, and based on the imputation instructions, a total of 4464 proteins were quantified, and 1838 were excluded from the analysis.

The QC tests were inspected for the imputed intensities used by the software to generate statistical comparisons. PCA plot of HPV-positive female and male groups showed some clustering between comparative groups, but also with some outliers (Figure 29A). The plot of imputed intensities indicates that some of the differences identified in this analysis are a result of the comparative variable (gender in HPV-positive individuals), as the PC1 (43.53%) and PC2 (12.98%) values are over 50% when combined (Figure 29A). There is little difference in the median imputed intensity across all samples in both groups, with generally consistent IQRs; however, there are several outliers above the maximum score within both comparative groups (Figure 29B). When comparing all samples in each group together, the CV plot suggested a lot of variation within the groups; however, the high number of outliers in imputed intensities of each sample may have skewed this result (Figure 29C). Therefore, we conclude a high level of correlation in imputed intensities of samples of each group, with a clear correlation of all HPV-positive male samples that inversely show low similarity to HPV-positive females (Figure 29D).

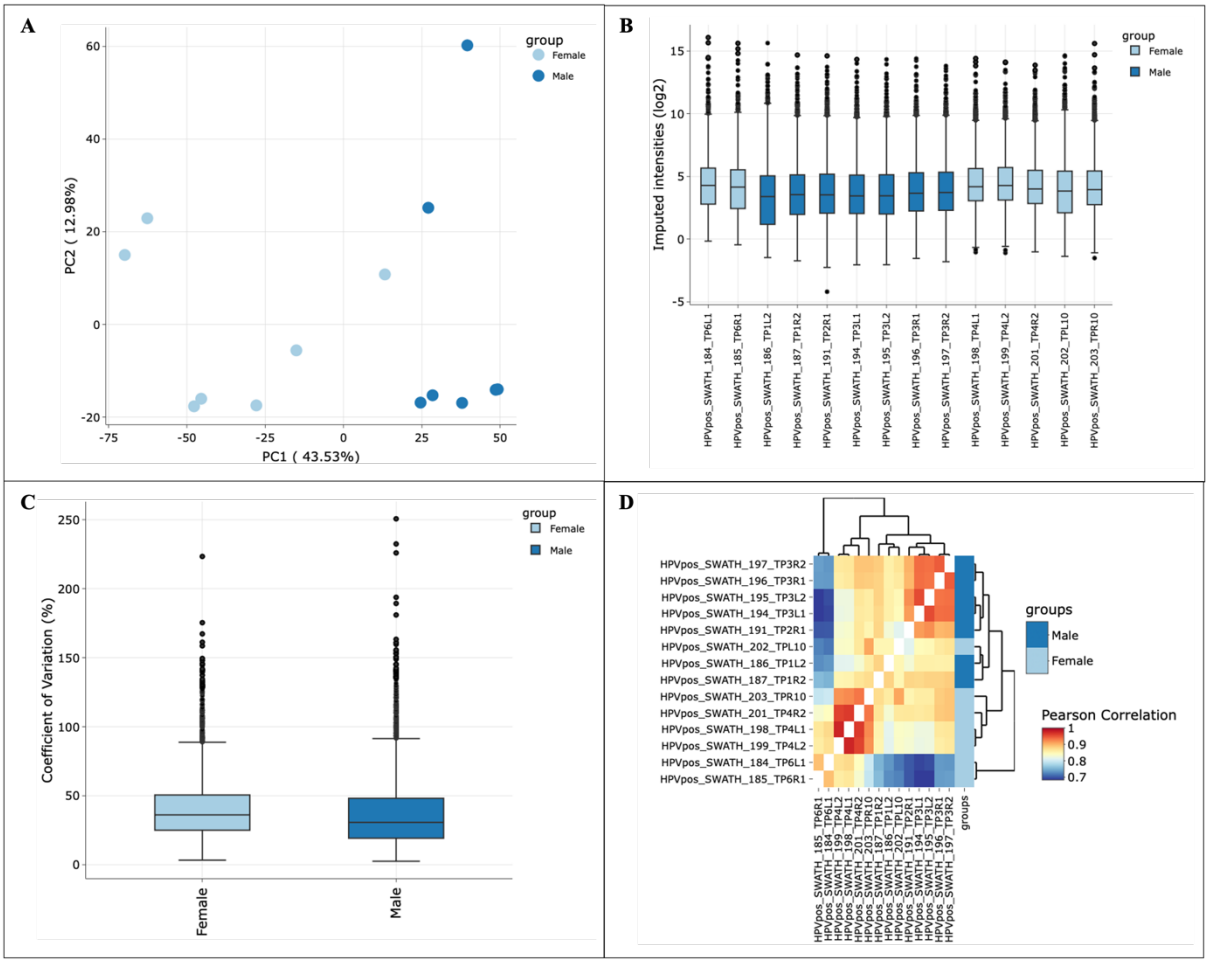


Figure 29. Quality control data generated by Amica software of differential abundance analysis of protein from HPV-positive male versus HPV-positive female non-cancerous tonsil tissue. Protein expression analysis was performed in Amica software (v 3.0.1) using the limma package for statistical analysis. This statistical analysis is based on the 4464 proteins quantified following normal imputation. **(A)** PCA plot of HPV-positive male and HPV-positive female imputed intensities. **(B)** Box plot of imputed intensities of each sample. **(C)** Coefficient of Variation plot of imputed intensities between comparative groups. **(D)** Pearson correlation plot to compare replicates of HPV-positive male and HPV-positive female samples. In all panels, HPV-positive male protein samples are represented in dark blue and HPV-positive female samples in light blue.

6.5.1. Differential protein expression of HPV-positive individuals based on their biological sex

We selected to analyse the differential protein expression statistics of HPV-positive male versus HPV-positive female, rather than the vice versa comparison, to reveal protein changes that might be responsible for the higher prevalence of oral HPV infections and disease in men. This direction of comparison between groups therefore means that any hits described here with a positive log_2FC have increased in the HPV-positive males in comparison to the HPV-positive female group, used as a reference, and the respective hits with a negative FC have decreased. The calculated adjusted p-values and logFC data were filtered in Microsoft Excel to indicate raw numbers of hits. We found 1376 hits with $p<0.05$ and 651 with $p<0.01$. With the addition of filtering for hits with a logFC ± 0.6 we found 1290 ($p<0.05$, logFC ± 0.6) and 646 ($p<0.01$, logFC ± 0.6) hits. In the HPV-positive versus HPV-negative comparison we used the p-values; however, given the large number of statistically significant adjusted p-values quantified in these data, we were able to use these adjusted values for determination of top protein hits for greater stringency.

All adjusted p-values and logFC data calculated by Amica for the total of 4464 quantified proteins were uploaded to VolcaNoseR2 for visualisation of top protein hits. The shape of the volcano plot of all data points, regardless of p-value/logFC thresholds for significance, indicated that most proteins had decreased expression in HPV-positive males with respect to HPV-positive females. Given the vast number of hits indicated by initial data analysis in Microsoft Excel, the stringency of significance thresholds were increased. A volcano plot was produced with the thresholds of $-\log_{10}$ adjusted p-value > 2 and $\log_2$ FC ± 1.5. The coloured dots represent the total 394 significantly changed hits in HPV-positive males compared to HPV-positive females (blue = decreased and red = increased) (Figure 30). We found only six hits that had increased using these parameters. The most significantly changed (in either positive or negative FC direction) are annotated in Figure 30. A summary of the adjusted p-value (significance), log_2FC (and direction of change), and the Manhattan distance of the data point corresponding to each top hit is shown in Table 18.

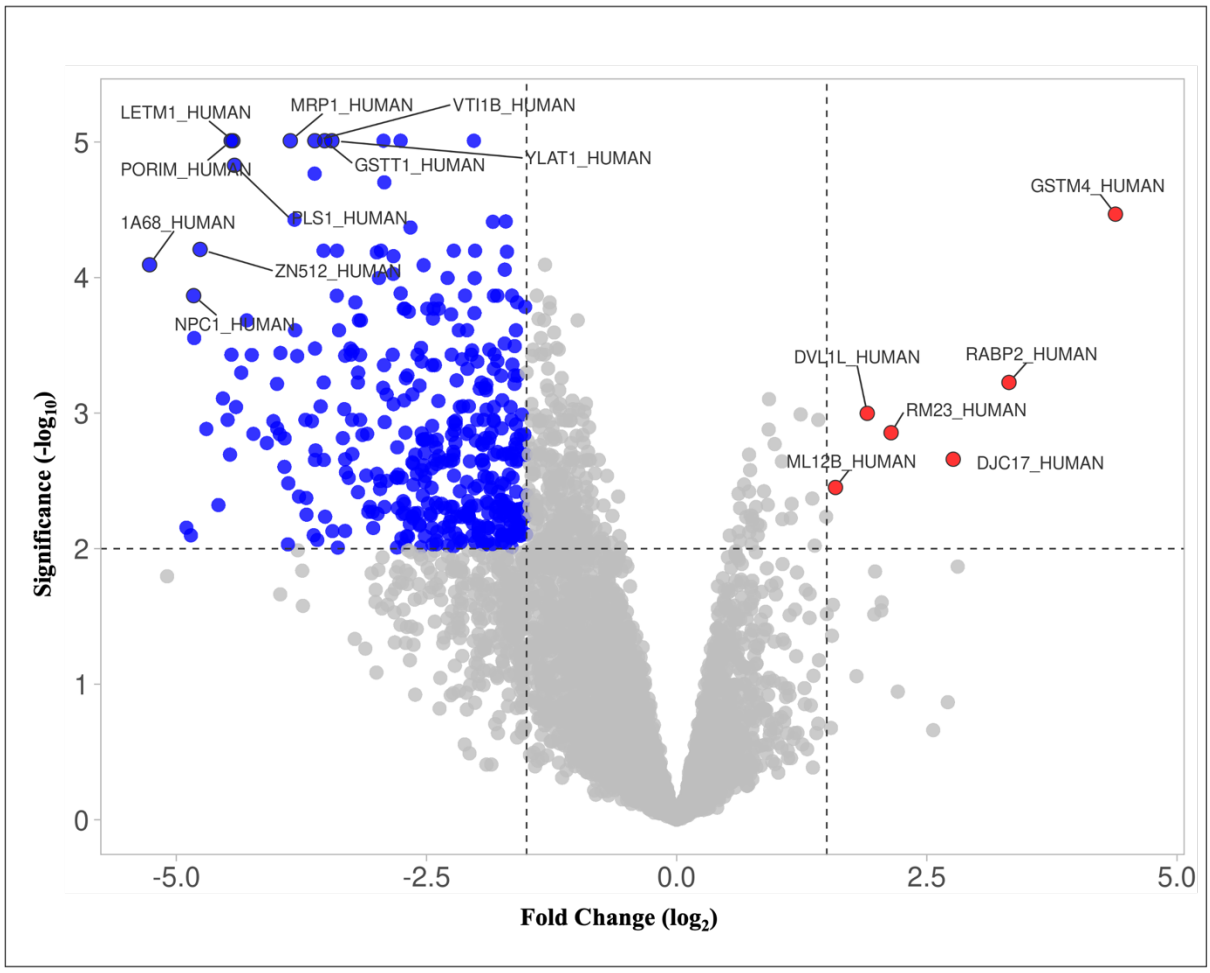


Figure 30. 394 significantly changed protein hits in HPV-positive males compared to HPV-positive females when filtered by a log_2FC ±1.5 and -log_{10} adj. p-value > 2. Log_2FC and adj. p-values were generated in a differential abundance analysis by online software Amica (v 3.0.1). Adj. p-values were converted to their -log_{10} equivalent in Microsoft Excel and -log_{10} adj.p-values, log_2FC, and protein IDs were uploaded to VolcaNoseR2 software. All data points are representative of the 5103 proteins within the dataset available for quantification. The proteins were filtered based on a -log_{10} p-value of 2 (equivalent to $p<0.01$) and a log_2FC ±1.5 (absolute FC ~2.8). Blue data points indicate a negative FC, with expression significantly lower in HPV-positive males compared to HPV-positive females. Red data points indicate a positive FC, with expression significantly higher in HPV-positive male samples compared to female. The top 6 increased (and significant) and top 10 decreased (and significant) hits are labelled with their UniProt IDs.

Table 18. Top six significantly increased and 10 decreased protein hits identified in HPV-positive male tonsil tissue compared to HPV-positive female tonsil tissue. The UniProt IDs and corresponding protein name for each hit is shown against the direction of expression change (e.g., increased = a positive FC in HPV-positive male tissues compared to HPV-positive female tonsil tissue). The significance shown is equal to a $-\log_{10}$(adj. p-value). These hits were selected based on their Manhattan distance, with the greatest distance considered most significant and then re-ordered as most significantly changed (by p-value) to least changed.

UniProtID	Protein Name	Change	Fold change (log2)	Significance	Manhattan distance
GSTM4_HUMAN	Glutathione S-transferase Mu 4	Increased	4.384694194	4.467876992	8.852571185999999
RABP2_HUMAN	Cellular retinoic acid-binding protein 2	Increased	3.321038009	3.226324702	6.547362711
DVL1L_HUMAN	Segment polarity protein dishevelled homolog DVL-1-like	Increased	1.904271508	2.998528933	4.902800441
RM23_HUMAN	Large ribosomal subunit protein uL23m	Increased	2.14293694	2.855298169	4.998235108999999
DJC17_HUMAN	DnaJ homolog subfamily C member 17	Increased	2.76383171	2.659809595	5.423641305
ML12B_HUMAN	Myosin regulatory light chain 12B	Increased	1.58709026	2.451515379	4.038605639
PORIM_HUMAN	Porimin	Decreased	-4.451948635	5.009704775	9.46165341
LETM1_HUMAN	Mitochondrial proton/calcium exchanger protein	Decreased	-4.434403184	5.009704775	9.444107959
MRP1_HUMAN	Multidrug resistance-associated protein 1	Decreased	-3.860851735	5.009704775	8.87055651
VTI1B_HUMAN	Vesicle transport through interaction with t-SNAREs homolog 1B	Decreased	-3.615473431	5.009704775	8.625178206000001
GSTT1_HUMAN	Glutathione S-transferase theta-1	Decreased	-3.517205874	5.009704775	8.526910649000001
YLAT1_HUMAN	Y+L amino acid transporter 1	Decreased	-3.445273592	5.009704775	8.454978367
PLS1_HUMAN	Phospholipid scramblase 1	Decreased	-4.418351975	4.828374209	9.246726184
ZN512_HUMAN	Zinc finger protein 512	Decreased	-4.762086206	4.207903986	8.969990192000001
1A68_HUMAN*	HLA class I histocompatibility antigen, A alpha chain	Decreased	-5.267222219	4.094538531	9.36176075
NPC1_HUMAN	NPC intracellular cholesterol transporter 1	Decreased	-4.826486752	3.866560097	8.693046849

* = indicates that the original UniProt ID identified by SWATH data searching in DIA-NN has now been merged with another UniProt ID entry name. In these instances, the protein name given is of that corresponding to the merged UniProt ID entry name.

6.5.2. Ontological analysis of differential protein expression data of HPV-positive male versus HPV-positive female non-cancerous tonsil tissue

UniProt IDs of proteins identified as significantly changed in DEA of HPV-positive males versus HPV-positive females non-cancerous tonsil tissue (N=394) were uploaded to PANTHER and searched against the *Homo sapiens* database. All IDs mapped to their corresponding gene (N=374) were further classified based on their protein classification, and any biological processes, molecular functions, and cellular components they are involved in, or carry out (Figure 31). As before, the nomenclature of reported data for any category of hits is as follows: X/total genes searched, % of total number of genes searched, % total hits for each category/classification.

Firstly, 374 genes were categorised into protein families, and based on the PANTHER database 84 were unable to be classified. Similarly, to the HPV-positive versus HPV-negative differentially expressed proteins, a large proportion of these protein hits were identified as part of the metabolic interconversion enzyme family (45/374, 12%, 15.5%). Another equally enriched protein family amongst these hits was transporter proteins (45/374, 12%, 15.5%). Again, only a small number of these proteins were found to be immune/defence (10/374, 2.7%, 3.4%). The remaining genes were categorised into 18 different families (Figure 31A).

Regarding the broad contribution of these proteins to any biological process/es, we found that most of these proteins perform cellular processes (224/374, 59.9% 35.9%); some examples of the most enriched of these processes described in the further breakdown of this category include cellular metabolic and cellular component organisation or biogenesis processes. In addition, many of these hits appear to be metabolic processes (95/374, 25.4%, 15.2%), biological regulation (93/374, 24.9%, 14.9%), and localisation (82/374, 21.9%, 13.1%). 111/374 were unclassified by PANTHER database by biological process searching and not included in Figure 31B.

We found that most of these proteins are molecularly involved in binding (124/374, 33.2%, 38.4%) and/or catalytic activities (98/374, 26.2%, 30.3%) (Figure 31C). The remaining enriched categories for molecular function were as identified for protein hits of HPV-positive versus HPV-negative tonsil tissue, with one additional group identified in this instance (cargo receptor activity; 1, 0.3%, 0.3%). 143 genes were unable to be classified by PANTHER for their molecular function. Finally, these proteins hits were categorised as a part of a cellular

anatomical entity and/or a protein containing complex when searched for their cellular component (Figure 31D), with 72 genes categorised as unclassified within this classification.

These data demonstrate that the protein hits identified as differentially expressed between HPV-positive males and HPV-positive females are broadly categorised into similar groups to the top hits for HPV-positive versus HPV-negative differential expression. We found that, again, the majority of these protein hits were metabolic interconversion enzymes. Whilst these proteins also have a range of biological process/es and molecular function/s, this ontological characterisation demonstrates that cellular processes and binding are the most frequent hit for each category, respectively.

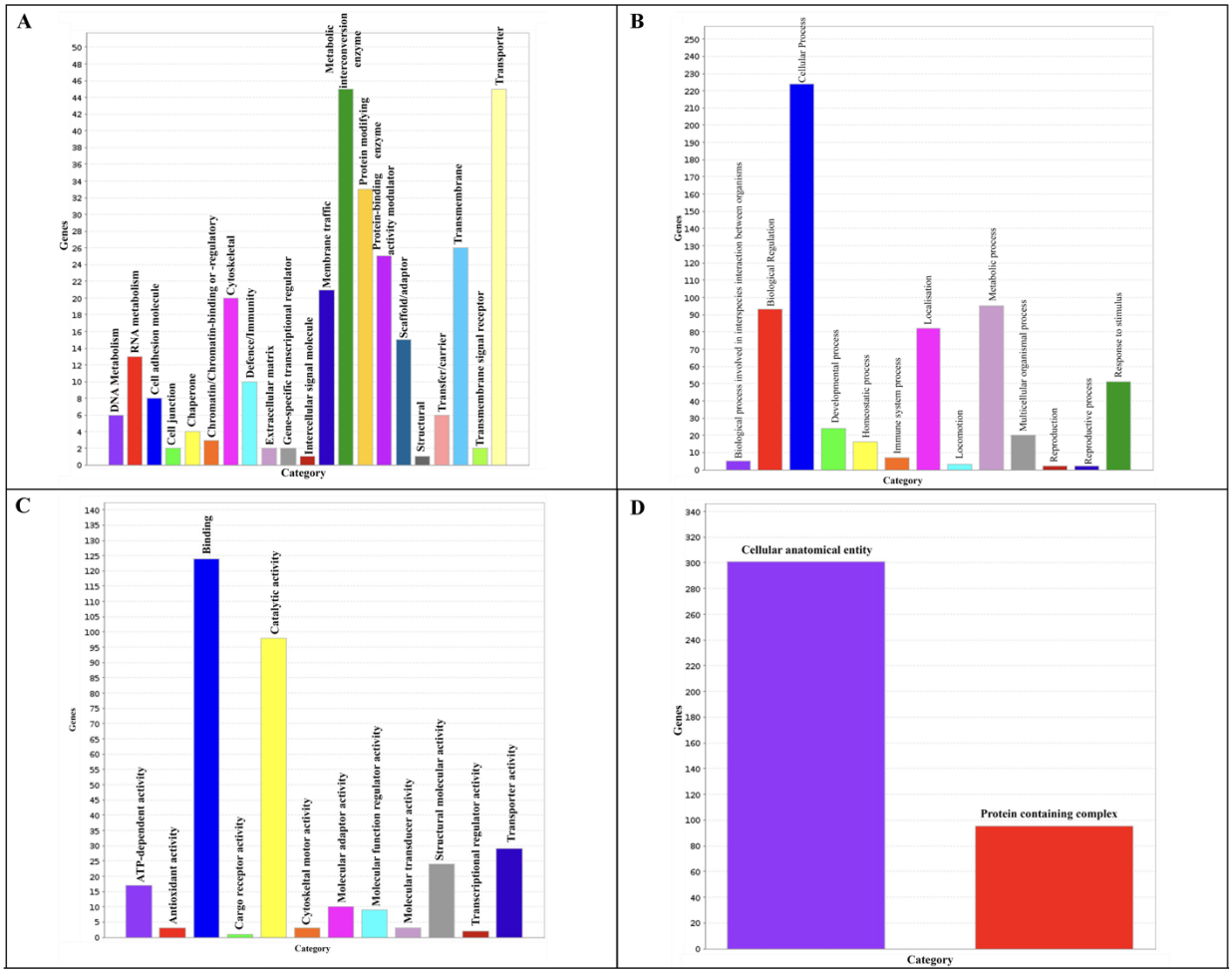


Figure 31. Ontological classification of proteins differentially expressed in non-cancerous tonsil tissue of HPV-positive males versus HPV-positive females. UniProt IDs of 394 protein hits identified as significantly differentially expressed in HPV-positive male versus HPV-positive female tonsil tissue were uploaded to PANTHER online software (version 19) and mapped to gene identities available in the database. 20 protein IDs in the search query were not found. The remaining 374 genes identified were examined based on their protein classification (A), biological process (B), molecular function (C), and cellular component (D). In all panels, the number of gene hits in each category is shown. In all analyses, genes that were considered unclassified in the PANTHER database in each category were removed; this equates to 84 genes in protein classification, 111 in biological process, 143 genes in molecular function, and 72 in cellular component.

6.5.3. Pathway enrichment analysis of differentially expressed proteins between HPV-positive males and females

To examine differences in how HPV influences host cell biology in males versus females; a pathway enrichment analysis was performed in Reactome software (version 90, https://reactome.org) to determine whether the panel of differentially expressed proteins significantly localised to broad functional networks and/or specific biological pathways. UniProt IDs of 394 protein hits identified as significantly differentially expressed in HPV-positive males compared to HPV-positive females were searched against the Reactome database to assess overrepresentation of these hits within networks and their individual pathways. 111 searched proteins were unidentified by the software and so were not included in this analysis.

The remaining 283 proteins appeared in all but one of the broad functional networks of pathways described in Reactome; we found none of these proteins mapped to any pathways of the 'digestion and absorption' network. Interestingly; in this instance, unlike in the HPV-positive versus HPV-negative pathway enrichment analysis, these protein hits did significantly map to some overall networks with FDR <0.05 as follows: haemostasis (FDR = 2.33e-03), transport of small molecules (FDR = 4.41e-03), and vesicle mediated transport (FDR = 3.04e-02). These data indicate that when filtered by overall network, these differentially expressed proteins in HPV-positive male compared to HPV-positive female tonsil tissue have the greatest impact on the overall processes of haemostasis, transport of small molecules, and vesicle mediated transport as broad networks of several pathways.

Next, the top overrepresented pathways within these protein hits were assessed to determine if significant association was found within any specific biological pathways (N=25) (Table 19). We found that the top 25 overrepresented pathways were part of the following broad functional networks: protein metabolism (N=9/25), RNA metabolism (N=6/25), disease (N=3/25), metabolism (N=2/25), cellular responses to stimuli (N=2/25), developmental biology (N=1/25), haemostasis (N=1/25), and immune system (N=1/25). These data show that the overall networks from which the top 25 pathways belong to are mostly not statistically significantly enriched themselves, with the exception of haemostasis. Instead, these top pathway hits represent clusters of proteins within specific pathways that would impact a precise biological process rather than an effect on the entire broad function of a network itself.

In this sub-analysis (HPV-positive males verses HPV-positive females) the top 25 pathways enriched within these protein hits were all statistically significant when adjusting for false discovery ($p<0.05$). As such, the direction of change of individual protein hits mapped to these pathways by differential protein expression analysis were consulted for all pathways to determine whether these data suggest broad promotion or inhibition of these biological pathways and reactions. Unsurprisingly, given the vast number of proteins identified as decreased in HPV-positive males, we found that all of the significantly changed proteins found within the top 25 overrepresented pathways were found to have decreased expression in HPV-positive male tonsil tissue (compared to HPV-positive female tissues), this suggests the change in protein expression inversely alters its canonical function.

The most overrepresented pathway of these protein hits was the major pathway of rRNA processing in the nucleolus and cytosol (p-value = 1.05e-11, FDR =8.35e-11) that results in the production of the 40S and 60S subunits of the eukaryotic ribosome. This pathway is comprised of 183 proteins and 29 of these were identified as significantly decreased within HPV-positive males compared to HPV-positive females. Some of the 29 proteins were involved either as essential or non-essential components or associates of the final ribosome, and contribute to multiple reactions within the overall ribosomal biogenesis process. In addition to this pathway, various sub-branches and similar hits all describing rRNA processing were identified as significantly enriched within the top 25 hits including rRNA processing in the nucleolus and cytosol and rRNA processing.

Overall, given the overrepresentation within, and negative impacts on ribosomal biogenesis as a pathway, it is unsurprising then these ribosomal proteins were present in many of the top significantly enriched pathways identified in these analyses. Several other pathways that are reliant upon ribosomal function were identified, and again, analysis of individual proteins found within these enriched pathways suggested that there was a negative impact on their progression as a result of significantly decreased expression of these ribosomal proteins/associated proteins within HPV-positive male tonsil tissue. Taken together, these data suggest that ribosomal biogenesis (rRNA processing) was most significantly impacted in HPV-positive males versus HPV-positive females and that other pathways that involve the ribosome are consequently impacted by the decreased expression of these proteins.

Table 19. 25 most enriched pathways within 394 significantly changed protein hits identified by DEA of HPV-positive male and HPV-positive female tonsil tissue. The 25 most significantly enriched biological pathways of which the searched protein hits are overrepresented within are ranked in order of their p-value; where calculated p-value is the same between pathways, they are further ranked on the number of entities (proteins) found within the pathway, and again on reactions found if entities found are also equal. The number of searched entities of the total proteins present within each pathway, and entities ratio (of the number of proteins within this pathway of the total reactome database) is shown alongside p-value and p-value adjusted for false discovery (FDR). For individual reactions within the pathway, the number of reactions in which the protein entities that were mapped to the pathway contribute to as a fraction of the total number of reactions within the pathway is shown alongside the reactions ratio (calculated as per entities ratio but for total reactions within reactome database).

Pathway name	Entities				Reactions	
	Found	Ratio	p-value	FDR*	Found	Ratio
Major pathway of rRNA processing in the nucleolus and cytosol	29/183	0.015	1.05e-11	8.35e-11	7/7	4.62e-04
SRP-dependent co-translational protein targeting to membrane	23/113	0.01	1.26e-11	8.35e-09	5/5	3.30e-04
Nonsense Mediated Decay (NMD) independent of the Exon Junction Complex (EJC)	21/96	0.008	2.72e-11	1.20e-08	1/1	6.59e-05
rRNA processing in the nucleus and cytosol	29/193	0.016	3.64e-11	1.21e-08	9/15	9.89e-04
Peptide chain elongation	20/90	0.008	6.17e-11	1.64e-08	4/5	3.30e-04
rRNA processing	29/203	0.017	1.17e-10	2.15e-08	9/21	0.001
Eukaryotic Translation Termination	20/94	0.008	1.31e-10	2.15e-08	3/5	3.30e-04
Selenocysteine synthesis	20/94	0.008	1.31e-10	2.15e-08	2/7	4.62e-04
Eukaryotic Translation Elongation	20/95	0.008	1.57e-10	2.30e-08	4/9	5.94e-04
Viral mRNA Translation	20/101	0.009	4.44e-10	5.77e-08	2/2	1.32e-04
Formation of a pool of free 40S subunits	20/102	0.009	5.25e-10	5.77e-08	4/16	0.001
Response of EIF2AK4 (GCN2) to amino acid deficiency	20/102	0.009	5.25e-10	5.77e-08	123/887	0.058
Nonsense-Mediated Decay	21/117	0.01	9.17e-10	8.62e-08	6/6	3.96e-04
Nonsense-Mediated Decay enhanced by the Exon Junction Complex	21/117	0.01	9.17e-10	8.62e-08	5/5	3.30e-04
L13a-mediated translational silencing of Ceruloplasmin expression	20/112	0.009	2.52e-09	2.22e-07	3/3	1.98e-04
Signalling by ROBO receptors	27/205	0.017	2.87e-09	2.25e-07	11/60	0.004
GTP hydrolysis and joining of the 60S ribosomal subunit	20/113	0.01	2.92e-09	2.25e-07	3/3	1.98e-04
Selenoamino acid metabolism	20/118	0.01	5.97e-09	4.35e-07	2/33	0.002
Cellular response to starvation	23/157	0.013	6.42e-09	4.43e-07	12/28	0.002
Influenza Infection	24/172	0.015	7.53e-09	4.72e-07	9/59	0.004
Eukaryotic Translation Initiation	20/120	0.01	7.86e-09	4.72e-07	13/21	0.001
Cap-dependent Translation Initiation	20/120	0.01	7.86e-09	4.72e-07	10/18	0.001
Influenza Viral RNA Transcription and Replication	22/152	0.013	1.70e-08	9.68e-07	4/14	9.23e-04
Basigin interactions	10/25	0.002	1.96e-08	1.08e-06	10/10	6.59e-04
Neutrophil degranulation	41/478	0.04	5.39e-08	2.86e-06	10/10	6.59e-04

6.6. Discussion

This chapter examined the proteome of HPV-positive tonsil tissues by DEA of SWATH-MS data; and to our knowledge, this study is the first to describe protein expression changes of the tonsil proteome in its entirety in HPV-positive non-cancerous tonsil tissues.

These data demonstrate that the respiratory electron transport chain was most enriched by proteins found to be decreased in HPV-positive tissue; and so we speculated that it is likely to be downregulated in the presence of HPV. This pathway is fundamental for energy production in human cells and its dysregulation has been identified as being pivotal to disease progression (Raimondi et al., 2020). Interestingly, in HPV-associated disease, the virus has been linked to manipulation of mitochondrial function to promote carcinogenesis (Cruz-Gregorio et al., 2020); and although we recognise that the tissues in this study are not cancerous, it is curious that we found dysregulation of the respiratory electron transport chain in non-cancerous tissues with HPV-DNA and this warrants additional investigation. Validation of our protein hits on a larger scale would determine whether this is a true consequence of HPV infection, and enable better understanding of the significance of this downregulation to viral pathogenesis in early infections, which is currently unknown.

Curious hits within proteins identified with significant differential expression within HPV-positive tonsil tissue suggest that there is a statistically relevant increase in some immune system proteins in HPV-positive tissues (including Cathepsin S, C7, and Lysozyme C). This was unsurprising as a viral infection would cause an immune response (Mueller and Rouse, 2008), and this could explain the increases present in HPV-positive tissue. However, some of these immune hits have been described with alternate roles in disease; for example, Cathepsin S is described in carcinogenesis through promotion of angiogenesis, amongst other mechanisms (McDowell et al., 2020), so it may have additional roles here. Nonetheless, the roles of these immune-related proteins in HPV pathogenesis are unclear but are interesting hits for exploration through further validation and investigation.

There were significant decreases in the expression of some proteins involved in cell death and genomic stability (i.e., biorientation of chromosomes in cell division protein 1-like 1 (Higgs et al., 2015)), that may elude to genomic instability in the presence of HPV in these tissues. Interestingly, some proteins with decreased expression were linked to the mitochondria, which itself was highlighted within these hits through pathway enrichment analysis. For example, one

mechanism of pro-apoptotic protein, BID (Wang et al., 1996), is its promotion of mitochondrial apoptosis (Li et al., 1998). However, for the analysis of both the top 10 increased and decreased hits, it is appreciated that our sample size is small and that an individual hit and a small panel of biomarkers would need further validation and investigation on a wider scale. Overall, exploration of the top 10 decreased hits provides additional evidence of the potential importance of decreased mitochondrial function in HPV-positive tonsil tissues.

The next analysis of these SWATH-MS data revealed there was a plethora of proteins with significantly changed expression within HPV-positive males compared to HPV-positive females, and that these proteins largely have decreased expression in males compared to females. Interestingly, the major pathway of rRNA processing in the nucleolus and cytosol was the most enriched pathway within these protein hits, itself a process often perturbed in various diseases, including cancer (Jiao et al., 2023). We do acknowledge that these findings are limited by the small sample size examined and that validation in a larger sample size is necessary in future work. Taken together, these data suggest there is a fundamental difference in the proteomic expression profile in HPV-positive males but that additional research with a larger sample size would be required to determine the impact on viral pathogenesis.

6.6.1. Conclusion

In this chapter, we present data that shows the proteome of HPV-positive in comparison to HPV-negative non-cancerous tonsil tissue. We found that in nearly all these hits, their expression had significantly decreased in HPV-positive compared to HPV-negative tissue. The most enriched pathway within HPV-positive tonsil tissue was respiratory electron transport chain, and this was negatively impacted by decreased protein expression. There is evidence provided for a difference in the proteome of HPV-positive males in comparison to HPV-positive females, with rRNA processing described as the most impacted biological process.

CHAPTER 7: Discussion.

7.0. Introduction

This project firstly sought to provide an updated estimate of the global prevalence of oral HPV within the healthy adult population. It then aimed to inform of HPV biology within the mouth through exploration of the prevalence of HPV in the non-cancerous oropharynx and scrutiny of the pathogenesis of the virus in HPV-positive individuals. Finally, it intended to describe a blueprint of protein expression in HPV-positive tonsil tissue to be used to elucidate how HPV is influencing host cell biology.

In this final chapter, I discuss the main findings of this project, how this research aligns with current published knowledge, and I acknowledge any limitations of the work with suggestions that would account for these in future studies. Finally, I outline the recommended next steps for continuation of the work presented in this thesis.

7.1. Global burden of oral HPV in the healthy adult population

Reports of the prevalence of HPV in the mouth of the healthy adult population are immensely varied; several studies indicate the burden of HPV within this anatomical region to be minimal (<1%) (Kumari et al., 2021, Napolitano et al., 2024), yet others provide evidence of a strikingly high prevalence of oral HPV (more than 50% of those tested) (Ribeiro et al., 2019). Therefore, in Chapter 4, we aimed to take into consideration these variances in oral HPV prevalence and determine the burden of HPV within the mouth of healthy adults globally using a systematic review and meta-analysis of the literature. It is of note that whilst a requirement for inclusion within this meta-analysis was that there should be no oral or oropharyngeal malignancy, it was not always clear the state of individual participants' health otherwise.

We included 85 studies that report on HPV prevalence from 84,346 individuals and determined an overall prevalence of 5.044% (95% CI 4.897-5.193%). We found that the global oral HPV prevalence was comparable to other existing meta-analyses that present a global oral HPV prevalence of ~4-8% (Tam et al., 2018, Mena et al., 2019).

Tam *et al*., 2018 provide the most recent global oral HPV prevalence estimate within the literature of 7.7% (Tam et al., 2018), a slightly higher oral HPV prevalence than we report in

our meta-analysis. In this study, the inclusion criteria was similar to our investigation as they too focussed on an entirely adult population; however, they only included studies up to 2017 (Tam et al., 2018). Therefore, this could suggest a decrease in global oral HPV prevalence from 2017 to 2024, when compared to our study. We pondered whether this was due to the use of alternate methods to transform data that could introduce discrepancies into the prevalence estimate (Lin and Xu, 2020); however, our studies both used the same method of data transformation (Freeman-Tukey double arcsine). Alternatively, the prevalence decrease could be representative of a genuine decrease in global oral HPV infection in adults, perhaps due to accumulative effects of HPV vaccination over a greater time period since the introduction of the programme in several countries.

Our meta-analysis of overall oral HPV prevalence suggest that to date, oral HPV is reasonably prevalent worldwide within the adult population whom present with no oral malignancies.

7.2. Oropharyngeal HPV prevalence in adult non-cancerous palatine tonsil tissues

To accomplish the aims of this project, tonsil tissues were categorised based on their HPV-status, through our modified qPCR screen that used L1 consensus primers; GP5+/GP6+, to scrutinise tonsil gDNA for the presence of several viral types (de Roda Husman et al., 1995). In Chapter 5, we present a 10.2% prevalence rate of HPV-DNA in the non-cancerous palatine tonsils of adult routine tonsillectomy patients recruited to the HPV clinical study (N=12/118; comprising 7 females, 4 males, and 1 unknown individual). This was approximately double the prevalence determined by our meta-analysis of oral HPV within healthy adults (5.044%), implying that oral HPV rates are, in fact, not slowing down.

Indeed, previous studies of HPV prevalence within the non-cancerous tonsils have described reasonably low rates at ~5% (Combes et al., 2017, Mujtaba et al., 2018, Rieth et al., 2018, Ahn et al., 2022); which is almost half the prevalence determined in this study. There are several explanations for the higher tonsillar HPV prevalence described within this work; for example, other individual studies of non-cancerous tonsil tissue are representative of persons/cohorts with distinct demographic factors to our sample that could influence the likelihood of HPV infection. For example, most existing studies have included under 18's in their samples, whereas this project consisted of an entirely adult cohort. Inclusion of juvenile individuals introduces a distinct participant demographic to the sample in regards their lifestyle characteristics that are risk factors for oral HPV (i.e., sexual behaviours/alcohol

consumption/possible smoking habits), and this may account for the higher prevalence observed here.

Interestingly, however, HPV has also been described as absent within non-cancerous palatine tonsil tissues on a few occasions. Palmer *et al*., 2014 provide evidence of this, and describe an HPV prevalence of 0% in a sample of non-cancerous tissues from tonsillectomy patients within the UK (Palmer et al., 2014). In this instance, both their inclusion of under 18's and that these tissues were collected 16-20 years ago mean this is likely not an accurate representation of modern-day tonsillar HPV prevalence in healthy adults. Still, even within entirely adult populations a 0% HPV prevalence in non-cancerous tonsil tissue is corroborated elsewhere (Ernster et al., 2009, Mordechai et al., 2019); but again, the findings of Ernster and colleagues at least are likely impacted by the use of older biospecimens that were collected pre-2002. Thus this low prevalence of HPV described in non-cancerous tonsil tissue, but high incidence of HPV-associated OPC (de Martel et al., 2017) (with over 90% HPV-16 positive (Gillison et al., 2015)), that are believed to arise within the tonsillar crypts (Kim et al., 2007), could be an indication that if HPV is able to establish an infection within the tonsils, there is a high chance it would progress to malignancy, and render the detection rate of HPV in tonsils of this study useful for investigation into viral pathogenesis in this region.

In our clinical study, as for our meta-analysis there is an uncertainty around participants' health, other than that they have no oral/oropharyngeal malignancy. Indeed, these tissues were confirmed to be non-malignant by a histopathologist at the hospital of collection; however, whilst no cancer was present the tissues are not entirely healthy and will have been removed for complaints such as chronic/acute tonsillitis or sleep apnoea (NHS, 2022). Due to the invasive nature of tonsillectomy surgery it would be unethical to remove these tissues without reason and so, this is unlikely to be addressed in future work of tonsillar HPV. Although, for the context of our tonsil data it is promising to see that other work has shown that oral HPV presence is not associated with tonsillitis compared to the tonsils of those without tonsillitis (p=0.111) (Ahn et al., 2022), and so, this may not have skewed our overall prevalence estimates. Nonetheless, the idea that tonsil tissue is not entirely healthy if it is being removed in an elective surgery still should be taken into consideration when examining these data; in particular, protein expression analysis at 7.7.

Another limitation to these data of HPV within the tonsils is that we were unable to analyse RNA isolated from 10% NBF fixed tissues (discussed in Chapter 2), and as a result whether HPV-positive individuals have an actively transcribing infection remains an outstanding question within these data.

7.3. Experimental factors that influence oral HPV prevalence

There are a multitude of factors that contribute to differences in the reports of oral HPV prevalence; thus, I here discuss how the research methodology (i.e., the viral types analysed for, and sampling method used) may impact individual reports, and provide an insight into the oropharyngeal/oral HPV prevalence estimates described in this study both through meta-analysis and within the non-cancerous tonsil tissue cohort.

7.3.1. Sample type for oral HPV prevalence investigation

One poignant factor that influences oral HPV prevalence estimates of individual studies is the sample material analysed. Within the field, the dominant testing methods used are a gargle/rinse and/or a swab/brushing of an area of the mouth, likely because they are less invasive than surgical resection of human tissue and can be collected in the absence of other complaints. Indeed, the prevalence determined within the HPV clinical study were based upon the non-cancerous tonsil tissues alone, whereas the meta-analysis combines several sampling methods to calculate overall prevalence estimates. Generally, HPV prevalence reports of HPV in the non-cancerous palatine tonsils are less frequent, and reported lower than within other regions of the mouth.

The use of sampling methods such as an oral swab, -rinse, -gargle, or saliva are often coupled with reports of highly prevalent oral HPV within the broad region of ~22-76% (Colon-López et al., 2014, Lam et al., 2016, Shigeishi et al., 2016, Auguste et al., 2017, Shigeishi et al., 2018, Jamieson et al., 2020, Visalli et al., 2021). Although, as expected given the range of populations reported throughout the literature, in some instances, oral HPV prevalence estimates of 0-5% are also described (D'Souza et al., 2009, Malerova et al., 2020, Wood et al., 2020, Güran et al., 2021, Ramesh et al., 2021, Riddell et al., 2022, Zhu et al., 2023a), more akin to those reported within the tonsils (Mordechai et al., 2019, Ahn et al., 2022). This broad range of HPV detection highlights the importance of the impact of other factors than sampling method upon oral HPV prevalence, as discussed later.

The importance of sampling method to the likelihood of detecting oral HPV is reinforced by the apparent lack of correlation of HPV presence in different samples taken from one person (Quabius et al., 2021), with evidence that HPV is more likely to be detected in the oral cavity than tonsil tissues of an HPV-positive healthy individual (Grün et al., 2017). Combes *et al*., 2017 further corroborate this, and demonstrate a discordance between sample type, and found that, of individuals whom tested HPV-positive by oral gargle, only a small percentage showed corresponding HPV-positivity in their tonsils (Combes et al., 2017).

This prevalent nature of HPV-DNA in oral samples could be due to recent participation in established oral HPV risk factors such as oral sex (Pickard et al., 2012, D'Souza et al., 2016); transient HPV-DNA that never becomes an established infection is more likely to be present in biomaterial of the mouth than within cells of the tonsil tissues. Indeed, HPV-positivity via DNA analysis does not indicate whether the virus is actively replicating in these prevalence studies (Quabius et al., 2021); and this can be addressed by additional investigation into the presence of HPV transcripts (RNA) and viral proteins, or longitudinal studies to determine active or persistent HPV infection, respectively.

Nonetheless, despite HPV often being reported as more prevalent within the oral cavity, we still considered palatine tonsil tissues as the most appropriate material to base the HPV clinical study (that describes oropharyngeal HPV prevalence in non-cancerous tonsil tissue) for the following reasons:

1. Curiously, oral squamous cell carcinomas (OSCCs) are less frequently HPV-positive than HPV-OPC (Fonsêca et al., 2023).
2. The incidence of HPV-OPC is rapidly increasing at a global rate of ~4% per year (Tim Waterboer, personal communication); with these tumours assumed to arise within the basal cells of the tonsillar crypts (Kim et al., 2007).

We had speculated that HPV-DNA detection may be rarer in tonsil tissues than within an oral swab of the mouth, but that prevalence and pathogenesis data of potentially fewer HPV infections in this region would be more informative for disease. Previous work in our laboratory showed HPV to be highly prevalent (22.79%) in the oral cavity of a university population, using buccal swabs as the sampling method (Whitton, 2020), and so it was not anticipated to achieve this level of HPV detection in the tonsil study – partially due to the sample type but also based on a broader inclusion criteria of any persons over 18 years old,

meaning there was a greater chance of a more diverse sample of not university students alone, which we proposed in our previous work to be reflective of increased engagement of 'risky' behaviours.

However, even though we are confident of the decision to base our HPV prevalence investigations on the tonsil tissues as most informative of viral pathogenesis, we appreciate that the use of resected tissue presents the ability to investigate HPV-status at only one point in time. Longitudinal studies of HPV within the oropharynx are useful to monitor for HPV persistence and can enable the analysis of factors involved in viral persistence that promote malignant changes. This is not possible once the tissue has been removed and this remains a caveat of studying tonsil tissues.

Overall, the sampling method appears to influence HPV prevalence, with oral HPV generally higher than reports of HPV in the tonsils. With this in mind, we had speculated that HPV prevalence would be higher in our global HPV estimate by meta-analysis as this includes several studies with a range of sampling methods; yet we found HPV within non-cancerous tonsil tissues to be almost twice as prevalent. However, the specimen investigated is only one factor that contributes to oral HPV prevalence and other factors have an influence on the prevalence of HPV within the mouth.

7.3.2. Viral types in oral HPV prevalence estimates

HPV prevalence investigations are often limited to just HPV-16, -18, or all HR-HPV types (Turner et al., 2011, Pattanshetty et al., 2014, Lucas-Roxburgh et al., 2015, Shigeishi et al., 2016, Stankiewicz Karita et al., 2020), and this is likely to influence overall estimates. The rationale for limiting the number of viral types examined for within a sample is that HR-HPVs; in particular HPV-16, are most influential to HPV-associated disease (Schache et al., 2016), but this could result in an underestimation of HPV within a cohort. The overall meta-analysis estimate I report in this study clearly demonstrates this: if only HR-HPV types had been included then the oral HPV prevalence calculated would have been halved, which indicates the significance of the presence of LR-HPV types in the adult population globally. Moreover, some studies show detection of LR-HPVs but no HR-HPVs, despite analysing for both (Kofler et al., 2020, Güran et al., 2021), which if their analysis had been limited to only HR-HPVs a 0% prevalence would have been reported in these instances. The exclusion of viral types in some

studies included in the meta-analysis will have impacted individual prevalences and, therefore, the overall rate I report.

With this in mind, consensus primers (GP5+/GP6+; (de Roda Husman et al., 1995)) were selected to allow for scrutiny of a broad range of viral types within the sample of non-cancerous tonsil tissues. However, one study from Turkey in fact did not find any HPV in their sample of non-cancerous tonsil tissue, despite using the same primer pair as this tonsillar study (Bayram et al., 2023). Instead, we demonstrate the presence of three viral types within these tissues; HPV-16, HPV-18, and pHR HPV-73. Given the dominance of HPV-16 in HPV-OPC, it has been suggested that this viral type alone would be most informative of HPV-induced malignant transformation in the oropharynx (Fakhry et al., 2020). However, as a result of the poor knowledge of oral HPV pathogenesis, we postulated that investigations into a broad range of viral types would be most informative of the impacts of HPV on the host cell biology, and possible interactions between viral types where co-infections were present. We were surprised to observe such few viral types within our tissues.

7.3.2.2. HPV viral types within the non-cancerous oropharynx and oral cavity

HPV viral types were characterised using the INNO-LiPA® HPV Genotyping Extra II Kit [Fujirebio], that is validated for HPV genotyping in cervical, urine, and FFPE samples, but has also been successfully used for research purposes within the healthy oral cavity (Koleśnik et al., 2022, de Oliveira et al., 2023, Rumianek et al., 2023). HPV was typable in the majority of HPV-positive individuals, indicating that the kit is also sensitive to detection of HPV within tonsillar gDNA isolated from fresh-frozen tissue. Indeed, as discussed later at 7.6., we suspect that this kit might also have increased sensitivity to detect HPV-DNA in highly fragmented DNA of formalin-fixed tonsil tissues. Here, we showed that viral types HPV-18, HPV-16, and HPV-73 were present in 83.3%, 50%, and 25% of HPV-positive individuals, respectively. Thus, in contrast to our own meta-analysis of global HR-HPV prevalence that suggests HR-HPV constitutes approximately half (2.1%) of the overall HPV prevalence (5%), we detected only HR or pHR in this sample of non-cancerous tonsil tissue.

The finding that HPV-18 is more prevalent than HPV-16 in the non-cancerous oral cavity/oropharynx is substantiated in other studies (Antonsson et al., 2014, Ribeiro et al., 2019), which suggests that HPV-16 might not be as significant within the healthy mouth as it is in HPV-associated disease. It is possible that an HPV-18 infection is less likely to persist to

cancer in this region and therefore is not as prevalent as HPV-16 in disease; and that whilst HPV-16 is not always the most dominant viral type, when these infections are present they are more likely to promote malignant transformation. However, independent to HPV-16, HPV-18 has been associated with HNSCC risk (Michaud et al., 2014). Curiously, in this sample of tonsil tissues, HPV-16 was always present in a dual infection with HPV-18. Whilst further research would provide an insight into the currently unknown significance of this within the head and neck, it is interesting that in HPV-associated cervical disease, the presence of a dual infection of HPV-16 and HPV-18 have been associated with more advanced disease than HPV-16 alone (Spinillo et al., 2020).

Contrastingly, in Chapter 4, we determined by a sub-comparison within the meta-analyses that males were three times more likely to have an oral HPV-16 infection; which is supported in several other studies (Gillison et al., 2012, Riddell et al., 2022, Giuliano et al., 2023). This statistic may explain the greater burden of HPV-associated disease within males, as HPV-16 dominates HPV-OPC cases (Gillison et al., 2015, Schache et al., 2016, Lechner et al., 2022). These data align with cancer statistics presented by Zumsteg *et al*., 2023 that show that men are at least three times more likely to have OPC than women (Zumsteg et al., 2023), and although not all oral HPV infections will develop into cancer, it implies that the dominance of HPV infection in men is precedent to HPV-associated disease by gender. To our knowledge, our meta-analysis provides the most recent global update of oral HPV-16 prevalence in males and females and confirms that healthy men are more likely to have HR HPV-16 in the mouth.

pHR HPV-73 was the only other viral type identified within these non-cancerous tonsil tissues, and this viral type has also been highlighted as a prevalent viral type in studies of pre-cancerous and cancerous lesions of the cervix (Quddus et al., 2014, Iacobone et al., 2022). Despite reports specifically of this viral type within the head and neck being sparse, Bräutigam *et al* interestingly also found HPV-73 to be prevalent here, and strikingly, present at a higher rate than HPV-18 in the oral cavity of females (Bräutigam et al., 2022). In cervical cancer, one study provides evidence of HPV-73 in a co-infection with HR HPV-18 (Quddus et al., 2014); similarly, our data show that in all cases of HPV-73 infection in non-cancerous tonsil tissue, HPV-18 was also present. Together these data suggest that HPV-73 warrants further investigation.

7.4. HPV pathogenesis in the non-cancerous oropharynx

It is well-established that HPV does not uniformly infect cells of the cervix (Burd, 2003). Our data demonstrate that this is also the case in HPV-positive non-cancerous tonsil tissues, and we show that where multiple sections of tissue were taken for analysis the amount of HPV detected by qPCR varies across an HPV-positive tissue.

The lack of studies of HPV within the healthy tonsils, and the generally low prevalence described here means there is little knowledge of the concordance of HPV infection between the left and right tonsil tissues. We postulated that HPV is likely to be detectable in both tonsils in the case of an oropharyngeal HPV infection, possibly as a result of an active viral infection in one tissue that re-infects the contralateral tonsil or that HPV is capable of uniformly infecting the tonsils. Within the sample of HPV-positive individuals, we found that HPV was mainly concordant in the left and right palatine tonsil tissues (N=10/12, 83.3%). A study from Rusan *et al*., 2015 oppose these data and show evidence of HPV infection in just one tonsil tissue of an HPV-positive individual (Rusan et al., 2015); however, this was only representative of one individuals who further was the only HPV-positive person in their analysis. Furthermore, not many studies examine both the left and right palatine tonsil tissues for oral HPV infection. In HPV-associated tonsillar cancer, disease rarely presents in both tissues, although there are some cases described (Shimizu et al., 2017, Theodoraki et al., 2017). Therefore, our study is amongst the initial data examining this aspect of oral HPV pathogenesis from a prospective point of view. Continued research would confirm our finding that HPV is almost entirely concordant in both tissues, and begin to elucidate whether this is clinically important.

7.5. Demographics and lifestyle characteristics of individuals with HPV-positive tonsil tissue/s

Adults that consented to the HPV clinical study were invited to complete a questionnaire that posed questions about their participation in lifestyle activities (i.e., sexual activity, alcohol consumption, smoking status) and demographic factors (i.e., ethnicity). Unfortunately, due to the small HPV-positive number within the sample of non-cancerous tonsil tissues, and the questions surrounding the true HPV-status of 10% NBF fixed tissues, multivariate analysis to determine statistical association of having oropharyngeal HPV and demographic/lifestyle characteristics were deemed inappropriate, but descriptive associations were made.

A caveat of associations made between HPV-positivity and risk factors for infection is that they are reliant on self-reported behaviours, and this underlies all studies of this nature in both

healthy individuals and those with HPV-HNCs. Unfortunately, there is little that can be done to combat this, except encouragement to participants to provide an accurate report of themselves in their self-reported questionnaire. To make statistical associations between risk factors and HPV prevalence a larger sample would be most suitable to account for some probable inaccurate reports of one's lifestyle behaviours/characteristics.

7.5.1. Gender, lifestyle practices, and oral HPV

The final composition of our sample of adult routine tonsillectomy patients included 30 males, 78 females, and 10 unknown individuals, and the majority of participants were aged 18-35 and female (N=67/118, 56.8%). The dominance of this demographic subset occurred by chance, with younger women being the most common elective tonsillectomy patient that consented to the clinical study during our sample collection period but this correlates with statistics that show that females are more likely to have tonsillectomy surgery (The Royal College of Surgeons of England, 2005, Wilson et al., 2023). The use of random sampling meant that all routine tonsillectomy patients over 18 years old were eligible to consent to the study, and so an even distribution within all demographic parameters (i.e., age, gender) would be highly unlikely. However, given that our overall oropharyngeal HPV prevalence within tonsil tissue is higher than most reports that examine the same sample type (Ernster et al., 2009, Palmer et al., 2014, Mordechai et al., 2019), and that males are generally considered to have an higher risk of oral HPV (Gillison et al., 2012), the high proportion of females included in our sample seems to have had little impact on prevalence estimates. One would assume the prevalence to be lower than average within a sample that has few male individuals; yet we did not find this to be the case.

In Chapter 5, we show that HPV-DNA was detected in tonsil tissue of 7 females and 4 males individuals, but that proportionally of the total number of either gender, HPV was more prevalent in either the left or right palatine tonsil of male individuals (4/30, 13.3%) than it was in females (7/78, 9%) but that this is non-statistically significant (p=0.49). The lack of statistical significance determined in these data will be due to the small number of HPV-positive individuals and/or that we did not have an even gender distribution within our sample with more females included. Nonetheless, the finding within our data that oropharyngeal HPV is more prevalent, albeit without significance, in males than females is contextually important as cases of HPV-OPC are a greater burden to men than women (Guo et al., 2022). The reasons for this burden remain unclear, and it is possible that males are simply more frequently infected

with tonsillar HPV, but that still raises the question, why are males more susceptible to oral HPV infection and disease? Given that we did not demonstrate a difference in prevalence that was of statistical value, it may be that prevalence rates within the oropharynx of healthy adults in the UK are similar, but that there are discrepancies in disease progression by gender.

The burden of oral HPV to men was also examined on a global scale within our meta-analysis, and it was determined that globally males were almost twice as likely to have an oral HPV infection than females (5.224% compared to 2.961%, respectively). This finding is corroborated in other meta-analyses (Shigeishi and Sugiyama, 2016, Tam et al., 2018), although not always significant (Mena et al., 2019). The findings of our meta-analysis extend those of the most recent study that includes data up to 2017, given that the association of oral HPV and males is shown to still be relevant up to the present year (2024). Overall, the meta-analysis data further support the extensive reports throughout the literature of the burden of oral HPV to men on a global scale, and emphasise this is in prevalence as well as in disease.

In contrast to the statistical association of male gender and oral HPV described on a global scale within our meta-analysis, it is not always the case that men are shown as more likely to have an oral HPV infection (Whitton et al., 2024). Some studies have shown that in their populations, oral HPV infection was proportionally higher in women than men (Gichki et al., 2012, Rajendra Santosh et al., 2019, Jamieson et al., 2020). However, these data are most likely dependent on the populations sampled; indeed, when sampling university populations, studies find similar rates of oral HPV in males and females (Sastre-Cantón et al., 2019, Morán-Torres et al., 2021, Whitton et al., 2024), probably due to greater participation in behaviours associated with oral HPV risk, which diminishes prevalence differences based on gender.

Nonetheless men are frequently described to have a higher rate of oral HPV infection than women within individual studies (Gillison et al., 2012, Giuliano et al., 2023, Yu et al., 2023). There are various explanations for this difference in HPV prevalence, but it is perhaps due to more frequent participation in activities that increase one's risk of acquiring an oral HPV infection; including risky sexual behaviours and/or a higher frequency of sexual partners (D'Souza et al., 2016, Osazuwa-Peters et al., 2019, Pauli et al., 2022). This was not the case for our HPV-positive sample of non-cancerous tonsil tissues, with similar reports of sexual behaviours reported by males and females. All but one female individual were sexually active and had indicated they had partaken in recent sexual intercourse (i.e., in the past year). In

addition, apart from 1 male individual, all HPV-positive persons indicated ≤10 total lifetime sexual partners, with 1-5 sexual partners the most common response; however, it is possible that this still increases risk of infection, as having a total of 3-5 lifetime sexual partners has shown to be linked to oral HPV in women (Bruno et al., 2023).

We appreciate that even though the questionnaire did ask for an indication of which sexual activities individuals had/currently participated in, it did not distinguish whether the number of total sexual partners were for oral, vaginal, or anal sex, and so, this is unknown for the HPV-positive participants in this study. However, regardless of this, both vaginal and oral sex have been associated with oral HPV infection (D'Souza et al., 2009, Cook et al., 2014, Shigeishi and Sugiyama, 2016, Bruno et al., 2023), and it is possible that the type of sex (including anal sex) is not important to overall oral HPV risk, as it has been shown there was no significant difference in participation in these activities and the presence of oral HPV (Pauli et al., 2022). However, given the frequency of reports associating oral sex with oral HPV infection, it may be useful to incorporate this distinction between number of partners to participation in each sexual activity within the questionnaire for further recruitment to the HPV clinical study.

Overall, the questionnaire data show no inconspicuous patterns of participation in lifestyle activities within our HPV-positive sample i.e., unusually high or low participation in other risk factors such as smoking (Gillison et al., 2012, Knight et al., 2016, Sonawane et al., 2017, Whitton et al., 2024) and/or alcohol consumption (Gillison et al., 2012). This suggests that it is unlikely that the frequency of participation in these lifestyle characteristics that has impacted the non-significant greater prevalence of oral HPV within men, given that there are no pronounced differences by these based on gender. This could be for several reasons but given the small sample of HPV-positive individuals, it is more likely that any trends in association between HPV-positivity and lifestyle characteristics would emerge with a larger sample size.

Of course, participation in lifestyle activities is only one explanation for the commonly perceived higher oral HPV prevalence in men. There are likely also fundamental differences in the pathogenesis of HPV within the mouth influenced by gender, or that HPV is able to influence host cell biology differently dependent on gender and this is explored within the proteome of HPV-positive individuals at 7.7. Moreover, it is possible that gender-specific hormones (i.e., androgens) are able to influence viral pathogenesis (Gillison et al., 2014).

7.5.2. HPV-positivity and vaccination status

Interestingly, in our tonsil study, we found that only 1/12 individuals reported having received the HPV vaccination and that 6/12 reported no history of HPV vaccination. However, 4/12 (33.3% of HPV-positive individuals) stated that they were unsure of their vaccination status. As suggested previously, the number of HPV-positive individuals determined in this work is also too small to decipher the significance of vaccination to HPV-status; indeed, it loosely shows that within the HPV-positive individuals, HPV vaccination is uncommon. However, as we are unsure of the true HPV-status of several of the 'HPV-negative' formalin-fixed DNA samples, statistical comparisons would be difficult to make from these data. Nonetheless, the proportion of HPV-positive individuals that indicated they were unsure whether they had received the vaccination highlights that educational interventions would be necessary to improve knowledge of the risks of HPV infection (particularly within the head and neck), that may encourage everyone eligible to receive the full HPV vaccination course. Evidence suggests that uptake of HPV vaccination is correlated with decreased oral HPV prevalence (Hirth et al., 2017, Mehanna et al., 2019a) and HPV-related disease (Falcaro et al., 2021, Adekanmbi et al., 2024), highlighting the need for strong vaccination coverage to protect against HPV-OPC.

7.6. HPV-negative or undetectable HPV in gDNA isolated from 10% NBF fixed tonsil tissues

Formalin fixation of tonsil tissues had an irrefutable impact on the ability to isolate nucleic acid of sufficient quality and quantity for molecular analyses (discussed in detail in Chapter 2). When investigating the prevalence of HPV within tonsillar gDNA, it became apparent that the ability to detect HPV-DNA in these samples had most likely been affected by the storage of tonsil tissues in 10% NBF. Analysis of melt curve plots showed that in some instances, gDNA samples from fixed tissues resulted in inconsistent results (i.e., products in the melt curve analysis that were not always present), and meant that the HPV-status determined in our consensus qPCR screen for formalin-fixed tissues was not reliable. This was unsurprising as housekeeping gene beta-actin had been difficult to detect, and given that these tissues were non-cancerous, if HPV-DNA was to be present it could be at a low copy number. In our overall prevalence estimate (10.2%), these formalin-fixed tissues were considered HPV-negative.

We propose a few reasons for these inconsistent data found in gDNA of 10% NBF fixed tissues:

1. The HPV-negative status assigned to gDNA samples from 10% NBF fixed tonsil tissues is a genuine occurrence.
2. Our HPV-DNA qPCR consensus screen was not sensitive enough to reliably detect HPV within 10% NBF fixed palatine tonsil tissues.
3. gDNA was too fragmented for detection of HPV infection, that would likely be low copy number if present.

The first possible explanation that all tonsil tissues fixed in 10% NBF were an entirely HPV-negative sample is one that should be considered. As mentioned, the detection of no HPV-DNA in non-cancerous palatine tonsil tissues has occurred in other studies (Ernster et al., 2009, Palmer et al., 2014, Mordechai et al., 2019), and so, we appreciate that it is not a certainty that any HPV-DNA would be detected in these 10% NBF fixed tissues. As a result, under normal circumstances whereby the melt curve plots were consistently and convincingly HPV-negative, as was the case with fresh-frozen HPV-negative samples, this could have been deemed a plausible result. However, we speculated that this was unlikely in this study for two reasons; firstly, that we had found oropharyngeal HPV to be highly prevalent in fresh-frozen tonsil specimens (N=12/48 individuals, 25%); and secondly, this was coupled with several gDNA samples of 10% NBF fixed tissues presenting with inconsistent melt curve plots.

Secondly, we then considered whether our modified consensus qPCR screen was sensitive enough to detect HPV in these tricky formalin-fixed samples. FFPE specimens have been used to investigate HPV in archival samples, and it is possible to detect HPV within these using PCR based methods (Lillsunde Larsson et al., 2015). However, Ernster and colleagues' study that had described a 0% prevalence in tonsils used FFPE specimens (Ernster et al., 2009) and it is possible that the formalin-fixed nature of these samples limited the ability to detect any virus. Indeed, it is well-established that FFPE specimens present a more difficult sample type to achieve data in any molecular analyses (Steiert et al., 2023), particularly if subjected to prolonged fixation times (Kocjan et al., 2016), and it has been shown that PCR primer pairs designed to bind shorter sequences of target genes are more appropriate for successful amplification (Lenze et al., 2012) due to potential DNA fragmentation effects of formalin fixation (Ferrer et al., 2007).

Indeed, we were aware of this from the outset and the primers that had been selected during HPV-DNA qPCR optimisation using fresh-frozen tonsillar gDNA; GP5+/GP6+ (Appendix D),

were still thought to be the best option for fragmented formalin-fixed tonsillar gDNA. These primers target a 150 bp region of the HPV L1 gene (de Roda Husman et al., 1995); much smaller than other commonly used primer pairs MY09/MY11 (450 bp) (Husnjak et al., 2000) and FAP59/FAP64 (480 bp) (Forslund et al., 1999), and even the original, unmodified, GP5/GP6 primers have been described as having sensitivity for HPV-DNA detection in FFPE specimens (Baay et al., 1996). Others have found that nested PCR offers a better HPV-DNA detection rate in FFPE specimens (Tawe et al., 2018), but as with the aforementioned study, this was found within a cancerous setting (OSCC) that would correlate with a higher viral load that would enable easier HPV detection, regardless of fixation.

Additional experiments (various master mixes and PCR approaches) were performed to investigate the sensitivity of our HPV-DNA qPCR assay for formalin-fixed tonsillar gDNA (Appendix D). An interesting finding from this was that INNO-LiPA® [Fujirebio] genotyping found one tonsillar gDNA sample HPV-16 positive, that had otherwise presented with 1 peak aligning to HPV-18 plasmid in just 1 technical replicate on a PCR plate. This was curious but SPF10 primers, that detect a 65 bp fragment of the L1 gene, have been shown consistently to be a primer set sensitive to HPV in FFPE samples (Alberizzi et al., 2014, Kocjan et al., 2016). This could not be investigated further, nor could all samples be retested due to cost and time limitations; but additional work to analyse the entire sample of formalin-fixed gDNAs would reveal whether this is a more sensitive method.

All studies describing HPV detection in fixed tissues use paraffin-embedded formalin-fixed samples; however, our tonsil tissues represent an unusual sample type, as they were formalin-fixed but not paraffin embedded. As such, the final explanation that we propose to be most likely to cause an entire HPV-negative status of the formalin-fixed gDNA samples is that prolonged periods in 10% NBF caused significant irreversible DNA fragmentation (Ferrer et al., 2007). Tonsil tissues were fixed in 10% NBF to render potential SARS-CoV-2 inactive (Jureka et al., 2020, Kumar et al., 2021), and all specimens collected up to March 2023 were immersed in 10% NBF for several months, before being transferred to 70% EtOH. Then, even following a protocol change to reduce the fixation time (IRAS Amendment 5.0), tissues could still be in 10% NBF for 96 hours before a transfer to 70% EtOH due to processing restrictions at hospitals. Given that PCR amplification of longer DNA fragments is inversely correlated with time fixation times (Vitošević et al., 2018); we anticipate that these tonsil gDNA samples are unlikely to behave as FFPE specimens as tissues had been subjected to prolonged fixation

periods and therefore, even though buffered formalin was used that should preserve DNA for longer (Ferrer et al., 2007, Vitošević et al., 2023), this was likely to have detrimental effects on DNA integrity. Therefore, we predict the true prevalence of this tonsil sample would be higher if 10% NBF had not been used as a storage method; our finding that one 10% NBF fixed sample was HPV-positive by INNO-LiPA® genotyping supports this.

Overall, we strongly advise, based on these data, that in future work formalin-fixation is not used for preservation of tissues that are going to be used for downstream molecular application. However, it is of note that one hospital (CRH) withdrew from the HPV clinical study based on the protocol amendment that permitted collection of fresh-frozen tonsil tissues only; and so, we appreciate that more difficult storage of tissues as fresh-frozen may lead to a lower frequency of tonsil tissues collected for the study due to additional complexities, and freezer space restrictions.

7.7. Protein expression blueprint of HPV in the non-cancerous oropharynx

HPV-positive OPCs have better prognosis than HPV-negative OPCs (Ang et al., 2010, Maxwell et al., 2016), and there is increasing evidence that when investigated as a cancer subset themselves, HPV-OPCs represent a molecularly diverse disease. Despite this, there is limited knowledge of HPV within the non-cancerous oropharynx; in particular, how the virus may influence host cell biology to yield different genotypes (i.e., HPV-positivity with or without p16 expression) that present cancer types suitable for alternate treatments. We speculated that HPV-positive tonsil tissue would have a changed proteome based on the presence of HPV, but that there would be also be differences in protein expression in HPV-positive tonsil tissue based on gender.

7.7.1. The proteome of non-cancerous HPV-positive tonsil tissue

In Chapter 6, we show a panel of proteins with mostly significantly decreased expression in HPV-positive compared to HPV-negative non-cancerous tonsil tissue. Interestingly, the respiratory electron transport pathway of the mitochondria was most overrepresented within these hits, and our data suggested that this process was most likely to be negatively modulated. Dysregulation of this pathway is frequently linked to carcinogenesis irrespective of HPV (Raimondi et al., 2020). However, in the context of these data and HPV-positivity, a previous study has provided evidence of the ability of the HPV-18 E2 protein to bind a range of mitochondrial proteins *in vitro* (Lai et al., 2013). Curiously, our data show two of their HPV-

18 E2 interacting proteins (cytochrome c oxidase subunit 2 and cytochrome C1), to be significantly decreased within HPV-positive tonsil tissue (themselves considered HPV-18 positive by genotyping). Despite mitochondria being involved in apoptosis, our data show downregulation of pro-apoptotic protein, BID (in individual hits rather than whole pathway enrichment) in HPV-positive tonsil tissue, suggesting apoptosis is being inhibited.

In other studies, HPV has been associated with mitochondrial function by various mechanisms in the context of malignancy (Cruz-Gregorio et al., 2020). Whilst the mitochondria appear to be an organelle targeted by HPV in some manner, to our knowledge this study is the first to suggest decreased expression of some proteins within complexes of the respiratory electron transport chain within non-cancerous tonsil tissue. Taken together, these data provide a potential mechanism of how HPV may influence host cell biology in non-cancerous tonsil tissue, but further research in additional tissues would confirm this and elucidate the importance to viral pathogenesis in early infections, which is currently unexplored.

Upregulation of immune proteins or interactors is an expected response to a viral infection (Mueller and Rouse, 2008), and we provide evidence of this trend with immune related proteins in the top 10 hits with increased expression within HPV-positive tonsil tissue. Firstly, we found higher expression of cathepsin S; a protease known to contribute to the process of MHC class II antigen presentation (Riese et al., 1996, Beers et al., 2005). Secondly, other immune proteins including C7 of the MAC (Xie et al., 2020), and also lysozyme C (Ragland and Criss, 2017). We speculate that their increased expression was in response to HPV presence in the tonsil tissues. Indeed, HPV-positivity in OPC is associated with a greater immune cell presence in their tumour microenvironment in comparison to their HPV-negative counterparts (Yan et al., 2024), and this discrepancy impacts prognosis (Gameiro et al., 2018).

Interestingly, cathepsin S protein has been implicated in cancer, particularly through enhancement of cancer hallmarks such as angiogenesis (Wang et al., 2006, McDowell et al., 2020). However, the significance of cathepsin S within cancer is debated, and with respect to HPV-positivity, there is evidence of its downregulation in HPV-positive cervical cell lines (Vazquez-Ortiz et al., 2005). However, in HNC, expression of this protein has been linked to a lower tumour mutational burden, but it is of note that these authors did not separate HNC based on HPV-status for this analysis (Liang et al., 2024). Nonetheless, given its potential role in carcinogenesis and disease advancement, cathepsin S should be further investigated as a

curious candidate that may enable detection of individuals at risk of HPV-induced transformation, although, we acknowledge that a larger sample size would be necessary to confirm this association in non-cancerous tonsil tissue.

An increased lysozyme protein expression has previously been highlighted as a marker of a persistent HPV infection within the oral cavity of healthy individuals (Haukioja et al., 2014). Our data support this, and extend this finding from saliva samples as reported by Haukioja and colleagues to the non-cancerous oropharynx, where we too show an increased expression of lysozyme type-C. Unfortunately, it is not possible to determine whether this association is correlated with viral persistence in the non-cancerous tonsil tissue of routine tonsillectomy patients as these tissues have been resected; thus, validation in oral/tonsillar swabs could be informative of this association in future work. In addition, we provide evidence of increased expression of C7 in non-cancerous HPV-positive tissues which supports previous data that suggests this protein is elevated in HPV-OPC in comparison to HPV-negative malignancies (Dickinson et al., 2020); however, whether this protein serves a tumorigenic or suppressive role in cancer is currently debated (Ying et al., 2016, Zhang et al., 2021). Despite the aforementioned presence of these immune-related proteins in disease or persistent oral HPV infection, their roles in HPV pathogenesis within the healthy oropharynx remain unclear and would require further investigation.

Whilst we suggest that the elevated expression of these immune proteins is a result of HPV within the tissues, we appreciate here that these tissues were removed for other complaints, including chronic tonsillitis that could too result in an altered immune environment (Todorović and Zvrko, 2013); but we still show an immune profile difference in the presence of an inflammatory background.

To determine differentially expressed proteins it is desirable to use the adj. p-values calculated by DEA (Jafari and Ansari-Pour, 2019). In this comparison, the adj. p-values gave rise to only 2 significantly changed hits. Whilst these proteins could be useful for potential biomarker candidates, this N number was not sufficient for the aims of this project. One explanation for this limited number of significant proteins filtered by adj. p-value is that the differential abundance calculation was performed on a small sample (8 versus 6) due to insufficient material from 10% NBF fixed tissues. Another possibility is that there may be little difference in the proteome of an HPV-positive tonsil tissue compared to an HPV-negative tissue, if the

HPV-positive specimen had a transient or early-stage infection. Indeed, our data support this hypothesis as the two comparative sample groups (HPV-positive versus HPV-negative) did display some clustering; as indicated by PCA plot of imputed intensities, but there were also some anomalies. This would be expected as if a HPV infection was present there would likely be a varied viral load between samples, and positivity across a tissue. Nonetheless, the p-value was used to generate a list of hits suitable for pathway enrichment analysis, but at a threshold of $p<0.01$ to increase the stringency.

Taken together, these data of HPV-positive versus HPV-negative tonsil protein expression demonstrate that there is a difference between the proteome of non-cancerous tonsils in the presence of HPV-DNA. Validation of this panel of proteins (N=222) in a larger sample of non-cancerous tissues would demonstrate whether these impacts on host cell biology described here are consistent. Indeed, improved understanding of the pathogenesis of the virus within the non-cancerous tonsil tissues is required before an accurate test for individuals at risk of HPV-induced malignant changes could be developed.

7.7.2. HPV-positive male versus HPV-positive female tonsil protein

In Chapter 6, we also provide a protein signature of HPV-positivity in the non-cancerous palatine tonsils of males compared to HPV-positive females. Strikingly, most proteins were downregulated in men. Interestingly, in this comparison the adj. p-value was suitable to generate a list of significantly changed proteins, and we outline a panel with mostly decreased expression (N=394) in HPV-positive males. Given that this adj. p-value takes into consideration multiple comparisons, these hits are less likely to contain false positives than our previous comparison (Jafari and Ansari-Pour, 2019) (HPV-positive versus HPV-negative tonsil tissue).

Curiously, our data demonstrate that rRNA processing within the nucleolus and cytosol was the most significantly enriched pathway within our panel of proteins differentially expressed in HPV-positive male tonsil tissue. Therefore unsurprisingly, protein translation (including individual pathways; initiation, elongation, and termination), were also modulated within these tissues. rRNA processing and protein translation are fundamental to normal cell biology, and their dysregulation is unsurprisingly a mechanism used to promote disease progression (Jiao et al., 2023). One example of viral manipulation of this is by SARS-CoV-2 that has been shown to negatively impact the production of the ribosome to promote its own infection (Yerlici et

al., 2024). We suggest that this is likely to have an impact on the host cell proteome, but the dysregulation of translation could have countless functional outcomes that would require additional scrutiny to understand any specific pathways that may promote HPV pathogenesis in these tissues, and could explain the differences between the burden of HPV in the oropharynx of males.

In considering this finding, HPV transcripts make use of host cell machinery for their translation (Kajitani and Schwartz, 2020, García et al., 2024); therefore, it is somewhat surprising that our data suggest that ribosomal biogenesis and protein translation are significantly overrepresented by proteins with decreased expression within HPV-positive male tonsil tissue. Studies from one research group show that the process of Cap-dependent translation is upregulated *in vitro* by E6 proteins of HPV-16 and -18 viral types (Spangle and Münger, 2010, Spangle et al., 2012). In our study, the HPV-positive tissues used for this DEA were all considered either HPV-16 and/or HPV-18 positive and curiously, we find Cap-dependent translation initiation to also be enriched within these protein hits but contrastingly, with decreased expression. Based on previous reports, our data are surprising given that HPV-associated disease within the oropharynx is more common within men than women. It is possible that these differences are due to the *ex vivo* nature of our study that utilises human tissue rather than cell lines. Moreover, previous studies into the regulation of this pathway by HPV do not consider discrepancies in gender.

To confirm these observations, in future work, it would be useful to perform a DEA of the tonsil proteome of HPV-negative males versus females. Indeed, it is likely there would be a difference in the tonsillar proteome between biological genders, regardless of viral infection. Therefore, analysis of HPV-negative individuals by gender would highlight proteins whose expression had changed due to HPV infection. Unfortunately, there were not sufficient HPV-negative samples taken from each male and female fresh-frozen tonsil tissues to enable this statistical comparison within this study.

Overall, these data suggest that the protein blueprint of HPV-positivity in the non-cancerous tonsil tissues of male individuals is significantly different to females. The clear difference in the proteome of HPV-positive individuals based on biological gender suggests changes in host cell biology, and such changes may provide an explanation for why HPV is prevalent within, and more likely to induce malignant transformation in the oropharynx of males. To our

knowledge the downregulation of processes involved in ribosomal biogenesis and protein translation has not been directly linked to HPV infection in non-cancerous tonsil tissue previously, and this knowledge could provide an insight into reasons for the discrepancies in male oral HPV infection and associated disease.

7.7.3. Classical markers for both HPV infection and HPV-OPC

p16 has historically been used as a marker for HPV (Benevolo et al., 2006), but its usefulness as an accurate marker for HPV in cancer of the oropharynx (Mena et al., 2022, Sichero et al., 2024), as well as non-malignant tissue has been disputed in several studies (Klingenberg et al., 2010). We examined p16 expression by IHC within our sample of HPV-positive tonsil tissues and showed no correlation between p16 expression and HPV-status. Interestingly, this discordance of HPV and p16 expression is pertinent in disease, and the recent finding of Mehanna and colleagues that p16 expression varies in HPV-OPC, and that the p16 status of these cancers impacts the prognostic outcomes (Mehanna et al., 2023) is significant to treatment modality selected. This is further demonstrated in our data generated from SWATH-MS analysis, in which p16 protein was not featured within differentially expressed proteins of HPV-positive versus HPV-negative tonsil tissue, supported by our histology. Thus our data together support the notion that p16 expression is an inappropriate marker of HPV within the oropharynx, and as such should not be used in the stratification of HPV-status in non-cancerous tonsil tissues; however, we accept here we do not know how advanced possible HPV infections are within our HPV-positive tonsil tissues.

Several years ago, a ‘protein signature’ of cells within the cervical SCJ that may represent a distinct population of stem cells thought to be the origin of HPV-induced cervical carcinogenesis were described, with a panel of five protein biomarkers specific to these cells including: CK7, AGR2, MMP7, CD63, and GDA (Herfs et al., 2012). The accuracy of these claims have been disputed, with suggestion that the use of embryonic tissue in their investigations is not a good representation of the environment within developed adult cells of the cervix (Regauer and Reich, 2021). However, these markers have been investigated to determine their relevance within OPC dependent on HPV-status. They were found collectively inappropriate as a ‘signature’ of HPV-OPC (Morbini et al., 2015). We show in our data that these proteins are not differentially expressed within HPV-positive non-cancerous tonsil tissues. Moreover, by IHC, we demonstrate that expression of one of these markers, CK7, is not noticeably different between HPV-positive and HPV-negative non-cancerous tonsil tissue.

Morbini and colleagues suggest that CK7 expression is a marker for active HPV infection, not just viral presence (Morbini et al., 2015), and we again acknowledge that the status of infection is not known for our non-cancerous samples.

7.7.4. General limitations and future direction

A consistent caveat throughout these proteomic data described here is the limited sample size available for statistical comparisons. Unfortunately, protein extracted from 10% NBF fixed tonsil tissue did not present with significant protein IDs following SWATH-MS analysis for DEA. This was surprising given that an extensive optimisation process had been performed prior to the final protein work that indicated our extraction protocol was suitable for SWATH-MS; however, this is likely a result of the negative correlation between protein yield and fixation time (Wolff et al., 2011), with most of these tonsil tissues having been 'over-fixed'. To conclude here, these data show that our recurring suggestion that 10% NBF fixation of tonsil tissue is unsuitable for molecular analysis also applies to this protein work. Indeed, this highlighted the significant impact that prolonged periods of tissue fixation have on the quality of extracted macromolecules, and we acknowledge that a larger fresh-frozen sample size would provide optimal results for DEA.

This study aimed to determine protein expression changes in HPV-positive tonsil tissue, and then within HPV-positive males, that would provide an insight into how HPV influences the host cell and any differences within men that may indicate reasons for the discrepancy in HPV infection by gender. With this in mind, these data focussed on an extensive list of dysregulated protein hits suitable for pathway enrichment analysis instead of limiting the search to a small number of hits (for example, under 10) to enable analysis of broad functional impacts. In future work, we propose that this protein blueprint of HPV-positivity is characterised and validated within additional tonsil tissues, and we hope that this would provide a signature of HPV infection that is 'at-risk' of malignant change that could be monitored.

These data of protein expression are gathered from HPV-positive non-cancerous tonsil tissue; and this sample type presents numerous benefits for analysis of viral pathogenesis (i.e., being the assumed site of HPV-OPC development, and the advantage of a significant amount of material available for multiple analyses in comparison to tonsillar/oral swabs). However, if a protein signature of HPV infection within the oropharynx was to be determined in future work, a long-term aim would be to translate this into a biomarker test that would allow identification

of individuals at risk of HPV-induced malignant change. Therefore, these proteins identified within tonsil tissue would need validation within other sampling means. As an accessible sampling method, an oral swab would be most suitable for a routine biomarker test; whereas, to collect a tonsillar swab whilst an individual was conscious would be uncomfortable and likely to induce a gag reflex, thus people would be less likely to participate in a testing scheme.

7.8. Limitation of this study

The limitations that apply to specific points of these data in this study are discussed throughout where relevant; however, here I refer back to a major caveat that has significantly impacted the data presented in this thesis.

The fixation of tonsil tissues was a decision made to keep researchers safe whilst working with human tissues during the Covid-19 pandemic. It was a decision that was entirely out of my hands but one that had detrimental effects on the quality and quantity of DNA and protein isolated from 10% NBF fixed tissues, of which constituted most of this sample (discussed within Chapter 2). In the context of these data this has meant that the true HPV-status of 10% NBF fixed tissues is unclear, and we do anticipate that the overall HPV prevalence within non-cancerous tonsil tissues would be higher had they have been stored as fresh-frozen tissue. Moreover, RNA isolation and analysis was completely unfeasible from these samples and so, RNA was not included in these data set. Finally, there was insufficient material for proteomic analysis, of which we strongly believe to be due to 10% NBF fixation, that resulted in a significantly decreased sample size for SWATH-MS analysis and therefore the findings of protein expression presented within HPV-positive tonsil tissues would need validating within a large cohort before characterisation of any specific biomarkers.

Overall, from the data presented in this thesis, it is strongly suggested that 10% NBF fixation of tissues is not used for inactivation of potential pathogens (i.e., SARS-CoV-2), or for tissue preservation, if molecular analyses is required downstream.

7.9. Conclusion

In this thesis, the prevalence and pathogenesis of oral HPV within healthy individuals is scrutinised and described. Firstly, through a systematic review and meta-analysis of existing literature within the adult population globally, an up-to-date (to 2024) estimate of oral HPV prevalence is provided. We show that oral HPV is reasonably prevalent globally (5.044%); but

most interestingly, in the context of HPV-associated disease, additional sub-analyses demonstrate males are twice as likely to have an oral HPV infection than women. HPV-16 is the most prevalent viral type in HPV-associated malignancy, and this is shown to be three times more prevalent in males, supporting the greater burden of disease within men.

An NIHR-portfolio registered clinical study was established that has recruited a sample of adult participants undergoing routine tonsillectomy surgery from 5 hospitals across the UK. These data show that HPV is prevalent within the non-cancerous tonsil tissues of this sample (10.2%); and this rate is much higher than the prevalence usually reported within this sample type. In regards to viral pathogenesis, this study demonstrated that HPV was largely concordant in both the left and right tonsil tissue if HPV was present, which opposes the situation within disease where HPV-positive tonsillar cancers usually present in one tissue. Strikingly, despite HPV-16 being most prevalent within HPV-associated disease, HR HPV-18 was most prevalent in these tissues and we show that no LR-HPV types were detected.

Despite the strong suggestion from this project, that formalin-fixation is inappropriate for tissue preservation if molecular analyses is required, a protocol to isolate DNA of sufficient quantity and quality for qPCR analysis was developed. However, given the inconsistent results discussed in melt curve analysis of formalin-fixed specimens; and that RNA and protein extracted from these tissues were also hugely impacted, we still suggest that 10% NBF is not used for fixation of tissues in future work.

To our knowledge, this thesis outlines a ‘protein expression blueprint’ indicative of HPV-DNA within non-cancerous tonsil tissues in a manner that has not previously been explored. Through comparison of the proteome of HPV-positive versus HPV-negative tissues, DEA revealed a plethora of proteins with mainly decreased expression in HPV-positive tonsils. Within the proteins with increased expression, immune function is common, which conforms to the presence of a viral infection. A sub-analysis of protein expression within HPV-positive individuals by gender demonstrated that HPV-positive males showed largely decreased expression within statistically changed hits; but curiously, that the main functions impacted is of rRNA processing and translation, which may help to describe why males are more likely to have an oral HPV infection and disease. This contribution to the knowledge and understanding of the natural history of HPV in the oropharynx is fundamental to improving outcomes for patients in the clinic.

We propose that additional work to validate and characterise proteins differentially expressed in HPV-positive tonsil tissue would enable the description of a ‘signature’ of HPV infection in non-cancerous tonsil tissues that could be translated in future to the clinic to help to identify individuals with an oral HPV infection at risk of future malignant change, or that may even represent novel treatment targets.

References

ADEKANMBI, V., SOKALE, I., GUO, F., NGO, J., HOANG, T. N., HSU, C. D., OLUYOMI, A. & BERENSON, A. B. 2024. Human Papillomavirus Vaccination and Human Papillomavirus-Related Cancer Rates. *JAMA Netw Open,* 7**,** e2431807.

ADLER, D., WALLACE, M., BENNIE, T., ABAR, B., SADEGHI, R., MEIRING, T., WILLIAMSON, A. L. & BEKKER, L. G. 2015. High risk human papillomavirus persistence among HIV-infected young women in South Africa. *Int J Infect Dis,* 33**,** 219-21.

AGILENT. 2021. *High Sensitivity RNA ScreenTape Assay for TapeStation Systems - Quick Guide* [Online]. Available: https://www.agilent.com/cs/library/usermanuals/public/HS-RNA_QuickGuide.pdf [Accessed 31st May 2024].

AHN, D., HEO, S. J., LEE, G. J., SOHN, J. H. & JEONG, J. Y. 2022. Prevalence and characteristics of tonsillar human papillomavirus infection in tumor-free patients undergoing tonsillectomy. *Auris Nasus Larynx,* 49**,** 229-234.

ALBERIZZI, P., SPINILLO, A., GARDELLA, B., CESARI, S. & SILINI, E. M. 2014. Evaluation of the HPV typing INNO-LiPA EXTRA assay on formalin-fixed paraffin-embedded cervical biopsy samples. *Journal of Clinical Virology,* 61**,** 535-539.

ANAYANNIS, N. V., SCHLECHT, N. F., BEN-DAYAN, M., SMITH, R. V., BELBIN, T. J., OW, T. J., BLAKAJ, D. M., BURK, R. D., LEONARD, S. M., WOODMAN, C. B., PARISH, J. L. & PRYSTOWSKY, M. B. 2018. Association of an intact E2 gene with higher HPV viral load, higher viral oncogene expression, and improved clinical outcome in HPV16 positive head and neck squamous cell carcinoma. *PLoS One,* 13**,** e0191581.

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

ANGELETTI, P. C., KIM, K., FERNANDES, F. J. & LAMBERT, P. F. 2002. Stable replication of papillomavirus genomes in Saccharomyces cerevisiae. *J Virol,* 76**,** 3350-8.

ANTONSSON, A., CORNFORD, M., PERRY, S., DAVIS, M., DUNNE, M. P. & WHITEMAN, D. C. 2014. Prevalence and risk factors for oral HPV infection in young Australians. *PLoS One,* 9**,** e91761.

ANTONSSON, A., DE SOUZA, M., WOOD, Z. C., CARROLL, A., VAN, K., PATERSON, L., PANDEYA, N. & WHITEMAN, D. C. 2021. Natural history of oral HPV infection: Longitudinal analyses in prospective cohorts from Australia. *Int J Cancer,* 148**,** 1964-1972.

APPLEBY, P., BERAL, V., BERRINGTON DE GONZÁLEZ, A., COLIN, D., FRANCESCHI, S., GOODILL, A., GREEN, J., PETO, J., PLUMMER, M. & SWEETLAND, S. 2006. Carcinoma of the cervix and tobacco smoking: collaborative reanalysis of individual data on 13,541 women with carcinoma of the cervix and 23,017 women without carcinoma of the cervix from 23 epidemiological studies. *Int J Cancer,* 118**,** 1481-95.

ARAÚJO, M. V., PINHEIRO, H. H., PINHEIRO JDE, J., QUARESMA, J. A., FUZII, H. T. & MEDEIROS, R. C. 2014. [Prevalence of human papillomavirus (HPV) in Belém, Pará State, Brazil, in the oral cavity of individuals without clinically diagnosable injuries]. *Cad Saude Publica,* 30**,** 1115-9.

ARIAS-PULIDO, H., PEYTON, C. L., JOSTE, N. E., VARGAS, H. & WHEELER, C. M. 2006. Human papillomavirus type 16 integration in cervical carcinoma in situ and in invasive cervical cancer. *J Clin Microbiol,* 44**,** 1755-62.

ARSA, L., SIRIPOON, T., TRACHU, N., FOYHIRUN, S., PANGPUNYAKULCHAI, D., SANPAPANT, S., JINAWATH, N., PATTARANUTAPORN, P., JINAWATH, A. & NGAMPHAIBOON, N. 2021. Discrepancy in p16 expression in patients with HPV-associated head and neck squamous cell carcinoma in Thailand: clinical characteristics and survival outcomes. *BMC Cancer,* 21**,** 504.

AUGUSTE, A., GAËTE, S., HERRMANN-STORCK, C., MICHINEAU, L., JOACHIM, C., DELOUMEAUX, J., DUFLO, S. & LUCE, D. 2017. Prevalence of oral HPV infection among healthy individuals and head and neck cancer cases in the French West Indies. *Cancer Causes Control,* 28**,** 1333-1340.

AYDIN, I., WEBER, S., SNIJDER, B., SAMPERIO VENTAYOL, P., KÜHBACHER, A., BECKER, M., DAY, P. M., SCHILLER, J. T., KANN, M., PELKMANS, L., HELENIUS, A. & SCHELHAAS, M. 2014. Large scale RNAi reveals the requirement of nuclear envelope breakdown for nuclear import of human papillomaviruses. *PLoS Pathog,* 10**,** e1004162.

AZÉMAR, M.-D., COMPERAT, E., RICHARD, F., CUSSENOT, O. & ROUPRÊT, M. 2011. Bladder recurrence after surgery for upper urinary tract urothelial cell carcinoma: Frequency, risk factors, and surveillance. *Urologic Oncology: Seminars and Original Investigations,* 29**,** 130-136.

BAAY, M. F., QUINT, W. G., KOUDSTAAL, J., HOLLEMA, H., DUK, J. M., BURGER, M. P., STOLZ, E. & HERBRINK, P. 1996. Comprehensive study of several general and type-specific primer pairs for detection of human papillomavirus DNA by PCR in paraffin-embedded cervical carcinomas. *J Clin Microbiol,* 34**,** 745-7.

BAGAITKAR, J., DEMUTH, D. R. & SCOTT, D. A. 2008. Tobacco use increases susceptibility to bacterial infection. *Tob Induc Dis,* 4**,** 12.

BALERMPAS, P., RÖDEL, F., KRAUSE, M., LINGE, A., LOHAUS, F., BAUMANN, M., TINHOFER, I., BUDACH, V., SAK, A., STUSCHKE, M., GKIKA, E., GROSU, A. L., ABDOLLAHI, A., DEBUS, J., STANGL, S., GANSWINDT, U., BELKA, C., PIGORSCH, S., MULTHOFF, G., COMBS, S. E., WELZ, S., ZIPS, D., LIM, S. Y., RÖDEL, C. & FOKAS, E. 2017. The PD-1/PD-L1 axis and human papilloma virus in patients with head and neck cancer after adjuvant chemoradiotherapy: A multicentre study of the German Cancer Consortium Radiation Oncology Group (DKTK-ROG). *Int J Cancer,* 141**,** 594-603.

BANERJEE, N. S., WANG, H. K., BROKER, T. R. & CHOW, L. T. 2011. Human papillomavirus (HPV) E7 induces prolonged G2 following S phase reentry in differentiated human keratinocytes. *J Biol Chem,* 286**,** 15473-82.

BARENDREGT, J. J., DOI, S. A., LEE, Y. Y., NORMAN, R. E. & VOS, T. 2013. Meta-analysis of prevalence. *Journal of Epidemiology and Community Health,* 67**,** 974.

BAYRAM, G., SIMSEK, T., HAZIR, S., OZEN, F. Z., ERDOGAN, M. M., SEYHAN, S., KAZAZ, H. & TEZCAN, E. 2023. High-risk human papillomavirus infection prevalence in non-malignant tonsillar tissue: A single-center cross-sectional study: Human papillomavirus infection prevalence. *Journal of Surgery and Medicine,* 7**,** 669-672.

BEACHLER, D. C., ABRAHAM, A. G., SILVERBERG, M. J., JING, Y., FAKHRY, C., GILL, M. J., DUBROW, R., KITAHATA, M. M., KLEIN, M. B., BURCHELL, A. N., KORTHUIS, P. T., MOORE, R. D. & D'SOUZA, G. 2014. Incidence and risk factors of HPV-related and HPV-unrelated Head and Neck Squamous Cell Carcinoma in HIV-infected individuals. *Oral Oncol,* 50**,** 1169-76.

BEACHLER, D. C., WEBER, K. M., MARGOLICK, J. B., STRICKLER, H. D., CRANSTON, R. D., BURK, R. D., WILEY, D. J., MINKOFF, H., REDDY, S., STAMMER, E. E., GILLISON, M. L. & D'SOUZA, G. 2012. Risk factors for oral HPV

infection among a high prevalence population of HIV-positive and at-risk HIV-negative adults. *Cancer Epidemiology Biomarkers and Prevention,* 21**,** 122-133.

BECERRA, S. P., DASS, C. R., YABE, T. & CRAWFORD, S. E. 2012. Pigment epithelium-derived factor: chemistry, structure, biology, and applications. *J Biomed Biotechnol,* 2012**,** 830975.

BEERS, C., BURICH, A., KLEIJMEER, M. J., GRIFFITH, J. M., WONG, P. & RUDENSKY, A. Y. 2005. Cathepsin S controls MHC class II-mediated antigen presentation by epithelial cells in vivo. *J Immunol,* 174**,** 1205-12.

BENEVOLO, M., MOTTOLESE, M., MARANDINO, F., VOCATURO, G., SINDICO, R., PIPERNO, G., MARIANI, L., SPERDUTI, I., CANALINI, P., DONNORSO, R. P. & VOCATURO, A. 2006. Immunohistochemical expression of p16(INK4a) is predictive of HR-HPV infection in cervical low-grade lesions. *Mod Pathol,* 19**,** 384-91.

BERENSON, A. B., HIRTH, J. M. & CHANG, M. 2022. Prevalence of Oral Human Papillomavirus Infection: Impact of Sex, Race/Ethnicity, and Vaccination Status. *Clin Infect Dis,* 74**,** 1230-1236.

BERNARD, B. A., BAILLY, C., LENOIR, M. C., DARMON, M., THIERRY, F. & YANIV, M. 1989. The human papillomavirus type 18 (HPV18) E2 gene product is a repressor of the HPV18 regulatory region in human keratinocytes. *J Virol,* 63**,** 4317-24.

BERNARD, H. U., BURK, R. D., CHEN, Z., VAN DOORSLAER, K., ZUR HAUSEN, H. & DE VILLIERS, E. M. 2010. Classification of papillomaviruses (PVs) based on 189 PV types and proposal of taxonomic amendments. *Virology,* 401**,** 70-9.

BETTAMPADI, D., VILLA, L. L., PONCE, E. L., SALMERON, J., SIRAK, B. A., ABRAHAMSEN, M., RATHWELL, J. A., REICH, R. R. & GIULIANO, A. R. 2020. Oral human papillomavirus prevalence and type distribution by country (Brazil, Mexico and the United States) and age among HPV infection in men study participants. *Int J Cancer,* 146**,** 3026-3033.

BIENKOWSKA-HABA, M., PATEL, H. D. & SAPP, M. 2009. Target cell cyclophilins facilitate human papillomavirus type 16 infection. *PLoS Pathog,* 5**,** e1000524.

BODILY, J. & LAIMINS, L. A. 2011. Persistence of human papillomavirus infection: keys to malignant progression. *Trends Microbiol,* 19**,** 33-9.

BODON, G., CHASSEFEYRE, R., PERNET-GALLAY, K., MARTINELLI, N., EFFANTIN, G., HULSIK, D. L., BELLY, A., GOLDBERG, Y., CHATELLARD-CAUSSE, C., BLOT, B., SCHOEHN, G., WEISSENHORN, W. & SADOUL, R. 2011. Charged multivesicular body protein 2B (CHMP2B) of the endosomal sorting complex required for transport-III (ESCRT-III) polymerizes into helical structures deforming the plasma membrane. *J Biol Chem,* 286**,** 40276-86.

BOUSARGHIN, L., TOUZÉ, A., SIZARET, P. Y. & COURSAGET, P. 2003. Human papillomavirus types 16, 31, and 58 use different endocytosis pathways to enter cells. *J Virol,* 77**,** 3846-50.

BOYER, S. N., WAZER, D. E. & BAND, V. 1996. E7 protein of human papilloma virus-16 induces degradation of retinoblastoma protein through the ubiquitin-proteasome pathway. *Cancer research,* 56**,** 4620-4624.

BRAAKHUIS, B. J., SNIJDERS, P. J., KEUNE, W. J., MEIJER, C. J., RUIJTER-SCHIPPERS, H. J., LEEMANS, C. R. & BRAKENHOFF, R. H. 2004. Genetic patterns in head and neck cancers that contain or lack transcriptionally active human papillomavirus. *J Natl Cancer Inst,* 96**,** 998-1006.

BRADBURY, R., INAGAKI, K., SINGH, G., AGANA, U., PATTERSON, K., MALLOCH, L., RODRIGUEZ, E., QVARNSTROM, Y. & HOBBS, C. V. 2023. A Pilot Comparison of Fixatives for Hookworm Real-time Polymerase Chain Reaction. *Am J Trop Med Hyg,* 108**,** 335-339.

BRÄUTIGAM, K., MEIER, S., MENEDER, S., PROPPE, L., STROSCHEIN, K., POLACK, S., KÖSTER, F., RODY, A. & BAUM, S. 2022. Distribution of HPV Subtypes in Diverse Anogenital and Oral Samples from Women and Correlation of Infections with Neoplasia of the Cervix. *Cancers (Basel),* 14.

BRONNIMANN, M. P., CALTON, C. M., CHIQUETTE, S. F., LI, S., LU, M., CHAPMAN, J. A., BRATTON, K. N., SCHLEGEL, A. M. & CAMPOS, S. K. 2016. Furin Cleavage of L2 during Papillomavirus Infection: Minimal Dependence on Cyclophilins. *J Virol,* 90**,** 6224-6234.

BROUWER, A. F., CAMPREDON, L. P., WALLINE, H. M., MARINELLI, B. M., GOUDSMIT, C. M., THOMAS, T. B., DELINGER, R. L., LAU, Y. K., ANDRUS, E. C., YOST, M. L., MCCLOSKEY, J. K., SULLIVAN, T. S., MORTENSEN, A. S., HUANG, S., MURPHY, K., CHENG, B., STANEK, K., NAIR, T., CAREY, T. E., MEZA, R. & EISENBERG, M. C. 2022. Prevalence and determinants of oral and cervicogenital HPV infection: Baseline analysis of the Michigan HPV and Oropharyngeal Cancer (MHOC) cohort study. *PLoS One,* 17**,** e0268104.

BRUNO, M. T., BOEMI, S., CARUSO, G., SGALAMBRO, F., FERLITO, S., CAVALLARO, A., SUDANO, M. C. & PALUMBO, M. 2023. Oral HPV Infection in Women with HPV-Positive Cervix Is Closely Related to Oral Sex. *Diagnostics (Basel),* 13.

BUI, T. C., MARKHAM, C. M., ROSS, M. W. & MULLEN, P. D. 2013. Examining the association between oral health and oral HPV infection. *Cancer Prev Res (Phila),* 6**,** 917-24.

BUI, T. C., TRAN, L. T. H., MARKHAM, C. M., HUYNH, T. T. T., TRAN, L. T., PHAM, V. T. T., TRAN, Q. M., HOANG, N. H., HWANG, L. Y. & STURGIS, E. M. 2015. Self-reported oral health, oral hygiene, and oral HPV infection in at-risk women in Ho Chi Minh City, Vietnam. *Oral surgery, oral medicine, oral pathology and oral radiology,* 120**,** 34-42.

BUMRUNGTHAI, S., EKALAKSANANAN, T., DUANGCHAI, D., LANPOL, P., PANYA, P., KATTIWONG, F., ACHARYA, S. & PIENTONG, C. 2019. Prevalence of human papillomavirus in oral rinse samples from healthy individuals in northern Thailand. *Journal of Oral Pathology & Medicine,* 48**,** 159-165.

BURD, E. M. 2003. Human papillomavirus and cervical cancer. *Clin Microbiol Rev,* 16**,** 1-17.

BURLEY, M., ROBERTS, S. & PARISH, J. L. 2020. Epigenetic regulation of human papillomavirus transcription in the productive virus life cycle. *Semin Immunopathol,* 42**,** 159-171.

CALO, W. A., LENNON, R. P., RUFFIN IV, M. T., KELLER, C., SPANOS, K., D'SOUZA, G. & KRASCHNEWSKI, J. L. 2022. Support for HPV vaccine school-entry requirements in the United States: The role of exemption policies. *Vaccine,* 40**,** 7426-7432.

CAMERON, R. L., KAVANAGH, K., PAN, J., LOVE, J., CUSCHIERI, K., ROBERTSON, C., AHMED, S., PALMER, T. & POLLOCK, K. G. 2016. Human Papillomavirus Prevalence and Herd Immunity after Introduction of Vaccination Program, Scotland, 2009-2013. *Emerg Infect Dis,* 22**,** 56-64.

CANFELL, K., CHESSON, H., KULASINGAM, S. L., BERKHOF, J., DIAZ, M. & KIM, J. J. 2012. Modeling preventative strategies against human papillomavirus-related disease in developed countries. *Vaccine,* 30 Suppl 5**,** F157-67.

CAO, S., WYLIE, K. M., WYCZALKOWSKI, M. A., KARPOVA, A., LEY, J., SUN, S., MASHL, R. J., LIANG, W.-W., WANG, X., JOHNSON, K., DIPERSIO, J. F., GAY, H., RATNER, L., CHEN, F., ADKINS, D. R. & DING, L. 2019. Dynamic host immune response in virus-associated cancers. *Communications Biology,* 2**,** 109.

CAPALDI, R. A., MALATESTA, F. & DARLEY-USMAR, V. M. 1983. Structure of cytochrome c oxidase. *Biochimica et Biophysica Acta (BBA) - Reviews on Bioenergetics,* 726**,** 135-148.

CASTELLANOS, M. R. & PAN, Q. 2016. Novel p53 therapies for head and neck cancer. *World Journal of Otorhinolaryngology-Head and Neck Surgery,* 2**,** 68-75.

CASTELO-BRANCO, C. & SOVERAL, I. 2014. The immune system and aging: a review. *Gynecological Endocrinology,* 30**,** 16-22.

CASTLE, P. E., RODRÍGUEZ, A. C., BURK, R. D., HERRERO, R., WACHOLDER, S., ALFARO, M., MORALES, J., GUILLEN, D., SHERMAN, M. E., SOLOMON, D. & SCHIFFMAN, M. 2009. Short term persistence of human papillomavirus and risk of cervical precancer and cancer: population based cohort study. *Bmj,* 339**,** b2569.

CATALÁN-CASTORENA, O., GARIBAY-CERDENARES, O. L., ILLADES-AGUIAR, B., RODRÍGUEZ-RUIZ, H. A., ZUBILLAGA-GUERRERO, M. I., LEYVA-VÁZQUEZ, M. A., ENCARNACIÓN-GUEVARA, S. & ALARCÓN-ROMERO, L. D. C. 2024. The role of HR-HPV integration in the progression of premalignant lesions into different cancer types. *Heliyon,* 10**,** e34999.

CAVENAGHI, V. B., GHOSN, E. J., CRUZ, N., ROSSI, L. M., DA SILVA, L., COSTA, H. O. & VILLA, L. L. 2013. Determination of HPV prevalence in oral/oropharyngeal mucosa samples in a rural district of São Paulo. *Braz J Otorhinolaryngol,* 79**,** 599-602.

CENTERS FOR DISEASE AND CONTROL AND PREVENTION 2011. National and state vaccination coverage among adolescents aged 13 through 17 years--United States, 2010. *MMWR Morb Mortal Wkly Rep,* 60**,** 1117-23.

CENTERS FOR DISEASE AND CONTROL AND PREVENTION 2024. HIV Surveillance Special Report: HIV Infection Risk, Prevention, and Testing Behaviors Among Men Who Have Sex With Men—National HIV Behavioral Surveillance, 19 U.S. Cities, 2023.

CENTERS FOR DISEASE CONTROL AND PREVENTION. 2021. *Human Papillomavirus (HPV) Vaccination: What Everyone Should Know* [Online]. Available: https://www.cdc.gov/vaccines/vpd/hpv/public/index.html#:~:text=Since%20HPV%20vaccination%20was%20first,adults%20are%20getting%20genital%20warts [Accessed 21st August 2023].

CERASUOLO, A., ANNUNZIATA, C., TORTORA, M., STARITA, N., STELLATO, G., GREGGI, S., MAGLIONE, M. G., IONNA, F., LOSITO, S., BOTTI, G., BUONAGURO, L., BUONAGURO, F. M. & TORNESELLO, M. L. 2017. Comparative analysis of HPV16 gene expression profiles in cervical and in oropharyngeal squamous cell carcinoma. *Oncotarget,* 8**,** 34070-34081.

CERQUEIRA, C., SAMPERIO VENTAYOL, P., VOGELEY, C. & SCHELHAAS, M. 2015. Kallikrein-8 Proteolytically Processes Human Papillomaviruses in the Extracellular Space To Facilitate Entry into Host Cells. *J Virol,* 89**,** 7038-52.

CHAND, V., JOHN, R., JAISWAL, N., JOHAR, S. S. & NAG, A. 2014. High-risk HPV16E6 stimulates hADA3 degradation by enhancing its SUMOylation. *Carcinogenesis,* 35**,** 1830-9.

CHATURVEDI, A. K., ANDERSON, W. F., LORTET-TIEULENT, J., CURADO, M. P., FERLAY, J., FRANCESCHI, S., ROSENBERG, P. S., BRAY, F. & GILLISON, M. L. 2013. Worldwide trends in incidence rates for oral cavity and oropharyngeal cancers. *J Clin Oncol,* 31**,** 4550-9.

CHATURVEDI, A. K., GRAUBARD, B. I., BROUTIAN, T., PICKARD, R. K., TONG, Z. Y., XIAO, W., KAHLE, L. & GILLISON, M. L. 2015. NHANES 2009-2012 Findings: Association of Sexual Behaviors with Higher Prevalence of Oral Oncogenic Human Papillomavirus Infections in U.S. Men. *Cancer Res,* 75**,** 2468-77.

CHATURVEDI, A. K., GRAUBARD, B. I., BROUTIAN, T., PICKARD, R. K. L., TONG, Z. Y., XIAO, W., KAHLE, L. & GILLISON, M. L. 2018. Effect of Prophylactic Human Papillomavirus (HPV) Vaccination on Oral HPV Infections Among Young Adults in the United States. *J Clin Oncol,* 36**,** 262-267.

CHATURVEDI, A. K., GRAUBARD, B. I., BROUTIAN, T., XIAO, W., PICKARD, R. K. L., KAHLE, L. & GILLISON, M. L. 2019. Prevalence of Oral HPV Infection in Unvaccinated Men and Women in the United States, 2009-2016. *Jama,* 322**,** 977-979.

CHATURVEDI, A. K., GRAUBARD, B. I., PICKARD, R. K., XIAO, W. & GILLISON, M. L. 2014. High-risk oral human papillomavirus load in the US population, National Health and Nutrition Examination Survey 2009-2010. *J Infect Dis,* 210**,** 441-7.

CHEUNG, J. L., CHEUNG, T. H., YU, M. Y. & CHAN, P. K. 2013. Virological characteristics of cervical cancers carrying pure episomal form of HPV16 genome. *Gynecol Oncol,* 131**,** 374-9.

CHEUNG, J. L., LO, K. W., CHEUNG, T. H., TANG, J. W. & CHAN, P. K. 2006. Viral load, E2 gene disruption status, and lineage of human papillomavirus type 16 infection in cervical neoplasia. *J Infect Dis,* 194**,** 1706-12.

CHO, H., KISHIKAWA, T., TOKITA, Y., SUZUKI, M., TAKEMOTO, N., HANAMOTO, A., FUKUSUMI, T., YAMAMOTO, M., FUJII, M., OHNO, Y. & INOHARA, H. 2020. Prevalence of human papillomavirus in oral gargles and tonsillar washings. *Oral Oncol,* 105**,** 104669.

CHOI, J. S., TANG, L., YU, A. J., SWANSON, M. S., CHAMBERS, T. N., KOKOT, N. C. & SINHA, U. K. 2020. Prevalence of oral human papillomavirus infections by race in the United States: An association with sexual behavior. *Oral Dis,* 26**,** 930-940.

CHUNG, C. H. & GILLISON, M. L. 2009. Human papillomavirus in head and neck cancer: its role in pathogenesis and clinical implications. *Clin Cancer Res,* 15**,** 6758-62.

CHUNG, J. Y., BRAUNSCHWEIG, T., WILLIAMS, R., GUERRERO, N., HOFFMANN, K. M., KWON, M., SONG, Y. K., LIBUTTI, S. K. & HEWITT, S. M. 2008. Factors in tissue handling and processing that impact RNA obtained from formalin-fixed, paraffin-embedded tissue. *J Histochem Cytochem,* 56**,** 1033-42.

COLON-LÓPEZ, V., QUIÑONES-AVILA, V., DEL TORO-MEJÍAS, L. M., REYES, K., RIVERA, M. E., NIEVES, K., SÁNCHEZ-VAZQUEZ, M. M., MARTÍNEZ-FERRER, M. & ORTIZ, A. P. 2014. Oral HPV infection in a clinic-based sample of Hispanic men. *BMC Oral Health,* 14**,** 7.

COMBES, J. D., DALSTEIN, V., GHEIT, T., CLIFFORD, G. M., TOMMASINO, M., CLAVEL, C., LACAU ST GUILY, J. & FRANCESCHI, S. 2017. Prevalence of human papillomavirus in tonsil brushings and gargles in cancer-free patients: The SPLIT study. *Oral Oncol,* 66**,** 52-57.

CONWAY, D. I., ROBERTSON, C., GRAY, H., YOUNG, L., MCDAID, L. M., WINTER, A. J., CAMPBELL, C., PAN, J., KAVANAGH, K., KEAN, S., BHATIA, R., CUBIE, H., CLARKSON, J. E., BAGG, J., POLLOCK, K. G. & CUSCHIERI, K. 2016. Human Papilloma Virus (HPV) Oral Prevalence in Scotland (HOPSCOTCH): A Feasibility Study in Dental Settings. *PLoS One,* 11**,** e0165847.

COOK, R. L., THOMPSON, E. L., KELSO, N. E., FRIARY, J., HOSFORD, J., BARKLEY, P., DODD, V. J., ABRAHAMSEN, M., AJINKYA, S., OBESSO, P. D., RASHID, M. H. & GIULIANO, A. R. 2014. Sexual behaviors and other risk factors for oral human papillomavirus infections in young women. *Sex Transm Dis,* 41**,** 486-92.

CROSBIE, E. J., EINSTEIN, M. H., FRANCESCHI, S. & KITCHENER, H. C. 2013. Human papillomavirus and cervical cancer. *Lancet,* 382**,** 889-99.

CRUZ-GREGORIO, A., ARANDA-RIVERA, A. K. & PEDRAZA-CHAVERRI, J. 2020. Human Papillomavirus-related Cancers and Mitochondria. *Virus Research,* 286**,** 198016.

D'SOUZA, G., AGRAWAL, Y., HALPERN, J., BODISON, S. & GILLISON, M. L. 2009. Oral sexual behaviors associated with prevalent oral human papillomavirus infection. *J Infect Dis,* 199**,** 1263-9.

D'SOUZA, G., CLEMENS, G., STRICKLER, H. D., WILEY, D. J., TROY, T., STRUIJK, L., GILLISON, M. & FAKHRY, C. 2020. Long-term Persistence of Oral HPV Over 7 Years of Follow-up. *JNCI Cancer Spectr,* 4**,** pkaa047.

D'SOUZA, G., KREIMER, A. R., VISCIDI, R., PAWLITA, M., FAKHRY, C., KOCH, W. M., WESTRA, W. H. & GILLISON, M. L. 2007. Case-control study of human papillomavirus and oropharyngeal cancer. *N Engl J Med,* 356**,** 1944-56.

D'SOUZA, G., TEWARI, S. R., TROY, T., WATERBOER, T., STRUIJK, L., CASTILLO, R., WRIGHT, H., SHEN, M., MILES, B., JOHANSSON, M., ROBBINS, H. A. & FAKHRY, C. 2023. Prevalence of oral and blood oncogenic human papillomavirus biomarkers among an enriched screening population: Baseline results of the MOUTH study. *Cancer,* 129**,** 2373-2384.

D'SOUZA, G., WENTZ, A., KLUZ, N., ZHANG, Y., SUGAR, E., YOUNGFELLOW, R. M., GUO, Y., XIAO, W. & GILLISON, M. L. 2016. Sex Differences in Risk Factors and Natural History of Oral Human Papillomavirus Infection. *J Infect Dis,* 213**,** 1893-6.

DAHLSTROM, K. R., BURCHELL, A. N., RAMANAKUMAR, A. V., RODRIGUES, A., TELLIER, P. P., HANLEY, J., COUTLÉE, F. & FRANCO, E. L. 2014. Sexual transmission of oral human papillomavirus infection among men. *Cancer Epidemiol Biomarkers Prev,* 23**,** 2959-64.

DALLA TORRE, D., BURTSCHER, D., EDLINGER, M., SÖLDER, E., WIDSCHWENDTER, A., RASSE, M. & PUELACHER, W. 2015. Comparison of the prevalence of human papilloma virus infection in histopathologically confirmed premalignant oral lesions and healthy oral mucosa by brush smear detection. *Oral Surg Oral Med Oral Pathol Oral Radiol,* 119**,** 333-9.

DALLA TORRE, D., BURTSCHER, D., SÖLDER, E., RASSE, M. & PUELACHER, W. 2019. The correlation between the quality of oral hygiene and oral HPV infection in adults: a prospective cross-sectional study. *Clin Oral Investig,* 23**,** 179-185.

DAVIDSON, C. L., RICHTER, K. L., VAN DER LINDE, M., COETSEE, J. & BOY, S. C. 2014. Prevalence of oral and oropharyngeal human papillomavirus in a sample of South African men: a pilot study. *S Afr Med J,* 104**,** 358-61.

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

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

DAY, P. M., THOMPSON, C. D., SCHOWALTER, R. M., LOWY, D. R. & SCHILLER, J. T. 2013. Identification of a role for the trans-Golgi network in human papillomavirus 16 pseudovirus infection. *J Virol,* 87**,** 3862-70.

DE LA GARZA-RAMOS, M. A., URRUTIA-BACA, V. H., URBINA-RIOS, C. S., GARCÍA-ROBAYO, D. A., TAMEZ-GUERRA, P. & GOMEZ-FLORES, R. 2020. Prevalence of human papillomavirus in the oral cavity of an indigenous community from Southwest México. *Infect Genet Evol,* 83**,** 104283.

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

DE OLIVEIRA, B. R., DINIZ, E. S. B. V., DOS SANTOS, K. C., CAETANO, K. A. A., MOTA, G., SADDI, V. A., RABELO-SANTOS, S. H., VILLA, L. L., VADDIPARTI, K., COOK, R. L., TELES, S. A. & CARNEIRO, M. 2023. Human Papillomavirus Positivity at 3 Anatomical Sites Among Transgender Women in Central Brazil. *Sex Transm Dis,* 50**,** 567-574.

DE RODA HUSMAN, A. M., WALBOOMERS, J. M., VAN DEN BRULE, A. J., MEIJER, C. J. & SNIJDERS, P. J. 1995. The use of general primers GP5 and GP6 elongated at their 3' ends with adjacent highly conserved sequences improves human papillomavirus detection by PCR. *J Gen Virol,* 76 (Pt 4)**,** 1057-62.

DE SOUZA, M. M. A., HARTEL, G., OLSEN, C. M., WHITEMAN, D. C. & ANTONSSON, A. 2023. Oral human papillomavirus (HPV) infection and HPV vaccination in an Australian cohort. *Int J Cancer,* 153**,** 417-426.

DE SOUZA, M. M. A., HARTEL, G., WHITEMAN, D. C. & ANTONSSON, A. 2018. Detection of oral HPV infection – Comparison of two different specimen collection methods and two HPV detection methods. *Diagnostic Microbiology and Infectious Disease,* 90**,** 267-271.

DELLA FERA, A. N., WARBURTON, A., COURSEY, T. L., KHURANA, S. & MCBRIDE, A. A. 2021. Persistent Human Papillomavirus Infection. *Viruses,* 13.

DEMICHEV, V., MESSNER, C. B., VERNARDIS, S. I., LILLEY, K. S. & RALSER, M. 2020. DIA-NN: neural networks and interference correction enable deep proteome coverage in high throughput. *Nat Methods,* 17**,** 41-44.

DEPARTMENT OF HEALTH. *Annual HPV vaccine coverage in England in 2009/2010* [Online]. Available: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/215800/dh_123826.pdf [Accessed 20th April 2022].

DICKINSON, A., SARASWAT, M., SYRJÄNEN, S., TOHMOLA, T., SILÉN, R., RANDÉN-BRADY, R., CARPÉN, T., HAGSTRÖM, J., HAGLUND, C., MATTILA, P., MÄKITIE, A., JOENVÄÄRÄ, S. & SILÉN, S. 2020. Comparing serum protein levels can aid in differentiating HPV-negative and -positive oropharyngeal squamous cell carcinoma patients. *PLoS One,* 15**,** e0233974.

DIDUSCH, S., MADERN, M., HARTL, M. & BACCARINI, M. 2022. amica: an interactive and user-friendly web-platform for the analysis of proteomics data. *BMC Genomics,* 23**,** 817.

DIGIUSEPPE, S., LUSZCZEK, W., KEIFFER, T. R., BIENKOWSKA-HABA, M., GUION, L. G. & SAPP, M. J. 2016. Incoming human papillomavirus type 16 genome resides in a vesicular compartment throughout mitosis. *Proceedings of the National Academy of Sciences,* 113**,** 6289-6294.

DODD, R. H., FREEMAN, M., DEKAJ, F., BAMFORTH, J., MIAH, A., SASIENI, P. & LOUIE, K. S. 2021. Awareness of the link between human papillomavirus and oral cancer in UK university students. *Prev Med,* 150**,** 106660.

DOORBAR, J. 2005. The papillomavirus life cycle. *J Clin Virol,* 32 Suppl 1**,** S7-15.

DOORBAR, J. 2013. The E4 protein; structure, function and patterns of expression. *Virology,* 445**,** 80-98.

DOORBAR, J., EGAWA, N., GRIFFIN, H., KRANJEC, C. & MURAKAMI, I. 2015. Human papillomavirus molecular biology and disease association. *Rev Med Virol,* 25 Suppl 1**,** 2-23.

DOORBAR, J., QUINT, W., BANKS, L., BRAVO, I. G., STOLER, M., BROKER, T. R. & STANLEY, M. A. 2012. The biology and life-cycle of human papillomaviruses. *Vaccine,* 30 Suppl 5**,** F55-70.

DRAKE, V. E., FAKHRY, C., WINDON, M. J., STEWART, C. M., AKST, L., HILLEL, A., CHIEN, W., HA, P., MILES, B., GOURIN, C. G., MANDAL, R., MYDLARZ, W. K., ROOPER, L., TROY, T., YAVVARI, S., WATERBOER, T., BRENNER, N., EISELE, D. W. & D'SOUZA, G. 2021. Timing, number, and type of sexual partners associated with risk of oropharyngeal cancer. *Cancer,* 127**,** 1029-1038.

EDELSTEIN, Z. R., SCHWARTZ, S. M., HAWES, S., HUGHES, J. P., FENG, Q., STERN, M. E., O'REILLY, S., LEE, S. K., FU XI, L. & KOUTSKY, L. A. 2012. Rates and determinants of oral human papillomavirus infection in young men. *Sex Transm Dis,* 39**,** 860-7.

EDWARDS, T. G., HELMUS, M. J., KOELLER, K., BASHKIN, J. K. & FISHER, C. 2013. Human papillomavirus episome stability is reduced by aphidicolin and controlled by DNA damage response pathways. *J Virol,* 87**,** 3979-89.

EGAWA, N. & DOORBAR, J. 2017. The low-risk papillomaviruses. *Virus Res,* 231**,** 119-127.

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

EGAWA, N., WANG, Q., GRIFFIN, H. M., MURAKAMI, I., JACKSON, D., MAHMOOD, R. & DOORBAR, J. 2017. HPV16 and 18 genome amplification show different E4-dependence, with 16E4 enhancing E1 nuclear accumulation and replicative efficiency via its cell cycle arrest and kinase activation functions. *PLoS Pathog,* 13**,** e1006282.

EHNS. 2024. *Make Sense Campaign, 16-21 September 2024 (European Head & Neck Society)* [Online]. Available: https://makesensecampaign.eu [Accessed 18th November 2024].

EICHTEN, A., WESTFALL, M., PIETENPOL, J. A. & MÜNGER, K. 2002. Stabilization and functional impairment of the tumor suppressor p53 by the human papillomavirus type 16 E7 oncoprotein. *Virology,* 295**,** 74-85.

EKANAYAKE WEERAMANGE, C., TANG, K. D., BARRERO, R. A., HARTEL, G., LIU, Z., LADWA, R., LANGTON-LOCKTON, J., FRAZER, I., KENNY, L., VASANI, S. & PUNYADEERA, C. 2023. Salivary micro RNAs as biomarkers for oropharyngeal cancer. *Cancer Med,* 12**,** 15128-15140.

ENERLY, E., FLINGTORP, R., CHRISTIANSEN, I. K., CAMPBELL, S., HANSEN, M., MYKLEBUST, T., WEIDERPASS, E. & NYGÅRD, M. 2019. An observational study comparing HPV prevalence and type distribution between HPV-vaccinated and -unvaccinated girls after introduction of school-based HPV vaccination in Norway. *PLoS One,* 14**,** e0223612.

ERNSTER, J. A., SCIOTTO, C. G., O'BRIEN, M. M., ROBINSON, L. J. & WILLSON, T. 2009. Prevalence of oncogenic human papillomavirus 16 and 18 in the palatine tonsils of the general adult population. *Arch Otolaryngol Head Neck Surg,* 135**,** 554-7.

ESQUENAZI, D., BUSSOLOTI FILHO, I., CARVALHO MDA, G. & BARROS, F. S. 2010. The frequency of human papillomavirus findings in normal oral mucosa of healthy people by PCR. *Braz J Otorhinolaryngol,* 76**,** 78-84.

EVANDE, R., RANA, A., BISWAS-FISS, E. E. & BISWAS, S. B. 2023. Protein-DNA Interactions Regulate Human Papillomavirus DNA Replication, Transcription, and Oncogenesis. *Int J Mol Sci,* 24.

EVANS, M. R., JAMES, C. D., LOUGHRAN, O., NULTON, T. J., WANG, X., BRISTOL, M. L., WINDLE, B. & MORGAN, I. M. 2017. An oral keratinocyte life cycle model identifies novel host genome regulation by human papillomavirus 16 relevant to HPV positive head and neck cancer. *Oncotarget,* 8**,** 81892-81909.

FAKHRY, C., FUNG, N., TEWARI, S. R. & D'SOUZA, G. 2020. Unique role of HPV16 in predicting oropharyngeal cancer risk more than other oncogenic oral HPV infections. *Oral Oncology,* 111**,** 104981.

FALCARO, M., CASTAÑON, A., NDLELA, B., CHECCHI, M., SOLDAN, K., LOPEZ-BERNAL, J., ELLISS-BROOKES, L. & SASIENI, P. 2021. The effects of the national HPV vaccination programme in England, UK, on cervical cancer and grade 3 cervical intraepithelial neoplasia incidence: a register-based observational study. *Lancet,* 398**,** 2084-2092.

FALCARO, M., SOLDAN, K., NDLELA, B. & SASIENI, P. 2024. Effect of the HPV vaccination programme on incidence of cervical cancer and grade 3 cervical intraepithelial neoplasia by socioeconomic deprivation in England: population based observational study. *Bmj,* 385**,** e077341.

FARSI, N. J., ROUSSEAU, M. C., SCHLECHT, N., CASTONGUAY, G., ALLISON, P., NGUYEN-TAN, P. F., SOULIÉRES, D., COUTLÉE, F., HIER, M., MADATHIL, S., FRANCO, E. L. & NICOLAU, B. 2017. Aetiological heterogeneity of head and neck squamous cell carcinomas: the role of human papillomavirus infections, smoking and alcohol. *Carcinogenesis,* 38**,** 1188-1195.

FAULKNER, G. C., KRAJEWSKI, A. S. & CRAWFORD, D. H. 2000. The ins and outs of EBV infection. *Trends in Microbiology,* 8**,** 185-189.

FELDSTEIN, O., GALI-ZAMIR, H., SCHEJTER, E., FEINBERG, T., YEHUDA-SHNAIDMAN, E., BORNSTEIN, J. & LEVY, T. 2023. High-risk HPV testing vs liquid-based cytology for cervical cancer screening among 25- to 30-year-old women: A historical cohort study. *Acta Obstet Gynecol Scand,* 102**,** 226-233.

FERGUSON, J., CAMPOS-LEÓN, K., PENTLAND, I., STOCKTON, J. D., GÜNTHER, T., BEGGS, A. D., GRUNDHOFF, A., ROBERTS, S., NOYVERT, B. & PARISH, J. L. 2021. The chromatin insulator CTCF regulates HPV18 transcript splicing and differentiation-dependent late gene expression. *PLoS Pathog,* 17**,** e1010032.

FERRER, I., ARMSTRONG, J., CAPELLARI, S., PARCHI, P., ARZBERGER, T., BELL, J., BUDKA, H., STRÖBEL, T., GIACCONE, G., ROSSI, G., BOGDANOVIC, N., FAKAI, P., SCHMITT, A., RIEDERERS, P., AL-SARRAJ, S., RAVID, R. & KRETZSCHMAR, H. 2007. Effects of formalin fixation, paraffin embedding, and time of storage on DNA preservation in brain tissue: a BrainNet Europe study. *Brain Pathol,* 17**,** 297-303.

FONSÊCA, T. C., JURAL, L. A., MARAÑÓN-VÁSQUEZ, G. A., MAGNO, M. B., ROZA, A., FERREIRA, D., MAIA, L. C., ROMAÑACH, M. J., AGOSTINI, M. & ABRAHÃO, A. C. 2023. Global prevalence of human papillomavirus-related oral and oropharyngeal squamous cell carcinomas: a systematic review and meta-analysis. *Clin Oral Investig,* 28**,** 62.

FORGAC, M. 2007. Vacuolar ATPases: rotary proton pumps in physiology and pathophysiology. *Nature Reviews Molecular Cell Biology,* 8**,** 917-929.

FORSLUND, O., ANTONSSON, A., NORDIN, P., STENQUIST, B. & GÖRAN HANSSON, B. 1999. A broad range of human papillomavirus types detected with a general PCR method suitable for analysis of cutaneous tumours and normal skin. *J Gen Virol,* 80 (Pt 9)**,** 2437-2443.

FORTES, H. R., VON RANKE, F. M., ESCUISSATO, D. L., ARAUJO NETO, C. A., ZANETTI, G., HOCHHEGGER, B., SOUZA, C. A. & MARCHIORI, E. 2017. Recurrent respiratory papillomatosis: A state-of-the-art review. *Respir Med,* 126**,** 116-121.

FRATTINI, M. G. & LAIMINS, L. A. 1994. Binding of the human papillomavirus E1 origin-recognition protein is regulated through complex formation with the E2 enhancer-binding protein. *Proc Natl Acad Sci U S A,* 91**,** 12398-402.

FRECKLETON, M.-A. 2023. Human Papillomaviruses as a Risk Factor for Oropharyngeal Squamous Cell Carcinoma. *In:* HARDING, J. J. (ed.) *Care of Head and Neck Cancer Patients for Dental Hygienists and Dental Therapists.* John Wiley & Sons.

FU, T. C., HUGHES, J. P., FENG, Q., HULBERT, A., HAWES, S. E., XI, L. F., SCHWARTZ, S. M., STERN, J. E., KOUTSKY, L. A. & WINER, R. L. 2015. Epidemiology of Human Papillomavirus Detected in the Oral Cavity and Fingernails of Mid-Adult Women. *Sex Transm Dis,* 42**,** 677-85.

FUESSEL HAWS, A. L., HE, Q., RADY, P. L., ZHANG, L., GRADY, J., HUGHES, T. K., STISSER, K., KONIG, R. & TYRING, S. K. 2004. Nested PCR with the PGMY09/11 and GP5(+)/6(+) primer sets improves detection of HPV DNA in cervical samples. *J Virol Methods,* 122**,** 87-93.

FUJIREBIO 2016. *INNO-LiPA HPV Genotyping Extra II Amp*.

GAMEIRO, S. F., GHASEMI, F., BARRETT, J. W., KOROPATNICK, J., NICHOLS, A. C., MYMRYK, J. S. & MALEKI VAREKI, S. 2018. Treatment-naïve HPV+ head and neck cancers display a T-cell-inflamed phenotype distinct from their HPV- counterparts that has implications for immunotherapy. *Oncoimmunology,* 7**,** e1498439.

GARCÍA, A., MALDONADO, G. & HERNÁNDEZ, G. 2024. Translational control of papillomavirus mRNAs in the spotlight. *Trends in Cell Biology,* 34**,** 703-706.

GERDES, F., TATSUTA, T. & LANGER, T. 2012. Mitochondrial AAA proteases — Towards a molecular understanding of membrane-bound proteolytic machines. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research,* 1823**,** 49-55.

GHEIT, T., MUWONGE, R., LUCAS, E., GALATI, L., ANANTHARAMAN, D., MCKAY-CHOPIN, S., MALVI, S. G., JAYANT, K., JOSHI, S., ESMY, P. O., PILLAI, M. R., BASU, P., SANKARANARAYANAN, R. & TOMMASINO, M. 2023. Impact of HPV vaccination on HPV-related oral infections. *Oral Oncol,* 136**,** 106244.

GICHKI, A. S., BUAJEEB, W., DOUNGUDOMDACHA, S. & KHOVIDHUNKIT, S. O. 2012. Detection of human papillomavirus in normal oral cavity in a group of Pakistani subjects using real-time PCR. *Asian Pac J Cancer Prev,* 13**,** 2299-304.

GILLISON, M. L., BROUTIAN, T., PICKARD, R. K., TONG, Z. Y., XIAO, W., KAHLE, L., GRAUBARD, B. I. & CHATURVEDI, A. K. 2012. Prevalence of oral HPV infection in the United States, 2009-2010. *Jama,* 307**,** 693-703.

GILLISON, M. L., CASTELLSAGUÉ, X., CHATURVEDI, A., GOODMAN, M. T., SNIJDERS, P., TOMMASINO, M., ARBYN, M. & FRANCESCHI, S. 2014. Eurogin Roadmap: comparative epidemiology of HPV infection and associated cancers of the head and neck and cervix. *Int J Cancer,* 134**,** 497-507.

GILLISON, M. L., CHATURVEDI, A. K., ANDERSON, W. F. & FAKHRY, C. 2015. Epidemiology of Human Papillomavirus-Positive Head and Neck Squamous Cell Carcinoma. *J Clin Oncol,* 33**,** 3235-42.

GILLISON, M. L., KOCH, W. M., CAPONE, R. B., SPAFFORD, M., WESTRA, W. H., WU, L., ZAHURAK, M. L., DANIEL, R. W., VIGLIONE, M., SYMER, D. E., SHAH, K. V. & SIDRANSKY, D. 2000. Evidence for a causal association between human papillomavirus and a subset of head and neck cancers. *J Natl Cancer Inst,* 92**,** 709-20.

GILLISON, M. L., TROTTI, A. M., HARRIS, J., EISBRUCH, A., HARARI, P. M., ADELSTEIN, D. J., JORDAN, R. C. K., ZHAO, W., STURGIS, E. M., BURTNESS, B., RIDGE, J. A., RINGASH, J., GALVIN, J., YAO, M., KOYFMAN, S. A., BLAKAJ, D. M., RAZAQ, M. A., COLEVAS, A. D., BEITLER, J. J., JONES, C. U., DUNLAP, N. E., SEAWARD, S. A., SPENCER, S., GALLOWAY, T. J., PHAN, J., DIGNAM, J.

J. & LE, Q. T. 2019. Radiotherapy plus cetuximab or cisplatin in human papillomavirus-positive oropharyngeal cancer (NRG Oncology RTOG 1016): a randomised, multicentre, non-inferiority trial. *Lancet,* 393**,** 40-50.

GIULIANO, A. R., FELSHER, M., WATERBOER, T., MIRGHANI, H., MEHANNA, H., ROBERTS, C., CHEN, Y. T., LYNAM, M., PEDRÓS, M., SANCHEZ, E., SIRAK, B., SURATI, S., ALEMANY, L., MORAIS, E. & PAVÓN, M. A. 2023. Oral Human Papillomavirus Prevalence and Genotyping Among a Healthy Adult Population in the US. *JAMA Otolaryngol Head Neck Surg,* 149**,** 783-795.

GONÇALVES, H. M., SILVA, J., PINTADO MAURY, I., TAVARES, A., CAMPOS, C., SOUSA, H., JACINTO, A., AGUIAR, P., CALDEIRA, L. & MEDEIROS, R. 2021. The prevalence and risk-factors of oral HPV DNA detection among HIV-infected men between men who have sex with men and heterosexual men. *Infect Dis (Lond),* 53**,** 19-30.

GONZÁLEZ-LOSA MDEL, R., MANZANO-CABRERA, L., RUEDA-GORDILLO, F., HERNÁNDEZ-SOLÍS, S. E. & PUERTO-SOLÍS, L. 2008. Low prevalence of high risk human papillomavirus in normal oral mucosa by hybrid capture 2. *Braz J Microbiol,* 39**,** 32-4.

GOODMAN, M. T., SARAIYA, M., THOMPSON, T. D., STEINAU, M., HERNANDEZ, B. Y., LYNCH, C. F., LYU, C. W., WILKINSON, E. J., TUCKER, T., COPELAND, G., PETERS, E. S., ALTEKRUSE, S. & UNGER, E. R. 2015. Human papillomavirus genotype and oropharynx cancer survival in the United States of America. *Eur J Cancer,* 51**,** 2759-67.

GOUGOUSIS, S., MOUCHTAROPOULOU, E., BESLI, I., VROCHIDIS, P., SKOUMPAS, I. & CONSTANTINIDIS, I. 2020. HPV-Related Oropharyngeal Cancer and Biomarkers Based on Epigenetics and Microbiome Profile. *Front Cell Dev Biol,* 8**,** 625330.

GOV UK. 2018. *HPV vaccination programme for men who have sex with men (MSM)* [Online]. Available: https://www.gov.uk/government/collections/hpv-vaccination-for-men-who-have-sex-with-men-msm-programme#:~:text=Contents&text=The%20national%20HPV%20vaccination%20programme,slide%20set%20for%20healthcare%20professionals. [Accessed 21st December 2024].

GOV UK. 2023. *HPV vaccination programme: changes from September 2023 letter* [Online]. Available: https://www.gov.uk/government/publications/hpv-vaccination-programme-changes-from-september-2023-letter/hpv-vaccination-programme-changes-from-september-2023-letter [Accessed 25th April 2025].

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

GRAHAM, S. V. 2017a. The human papillomavirus replication cycle, and its links to cancer progression: a comprehensive review. *Clin Sci (Lond),* 131**,** 2201-2221.

GRAHAM, S. V. 2017b. Keratinocyte Differentiation-Dependent Human Papillomavirus Gene Regulation. *Viruses,* 9.

GRÜN, N., MBUYA, W., TERNHAG, A., RAMQVIST, T., AHLBERG, A., JANGARD, M., DALIANIS, T. & HAMMARSTEDT-NORDENVALL, L. 2017. Human papillomavirus prevalence in mouthwashes of patients undergoing tonsillectomy shows dominance of HPV69, without the corresponding finding in the tonsils. *Infect Dis (Lond),* 49**,** 588-593.

GUDLEVICIENE, Z., KANOPIENE, D., STUMBRYTE, A., BAUSYTE, R., KIRVELAITIS, E., SIMANAVICIENE, V. & ZVIRBLIENE, A. 2015. Integration of human papillomavirus type 16 in cervical cancer cells. *Open Med (Wars),* 10**,** 1-7.

GUO, F., CHANG, M., SCHOLL, M., MCKINNON, B. & BERENSON, A. B. 2022. Trends in Oropharyngeal Cancer Incidence Among Adult Men and Women in the United States From 2001 to 2018. *Front Oncol,* 12**,** 926555.

GUPTA, A., PERKINS, R. B., ORTEGA, G., FELDMAN, S. & VILLA, A. 2019. Barrier use during oro-genital sex and oral Human Papillomavirus prevalence: Analysis of NHANES 2009–2014. *Oral Diseases,* 25**,** 609-616.

GÜRAN, M., NEOPHYTOU, P. I., ŞANLITÜRK, G., DEMIRDÖV, D. K., AL-SHARAKHI, M., HIKAL, Y., HAIDAR, M., KUNT, K. S., SELEVANY, H. & GÜRKAN, C. 2021. A Survey On prevalence, Risk Factors and Knowledge for Oral Human Papillomavirus Infections Among University Students. *Future Virology,* 16**,** 475-489.

HANG, D., LIU, F., LIU, M., HE, Z., SUN, M., LIU, Y., LI, J., PAN, Y., NING, T. & GUO, C. 2014a. Oral human papillomavirus infection and its risk factors among 5,410 healthy adults in China, 2009–2011. *Cancer Epidemiology and Prevention Biomarkers,* 23**,** 2101-2110.

HANG, D., LIU, F., LIU, M., HE, Z., SUN, M., LIU, Y., LI, J., PAN, Y., NING, T., GUO, C., LIANG, Y., XU, R., ZHANG, L., CAI, H. & KE, Y. 2014b. Oral human papillomavirus infection and its risk factors among 5,410 healthy adults in China, 2009-2011. *Cancer Epidemiol Biomarkers Prev,* 23**,** 2101-10.

HASHIBE, M., BRENNAN, P., CHUANG, S. C., BOCCIA, S., CASTELLSAGUE, X., CHEN, C., CURADO, M. P., DAL MASO, L., DAUDT, A. W., FABIANOVA, E., FERNANDEZ, L., WÜNSCH-FILHO, V., FRANCESCHI, S., HAYES, R. B., HERRERO, R., KELSEY, K., KOIFMAN, S., LA VECCHIA, C., LAZARUS, P., LEVI, F., LENCE, J. J., MATES, D., MATOS, E., MENEZES, A., MCCLEAN, M. D., MUSCAT, J., ELUF-NETO, J., OLSHAN, A. F., PURDUE, M., RUDNAI, P., SCHWARTZ, S. M., SMITH, E., STURGIS, E. M., SZESZENIA-DABROWSKA, N., TALAMINI, R., WEI, Q., WINN, D. M., SHANGINA, O., PILARSKA, A., ZHANG, Z. F., FERRO, G., BERTHILLER, J. & BOFFETTA, P. 2009. Interaction between tobacco and alcohol use and the risk of head and neck cancer: pooled analysis in the International Head and Neck Cancer Epidemiology Consortium. *Cancer Epidemiol Biomarkers Prev,* 18**,** 541-50.

HAUKIOJA, A., ASUNTA, M., SÖDERLING, E. & SYRJÄNEN, S. 2014. Persistent oral human papillomavirus infection is associated with smoking and elevated salivary immunoglobulin G concentration. *Journal of Clinical Virology,* 61**,** 101-106.

HEARNDEN, V., MURDOCH, C., D'APICE, K., DUTHIE, S., HAYWARD, N. J. & POWERS, H. J. 2018. Oral human papillomavirus infection in England and associated risk factors: a case-control study. *BMJ Open,* 8**,** e022497.

HERFS, M., YAMAMOTO, Y., LAURY, A., WANG, X., NUCCI, M. R., MCLAUGHLIN-DRUBIN, M. E., MÜNGER, K., FELDMAN, S., MCKEON, F. D., XIAN, W. & CRUM, C. P. 2012. A discrete population of squamocolumnar junction cells implicated in the pathogenesis of cervical cancer. *Proc Natl Acad Sci U S A,* 109**,** 10516-21.

HERRERO, R., QUINT, W., HILDESHEIM, A., GONZALEZ, P., STRUIJK, L., KATKI, H. A., PORRAS, C., SCHIFFMAN, M., RODRIGUEZ, A. C., SOLOMON, D., JIMENEZ, S., SCHILLER, J. T., LOWY, D. R., VAN DOORN, L. J., WACHOLDER, S. & KREIMER, A. R. 2013. Reduced prevalence of oral human papillomavirus (HPV) 4 years after bivalent HPV vaccination in a randomized clinical trial in Costa Rica. *PLoS One,* 8**,** e68329.

HERZUM, A., CICCARESE, G., DRAGO, F., PASTORINO, A., DEZZANA, M., MAVILIA, M. G., SOLA, S., COPELLO, F. & PARODI, A. 2022. Cervical, oral and anal Human papillomavirus infection in women attending the Dermatology Unit of a regional reference hospital in Genoa, Italy: a prevalence study. *J Prev Med Hyg,* 63**,** E415-e419.

HESS, K. L., CREPAZ, N., ROSE, C., PURCELL, D. & PAZ-BAILEY, G. 2017. Trends in Sexual Behavior Among Men Who have Sex with Men (MSM) in High-Income Countries, 1990-2013: A Systematic Review. *AIDS Behav,* 21**,** 2811-2834.

HIGGS, M. R., REYNOLDS, J. J., WINCZURA, A., BLACKFORD, A. N., BOREL, V., MILLER, E. S., ZLATANOU, A., NIEMINUSZCZY, J., RYAN, E. L., DAVIES, N. J., STANKOVIC, T., BOULTON, S. J., NIEDZWIEDZ, W. & STEWART, G. S. 2015. BOD1L Is Required to Suppress Deleterious Resection of Stressed Replication Forks. *Mol Cell,* 59**,** 462-77.

HIRTH, J. M., CHANG, M. & RESTO, V. A. 2017. Prevalence of oral human papillomavirus by vaccination status among young adults (18-30years old). *Vaccine,* 35**,** 3446-3451.

HONG, A., ZHANG, X., JONES, D., VEILLARD, A.-S., ZHANG, M., MARTIN, A., LYONS, J. G., LEE, C. S. & ROSE, B. 2016a. Relationships between p53 mutation, HPV status and outcome in oropharyngeal squamous cell carcinoma. *Radiotherapy and Oncology,* 118**,** 342-349.

HONG, A. M., FERGUSON, P., DODDS, T., JONES, D., LI, M., YANG, J. & SCOLYER, R. A. 2019. Significant association of PD-L1 expression with human papillomavirus positivity and its prognostic impact in oropharyngeal cancer. *Oral Oncol,* 92**,** 33-39.

HONG, A. M., VILAIN, R. E., ROMANES, S., YANG, J., SMITH, E., JONES, D., SCOLYER, R. A., LEE, C. S., ZHANG, M. & ROSE, B. 2016b. PD-L1 expression in tonsillar cancer is associated with human papillomavirus positivity and improved survival: implications for anti-PD1 clinical trials. *Oncotarget,* 7**,** 77010-77020.

HONG, D., LIU, J., HU, Y., LU, X., LI, B., LI, Y., HU, D., LU, W., XIE, X. & CHENG, X. 2017. Viral E6 is overexpressed via high viral load in invasive cervical cancer with episomal HPV16. *BMC Cancer,* 17**,** 136.

HORMONE HEALTH. 2024. *Cervical Cancer Prevention Week 2024* [Online]. Available: https://hormonehealth.co.uk/cervical-cancer-prevention-week-2024 [Accessed 11th November 2024].

HOY, D., BROOKS, P., WOOLF, A., BLYTH, F., MARCH, L., BAIN, C., BAKER, P., SMITH, E. & BUCHBINDER, R. 2012. Assessing risk of bias in prevalence studies: modification of an existing tool and evidence of interrater agreement. *J Clin Epidemiol,* 65**,** 934-9.

HUI, A. B., LIN, A., XU, W., WALDRON, L., PEREZ-ORDONEZ, B., WEINREB, I., SHI, W., BRUCE, J., HUANG, S. H., O'SULLIVAN, B., WALDRON, J., GULLANE, P., IRISH, J. C., CHAN, K. & LIU, F. F. 2013. Potentially prognostic miRNAs in HPV-associated oropharyngeal carcinoma. *Clin Cancer Res,* 19**,** 2154-62.

HULL, R., MBELE, M., MAKHAFOLA, T., HICKS, C., WANG, S. M., REIS, R. M., MEHROTRA, R., MKHIZE-KWITSHANA, Z., KIBIKI, G., BATES, D. O. & DLAMINI, Z. 2020. Cervical cancer in low and middle-income countries. *Oncol Lett,* 20**,** 2058-2074.

HUSNJAK, K., GRCE, M., MAGDIĆ, L. & PAVELIĆ, K. 2000. Comparison of five different polymerase chain reaction methods for detection of human papillomavirus in cervical cell specimens. *J Virol Methods,* 88**,** 125-34.

IACOBONE, A. D., BOTTARI, F., GUERRIERI, M. E., VIDAL URBINATI, A. M., GHIONI, M., SPOLTI, N., PINO, I., PASSERINI, R., DI PACE, R. C., FRANCHI, D. & PRETI, E. P. 2022. The Potential Impact of High-Risk Human Papillomavirus-Negative Cervical Intraepithelial Neoplasia 2+ on Primary Human Papillomavirus Screening. *Am J Clin Pathol,* 157**,** 130-135.

IFHNOS. 2016. *World Head and Neck Cancer Day - International Federation of Head and Neck Oncologic Societies* [Online]. Available: https://ifhnos.net/whncd/index/ [Accessed 18th November 2024].

INTROCASO, C. E., DUNNE, E. F., HARIRI, S., PANICKER, G., UNGER, E. R. & MARKOWITZ, L. E. 2014. Prevaccine era human papillomavirus types 6, 11, 16 and 18 seropositivity in the U.S.A., National Health and Nutrition Examination Surveys, 2003-2006. *Sex Transm Infect,* 90**,** 505-8.

IWAI, Y., ISHIDA, M., TANAKA, Y., OKAZAKI, T., HONJO, T. & MINATO, N. 2002. Involvement of PD-L1 on tumor cells in the escape from host immune system and tumor immunotherapy by PD-L1 blockade. *Proc Natl Acad Sci U S A,* 99**,** 12293-7.

JAFARI, M. & ANSARI-POUR, N. 2019. Why, When and How to Adjust Your P Values? *Cell J,* 20**,** 604-607.

JAMIESON, L. M., ANTONSSON, A., GARVEY, G., JU, X., SMITH, M., LOGAN, R. M., JOHNSON, N. W., HEDGES, J., SETHI, S., DUNBAR, T., LEANE, C., HILL, I., BROWN, A., RODER, D., DE SOUZA, M. & CANFELL, K. 2020. Prevalence of Oral Human Papillomavirus Infection Among Australian Indigenous Adults. *JAMA Netw Open,* 3**,** e204951.

JCVI. 2008. *JCVI statement on Human papillomavirus vaccines to protect against cervical cancer* [Online]. Available: https://webarchive.nationalarchives.gov.uk/ukgwa/20120907095410/http:/www.dh.gov.uk/ab/JCVI/DH_094744 [Accessed 20th August 2023].

JEON, S. & LAMBERT, P. F. 1995. Integration of human papillomavirus type 16 DNA into the human genome leads to increased stability of E6 and E7 mRNAs: implications for cervical carcinogenesis. *Proc Natl Acad Sci U S A,* 92**,** 1654-8.

JHA, S., VANDE POL, S., BANERJEE, N. S., DUTTA, A. B., CHOW, L. T. & DUTTA, A. 2010. Destabilization of TIP60 by human papillomavirus E6 results in attenuation of TIP60-dependent transcriptional regulation and apoptotic pathway. *Mol Cell,* 38**,** 700-11.

JIANG, M., XIE, X., ZHU, X., JIANG, S., MILENKOVIC, D., MISIC, J., SHI, Y., TANDUKAR, N., LI, X., ATANASSOV, I., JENNINGER, L., HOBERG, E., ALBARRAN-GUTIERREZ, S., SZILAGYI, Z., MACAO, B., SIIRA, S. J., CARELLI, V., GRIFFITH, J. D., GUSTAFSSON, C. M., NICHOLLS, T. J., FILIPOVSKA, A., LARSSON, N. G. & FALKENBERG, M. 2021. The mitochondrial single-stranded DNA binding protein is essential for initiation of mtDNA replication. *Sci Adv,* 7.

JIAO, L., LIU, Y., YU, X.-Y., PAN, X., ZHANG, Y., TU, J., SONG, Y.-H. & LI, Y. 2023. Ribosome biogenesis in disease: new players and therapeutic targets. *Signal Transduction and Targeted Therapy,* 8**,** 15.

JOHNSON, K. M., KINES, R. C., ROBERTS, J. N., LOWY, D. R., SCHILLER, J. T. & DAY, P. M. 2009. Role of heparan sulfate in attachment to and infection of the murine female genital tract by human papillomavirus. *J Virol,* 83**,** 2067-74.

JONES, D. L., ALANI, R. M. & MÜNGER, K. 1997a. The human papillomavirus E7 oncoprotein can uncouple cellular differentiation and proliferation in human keratinocytes by abrogating p21Cip1-mediated inhibition of cdk2. *Genes Dev,* 11**,** 2101-11.

JONES, D. L., THOMPSON, D. A. & MÜNGER, K. 1997b. Destabilization of the RB tumor suppressor protein and stabilization of p53 contribute to HPV type 16 E7-induced apoptosis. *Virology,* 239**,** 97-107.

JUREKA, A. S., SILVAS, J. A. & BASLER, C. F. 2020. Propagation, Inactivation, and Safety Testing of SARS-CoV-2. *Viruses,* 12.

KAJITANI, N., SATSUKA, A., KAWATE, A. & SAKAI, H. 2012. Productive Lifecycle of Human Papillomaviruses that Depends Upon Squamous Epithelial Differentiation. *Front Microbiol,* 3**,** 152.

KAJITANI, N. & SCHWARTZ, S. 2020. Role of Viral Ribonucleoproteins in Human Papillomavirus Type 16 Gene Expression. *Viruses,* 12.
KALANTARI, M., BLENNOW, E., HAGMAR, B. & JOHANSSON, B. 2001. Physical state of HPV16 and chromosomal mapping of the integrated form in cervical carcinomas. *Diagn Mol Pathol,* 10**,** 46-54.
KALANTARI, M., KARLSEN, F., KRISTENSEN, G., HOLM, R., HAGMAR, B. & JOHANSSON, B. 1998. Disruption of the E1 and E2 reading frames of HPV 16 in cervical carcinoma is associated with poor prognosis. *Int J Gynecol Pathol,* 17**,** 146-53.
KANDUC, D. 2002. Translational regulation of human papillomavirus type 16 E7 mRNA by the peptide SEQIKA, shared by rabbit alpha(1)-globin and human cytokeratin 7. *J Virol,* 76**,** 7040-8.
KAROLINSKA INSTITUTET. 2023. *International Human Papillomavirus (HPV) Reference Center* [Online]. Available: https://www.hpvcenter.se/human_reference_clones/ [Accessed 20th April 2022].
KAVANAGH, K., POLLOCK, K. G., CUSCHIERI, K., PALMER, T., CAMERON, R. L., WATT, C., BHATIA, R., MOORE, C., CUBIE, H., CRUICKSHANK, M. & ROBERTSON, C. 2017. Changes in the prevalence of human papillomavirus following a national bivalent human papillomavirus vaccination programme in Scotland: a 7-year cross-sectional study. *Lancet Infect Dis,* 17**,** 1293-1302.
KERO, K., RAUTAVA, J., SYRJÄNEN, K., GRENMAN, S. & SYRJÄNEN, S. 2012. Oral mucosa as a reservoir of human papillomavirus: point prevalence, genotype distribution, and incident infections among males in a 7-year prospective study. *Eur Urol,* 62**,** 1063-70.
KHAYARGOLI, P., NIYIBIZI, J., MAYRAND, M. H., AUDIBERT, F., MONNIER, P., BRASSARD, P., LAPORTE, L., LACAILLE, J., ZAHREDDINE, M., BÉDARD, M. J., GIRARD, I., FRANCOEUR, D., CARCELLER, A. M., LACROIX, J., FRASER, W., COUTLÉE, F. & TROTTIER, H. 2023. Human Papillomavirus Transmission and Persistence in Pregnant Women and Neonates. *JAMA Pediatr,* 177**,** 684-692.
KIM, J. J., BURGER, E. A., SY, S. & CAMPOS, N. G. 2017. Optimal Cervical Cancer Screening in Women Vaccinated Against Human Papillomavirus. *J Natl Cancer Inst,* 109.
KIM, S. H., JUHNN, Y. S., KANG, S., PARK, S. W., SUNG, M. W., BANG, Y. J. & SONG, Y. S. 2006. Human papillomavirus 16 E5 up-regulates the expression of vascular endothelial growth factor through the activation of epidermal growth factor receptor, MEK/ ERK1,2 and PI3K/Akt. *Cell Mol Life Sci,* 63**,** 930-8.
KIM, S. H., KOO, B. S., KANG, S., PARK, K., KIM, H., LEE, K. R., LEE, M. J., KIM, J. M., CHOI, E. C. & CHO, N. H. 2007. HPV integration begins in the tonsillar crypt and leads to the alteration of p16, EGFR and c-myc during tumor formation. *Int J Cancer,* 120**,** 1418-25.
KINES, R. C., THOMPSON, C. D., LOWY, D. R., SCHILLER, J. T. & DAY, P. M. 2009. The initial steps leading to papillomavirus infection occur on the basement membrane prior to cell surface binding. *Proc Natl Acad Sci U S A,* 106**,** 20458-63.
KLINGELHUTZ, A. J., FOSTER, S. A. & MCDOUGALL, J. K. 1996. Telomerase activation by the E6 gene product of human papillomavirus type 16. *Nature,* 380**,** 79-82.
KLINGENBERG, B., HAFKAMP, H. C., HAESEVOETS, A., MANNI, J. J., SLOOTWEG, P. J., WEISSENBORN, S. J., KLUSSMANN, J. P. & SPEEL, E. J. 2010. p16 INK4A overexpression is frequently detected in tumour-free tonsil tissue without association with HPV. *Histopathology,* 56**,** 957-67.
KNAPPE, M., BODEVIN, S., SELINKA, H. C., SPILLMANN, D., STREECK, R. E., CHEN, X. S., LINDAHL, U. & SAPP, M. 2007. Surface-exposed amino acid residues of

HPV16 L1 protein mediating interaction with cell surface heparan sulfate. *J Biol Chem,* 282**,** 27913-22.
KNIGHT, G. & ROBERTS, B. 2020. Awareness of oral and genital human papillomavirus (HPV) infection in young adolescents prior to gender-neutral vaccination. *BMJ Sex Reprod Health*.
KNIGHT, G. L., NEEDHAM, L., WARD, D. & ROBERTS, S. 2016. Pilot study investigating the prevalence of oral Human Papilloma Viral (HPV) infection in young adults. *Public Health,* 132**,** 105-7.
KOCJAN, B. J., HOŠNJAK, L. & POLJAK, M. 2016. Detection of alpha human papillomaviruses in archival formalin-fixed, paraffin-embedded (FFPE) tissue specimens. *Journal of Clinical Virology,* 76**,** S88-S97.
KOFLER, B., LAIMER, J., BRUCKMOSER, E., STEINBICHLER, T. B., RUNGE, A., SCHARTINGER, V. H., VON LAER, D. & BORENA, W. 2020. The role of hpv and non-hpv sexually transmitted infections in patients with oropharyngeal carcinoma: A case control study. *Cancers,* 12.
KOFOED, K., NORRBOM, C., FORSLUND, O., MØLLER, C., FRØDING, L. P., PEDERSEN, A. E., MARKAUSKAS, A., BLOMBERG, M., BAUMGARTNER-NIELSEN, J., MADSEN, J. T., STRAUSS, G., MADSEN, K. G. & SAND, C. 2015. Low prevalence of oral and nasal human papillomavirus in employees performing CO2-laser evaporation of genital warts or loop electrode excision procedure of cervical dysplasia. *Acta Derm Venereol,* 95**,** 173-6.
KOLA-PALMER, S. & DHINGRA, K. 2020. Awareness and knowledge of human papilloma virus in UK women aged 25 years and over: Results from a cross-sectional internet-based survey. *Eur J Cancer Care (Engl),* 29**,** e13181.
KOLEŚNIK, M., STĘPIEŃ, E. & POLZ-DACEWICZ, M. 2022. Prevalence of Human Papillomavirus (HPV) in the Oral Cavity of a Healthy Population in South-Eastern Poland. *Int J Environ Res Public Health,* 19.
KOSHIOL, J., LINDSAY, L., PIMENTA, J. M., POOLE, C., JENKINS, D. & SMITH, J. S. 2008. Persistent Human Papillomavirus Infection and Cervical Neoplasia: A Systematic Review and Meta-Analysis. *American Journal of Epidemiology,* 168**,** 123-137.
KRAWCZYK, E., SUPRYNOWICZ, F. A., LIU, X., DAI, Y., HARTMANN, D. P., HANOVER, J. & SCHLEGEL, R. 2008. Koilocytosis: a cooperative interaction between the human papillomavirus E5 and E6 oncoproteins. *Am J Pathol,* 173**,** 682-8.
KREIMER, A. R., ALBERG, A. J., DANIEL, R., GRAVITT, P. E., VISCIDI, R., GARRETT, E. S., SHAH, K. V. & GILLISON, M. L. 2004. Oral human papillomavirus infection in adults is associated with sexual behavior and HIV serostatus. *J Infect Dis,* 189**,** 686-98.
KREIMER, A. R., BHATIA, R. K., MESSEGUER, A. L., GONZÁLEZ, P., HERRERO, R. & GIULIANO, A. R. 2010. Oral human papillomavirus in healthy individuals: a systematic review of the literature. *Sex Transm Dis,* 37**,** 386-91.
KREIMER, A. R., VILLA, A., NYITRAY, A. G., ABRAHAMSEN, M., PAPENFUSS, M., SMITH, D., HILDESHEIM, A., VILLA, L. L., LAZCANO-PONCE, E. & GIULIANO, A. R. 2011. The epidemiology of oral HPV infection among a multinational sample of healthy men. *Cancer Epidemiol Biomarkers Prev,* 20**,** 172-82.
KROOK, M. A., REESER, J. W., ERNST, G., BARKER, H., WILBERDING, M., LI, G., CHEN, H.-Z. & ROYCHOWDHURY, S. 2021. Fibroblast growth factor receptors in cancer: genetic alterations, diagnostics, therapeutic targets and mechanisms of resistance. *British Journal of Cancer,* 124**,** 880-892.
KRSEK, A., BATICIC, L., SOTOSEK, V. & BRAUT, T. 2024. The Role of Biomarkers in HPV-Positive Head and Neck Squamous Cell Carcinoma: Towards Precision Medicine. *Diagnostics (Basel),* 14.

KUMAR, R., KOLLOLI, A. & SUBBIAN, S. 2021. Inactivation and Elimination of SARS-CoV-2 in Biosamples Using Simple Fixatives and Ultrafiltration. *Methods Protoc,* 4.

KUMARI, B., ALI, A., SHARMA, A. K., GEHLOT, A. & SHARMA, S. 2021. Non-invasive saliva-based screening of high-risk Human Papilloma Virus 16 and 18 in healthy young adults and creating awareness about its vaccination. *J Family Med Prim Care,* 10**,** 387-391.

KUROSE, K., TERAI, M., SOEDARSONO, N., RABELLO, D., NAKAJIMA, Y., BURK, R. D. & TAKAGI, M. 2004. Low prevalence of HPV infection and its natural history in normal oral mucosa among volunteers on Miyako Island, Japan. *Oral Surg Oral Med Oral Pathol Oral Radiol Endod,* 98**,** 91-6.

LAI, D., TAN, C. L., GUNARATNE, J., QUEK, L. S., NEI, W., THIERRY, F. & BELLANGER, S. 2013. Localization of HPV-18 E2 at mitochondrial membranes induces ROS release and modulates host cell metabolism. *PLoS One,* 8**,** e75625.

LAJER, C. B., GARNÆS, E., FRIIS-HANSEN, L., NORRILD, B., THERKILDSEN, M. H., GLUD, M., ROSSING, M., LAJER, H., SVANE, D., SKOTTE, L., SPECHT, L., BUCHWALD, C. & NIELSEN, F. C. 2012. The role of miRNAs in human papilloma virus (HPV)-associated cancers: bridging between HPV-related head and neck cancer and cervical cancer. *Br J Cancer,* 106**,** 1526-34.

LAM, J. O., BREAM, J. H., SUGAR, E. A., COLES, C. L., WEBER, K. M., BURK, R. D., WILEY, D. J., CRANSTON, R. D., REDDY, S., MARGOLICK, J. B., STRICKLER, H. D., WENTZ, A., JACOBSON, L., GUO, Y., XIAO, W., GILLISON, M. L. & D'SOUZA, G. 2016. Association of serum cytokines with oral HPV clearance. *Cytokine,* 83**,** 85-91.

LANG KUHS, K. A., GONZALEZ, P., STRUIJK, L., CASTRO, F., HILDESHEIM, A., VAN DOORN, L. J., RODRIGUEZ, A. C., SCHIFFMAN, M., QUINT, W., LOWY, D. R., PORRAS, C., DELVECCHIO, C., KATKI, H. A., JIMENEZ, S., SAFAEIAN, M., SCHILLER, J., SOLOMON, D., WACHOLDER, S., HERRERO, R. & KREIMER, A. R. 2013. Prevalence of and risk factors for oral human papillomavirus among young women in Costa Rica. *J Infect Dis,* 208**,** 1643-52.

LAPRISE, C., MADATHIL, S. A., SCHLECHT, N. F., CASTONGUAY, G., SOULIÈRES, D., NGUYEN-TAN, P. F., ALLISON, P., COUTLÉE, F., HIER, M., ROUSSEAU, M. C., FRANCO, E. L. & NICOLAU, B. 2017. Human papillomavirus genotypes and risk of head and neck cancers: Results from the HeNCe Life case-control study. *Oral Oncology,* 69**,** 56-61.

LECHNER, M., LIU, J., MASTERSON, L. & FENTON, T. R. 2022. HPV-associated oropharyngeal cancer: epidemiology, molecular biology and clinical management. *Nat Rev Clin Oncol***,** 1-22.

LEE, D., KIM, H. Z., JEONG, K. W., SHIM, Y. S., HORIKAWA, I., BARRETT, J. C. & CHOE, J. 2002. Human papillomavirus E2 down-regulates the human telomerase reverse transcriptase promoter. *J Biol Chem,* 277**,** 27748-56.

LEECHANACHAI, P., BANKS, L., MOREAU, F. & MATLASHEWSKI, G. 1992. The E5 gene from human papillomavirus type 16 is an oncogene which enhances growth factor-mediated signal transduction to the nucleus. *Oncogene,* 7**,** 19-25.

LENOUVEL, D., GONZÁLEZ-MOLES, M. Á., RUIZ-ÁVILA, I., GONZALEZ-RUIZ, L., GONZALEZ-RUIZ, I. & RAMOS-GARCÍA, P. 2020. Prognostic and clinicopathological significance of PD-L1 overexpression in oral squamous cell carcinoma: A systematic review and comprehensive meta-analysis. *Oral Oncology,* 106**,** 104722.

LENZE, D., MÜLLER, H.-H. & HUMMEL, M. 2012. Considerations for the use of formalin-fixed and paraffin-embedded tissue specimens for clonality analysis. *Journal of Hematopathology,* 5**,** 27-34.
LI, H., ZHU, H., XU, C. J. & YUAN, J. 1998. Cleavage of BID by caspase 8 mediates the mitochondrial damage in the Fas pathway of apoptosis. *Cell,* 94**,** 491-501.
LIANG, S., DANG, B., CHEN, S. & MI, H. 2024. Prognostic value and immunological role of cathepsin S gene in pan-cancer. *Oncol Lett,* 27**,** 41.
LICITRA, L., PERRONE, F., BOSSI, P., SUARDI, S., MARIANI, L., ARTUSI, R., OGGIONNI, M., ROSSINI, C., CANTÙ, G., SQUADRELLI, M., QUATTRONE, P., LOCATI, L. D., BERGAMINI, C., OLMI, P., PIEROTTI, M. A. & PILOTTI, S. 2006. High-risk human papillomavirus affects prognosis in patients with surgically treated oropharyngeal squamous cell carcinoma. *J Clin Oncol,* 24**,** 5630-6.
LIEDERBACH, E., KYRILLOS, A., WANG, C. H., LIU, J. C., STURGIS, E. M. & BHAYANI, M. K. 2017. The national landscape of human papillomavirus-associated oropharynx squamous cell carcinoma. *Int J Cancer,* 140**,** 504-512.
LILLSUNDE LARSSON, G., CARLSSON, J., KARLSSON, M. G. & HELENIUS, G. 2015. Evaluation of HPV Genotyping Assays for Archival Clinical Samples. *The Journal of Molecular Diagnostics,* 17**,** 293-301.
LIM, Y. X., MIERZWA, M. L., SARTOR, M. A. & D'SILVA, N. J. 2023. Clinical, morphologic and molecular heterogeneity of HPV-associated oropharyngeal cancer. *Oncogene,* 42**,** 2939-2955.
LIN, L. & XU, C. 2020. Arcsine-based transformations for meta-analysis of proportions: Pros, cons, and alternatives. *Health Sci Rep,* 3**,** e178.
LIN, X., KANG, K., CHEN, P., ZENG, Z., LI, G., XIONG, W., YI, M. & XIANG, B. 2024. Regulatory mechanisms of PD-1/PD-L1 in cancers. *Mol Cancer,* 23**,** 108.
LIU, J., ZHANG, M., ROSE, B., VEILLARD, A. S., JONES, D., ZHANG, X., SOON LEE, C., MILROSS, C. & HONG, A. 2015. Ki67 Expression has Prognostic Significance in Relation to Human Papillomavirus Status in Oropharyngeal Squamous Cell Carcinoma. *Ann Surg Oncol,* 22**,** 1893-900.
LU, Y., XIE, Z., LUO, G., YAN, H., QIAN, H.-Z., FU, L., WANG, B., HUANG, R., CAO, F., LIN, H., YOU, R., TAN, L., YU, T., CHEN, M., LI, C., LIU, X., LEI, W. & ZOU, H. 2022. Global burden of oropharyngeal cancer attributable to human papillomavirus by anatomical subsite and geographic region. *Cancer Epidemiology,* 78**,** 102140.
LUCAS-ROXBURGH, R., BENSCHOP, J., DUNOWSKA, M. & PERROTT, M. 2015. Prevalence of human papillomaviruses in the mouths of New Zealand women. *N Z Med J,* 128**,** 45-52.
LUDWIG, C., GILLET, L., ROSENBERGER, G., AMON, S., COLLINS, B. C. & AEBERSOLD, R. 2018. Data-independent acquisition-based SWATH-MS for quantitative proteomics: a tutorial. *Mol Syst Biol,* 14**,** e8126.
LUPATO, V., HOLZINGER, D., HÖFLER, D., MENEGALDO, A., GIORGI ROSSI, P., DEL MISTRO, A., DA MOSTO, M. C., PAWLITA, M. & BOSCOLO-RIZZO, P. 2017. Prevalence and Determinants of Oral Human Papillomavirus Infection in 500 Young Adults from Italy. *PLoS One,* 12**,** e0170091.
LYFORD-PIKE, S., PENG, S., YOUNG, G. D., TAUBE, J. M., WESTRA, W. H., AKPENG, B., BRUNO, T. C., RICHMON, J. D., WANG, H., BISHOP, J. A., CHEN, L., DRAKE, C. G., TOPALIAN, S. L., PARDOLL, D. M. & PAI, S. I. 2013. Evidence for a role of the PD-1:PD-L1 pathway in immune resistance of HPV-associated head and neck squamous cell carcinoma. *Cancer Res,* 73**,** 1733-41.
MACHADO, A. P., GATTO DE ALMEIDA, F., BONIN, C. M., MARTINS PRATA, T. T., SOBRINHO ÁVILLA, L., JUNQUEIRA PADOVANI, C. T., TEIXEIRA FERREIRA,

A. M., DOS SANTOS FERNANDES, C. E. & TOZETTI, I. A. 2014. Presence of highly oncogenic human papillomavirus in the oral mucosa of asymptomatic men. *Braz J Infect Dis,* 18**,** 266-70.

MACHALEK, D. A., CHOW, E. P., GARLAND, S. M., WIGAN, R., CORNALL, A. M., FAIRLEY, C. K., KALDOR, J. M., HOCKING, J. S., WILLIAMS, H., MCNULTY, A., BELL, C., MARSHALL, L., OOI, C., CHEN, M. Y. & TABRIZI, S. N. 2017. Human Papillomavirus Prevalence in Unvaccinated Heterosexual Men After a National Female Vaccination Program. *J Infect Dis,* 215**,** 202-208.

MAGGIRWAR, S. B. & KHALSA, J. H. 2021. The Link between Cannabis Use, Immune System, and Viral Infections. *Viruses,* 13.

MALEROVA, S., HEJTMANKOVA, A., HAMSIKOVA, E., SALAKOVA, M., SMAHELOVA, J., KLOZAR, J. & TACHEZY, R. 2020. Prevalence and Risk Factors for Oral HPV in Healthy Population, in Central Europe. *Anticancer Res,* 40**,** 1597-1604.

MALLEN-ST CLAIR, J., ALANI, M., WANG, M. B. & SRIVATSAN, E. S. 2016. Human papillomavirus in oropharyngeal cancer: The changing face of a disease. *Biochim Biophys Acta,* 1866**,** 141-150.

MALLENDER, R. 2024. *The Pathogenesis of Human Papillomavirus (HPV) in Cancer of the Oropharynx.* Doctor of Philosophy, University of Derby.

MARAIS, D. J., SAMPSON, C., JEFTHA, A., DHAYA, D., PASSMORE, J. A. S., DENNY, L., RYBICKI, E. P., VAN DER WALT, E., STEPHEN, L. X. G. & WILLIAMSON, A. L. 2006. More men than women make mucosal IgA antibodies to human papillomavirus type 16 (HPV-16) and HPV-18: A study of oral HPV and oral HPV antibodies in a normal healthy population. *BMC Infectious Diseases,* 6.

MARKOWITZ, L. E., DUNNE, E. F., SARAIYA, M., CHESSON, H. W., CURTIS, C. R., GEE, J., BOCCHINI, J. A., JR. & UNGER, E. R. 2014. Human papillomavirus vaccination: recommendations of the Advisory Committee on Immunization Practices (ACIP). *MMWR Recomm Rep,* 63**,** 1-30.

MARKOWITZ, L. E., HARIRI, S., LIN, C., DUNNE, E. F., STEINAU, M., MCQUILLAN, G. & UNGER, E. R. 2013. Reduction in human papillomavirus (HPV) prevalence among young women following HPV vaccine introduction in the United States, National Health and Nutrition Examination Surveys, 2003-2010. *J Infect Dis,* 208**,** 385-93.

MARKOWITZ, L. E., LIU, G., HARIRI, S., STEINAU, M., DUNNE, E. F. & UNGER, E. R. 2016. Prevalence of HPV After Introduction of the Vaccination Program in the United States. *Pediatrics,* 137**,** e20151968.

MARKOWITZ, L. E., STERNBERG, M., DUNNE, E. F., MCQUILLAN, G. & UNGER, E. R. 2009. Seroprevalence of Human Papillomavirus Types 6, 11, 16, and 18 in the United States: National Health and Nutrition Examination Survey 2003–2004. *The Journal of Infectious Diseases,* 200**,** 1059-1067.

MAUFORT, J. P., SHAI, A., PITOT, H. C. & LAMBERT, P. F. 2010. A role for HPV16 E5 in cervical carcinogenesis. *Cancer Res,* 70**,** 2924-31.

MAXWELL, J. H., GRANDIS, J. R. & FERRIS, R. L. 2016. HPV-Associated Head and Neck Cancer: Unique Features of Epidemiology and Clinical Management. *Annu Rev Med,* 67**,** 91-101.

MAZUL, A. L., NAIK, A. N., ZHAN, K. Y., STEPAN, K. O., OLD, M. O., KANG, S. Y., NAKKEN, E. R. & PURAM, S. V. 2021. Gender and race interact to influence survival disparities in head and neck cancer. *Oral Oncology,* 112**,** 105093.

MCBRIDE, A. A. 2013. The Papillomavirus E2 proteins. *Virology,* 445**,** 57-79.

MCDOWELL, S. H., GALLAHER, S. A., BURDEN, R. E. & SCOTT, C. J. 2020. Leading the invasion: The role of Cathepsin S in the tumour microenvironment. *Biochim Biophys Acta Mol Cell Res,* 1867**,** 118781.

MCINTOSH, P. B., MARTIN, S. R., JACKSON, D. J., KHAN, J., ISAACSON, E. R., CALDER, L., RAJ, K., GRIFFIN, H. M., WANG, Q., LASKEY, P., ECCLESTON, J. F. & DOORBAR, J. 2008. Structural analysis reveals an amyloid form of the human papillomavirus type 16 E1--E4 protein and provides a molecular basis for its accumulation. *J Virol,* 82**,** 8196-203.

MCINTOSH, R. D., ANDRUS, E. C., WALLINE, H. M., SANDLER, C. B., GOUDSMIT, C. M., MORAVEK, M. B., STROUMSA, D., KATTARI, S. K. & BROUWER, A. F. 2024. Prevalence and Determinants of Cervicovaginal, Oral, and Anal Human Papillomavirus Infection in a Population of Transgender and Gender Diverse People Assigned Female at Birth. *LGBT Health*.

MCINTYRE, J. B., WU, J. S., CRAIGHEAD, P. S., PHAN, T., KÖBEL, M., LEES-MILLER, S. P., GHATAGE, P., MAGLIOCCO, A. M. & DOLL, C. M. 2013. PIK3CA mutational status and overall survival in patients with cervical cancer treated with radical chemoradiotherapy. *Gynecologic Oncology,* 128**,** 409-414.

MEDAWAR, P. B. 1941. III.—THE RATE OF PENETRATION OF FIXATIVES. *Journal of the Royal Microscopical Society,* 61**,** 46-57.

MEHANNA, H., BRYANT, T. S., BABRAH, J., LOUIE, K., BRYANT, J. L., SPRUCE, R. J., BATIS, N., OLALEYE, O., JONES, J., STRUIJK, L., MOLIJN, A., VORSTERS, A., ROSILLON, D., TAYLOR, S. & D'SOUZA, G. 2019a. Human Papillomavirus (HPV) Vaccine Effectiveness and Potential Herd Immunity for Reducing Oncogenic Oropharyngeal HPV-16 Prevalence in the United Kingdom: A Cross-sectional Study. *Clin Infect Dis,* 69**,** 1296-1302.

MEHANNA, H., BRYANT, T. S., BABRAH, J., LOUIE, K., BRYANT, J. L., SPRUCE, R. J., BATIS, N., OLALEYE, O., JONES, J., STRUIJK, L., MOLIJN, A., VORSTERS, A., ROSILLON, D., TAYLOR, S. & D'SOUZA, G. 2019b. Human Papillomavirus (HPV) Vaccine Effectiveness and Potential Herd Immunity for Reducing Oncogenic Oropharyngeal HPV-16 Prevalence in the United Kingdom: A Cross-sectional Study. *Clinical Infectious Diseases,* 69**,** 1296-1302.

MEHANNA, H., ROBINSON, M., HARTLEY, A., KONG, A., FORAN, B., FULTON-LIEUW, T., DALBY, M., MISTRY, P., SEN, M., O'TOOLE, L., AL BOOZ, H., DYKER, K., MOLERON, R., WHITAKER, S., BRENNAN, S., COOK, A., GRIFFIN, M., AYNSLEY, E., ROLLES, M., DE WINTON, E., CHAN, A., SRINIVASAN, D., NIXON, I., GRUMETT, J., LEEMANS, C. R., BUTER, J., HENDERSON, J., HARRINGTON, K., MCCONKEY, C., GRAY, A. & DUNN, J. 2019c. Radiotherapy plus cisplatin or cetuximab in low-risk human papillomavirus-positive oropharyngeal cancer (De-ESCALaTE HPV): an open-label randomised controlled phase 3 trial. *Lancet,* 393**,** 51-60.

MEHANNA, H., TABERNA, M., VON BUCHWALD, C., TOUS, S., BROOKS, J., MENA, M., MOREY, F., GRØNHØJ, C., RASMUSSEN, J. H., GARSET-ZAMANI, M., BRUNI, L., BATIS, N., BRAKENHOFF, R. H., LEEMANS, C. R., BAATENBURG DE JONG, R. J., KLUSSMANN, J. P., WUERDEMANN, N., WAGNER, S., DALIANIS, T., MARKLUND, L., MIRGHANI, H., SCHACHE, A., JAMES, J. A., HUANG, S. H., O'SULLIVAN, B., NANKIVELL, P., BROGLIE, M. A., HOFFMANN, M., QUABIUS, E. S. & ALEMANY, L. 2023. Prognostic implications of p16 and HPV discordance in oropharyngeal cancer (HNCIG-EPIC-OPC): a multicentre, multinational, individual patient data analysis. *Lancet Oncol,* 24**,** 239-251.

MEITES, E., WINER, R. L., NEWCOMB, M. E., GORBACH, P. M., QUEREC, T. D., RUDD, J., COLLINS, T., LIN, J., MOORE, J., REMBLE, T., SWANSON, F., FRANZ, J., BOLAN, R. K., GOLDEN, M. R., MUSTANSKI, B., CROSBY, R. A., UNGER, E. R. & MARKOWITZ, L. E. 2020. Vaccine effectiveness against prevalent anal and oral human papillomavirus infection among men who have sex with men - United States, 2016-2018. *Journal of Infectious Diseases,* 222**,** 2052-2060.

MENA, M., TABERNA, M., MONFIL, L., ARBYN, M., DE SANJOSÉ, S., BOSCH, F. X., ALEMANY, L. & BRUNI, L. 2019. Might Oral Human Papillomavirus (HPV) Infection in Healthy Individuals Explain Differences in HPV-Attributable Fractions in Oropharyngeal Cancer? A Systematic Review and Meta-analysis. *J Infect Dis,* 219**,** 1574-1585.

MENA, M., WANG, X., TOUS, S., QUIROS, B., CLAVERO, O., ALEJO, M., MOREY, F., TABERNA, M., LEON VINTRO, X., LLOVERAS RUBIO, B., ALOS, L., MEHANNA, H., QUINT, W., PAWLITA, M., TOMMASINO, M., PAVÓN, M. A., MUÑOZ, N., DE SANJOSE, S., BOSCH, F. X., ALEMANY, L., ON BEHALF OF THE ICO INTERNATIONAL HPV IN, H. & NECK CANCER STUDY, G. 2022. Concordance of p16(INK4a) and E6*I mRNA among HPV-DNA-Positive Oropharyngeal, Laryngeal, and Oral Cavity Carcinomas from the ICO International Study. *Cancers (Basel),* 14.

MERCER, C. H., TANTON, C., PRAH, P., ERENS, B., SONNENBERG, P., CLIFTON, S., MACDOWALL, W., LEWIS, R., FIELD, N., DATTA, J., COPAS, A. J., PHELPS, A., WELLINGS, K. & JOHNSON, A. M. 2013. Changes in sexual attitudes and lifestyles in Britain through the life course and over time: findings from the National Surveys of Sexual Attitudes and Lifestyles (Natsal). *Lancet,* 382**,** 1781-94.

MERCK. 2024. *GenElute™ RNA/DNA/Protein Purification Kit* [Online]. Available: https://www.sigmaaldrich.com/GB/en/technical-documents/protocol/genomics/dna-and-rna-purification/rna-dna-protein-purification-kit [Accessed 31st May 2024].

MI, H. & THOMAS, P. 2009. PANTHER pathway: an ontology-based pathway database coupled with data analysis tools. *Methods Mol Biol,* 563**,** 123-40.

MICHAELS, L. 1987. Normal Anatomy, Histology; Inflammatory Diseases. *Ear, Nose and Throat Histopathology.* London: Springer London.

MICHAUD, D. S., LANGEVIN, S. M., ELIOT, M., NELSON, H. H., PAWLITA, M., MCCLEAN, M. D. & KELSEY, K. T. 2014. High-risk HPV types and head and neck cancer. *Int J Cancer,* 135**,** 1653-61.

MILAD, N., FANTAUZZI, M. F., MCGRATH, J. J. C., CASS, S. P., THAYAPARAN, D., WANG, P., AFKHAMI, S., AGUIAR, J. A., ASK, K., DOXEY, A. C., STAMPFLI, M. R. & HIROTA, J. A. 2023. Cannabis smoke suppresses antiviral immune responses to influenza A in mice. *ERJ Open Research,* 9**,** 00219-2023.

MINHAS, S., KASHIF, M., REHMAN, Z., IDREES, M. & ANSARI, F. 2023. Molecular characterization and risk factors of oral high-risk human papillomavirus among females in Punjab, Pakistan. *J Med Virol,* 95**,** e27893.

MIYAMOTO, Y., YAMADA, K. & YONEDA, Y. 2016. Importin α: a key molecule in nuclear transport and non-transport functions. *J Biochem,* 160**,** 69-75.

MOODY, C. 2017. Mechanisms by which HPV Induces a Replication Competent Environment in Differentiating Keratinocytes. *Viruses,* 9.

MORÁN-TORRES, A., PAZOS-SALAZAR, N. G., TÉLLEZ-LORENZO, S., JIMÉNEZ-LIMA, R., LIZANO, M., REYES-HERNÁNDEZ, D. O., MARIN-AQUINO, J. J. & MANZO-MERINO, J. 2021. HPV oral and oropharynx infection dynamics in young population. *Braz J Microbiol,* 52**,** 1991-2000.

MORBINI, P., CAPELLO, G. L., ALBERIZZI, P., BENAZZO, M., PAGLINO, C., COMOLI, P. & PEDRAZZOLI, P. 2015. Markers of squamocolumnar junction cells in normal tonsils and oropharyngeal cancer with and without HPV infection. *Histol Histopathol,* 30**,** 833-9.

MORDECHAI, R. A., STEINBERG, S., APEL-SARID, L., SHAOUL, E., ROZEN, S. Z., KHOURY, L. & RONEN, O. 2019. Detection of high-risk human papillomavirus in the tonsils of galilee region adults and young adults undergoing tonsillectomy. *Eur Arch Otorhinolaryngol,* 276**,** 2865-2871.

MOSMANN, J. P., ZAYAS, S., KIGUEN, A. X., VENEZUELA, R. F., ROSATO, O. & CUFFINI, C. G. 2021. Human papillomavirus and Chlamydia trachomatis in oral and genital mucosa of women with normal and abnormal cervical cytology. *BMC Infect Dis,* 21**,** 422.

MOURAD, M., JETMORE, T., JATEGAONKAR, A. A., MOUBAYED, S., MOSHIER, E. & URKEN, M. L. 2017. Epidemiological Trends of Head and Neck Cancer in the United States: A SEER Population Study. *J Oral Maxillofac Surg,* 75**,** 2562-2572.

MOURAD OUZZANI, H. H., ZBYS FEDOROWICZ, AND AHMED ELMAGARMID 2016. Rayyan---a web and mobile app for systematic reviews. *Systematic Reviews,* 5**,** 210.

MSD. 2024. *GARDASIL® 9 (Human Papillomavirus 9-valent Vaccine) (Recombinant, adsorbed)* [Online]. Available: https://www.msdconnect.co.uk/training-and-resources/gardasil-9/ [Accessed 20th December 2024].

MUELLER, S. N. & ROUSE, B. T. 2008. 27 - Immune responses to viruses. *In:* RICH, R. R., FLEISHER, T. A., SHEARER, W. T., SCHROEDER, H. W., FREW, A. J. & WEYAND, C. M. (eds.) *Clinical Immunology (Third Edition).* Edinburgh: Mosby.

MUJTABA, H., WANG, Y., DUAN, Y., CAO, M., ZHANG, N., BATOOL, I., MURTAZA, A., CHEN, X. & WANG, Y. 2018. Human papillomavirus in tonsillectomy specimen from China and Pakistan - Prevalence and genotype distribution. *Pathol Res Pract,* 214**,** 1713-1718.

MUÑOZ, N., BOSCH, F. X., DE SANJOSÉ, S., HERRERO, R., CASTELLSAGUÉ, X., SHAH, K. V., SNIJDERS, P. J. & MEIJER, C. J. 2003. Epidemiologic classification of human papillomavirus types associated with cervical cancer. *N Engl J Med,* 348**,** 518-27.

NAPOLITANO, F., ANGELILLO, S., BIANCO, A., DI GIUSEPPE, G., DI ONOFRIO, V., LICATA, F., LIGUORI, G., NOBILE, C. G. A., PAVIA, M., PELULLO, C. P., ZITO MARINO, F. & ANGELILLO, I. F. 2024. Genital and Oral HPV Geno-Prevalence Measured through Urine and Saliva Samples in Young Adults in Italy. *Vaccines (Basel),* 12.

NASCIMENTO, A., NOCETTI, M. C., LUGO, L. Z. A., JACOB, C. M. B., MACHADO, A. P., PADOVANI, C. T. J., FERREIRA, A. M. T., FERNANDES, C. & TOZETTI, I. A. 2019. Oncogenic high-risk human papillomavirus in patients with full denture. *Braz Oral Res,* 33**,** e091.

NATIONAL CANCER INSTITUTE. 2020. *ACS's Updated Cervical Cancer Screening Guidelines Explained* [Online]. Available: https://www.cancer.gov/news-events/cancer-currents-blog/2020/cervical-cancer-screening-hpv-test-guideline [Accessed 20th August 2023].

NHS. 2022. *Tonsillectomy Costing approach learning extension (formerly CA1)* [Online]. Available: https://www.england.nhs.uk/wp-content/uploads/2021/03/1304_ACG_LE_Tonsillectomy-CA1_v1_mar22.pdf [Accessed 25th April 2025].

NHS INFORM - SCOTLAND. 2023. *Cervical screening (smear test)* [Online]. Available: https://www.nhsinform.scot/healthy-living/screening/cervical-screening-smear-test [Accessed 20th August 2023].
NHS UK. 2021. *Head and neck cancer* [Online]. Available: https://www.nhs.uk/conditions/head-and-neck-cancer/ [Accessed 13th November 2024].
NOMURA, T., NAZMUL, T., YOSHIMOTO, R., HIGASHIURA, A., ODA, K. & SAKAGUCHI, T. 2021. Ethanol Susceptibility of SARS-CoV-2 and Other Enveloped Viruses. *Biocontrol Sci,* 26**,** 177-180.
NORTH, A. L. & NICCOLAI, L. M. 2016. Human Papillomavirus Vaccination Requirements in US Schools: Recommendations for Moving Forward. *Am J Public Health,* 106**,** 1765-70.
NULTON, T. J., OLEX, A. L., DOZMOROV, M., MORGAN, I. M. & WINDLE, B. 2017. Analysis of The Cancer Genome Atlas sequencing data reveals novel properties of the human papillomavirus 16 genome in head and neck squamous cell carcinoma. *Oncotarget,* 8**,** 17684-17699.
OH, H. Y., KIM, M. K., SEO, S., LEE, D. O., CHUNG, Y. K., LIM, M. C., KIM, J., LEE, C. W. & PARK, S. 2015. Alcohol consumption and persistent infection of high-risk human papillomavirus. *Epidemiol Infect,* 143**,** 1442-50.
OLIVEIRA, L. R. & RIBEIRO-SILVA, A. 2011. Prognostic significance of immunohistochemical biomarkers in oral squamous cell carcinoma. *International Journal of Oral and Maxillofacial Surgery,* 40**,** 298-307.
OLIVER, S. E., GORBACH, P. M., GRATZER, B., STEINAU, M., COLLINS, T., PARRISH, A., KERNDT, P. R., CROSBY, R. A., UNGER, E. R., MARKOWITZ, L. E. & MEITES, E. 2018. Risk Factors for Oral Human Papillomavirus Infection Among Young Men Who Have Sex With Men-2 Cities, United States, 2012-2014. *Sex Transm Dis,* 45**,** 660-665.
OLIVER, S. E., UNGER, E. R., LEWIS, R., MCDANIEL, D., GARGANO, J. W., STEINAU, M. & MARKOWITZ, L. E. 2017. Prevalence of Human Papillomavirus Among Females After Vaccine Introduction-National Health and Nutrition Examination Survey, United States, 2003-2014. *J Infect Dis,* 216**,** 594-603.
ORTIZ, A. P., GONZÁLEZ, D., RAMOS, J., MUÑOZ, C., REYES, J. C. & PÉREZ, C. M. 2018a. Association of marijuana use with oral HPV infection and periodontitis among Hispanic adults: Implications for oral cancer prevention. *J Periodontol,* 89**,** 540-548.
ORTIZ, A. P., GONZÁLEZ, D., VIVALDI-OLIVER, J., CASTAÑEDA, M., RIVERA, V., DÍAZ, E., CENTENO, H., MUÑOZ, C., PALEFSKY, J., JOSHIPURA, K. & PÉREZ, C. M. 2018b. Periodontitis and oral human papillomavirus infection among Hispanic adults. *Papillomavirus Res,* 5**,** 128-133.
OSAZUWA-PETERS, N., ADJEI BOAKYE, E., ROHDE, R. L., GANESH, R. N., MOIYADI, A. S., HUSSAINI, A. S. & VARVARES, M. A. 2019. Understanding of risk factors for the human papillomavirus (HPV) infection based on gender and race. *Scientific reports,* 9**,** 297.
OTALI, D., HE, Q., STOCKARD, C. R. & GRIZZLE, W. E. 2013. Preservation of immunorecognition by transferring cells from 10% neutral buffered formalin to 70% ethanol. *Biotech Histochem,* 88**,** 170-80.
OTT, C., ROSS, K., STRAUB, S., THIEDE, B., GÖTZ, M., GOOSMANN, C., KRISCHKE, M., MUELLER, M. J., KROHNE, G., RUDEL, T. & KOZJAK-PAVLOVIC, V. 2012. Sam50 functions in mitochondrial intermembrane space bridging and biogenesis of respiratory complexes. *Mol Cell Biol,* 32**,** 1173-88.

OZBUN, M. A. 2019. Extracellular events impacting human papillomavirus infections: Epithelial wounding to cell signaling involved in virus entry. *Papillomavirus Res,* 7**,** 188-192.

PADERNO, A., PETRELLI, F., LORINI, L., CAPRIOTTI, V., GURIZZAN, C. & BOSSI, P. 2024. The predictive role of PD-L1 in head and neck cancer: A systematic review and meta-analysis. *Oral Oncology,* 153**,** 106799.

PALMER, E., NEWCOMBE, R. G., GREEN, A. C., KELLY, C., NOEL GILL, O., HALL, G., FIANDER, A. N., PIROTTE, E., HIBBITTS, S. J., HOMER, J. & POWELL, N. G. 2014. Human papillomavirus infection is rare in nonmalignant tonsil tissue in the UK: implications for tonsil cancer precursor lesions. *Int J Cancer,* 135**,** 2437-43.

PALMER, T., WALLACE, L., POLLOCK, K. G., CUSCHIERI, K., ROBERTSON, C., KAVANAGH, K. & CRUICKSHANK, M. 2019. Prevalence of cervical disease at age 20 after immunisation with bivalent HPV vaccine at age 12-13 in Scotland: retrospective population study. *Bmj,* 365**,** l1161.

PAPANICOLAOU, G. N. & TRAUT, H. F. 1997. The diagnostic value of vaginal smears in carcinoma of the uterus. 1941. *Arch Pathol Lab Med,* 121**,** 211-24.

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

PATEL, D., HUANG, S. M., BAGLIA, L. A. & MCCANCE, D. J. 1999. The E6 protein of human papillomavirus type 16 binds to and inhibits co-activation by CBP and p300. *Embo j,* 18**,** 5061-72.

PATTANSHETTY, S., KOTRASHETTI, V. S., NAYAK, R., BHAT, K., SOMANNAVAR, P. & BABJI, D. 2014. PCR based detection of HPV 16 and 18 genotypes in normal oral mucosa of tobacco users and non-users. *Biotech Histochem,* 89**,** 433-9.

PAULI, S., KOPS, N. L., BESSEL, M., LINA VILLA, L., MORENO ALVES SOUZA, F., MENDES PEREIRA, G. F., NEVES HUGO, F. & WENDLAND, E. 2022. Sexual practices and HPV infection in unvaccinated young adults. *Sci Rep,* 12**,** 12385.

PERRY, M. E. 1994. The specialised structure of crypt epithelium in the human palatine tonsil and its functional significance. *J Anat,* 185 (Pt 1)**,** 111-27.

PETROSKY, E., BOCCHINI, J. A., JR., HARIRI, S., CHESSON, H., CURTIS, C. R., SARAIYA, M., UNGER, E. R. & MARKOWITZ, L. E. 2015. Use of 9-valent human papillomavirus (HPV) vaccine: updated HPV vaccination recommendations of the advisory committee on immunization practices. *MMWR Morb Mortal Wkly Rep,* 64**,** 300-4.

PETROVA, P. S., RAIBEKAS, A., PEVSNER, J., VIGO, N., ANAFI, M., MOORE, M. K., PEAIRE, A. E., SHRIDHAR, V., SMITH, D. I. & KELLY, J. 2003. MANF: a new mesencephalic, astrocyte-derived neurotrophic factor with selectivity for dopaminergic neurons. *Journal of Molecular Neuroscience,* 20**,** 173-187.

PETT, M. & COLEMAN, N. 2007. Integration of high-risk human papillomavirus: a key event in cervical carcinogenesis? *J Pathol,* 212**,** 356-67.

PICKARD, R. K., XIAO, W., BROUTIAN, T. R., HE, X. & GILLISON, M. L. 2012. The prevalence and incidence of oral human papillomavirus infection among young men and women, aged 18-30 years. *Sex Transm Dis,* 39**,** 559-66.

PIERCE CAMPBELL, C. M., KREIMER, A. R., LIN, H. Y., FULP, W., O'KEEFE, M. T., INGLES, D. J., ABRAHAMSEN, M., VILLA, L. L., LAZCANO-PONCE, E. & GIULIANO, A. R. 2015. Long-term persistence of oral human papillomavirus type 16: the HPV Infection in Men (HIM) study. *Cancer Prev Res (Phila),* 8**,** 190-6.

PIM, D., COLLINS, M. & BANKS, L. 1992. Human papillomavirus type 16 E5 gene stimulates the transforming activity of the epidermal growth factor receptor. *Oncogene,* 7**,** 27-32.

PIM, D., MASSIMI, P. & BANKS, L. 1997. Alternatively spliced HPV-18 E6* protein inhibits E6 mediated degradation of p53 and suppresses transformed cell growth. *Oncogene,* 15**,** 257-64.

PINGALI, C., YANKEY, D., ELAM-EVANS, L. D., MARKOWITZ, L. E., WILLIAMS, C. L., FREDUA, B., MCNAMARA, L. A., STOKLEY, S. & SINGLETON, J. A. 2021. National, Regional, State, and Selected Local Area Vaccination Coverage Among Adolescents Aged 13-17 Years - United States, 2020. *MMWR Morb Mortal Wkly Rep,* 70**,** 1183-1190.

PIRAMI, L., GIACHÈ, V. & BECCIOLINI, A. 1997. Analysis of HPV16, 18, 31, and 35 DNA in pre-invasive and invasive lesions of the uterine cervix. *J Clin Pathol,* 50**,** 600-4.

POLESEL, J., MENEGALDO, A., TIRELLI, G., GIACOMARRA, V., GUERRIERI, R., BABOCI, L., CASAROTTO, M., LUPATO, V., FANETTI, G., BOSCOLO-RIZZO, P. & FRATTA, E. 2021. Prognostic Significance of PD-L1 Expression In Patients With Primary Oropharyngeal Squamous Cell Carcinoma: A Meta-Analysis. *Front Oncol,* 11**,** 787864.

PRIMER DESIGN 2017. *Copy number standard curve for absolute quantification by real-time PCR*.

PUBLIC HEALTH ENGLAND. 2019. *NHS public health functions agreement 2019-20* [Online]. Available: https://www.england.nhs.uk/wp-content/uploads/2017/04/Service-Specification-No.11-Human-Papillomavirus-Virus-v2-080819.pdf [Accessed 20th April 2022].

PUBLIC HEALTH ENGLAND. 2020a. *Human papillomavirus (HPV) vaccination coverage in adolescent females and males in England: academic year 2019 to 2020* [Online]. Available: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/927694/hpr1920_HPV-vc.pdf [Accessed 20th April 2022].

PUBLIC HEALTH ENGLAND. 2020b. *Topic 1: the NHS Cervical Screening Programme (NHSCSP)* [Online]. Available: https://www.gov.uk/government/publications/cervical-screening-cervical-sample-taker-training/topic-1-the-nhs-cervical-screening-programme-nhs-csp#:~:text=3.1%20History%20of%20the%20cervical%20screening%20programme&text=1960s%3A%20cervical%20screening%20began,of%20inadequate%20cervical%20screening%20results [Accessed 20th April 2022].

PUBLIC HEALTH ENGLAND. 2021a. *Cervical Screening Saves Lives - evaluating the national campaign* [Online]. Available: https://phescreening.blog.gov.uk/2021/09/28/cervical-screening-saves-lives/ [Accessed 20th August 2023].

PUBLIC HEALTH ENGLAND. 2021b. *Cervical screening: programme overview* [Online]. Available: https://www.gov.uk/guidance/cervical-screening-programme-overview [Accessed 20th August 2023].

PUBLIC HEALTH ENGLAND. 2022a. *Cervical screening uptake* [Online]. Available: https://www.gov.uk/government/publications/be-clear-on-cancer-first-national-cervical-screening-saves-lives-campaign/cervical-screening-uptake [Accessed 20th August 2023].
PUBLIC HEALTH ENGLAND. 2022b. *Human papillomavirus (HPV) vaccination coverage in adolescents in England: 2021-2022* [Online]. Available: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1126762/hpr1322-HPV2.pdf [Accessed 20th August 2023].
PUBLIC HEALTH WALES. *What is Cervical Screening?* [Online]. Available: https://phw.nhs.wales/services-and-teams/cervical-screening-wales/what-is-cervical-screening/ [Accessed 20th August 2023].
PYEON, D., PEARCE, S. M., LANK, S. M., AHLQUIST, P. & LAMBERT, P. F. 2009. Establishment of human papillomavirus infection requires cell cycle progression. *PLoS Pathog,* 5**,** e1000318.
QIAGEN 2009. *TissueLyser LT Handbook*.
QIAGEN 2020. *AllPrep® DNA/RNA FFPE Handbook*.
QIAGEN 2023. *DNeasy® Blood & Tissue Handbook*.
QUABIUS, E. S., TRIBIUS, S., HEINRICHS, A., HAASER, D., KÜHNEL, A., LAUDIEN, M., HOPPE, F., MLYNSKI, R., AMBROSCH, P. & HOFFMANN, M. 2021. HPV DNA/RNA detection in various oral and oropharyngeal biomaterials identifies active HPV infections also in non-neoplastic tonsils. *Transl Oncol,* 14**,** 101002.
QUDDUS, M. R., MANNA, P., SUNG, C. J., KERLEY, S., STEINHOFF, M. M. & LAWRENCE, W. D. 2014. Prevalence, distribution, and viral burden of all 15 high-risk human papillomavirus types in adenosquamous carcinoma of the uterine cervix: a multiplex real-time polymerase chain reaction-based study. *Hum Pathol,* 45**,** 303-9.
RAGIN, C. C., WHEELER, V. W., WILSON, J. B., BUNKER, C. H., GOLLIN, S. M., PATRICK, A. L. & TAIOLI, E. 2007. Distinct distribution of HPV types among cancer-free Afro-Caribbean women from Tobago. *Biomarkers,* 12**,** 510-22.
RAGLAND, S. A. & CRISS, A. K. 2017. From bacterial killing to immune modulation: Recent insights into the functions of lysozyme. *PLoS Pathog,* 13**,** e1006512.
RAIMONDI, V., CICCARESE, F. & CIMINALE, V. 2020. Oncogenic pathways and the electron transport chain: a dangeROS liaison. *Br J Cancer,* 122**,** 168-181.
RAJENDRA SANTOSH, A. B., CHRISTIAN, N. A., JONES, T., THOMS-RODRIGUEZ, C. A., CONDAPPA, A., THOMPSON, T., PINKNEY, J., BARTON, E. N. & LINDO, J. 2019. Molecular epidemiology of human papillomavirus genotypes in oral rinses from HIV-positive and HIV-negative Jamaican patients. *J Investig Clin Dent,* 10**,** e12365.
RAMAN, N., WEIR, E. & MÜLLER, S. 2016. The AAA ATPase MDN1 Acts as a SUMO-Targeted Regulator in Mammalian Pre-ribosome Remodeling. *Mol Cell,* 64**,** 607-615.
RAMESH, P. S., KRISHNAMURTHY, S., SHRESTHA, S., NATARAJ, S. M. & DEVEGOWDA, D. 2021. Knowledge, awareness and prevalence of Human Papillomavirus among local University students and Healthcare workers in South India: A cross-sectional study. *Clinical Epidemiology and Global Health,* 12**,** 100839.
RAMPIAS, T., SASAKI, C., WEINBERGER, P. & PSYRRI, A. 2009. E6 and E7 gene silencing and transformed phenotype of human papillomavirus 16–positive oropharyngeal cancer cells. *Journal of the National Cancer Institute,* 101**,** 412-423.
RASMUSSEN, J. H., GRØNHØJ, C., HÅKANSSON, K., FRIBORG, J., ANDERSEN, E., LELKAITIS, G., KLUSSMANN, J. P., WITTEKINDT, C., WAGNER, S., VOGELIUS, I. R. & VON BUCHWALD, C. 2019. Risk profiling based on p16 and HPV DNA more accurately predicts location of disease relapse in patients with oropharyngeal squamous cell carcinoma. *Ann Oncol,* 30**,** 629-636.

RAUTAVA, J., WILLBERG, J., LOUVANTO, K., WIDEMAN, L., SYRJÄNEN, K., GRÉNMAN, S. & SYRJÄNEN, S. 2012. Prevalence, genotype distribution and persistence of human papillomavirus in oral mucosa of women: a six-year follow-up study. *PLoS One,* 7**,** e42171.

RAYMOND, S., LIU, B. & TAIOLI, E. 2018. Racial differences in the prevalence of oncogenic oral human papillomavirus infection types in the United States adult population. *Head Neck,* 40**,** 2219-2227.

REGAUER, S. & REICH, O. 2021. The origin of Human Papillomavirus (HPV) - induced cervical squamous cancer. *Curr Opin Virol,* 51**,** 111-118.

REN, S., GAYKALOVA, D. A., GUO, T., FAVOROV, A. V., FERTIG, E. J., TAMAYO, P., CALLEJAS-VALERA, J. L., ALLEVATO, M., GILARDI, M., SANTOS, J., FUKUSUMI, T., SAKAI, A., ANDO, M., SADAT, S., LIU, C., XU, G., FISCH, K. M., WANG, Z., MOLINOLO, A. A., GUTKIND, J. S., IDEKER, T., KOCH, W. M. & CALIFANO, J. A. 2020. HPV E2, E4, E5 drive alternative carcinogenic pathways in HPV positive cancers. *Oncogene,* 39**,** 6327-6339.

RIBEIRO, M. G. M., MARCOLINO, L. D., RAMOS, B. R. A., TRENTO, C. L., ARAUJO, K., SILVA, J. R. S., JAIN, S., GURGEL, R. Q., SILVA, M. G. & DOLABELLA, S. S. 2019. Elevated prevalence of high-risk HPV in healthy oral mucosa of users and nonusers of drugs in Northeastern Brazil. *Braz Oral Res,* 33**,** e031.

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

RICKFORD, R., ROGERS, M., HALLIDAY, A., LAMPTEY, P. & KOLA-PALMER, S. 2023. Attitudes to reducing cervical screening frequency among UK women: A qualitative analysis. *Psychooncology,* 32**,** 721-729.

RIDDELL, J. T., BROUWER, A. F., WALLINE, H. M., CAMPREDON, L. P., MEZA, R., EISENBERG, M. C., ANDRUS, E. C., DELINGER, R. L., YOST, M. L., MCCLOSKEY, J. K., THOMAS, T. B., HUANG, S., FERRIS, R. L., SHIN, D. M., FAKHRY, C., OW, T., LI, D., BERLOT, A., CAREY, T. E. & SCHLECHT, N. F. 2022. Oral human papillomavirus prevalence, persistence, and risk-factors in HIV-positive and HIV-negative adults. *Tumour Virus Res,* 13**,** 200237.

RIESE, R. J., WOLF, P. R., BRÖMME, D., NATKIN, L. R., VILLADANGOS, J. A., PLOEGH, H. L. & CHAPMAN, H. A. 1996. Essential role for cathepsin S in MHC class II-associated invariant chain processing and peptide loading. *Immunity,* 4**,** 357-66.

RIETH, K. K. S., GILL, S. R., LOTT-LIMBACH, A. A., MERKLEY, M. A., BOTERO, N., ALLEN, P. D. & MILLER, M. C. 2018. Prevalence of High-Risk Human Papillomavirus in Tonsil Tissue in Healthy Adults and Colocalization in Biofilm of Tonsillar Crypts. *JAMA Otolaryngol Head Neck Surg,* 144**,** 231-237.

ROB, F., HUGO, J., SALÁKOVÁ, M., ŠMAHELOVÁ, J., GKALPAKIOTIS, S., BOHÁČ, P. & TACHEZY, R. 2022. Prevalence of genital and oral human papillomavirus infection among psoriasis patients on biologic therapy. *Dermatol Ther,* 35**,** e15735.

ROBERTS, S., EVANS, D., MEHANNA, H. & PARISH, J. L. 2019. Modelling human papillomavirus biology in oropharyngeal keratinocytes. *Philos Trans R Soc Lond B Biol Sci,* 374**,** 20180289.

ROSEN, B. J., WALTER, L., GILMAN, R. H., CABRERRA, L., GRAVITT, P. E. & MARKS, M. A. 2016. Prevalence and correlates of oral human papillomavirus infection among healthy males and females in Lima, Peru. *Sex Transm Infect,* 92**,** 149-54.

ROUSSEAU, P. & AUTEXIER, C. 2015. Telomere biology: Rationale for diagnostics and therapeutics in cancer. *RNA Biol,* 12**,** 1078-82.

RUMIANEK, B., HILLMAN, R., JEOFFREYS, N., GHAZANFAR, S., SCHIFTER, M. & AJWANI, S. 2023. Prevalence of oral HPV in the adult sample population in Sydney. *Aust Dent J,* 68**,** 19-25.

RUSAN, M., KLUG, T. E., HENRIKSEN, J. J., BONDE, J. H., FUURSTED, K. & OVESEN, T. 2015. Prevalence of tonsillar human papillomavirus infections in Denmark. *Eur Arch Otorhinolaryngol,* 272**,** 2505-12.

SAFAEIAN, M., PORRAS, C., SCHIFFMAN, M., RODRIGUEZ, A. C., WACHOLDER, S., GONZALEZ, P., QUINT, W., VAN DOORN, L. J., SHERMAN, M. E., XHENSEVAL, V., HERRERO, R. & HILDESHEIM, A. 2010. Epidemiological study of anti-HPV16/18 seropositivity and subsequent risk of HPV16 and -18 infections. *J Natl Cancer Inst,* 102**,** 1653-62.

SÁNCHEZ-SILES, M., REMEZAL-SOLANO, M., LÓPEZ-LÓPEZ, A. M. & CAMACHO-ALONSO, F. 2020. Prevalence of human papillomavirus in the saliva of sexually active women with cervical intraepithelial neoplasias. *Medicina Oral Patologia Oral y Cirugia Bucal,* 25**,** e195-e204.

SANTOS, A. J., NOGUEIRA, C., ORTEGA-BELLIDO, M. & MALHOTRA, V. 2016. TANGO1 and Mia2/cTAGE5 (TALI) cooperate to export bulky pre-chylomicrons/VLDLs from the endoplasmic reticulum. *J Cell Biol,* 213**,** 343-54.

SASTRE-CANTÓN, M., PÉREZ-VILAR, S., VILATA-CORELL, J. J. & DÍEZ-DOMINGO, J. 2019. Prevalence of oral human papillomavirus infection among university students in Valencia, Spain. *Vaccine,* 37**,** 6276-6281.

SCARTH, J. A., PATTERSON, M. R., MORGAN, E. L. & MACDONALD, A. 2021. The human papillomavirus oncoproteins: a review of the host pathways targeted on the road to transformation. *J Gen Virol,* 102.

SCHACHE, A. G., POWELL, N. G., CUSCHIERI, K. S., ROBINSON, M., LEARY, S., MEHANNA, H., RAPOZO, D., LONG, A., CUBIE, H., JUNOR, E., MONAGHAN, H., HARRINGTON, K. J., NUTTING, C. M., SCHICK, U., LAU, A. S., UPILE, N., SHEARD, J., BROUGHAM, K., WEST, C. M., OGUEJIOFOR, K., THOMAS, S., NESS, A. R., PRING, M., THOMAS, G. J., KING, E. V., MCCANCE, D. J., JAMES, J. A., MORAN, M., SLOAN, P., SHAW, R. J., EVANS, M. & JONES, T. M. 2016. HPV-Related Oropharynx Cancer in the United Kingdom: An Evolution in the Understanding of Disease Etiology. *Cancer Res,* 76**,** 6598-6606.

SCHEFFNER, M., NUBER, U. & HUIBREGTSE, J. M. 1995. Protein ubiquitination involving an E1-E2-E3 enzyme ubiquitin thioester cascade. *Nature,* 373**,** 81-3.

SCHEFFNER, M., WERNESS, B. A., HUIBREGTSE, J. M., LEVINE, A. J. & HOWLEY, P. M. 1990. The E6 oncoprotein encoded by human papillomavirus types 16 and 18 promotes the degradation of p53. *Cell,* 63**,** 1129-36.

SCHELHAAS, M., SHAH, B., HOLZER, M., BLATTMANN, P., KÜHLING, L., DAY, P. M., SCHILLER, J. T. & HELENIUS, A. 2012. Entry of human papillomavirus type 16 by actin-dependent, clathrin- and lipid raft-independent endocytosis. *PLoS Pathog,* 8**,** e1002657.

SCHILLER, J. T., DAY, P. M. & KINES, R. C. 2010. Current understanding of the mechanism of HPV infection. *Gynecol Oncol,* 118**,** S12-7.

SCHNEIDER, A. 1993. Pathogenesis of genital HPV infection. *Genitourin Med,* 69**,** 165-73.

SCHWARZER, G., CHEMAITELLY, H., ABU-RADDAD, L. J. & RÜCKER, G. 2019. Seriously misleading results using inverse of Freeman-Tukey double arcsine transformation in meta-analysis of single proportions. *Res Synth Methods,* 10**,** 476-483.

SELINKA, H. C., FLORIN, L., PATEL, H. D., FREITAG, K., SCHMIDTKE, M., MAKAROV, V. A. & SAPP, M. 2007. Inhibition of transfer to secondary receptors by

heparan sulfate-binding drug or antibody induces noninfectious uptake of human papillomavirus. *J Virol,* 81**,** 10970-80.

SETHI, S., JU, X., ANTONSSON, A., CANFELL, K., SMITH, M. A., GARVEY, G., HEDGES, J. & JAMIESON, L. 2022. Oral HPV Infection among Indigenous Australians; Incidence, Persistence, and Clearance at 12-Month Follow-up. *Cancer Epidemiol Biomarkers Prev,* 31**,** 604-613.

SHABIHKHANI, M., LUCEY, G. M., WEI, B., MARENINOV, S., LOU, J. J., VINTERS, H. V., SINGER, E. J., CLOUGHESY, T. F. & YONG, W. H. 2014. The procurement, storage, and quality assurance of frozen blood and tissue biospecimens in pathology, biorepository, and biobank settings. *Clin Biochem,* 47**,** 258-66.

SHAW, P. H. 1996. The Role of p53 in Cell Cycle Regulation. *Pathology - Research and Practice,* 192**,** 669-675.

SHEIKH, S., BIUNDO, E., COURCIER, S., DAMM, O., LAUNAY, O., MAES, E., MARCOS, C., MATTHEWS, S., MEIJER, C., POSCIA, A., POSTMA, M., SAKA, O., SZUCS, T. & BEGG, N. 2018. A report on the status of vaccination in Europe. *Vaccine,* 36**,** 4979-4992.

SHERMAN, S. M. & NAILER, E. 2018. Attitudes towards and knowledge about Human Papillomavirus (HPV) and the HPV vaccination in parents of teenage boys in the UK. *PLoS One,* 13**,** e0195801.

SHETH, S., FARQUHAR, D. R., LENZE, N. R., MAZUL, A., BRENNAN, P., ANANTHARAMAN, D., ABEDI-ARDEKANI, B., ZEVALLOS, J. P., HAYES, D. N. & OLSHAN, F. 2021. Decreased overall survival in black patients with HPV-associated oropharyngeal cancer. *American Journal of Otolaryngology,* 42**,** 102780.

SHIBA, T. L. & CHHETRI, D. K. 2019. Chapter 3 - Oropharynx. *In:* CHHETRI, D. K. & DEWAN, K. (eds.) *Dysphagia Evaluation and Management in Otolaryngology.* Elsevier.

SHIGEISHI, H. & SUGIYAMA, M. 2016. Risk Factors for Oral Human Papillomavirus Infection in Healthy Individuals: A Systematic Review and Meta-Analysis. *J Clin Med Res,* 8**,** 721-9.

SHIGEISHI, H., SUGIYAMA, M., OHTA, K., RAHMAN, M. Z. & TAKECHI, M. 2016. Higher prevalence and gene amplification of HPV16 in oropharynx as compared to oral cavity. *J Appl Oral Sci,* 24**,** 397-403.

SHIGEISHI, H., SUGIYAMA, M., OHTA, K., YOKOYAMA, S., SAKUMA, M., MUROZUMI, H., KATO, H. & TAKECHI, M. 2018. High HPV16 E6 viral load in the oral cavity is associated with an increased number of bacteria: A preliminary study. *Biomedical Reports,* 8**,** 59-64.

SHIMIZU, F., OKAMI, K., EBISUMOTO, K., MAKI, D., SAKAI, A., OGURA, G., NAKAMURA, N. & IIDA, M. 2017. Synchronous HPV-Related Cancer of Bilateral Tonsils Detected Using Transoral Endoscopic Examination with Narrow-Band Imaging. *Case Rep Otolaryngol,* 2017**,** 9647010.

SIAS, C., SALICHOS, L., LAPA, D., DEL NONNO, F., BAIOCCHINI, A., CAPOBIANCHI, M. R. & GARBUGLIA, A. R. 2019. Alpha, Beta, gamma human PapillomaViruses (HPV) detection with a different sets of primers in oropharyngeal swabs, anal and cervical samples. *Virol J,* 16**,** 27.

SICHERO, L., GONÇALVES, M. G., BETTONI, F., COSER, E. M., MOTA, G., NUNES, R. A. L., MERCANTE, A. M. D. C., NATALINO, R., UNO, M., FERREIRA ALVES, M. J., MATOS, L. L., KOWALSKI, L. P., KULCSAR, M. A. V., DE ALVARENGA, G. F., HÖFLER, D., SCHROEDER, L., WATERBOER, T., TOMMASINO, M. & VILLA, L. L. 2024. Detection of serum biomarkers of HPV-16 driven oropharynx and oral cavity cancer in Brazil. *Oral Oncology,* 149**,** 106676.

SIMONSON, S. J., DIFILIPPANTONIO, M. J. & LAMBERT, P. F. 2005. Two distinct activities contribute to human papillomavirus 16 E6's oncogenic potential. *Cancer Res,* 65**,** 8266-73.

SINGH, D., VIGNAT, J., LORENZONI, V., ESLAHI, M., GINSBURG, O., LAUBY-SECRETAN, B., ARBYN, M., BASU, P., BRAY, F. & VACCARELLA, S. 2023. Global estimates of incidence and mortality of cervical cancer in 2020: a baseline analysis of the WHO Global Cervical Cancer Elimination Initiative. *Lancet Glob Health,* 11**,** e197-e206.

SKOCZYŃSKI, M., GOŹDZICKA-JÓZEFIAK, A. & KWAŚNIEWSKA, A. 2015. The prevalence of human papillomavirus between the neonates and their mothers. *BioMed Research International,* 2015.

SMITH, E. M., RITCHIE, J. M., YANKOWITZ, J., WANG, D., TUREK, L. P. & HAUGEN, T. H. 2004. HPV prevalence and concordance in the cervix and oral cavity of pregnant women. *Infectious Disease in Obstetrics and Gynecology,* 12**,** 45-56.

SMITH, J. A., WHITE, E. A., SOWA, M. E., POWELL, M. L., OTTINGER, M., HARPER, J. W. & HOWLEY, P. M. 2010. Genome-wide siRNA screen identifies SMCX, EP400, and Brd4 as E2-dependent regulators of human papillomavirus oncogene expression. *Proc Natl Acad Sci U S A,* 107**,** 3752-7.

SNIJDERS, P. J., STEENBERGEN, R. D., HEIDEMAN, D. A. & MEIJER, C. J. 2006. HPV-mediated cervical carcinogenesis: concepts and clinical implications. *J Pathol,* 208**,** 152-64.

SONAWANE, K., SUK, R., CHIAO, E. Y., CHHATWAL, J., QIU, P., WILKIN, T., NYITRAY, A. G., SIKORA, A. G. & DESHMUKH, A. A. 2017. Oral Human Papillomavirus Infection: Differences in Prevalence Between Sexes and Concordance With Genital Human Papillomavirus Infection, NHANES 2011 to 2014. *Ann Intern Med,* 167**,** 714-724.

SOPORI, M. L. & KOZAK, W. 1998. Immunomodulatory effects of cigarette smoke. *Journal of Neuroimmunology,* 83**,** 148-156.

SPANGLE, J. M., GHOSH-CHOUDHURY, N. & MUNGER, K. 2012. Activation of cap-dependent translation by mucosal human papillomavirus E6 proteins is dependent on the integrity of the LXXLL binding motif. *J Virol,* 86**,** 7466-72.

SPANGLE, J. M. & MÜNGER, K. 2010. The human papillomavirus type 16 E6 oncoprotein activates mTORC1 signaling and increases protein synthesis. *J Virol,* 84**,** 9398-407.

SPANOS, W. C., HOOVER, A., HARRIS, G. F., WU, S., STRAND, G. L., ANDERSON, M. E., KLINGELHUTZ, A. J., HENDRIKS, W., BOSSLER, A. D. & LEE, J. H. 2008. The PDZ binding motif of human papillomavirus type 16 E6 induces PTPN13 loss, which allows anchorage-independent growth and synergizes with ras for invasive growth. *J Virol,* 82**,** 2493-500.

SPINILLO, A., DOMINONI, M., BOSCHI, A. C., SOSSO, C., FIANDRINO, G., CESARI, S. & GARDELLA, B. 2020. Clinical Significance of the Interaction between Human Papillomavirus (HPV) Type 16 and Other High-Risk Human Papillomaviruses in Women with Cervical Intraepithelial Neoplasia (CIN) and Invasive Cervical Cancer. *J Oncol,* 2020**,** 6508180.

SPODEN, G., FREITAG, K., HUSMANN, M., BOLLER, K., SAPP, M., LAMBERT, C. & FLORIN, L. 2008. Clathrin- and caveolin-independent entry of human papillomavirus type 16--involvement of tetraspanin-enriched microdomains (TEMs). *PLoS One,* 3**,** e3313.

SPODEN, G., KÜHLING, L., CORDES, N., FRENZEL, B., SAPP, M., BOLLER, K., FLORIN, L. & SCHELHAAS, M. 2013. Human papillomavirus types 16, 18, and 31 share similar endocytic requirements for entry. *J Virol,* 87**,** 7765-73.

STANKIEWICZ KARITA, H. C., MAGARET, A., HUANG, M. L., JEROME, K. R., FENG, Q. & WALD, A. 2020. Quantitative Oral HPV16 and HPV18 Detection in Persons Attending Dental Clinics. *Sex Transm Dis,* 47**,** 100-104.

STEIERT, T. A., PARRA, G., GUT, M., ARNOLD, N., TROTTA, J. R., TONDA, R., MOUSSY, A., GERBER, Z., ABUJA, P. M., ZATLOUKAL, K., RÖCKEN, C., FOLSERAAS, T., GRIMSRUD, M. M., VOGEL, A., GOEPPERT, B., ROESSLER, S., HINZ, S., SCHAFMAYER, C., ROSENSTIEL, P., DELEUZE, J. F., GUT, I. G., FRANKE, A. & FORSTER, M. 2023. A critical spotlight on the paradigms of FFPE-DNA sequencing. *Nucleic Acids Res,* 51**,** 7143-7162.

SUEHIRO, T. T., DAMKE, G. M. Z. F., DAMKE, E., DE AZEVEDO RAMOS, P. L. R., DE ANDRADE PEREIRA SILVA, M., PELLOSO, S. M., HUH, W. K., FRANCO, R. A. F., DA SILVA, V. R. S., SCARINCI, I. C. & CONSOLARO, M. E. L. 2020. Cervical and oral human papillomavirus infection in women living with human immunodeficiency virus (HIV) and matched HIV-negative controls in Brazil. *Infectious Agents and Cancer,* 15.

SZABO, G. 1997. Alcohol's contribution to compromised immunity. *Alcohol Health Res World,* 21**,** 30-41.

SZABO, G. & SAHA, B. 2015. Alcohol's Effect on Host Defense. *Alcohol Res,* 37**,** 159-70.

SZYMONOWICZ, K. A. & CHEN, J. 2020. Biological and clinical aspects of HPV-related cancers. *Cancer Biol Med,* 17**,** 864-878.

TAM, S., FU, S., XU, L., KRAUSE, K. J., LAIRSON, D. R., MIAO, H., STURGIS, E. M. & DAHLSTROM, K. R. 2018. The epidemiology of oral human papillomavirus infection in healthy populations: A systematic review and meta-analysis. *Oral Oncol,* 82**,** 91-99.

TAWE, L., GROVER, S., NARASIMHAMURTHY, M., MOYO, S., GASEITSIWE, S., KASVOSVE, I. & PAGANOTTI, G. M. 2018. Molecular detection of human papillomavirus (HPV) in highly fragmented DNA from cervical cancer biopsies using double-nested PCR. *MethodsX,* 5**,** 569-578.

TERAI, M., HASHIMOTO, K., YODA, K. & SATA, T. 1999. High prevalence of human papillomaviruses in the normal oral cavity of adults. *Oral Microbiology and Immunology,* 14**,** 201-205.

THAIN, A., JENKINS, O., CLARKE, A. R. & GASTON, K. 1996. CpG methylation directly inhibits binding of the human papillomavirus type 16 E2 protein to specific DNA sequences. *J Virol,* 70**,** 7233-5.

THAKRAL, A., LEE, J. J. W., HOU, T., HUENIKEN, K., DUDDING, T., GORMLEY, M., VIRANI, S., OLSHAN, A., DIERGAARDE, B., NESS, A. R., WATERBOER, T., SMITH-BYRNE, K., BRENNAN, P., HAYES, D. N., SANDERSON, E., BROWN, M. C., HUANG, S., BRATMAN, S. V., SPREAFICO, A., DE ALMEIDA, J., DAVIES, J. C., BIERUT, L., MACFARLANE, G. J., LAGIOU, P., LAGIOU, A., POLESEL, J., AGUDO, A., ALEMANY, L., AHRENS, W., HEALY, C. M., CONWAY, D. I., NYGARD, M., CANOVA, C., HOLCATOVA, I., RICHIARDI, L., ZNAOR, A., GOLDSTEIN, D. P., HUNG, R. J., XU, W., LIU, G. & ESPIN-GARCIA, O. 2024. Smoking and alcohol by HPV status in head and neck cancer: a Mendelian randomization study. *Nature Communications,* 15**,** 7835.

THANASAS, I., LAVRANOS, G., GKOGKOU, P. & PARASKEVIS, D. 2020. Understanding of Young Adolescents About HPV Infection: How Health Education Can Improve Vaccination Rate. *J Cancer Educ,* 35**,** 850-859.

THE ROYAL COLLEGE OF SURGEONS OF ENGLAND 2005. National Prospective Tonsillectomy Audit.

THEODORAKI, M. N., VEIT, J. A., HOFFMANN, T. K. & GREVE, J. 2017. Synchronous bilateral tonsil carcinoma: case presentation and review of the literature. *Infect Agent Cancer,* 12**,** 38.

THERMOFISHER SCIENTIFIC. 2012. *Interpretation of Nucleic Acid 260/280 Ratios* [Online]. Available: https://assets.thermofisher.com/TFS-Assets/CAD/Product-Bulletins/T123-NanoDrop-Lite-Interpretation-of-Nucleic-Acid-260-280-Ratios.pdf [Accessed 30th May 2024].

THERMOFISHER SCIENTIFIC. 2016. *TRIzol Reagent User Guide, Invitrogen* [Online]. Available: https://tools.thermofisher.com/content/sfs/manuals/trizol_reagent.pdf [Accessed 25th April 2024].

THERMOFISHER SCIENTIFIC. 2022a. *PowerTrack™ SYBR™ Green Master Mix - USER GUIDE* [Online]. Available: https://www.thermofisher.com/document-connect/document-connect.html?url=https://assets.thermofisher.com/TFS-Assets%2FLSG%2Fmanuals%2FMAN0018825_PowerTrackSYBRGreenMasterMix_UG.pdf [Accessed 2nd July 2024].

THERMOFISHER SCIENTIFIC 2022b. *Qubit™ Protein Assay Kits*.

THOMAS, M., PIM, D. & BANKS, L. 1999. The role of the E6-p53 interaction in the molecular pathogenesis of HPV. *Oncogene,* 18**,** 7690-700.

THOMAS, P. D. 2017. The Gene Ontology and the Meaning of Biological Function. *Methods Mol Biol,* 1446**,** 15-24.

THOMPSON, D. L., DOUGLAS, J. M., JR., FOSTER, M., HAGENSEE, M. E., DIGUISEPPI, C., BARÓN, A. E., CAMERON, J. E., SPENCER, T. C., ZENILMAN, J., MALOTTE, C. K., BOLAN, G., KAMB, M. L. & PETERMAN, T. A. 2004. Seroepidemiology of infection with human papillomavirus 16, in men and women attending sexually transmitted disease clinics in the United States. *J Infect Dis,* 190**,** 1563-74.

TODOROVIĆ, M. M. & ZVRKO, E. Z. 2013. Immunoregulatory cytokines and chronic tonsillitis. *Bosn J Basic Med Sci,* 13**,** 230-6.

TRAUT, H. F. & PAPANICOLAOU, G. N. 1943. Cancer of the Uterus: The Vaginal Smear in Its Diagnosis. *Cal West Med,* 59**,** 121-2.

TSIKIS, S., HOEFER, L., BETHIMOUTIS, G., NICOLAIDOU, E., PAPARIZOS, V., ANTONIOU, C., CHARDALIAS, L., STAVROPOULOS, G. E., SHARMA, S., LONG, B. C., MUJACIC, I., KADRI, S., SEGAL, J. P., FURTADO, L. V., SCHNEIDER, J. & CHARNOT-KATSIKAS, A. 2018. Risk factors, prevalence, and site concordance of human papillomavirus in high-risk Greek men. *Eur J Cancer Prev,* 27**,** 514-520.

TURNER, D. O., WILLIAMS-COCKS, S. J., BULLEN, R., CATMULL, J., FALK, J., MARTIN, D., MAUER, J., BARBER, A. E., WANG, R. C., GERSTENBERGER, S. L. & KINGSLEY, K. 2011. High-risk human papillomavirus (HPV) screening and detection in healthy patient saliva samples: a pilot study. *BMC Oral Health,* 11**,** 28.

UK HEALTH SECURITY AGENCY. 2019. *HPV vaccination programme* [Online]. Available: https://www.gov.uk/government/collections/hpv-vaccination-programme [Accessed 20th April 2022].

VANDE POL, S. B. & KLINGELHUTZ, A. J. 2013. Papillomavirus E6 oncoproteins. *Virology,* 445**,** 115-37.

VATS, A., SKRABAR, N., DEL SAL, G. & BANKS, L. 2022. Loss of the E6AP Ubiquitin Ligase Induces p53-Dependent Phosphorylation of Human Papillomavirus 18 E6 in Cells Derived from Cervical Cancer. *J Virol,* 96**,** e0150321.

VAZQUEZ-ORTIZ, G., PINA-SANCHEZ, P., VAZQUEZ, K., DUENAS, A., TAJA, L., MENDOZA, P., GARCIA, J. A. & SALCEDO, M. 2005. Overexpression of cathepsin F, matrix metalloproteinases 11 and 12 in cervical cancer. *BMC Cancer,* 5**,** 68.

VERNON, S. D., UNGER, E. R., MILLER, D. L., LEE, D. R. & REEVES, W. C. 1997. Association of human papillomavirus type 16 integration in the E2 gene with poor disease-free survival from cervical cancer. *Int J Cancer,* 74**,** 50-6.

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

VISALLI, G., CURRÒ, M., FACCIOLÀ, A., RISO, R., MONDELLO, P., LAGANÀ, P., DI PIETRO, A., PICERNO, I. & SPATARO, P. 2016. Prevalence of human papillomavirus in saliva of women with HPV genital lesions. *Infectious Agents and Cancer,* 11.

VISALLI, G., DI PIETRO, A., CURRÒ, M., PRUITI CIARELLO, M., D'ANDREA, F., NUNNARI, G., PELLICANÒ, G. F. & FACCIOLÀ, A. 2021. How Much Does HIV Positivity Affect the Presence of Oral HPV? A Molecular Epidemiology Survey. *Int J Environ Res Public Health,* 18.

VITOŠEVIĆ, K., TODOROVIĆ, D., SLOVIĆ, Ž., VARLJEN, T., RADAKOVIĆ, I., RADOJEVIĆ, D., ČANOVIĆ, V. & TODOROVIĆ, M. 2023. The quality of DNA isolated from autopsy formalin-fixed and formalin-fixed paraffin-embedded tissues: study of 1662 samples. *Mol Biol Rep,* 50**,** 6323-6336.

VITOŠEVIĆ, K., TODOROVIĆ, M., VARLJEN, T., SLOVIĆ, Ž., MATIĆ, S. & TODOROVIĆ, D. 2018. Effect of formalin fixation on pcr amplification of DNA isolated from healthy autopsy tissues. *Acta Histochemica,* 120**,** 780-788.

VOJTECHOVA, Z., SABOL, I., SALAKOVA, M., SMAHELOVA, J., ZAVADIL, J., TUREK, L., GREGA, M., KLOZAR, J., PROCHAZKA, B. & TACHEZY, R. 2016. Comparison of the miRNA profiles in HPV-positive and HPV-negative tonsillar tumors and a model system of human keratinocyte clones. *BMC Cancer,* 16**,** 382.

WALKER, J. E., OLIVER, J. C., STEWART, A. M., BEH, S. T., ARCE, R. A., GLASS, M. J., VARGAS, D. E., QIJI, S. H., INTORCIA, A. J., BORJA, C. I., CLINE, M. P., HEMMINGSEN, S. J., KRUPP, A. N., MCHATTIE, R. D., MARINER, M. R., LORENZINI, I., ASLAM, S., TREMBLAY, C., BEACH, T. G. & SERRANO, G. E. 2023. Measuring Up: A Comparison of TapeStation 4200 and Bioanalyzer 2100 as Measurement Tools for RNA Quality in Postmortem Human Brain Samples. *Int J Mol Sci,* 24.

WANG, B., SUN, J., KITAMOTO, S., YANG, M., GRUBB, A., CHAPMAN, H. A., KALLURI, R. & SHI, G. P. 2006. Cathepsin S controls angiogenesis and tumor growth via matrix-derived angiogenic factors. *J Biol Chem,* 281**,** 6020-9.

WANG, H. K., DUFFY, A. A., BROKER, T. R. & CHOW, L. T. 2009. Robust production and passaging of infectious HPV in squamous epithelium of primary human keratinocytes. *Genes Dev,* 23**,** 181-94.

WANG, K., YIN, X. M., CHAO, D. T., MILLIMAN, C. L. & KORSMEYER, S. J. 1996. BID: a novel BH3 domain-only death agonist. *Genes Dev,* 10**,** 2859-69.

WANG, M. B., LIU, I. Y., GORNBEIN, J. A. & NGUYEN, C. T. 2015. HPV-Positive Oropharyngeal Carcinoma: A Systematic Review of Treatment and Prognosis. *Otolaryngol Head Neck Surg,* 153**,** 758-69.

WANG, Y., CHEN, F., QU, W., GONG, Y., WANG, Y., CHEN, L., ZHOU, Q., MO, J., ZHANG, H., LIN, L., BI, T., WANG, X., GU, J., LI, Y. & SUI, L. 2024. Alternative splicing in the genome of HPV and its regulation. *Front Cell Infect Microbiol,* 14**,** 1443868.

WEINBERGER, P. M., YU, Z., HAFFTY, B. G., KOWALSKI, D., HARIGOPAL, M., BRANDSMA, J., SASAKI, C., JOE, J., CAMP, R. L., RIMM, D. L. & PSYRRI, A.

2006. Molecular classification identifies a subset of human papillomavirus--associated oropharyngeal cancers with favorable prognosis. *J Clin Oncol,* 24**,** 736-47.

WESTRA, W. H., TAUBE, J. M., POETA, M. L., BEGUM, S., SIDRANSKY, D. & KOCH, W. M. 2008. Inverse relationship between human papillomavirus-16 infection and disruptive p53 gene mutations in squamous cell carcinoma of the head and neck. *Clin Cancer Res,* 14**,** 366-9.

WHITTON, A. 2020. *Oral Human Papillomavirus (HPV) Prevalence and Abundance in the UK Young Adult Population.* Doctor of Philosophy, University of Derby.

WHITTON, A. F., KNIGHT, G. L. & MARSH, E. K. 2024. Risk factors associated with oral Human Papillomavirus (HPV) prevalence within a young adult population. *BMC Public Health,* 24**,** 1485.

WILSON, J. A., O'HARA, J., FOUWEATHER, T., HOMER, T., STOCKEN, D. D., VALE, L., HAIGHTON, C., ROUSSEAU, N., WILSON, R., MCSWEENEY, L., WILKES, S., MORRISON, J., MACKENZIE, K., AH-SEE, K., CARRIE, S., HOPKINS, C., HOWE, N., HUSSAIN, M., MEHANNA, H., RAINE, C., SULLIVAN, F., VON WILAMOWITZ-MOELLENDORFF, A. & TEARE, M. D. 2023. Conservative management versus tonsillectomy in adults with recurrent acute tonsillitis in the UK (NATTINA): a multicentre, open-label, randomised controlled trial. *Lancet,* 401**,** 2051-2059.

WIMARDHANI, Y. S., SASANTI, H., WARDHANY, I. I., SARSITO, A. S., PRADONO, S. A., SUBITA, G. P., SOEGYANTO, A. I., RAHMAYANTI, F., CHAMUSRI, N. & IAMAROON, A. 2015. Saliva-Based Screening of High-Risk Human Papillomavirus Strains: Detection in Female Indonesian and Thai Dental Students. *Asian Pacific journal of cancer prevention : APJCP,* 16**,** 5525-5529.

WINDON, M. J., WATERBOER, T., HILLEL, A. T., CHIEN, W., BEST, S., STEWART, C., AKST, L., TROY, T., BENDER, N., MILES, B., RYAN, W. R., MANDAL, R., PITMAN, K., EISELE, D. W., FAKHRY, C. & D'SOUZA, G. 2019. Sex differences in HPV immunity among adults without cancer. *Hum Vaccin Immunother,* 15**,** 1935-1941.

WOLFF, C., SCHOTT, C., PORSCHEWSKI, P., REISCHAUER, B. & BECKER, K. F. 2011. Successful protein extraction from over-fixed and long-term stored formalin-fixed tissues. *PLoS One,* 6**,** e16353.

WONG, M. C. S., VLANTIS, A. C., LIANG, M., WONG, P. Y., HO, W. C. S., BOON, S. S., SZE, R. K. H., LEUNG, C., CHAN, P. K. S. & CHEN, Z. 2018. Prevalence and Epidemiologic Profile of Oral Infection with Alpha, Beta, and Gamma Papillomaviruses in an Asian Chinese Population. *J Infect Dis,* 218**,** 388-397.

WONGJAMPA, W., NAKAHARA, T., TANAKA, K., YUGAWA, T., EKALAKSANANAN, T., KLEEBKAOW, P., GOSHIMA, N., KIYONO, T. & PIENTONG, C. 2023. An in vitro carcinogenesis model for cervical cancer harboring episomal form of HPV16. *PLoS One,* 18**,** e0281069.

WOOD, N. H., MAKUA, K. S., LEBELO, R. L., REDZIC, N., BENOY, I., VANDERVEKEN, O. M. & BOGERS, J. 2020. Human Papillomavirus Prevalence in Oral and Oropharyngeal Rinse and Gargle Specimens of Dental Patients and of an HIV-Positive Cohort from Pretoria, South Africa. *Advances in Virology,* 2020.

WOODS, R. S. R., KEEGAN, H., WHITE, C., TEWARI, P., TONER, M., KENNEDY, S., O'REGAN, E. M., MARTIN, C. M., TIMON, C. V. I. & O'LEARY, J. J. 2017. Cytokeratin 7 in Oropharyngeal Squamous Cell Carcinoma: A Junctional Biomarker for Human Papillomavirus-Related Tumors. *Cancer Epidemiol Biomarkers Prev,* 26**,** 702-710.

WORLD HEALTH ORGANIZATION. 2023. *World Health Organization - Immunization dashboard, European Region* [Online]. Available:

https://immunizationdata.who.int/dashboard/regions/european-region [Accessed 20th November 2024].
WORLD HEALTH ORGANIZATION. 2024a. *Cervical Cancer* [Online]. Available: https://www.who.int/health-topics/cervical-cancer#tab=tab_1 [Accessed 10th November 2024].
WORLD HEALTH ORGANIZATION. 2024b. *Cervical Cancer Elimination Initiative* [Online]. Available: https://www.who.int/initiatives/cervical-cancer-elimination-initiative [Accessed 19th November 2024].
WORLD HEALTH ORGANIZATION. 2024c. *Coronavirus disease (COVID-19)* [Online]. Available: https://www.who.int/health-topics/coronavirus#tab=tab_1 [Accessed 31st May 2024].
WORLD HEALTH ORGANIZATION. 2024d. *IARC Monographs on the Identification of Carcinogenic Hazards to Humans* [Online]. Available: https://monographs.iarc.who.int/list-of-classifications [Accessed 11th November 2024].
WORLD HEALTH ORGANIZATION. 2024e. *Immunization coverage* [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/immunization-coverage [Accessed 10th November 2024].
WOTMAN, M., HERMAN, S. W., COSTANTINO, P. & THAM, T. 2020. The Prognostic Role of Programmed Death-Ligand 1 in Nasopharyngeal Carcinoma. *The Laryngoscope,* 130**,** 2598-2606.
WU, E. W., CLEMENS, K. E., HECK, D. V. & MÜNGER, K. 1993. The human papillomavirus E7 oncoprotein and the cellular transcription factor E2F bind to separate sites on the retinoblastoma tumor suppressor protein. *J Virol,* 67**,** 2402-7.
WU, S. Y., LEE, A. Y., HOU, S. Y., KEMPER, J. K., ERDJUMENT-BROMAGE, H., TEMPST, P. & CHIANG, C. M. 2006. Brd4 links chromatin targeting to HPV transcriptional silencing. *Genes Dev,* 20**,** 2383-96.
XIE, C. B., JANE-WIT, D. & POBER, J. S. 2020. Complement Membrane Attack Complex: New Roles, Mechanisms of Action, and Therapeutic Targets. *Am J Pathol,* 190**,** 1138-1150.
XING, B., GUO, J., SHENG, Y., WU, G. & ZHAO, Y. 2020. Human Papillomavirus-Negative Cervical Cancer: A Comprehensive Review. *Front Oncol,* 10**,** 606335.
YAMAGUCHI, F., UMEDA, Y., SHIMAMOTO, S., TSUCHIYA, M., TOKUMITSU, H., TOKUDA, M. & KOBAYASHI, R. 2012. S100 proteins modulate protein phosphatase 5 function: a link between CA2+ signal transduction and protein dephosphorylation. *J Biol Chem,* 287**,** 13787-98.
YAN, S., ZHANG, X., LIN, Q., DU, M., LI, Y., HE, S., CHEN, J., LI, X., BEI, J., CHEN, S. & SONG, M. 2024. Deciphering the interplay of HPV infection, MHC-II expression, and CXCL13(+) CD4(+) T cell activation in oropharyngeal cancer: implications for immunotherapy. *Cancer Immunol Immunother,* 73**,** 206.
YANG, E. J., QUICK, M. C., HANAMORNROONGRUANG, S., LAI, K., DOYLE, L. A., MCKEON, F. D., XIAN, W., CRUM, C. P. & HERFS, M. 2015. Microanatomy of the cervical and anorectal squamocolumnar junctions: a proposed model for anatomical differences in HPV-related cancer risk. *Mod Pathol,* 28**,** 994-1000.
YERLICI, V. T., ASTORI, A., KEJIOU, N. S., JORDAN, C. A., KHOSRAVIANI, N., CHAN, J. N. Y., HAKEM, R., RAUGHT, B., PALAZZO, A. F. & MEKHAIL, K. 2024. SARS-CoV-2 targets ribosomal RNA biogenesis. *Cell Reports,* 43**,** 113891.
YING, L., ZHANG, F., PAN, X., CHEN, K., ZHANG, N., JIN, J., WU, J., FENG, J., YU, H., JIN, H. & SU, D. 2016. Complement component 7 (C7), a potential tumor suppressor, is correlated with tumor progression and prognosis. *Oncotarget,* 7**,** 86536-86546.

YOSHIMATSU, Y., NAKAHARA, T., TANAKA, K., INAGAWA, Y., NARISAWA-SAITO, M., YUGAWA, T., OHNO, S. I., FUJITA, M., NAKAGAMA, H. & KIYONO, T. 2017. Roles of the PDZ-binding motif of HPV 16 E6 protein in oncogenic transformation of human cervical keratinocytes. *Cancer Sci,* 108**,** 1303-1309.

YU, L., MAJERCIAK, V. & ZHENG, Z. M. 2022. HPV16 and HPV18 Genome Structure, Expression, and Post-Transcriptional Regulation. *Int J Mol Sci,* 23.

YU, S., ZHU, Y., HE, H., HU, Y., ZHU, X., DIAO, W., LI, S., SHAN, G. & CHEN, X. 2023. Prevalence and risk factors of oral human papillomavirus infection among 4212 healthy adults in Hebei, China. *BMC Infect Dis,* 23**,** 773.

ZARGARAN, M. & AZIZI JALILIAN, F. 2020. Prevalence and risk factors of oral HPV infection among a healthy Iranian population; Kurdistan, West of Iran. *J Med Virol,* 92**,** 3821-3829.

ZHANG, H., ZHAO, Y., LIU, X., FU, L., GU, F. & MA, Y. 2021. High Expression of Complement Component C7 Indicates Poor Prognosis of Breast Cancer and Is Insensitive to Taxane-Anthracycline Chemotherapy. *Front Oncol,* 11**,** 724250.

ZHENG, Z. M. & BAKER, C. C. 2006. Papillomavirus genome structure, expression, and post-transcriptional regulation. *Front Biosci,* 11**,** 2286-302.

ZHOU, L., QIU, Q., ZHOU, Q., LI, J., YU, M., LI, K., XU, L., KE, X., XU, H., LU, B., WANG, H., LU, W., LIU, P. & LU, Y. 2022. Long-read sequencing unveils high-resolution HPV integration and its oncogenic progression in cervical cancer. *Nature Communications,* 13**,** 2563.

ZHU, Y., HE, H., ZHU, X., HU, Y., YU, S., DIAO, W., LI, S., SHAN, G. & CHEN, X. 2023a. Epidemiology of Oral Human Papillomavirus Infection among Diverse Chinese Adults in Typical Areas of China: Findings from the DLCC Study. *Cancer Epidemiol Biomarkers Prev,* 32**,** 1635-1643.

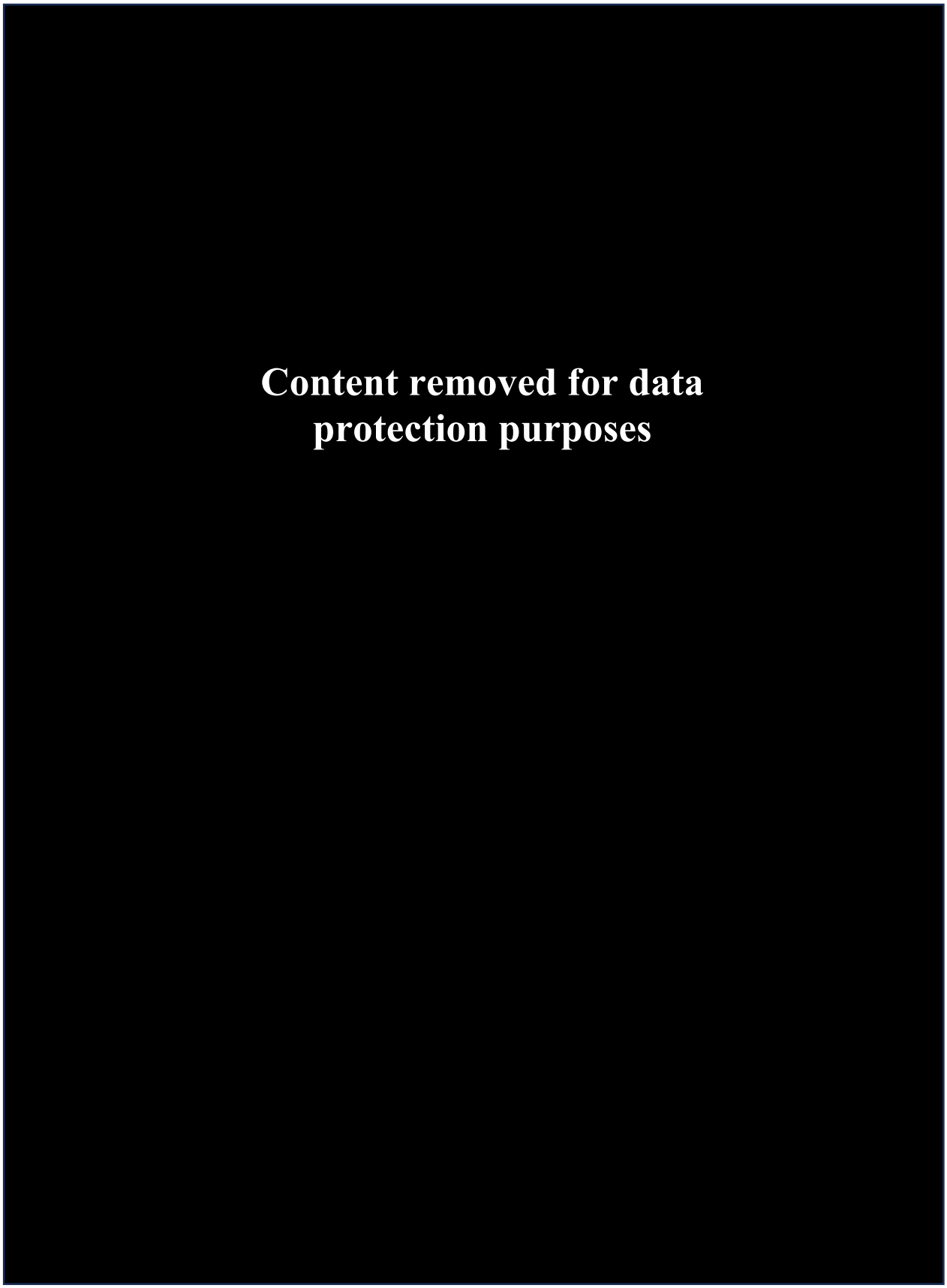
ZHU, Y., ZHU, X., DIAO, W., LIANG, Z., GAO, Z. & CHEN, X. 2023b. Correlation of immune makers with HPV 16 infections and the prognosis in oropharyngeal squamous cell carcinoma. *Clinical Oral Investigations,* 27**,** 1423-1433.

ZUMSTEG, Z. S., LUU, M., ROSENBERG, P. S., ELROD, J. K., BRAY, F., VACCARELLA, S., GAY, C., LU, D. J., CHEN, M. M., CHATURVEDI, A. K. & GOODMAN, M. T. 2023. Global epidemiologic patterns of oropharyngeal cancer incidence trends. *J Natl Cancer Inst,* 115**,** 1544-1554.

ZUR HAUSEN, H. 1977. Human papillomaviruses and their possible role in squamous cell carcinomas. *Current topics in microbiology and immunology***,** 1-30.

APPENDICES

Appendix A: NIHR-portfolio HPV clinical study documentation



Content removed for data protection purposes

Content removed for data protection purposes

A Research Ethics Committee established by the Health Research Authority

Figure 32. HRA approval from the Research Ethics Committee (REC) for the HPV clinical study.
REC reference number – 16/SW/0079.

Research Study: Investigation in to the infectivity levels of Human Papillomavirus (HPV) in tonsillar tissue

The University of Derby, together with the University Hospitals of Derby and Burton, Leicester, Nottingham, and Princess Royal Hospital (Telford), and Liverpool University Hospitals, are conducting a research study looking at the levels of infection of Human Papillomavirus (HPV)/Merkel cell polyomavirus in adults. We would like to invite you to participate in this study. Joining this study is entirely up to you. This information sheet is to explain the reasons for the study, to help you decide if you would like to take part and answer some common questions you may have.

HPV is a virus that has been shown to cause a range of cancers, including cervical cancer, mouth/neck cancer, and it has been linked with lung cancer. Both men and women can become infected with HPV and infection with this virus is very common: about 80% of the population are exposed to the virus before they reach 25 years old. Most people naturally clear the infection after 18 months, but the immune system of some individuals is not able to kill the virus, and these people remain infected for long periods of time. It is this long-term infection with HPV that puts a person at risk of developing a HPV-related cancer.

Long term HPV infection is the main cause of cervical cancer. In 2009 the NHS introduced the HPV vaccination programme for 12-13 year old girls to prevent them from getting infected with HPV. Following a decision in 2018, boys are now vaccinated against HPV across the UK.

The reason why men are getting more HPV-positive mouth/neck cancer is not currently understood. Therefore the purpose of this research study is to determine the levels of oral HPV infection in adults to investigate whether men are getting more oral HPV infections and/or a worse HPV infection, when compared to women. It will also investigate how many people are infected with HPV in the mouth. The study also aims to examine if there are particular regions within the mouth which appear to be more prone to getting an HPV infection, so that these areas can be more closely monitored by health care specialists for the early signs of cancer in the future.

The study will compare samples taken from adults undergoing tonsillectomies to investigate if there are any differences in viral infection in their tonsils and saliva. No personal information will be reviewed, and data will be anonymised.

The only patients eligible for this study are patients who are due to be undergoing a tonsillectomy at the Hospitals of Derby and Burton, or Leicester, or Telford, or Nottingham, or Liverpool are aged over 18, and are able to give informed consent to enter the study. Participation in this study is completely voluntary, and participants can withdraw at any stage. There are no detrimental effects to your treatment should you choose to decline entry or withdraw from the study at a later stage. If you choose to participate, you will not need to attend the hospital for any additional visits.

We will be analysing two types of samples at the University of Derby. The samples will be tonsil tissue and saliva samples. Participants will be having their tonsils removed in theatre, therefore there are no additional samples taken from your tonsils. To participate in this study you also need to supply a mouth swab before your tonsillectomy.

We will also ask you to complete a short, pseudoanonymised questionnaire about health risks related to HPV infection. This questionnaire will only be reviewed by the researching team. Once you have donated your saliva samples, completed the questionnaire, and had your tonsillectomy there would be no further participation in this study. We expect a total of 250-300 participants to enter the study until 2025. We do not see any increased risk for any patients that enter this study. There is no benefit for participants who enter into this study, however it would help to advance research in this field. There are no additional risks from entering the trial, other than those relating to your tonsillectomy that you are already undertaking. We will not inform you of the results of any tests carried out, but if you provide us with your email/postal address, we will notify you of the summary outcomes of the study.

Content removed for data protection purposes

withdraw at any time, and please contact us at this address if you have any further concerns or problems.

To link your consent form to your donated samples and questionnaire you need to supply a unique identification number. This should consist of the first 4 digits of your date of birth and last 4 digits of your phone number.

Content removed for data protection purposes

Please generate your unique identification number

_ _ _ _ _ _ _ _

UNIQUE IDENTIFICATION NUMBER: _ _ _ _ _ _ _ _

Figure 33. Letter of Invitation for HPV clinical study (v 5.3). Eligible participants were provided with a letter of invitation prior to their scheduled tonsillectomy surgery.

Research Study: Investigation into rates of Human Papillomavirus (HPV) infection in the mouth.

The University of Derby, and the University Hospitals of Derby and Burton, Leicester, Nottingham, and Princess Royal Hospital (Telford), and Liverpool University Hospitals, are conducting a research study looking at the levels of infection of Human Papillomavirus (HPV)/Merkel cell polyomavirus in adults. We would like to invite you to participate in this study. Joining this study is entirely up to you. This information sheet intends to explain the reasons for the study in order to answer common questions, and help you decide if you would like to take part. A member of our team will be available to address any further questions you may have, and please do ask if anything is unclear.

HPV is a virus that has been shown to cause a range of cancers, including cervical cancer and mouth/neck cancer. It may also be linked to lung cancer. Both men and women can become infected with HPV. Infection with HPV is very common, and about 80% of the population are exposed to the virus before they reach 25 years old. Most people naturally clear the infection after about 18 months, but the immune system of some individuals is not able to kill the virus, and these people remain infected for long periods of time. It is this long-term infection with HPV that puts a person at risk of developing a HPV related cancer.

Long term HPV infection is the main cause of cervical cancer. In 2009 the NHS introduced the HPV vaccination programme for 12-13 year old girls to prevent them from becoming infected with HPV. Following a decision in 2018, boys are now vaccinated against HPV across the UK.

The reason why men are getting more HPV related mouth/neck cancer is not currently understood. Therefore the purpose of this research study is to determine the levels of oral HPV infection in adults. We will investigate whether men are getting more oral HPV infection and/or a worse HPV infection when compared to women. It will also investigate how many people are infected with HPV. The study also aims to see if there are particular regions of the mouth that are more prone to HPV infection, such as the tonsil (where most oral HPV-related cancers are found). Potentially these areas may be more closely monitored by health care specialists for the early signs of cancer in the future.

The study will compare samples taken from adults undergoing tonsillectomies, to investigate if there are any differences in viral infection in their tonsils and saliva. All data handled will be pseudoanonymised. We are expecting a total of 250-300 participants in this study.

The study is being carried out at the Hospitals of Derby and Burton, Leicester, Nottingham, Telford, and Liverpool. The research team are comprised of medical doctors from the hospitals, together with research scientists from the University of Derby. None of the research team are being paid to conduct the study, but the research is being funded by the University of Derby, The Oracle Cancer Trust, and the Rosetrees Trust. The findings of this investigation will be published in a relevant virology and/or public health research journal(s). All research in the NHS is reviewed by an independent group called a Research Ethics Committee to protect participants' interests. This study has been reviewed by the South West Exeter Research Ethics Committee. This study is registered with ClinicalTrials.gov and our results will be published there to allow you to access them if you wish; ClinicalTrials.Gov Identifier: NCT02841631. You will also receive a letter or email (if you provide us with these details on the consent form) after the study is complete summarising the results of the study.

Who can participate in this study?

Patients who are eligible for this study are patients that are due to undergo a tonsillectomy at the Hospitals of Derby and Burton, Leicester, Nottingham, Telford, or Liverpool are **aged over 18,** and are able to consent to enter the study. This study is voluntary and participants can withdraw at any stage from the study. There will be no impact on or change to your treatment if you choose to decline entry or withdraw from this study at a later stage.

Potential applicants should allow themselves sufficient time to consider participating in this study. You do not have to agree to participate today. You are welcome to discuss your possible participation in this trial with the research team, General Practitioner, family or friends. If you do express a possible interest in participating in this study, but wish to have some time to consider, we will plan to discuss your participation at a future appointment.

Content removed for data protection purposes

What participation in this study will involve

We plan to recruit a total of 250-300 participants for the study over 6 years between 2019 and 2025. Evidence suggests there is no increased risk for any patients that enter this study. There is no additional treatment benefit for participants who enter into this study, however it would help to advance our knowledge of oral HPV and its link to head and neck cancer. There are no disadvantages from not taking part.

There are three steps to complete to be enrolled in this study:

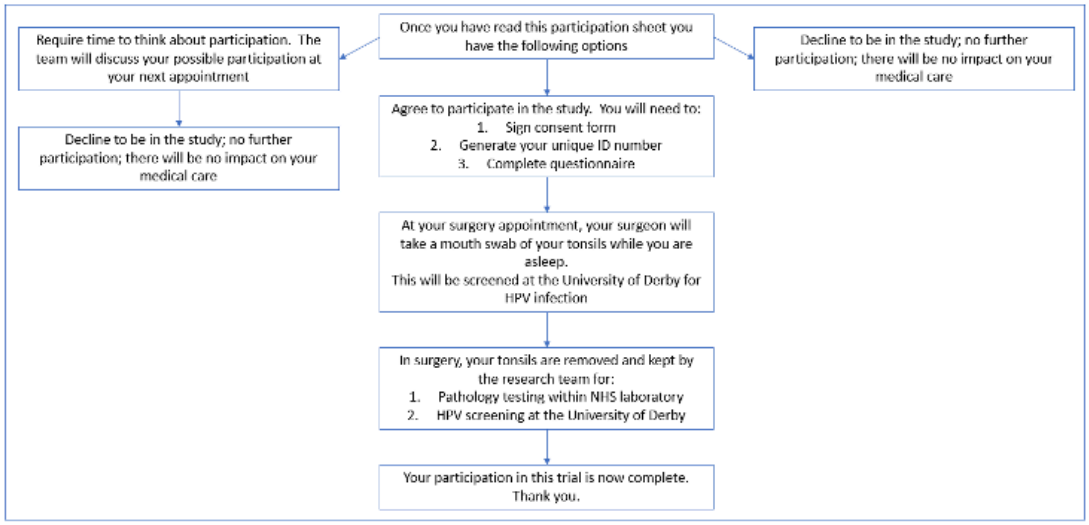
1. Complete a questionnaire related to lifestyle choices (version 5.3 dated 21/05/2024)
2. A sample of saliva will be taken via a mouth swab
3. Give informed consent for the tonsillar tissue that is removed during your tonsillectomy to be analysed for HPV

Some of the questions you will be asked to complete on the questionnaire will be related to lifestyle choices that have been shown to affect HPV infection, which include sexual activities, smoking, and drinking. There will be no questions related to your personal data, such as name or address etc. You will be given time to complete this questionnaire in private, and the questionnaire will be linked to consent forms and your donated samples by your unique identification number. The questionnaire should only take around 5 minutes to complete. There are no additional visits needed to enter this study.

The saliva samples that are taken via mouth swab and will be analysed for HPV at the University of Derby.

There are no additional risks to any participants entering the study. The staff involved in this study will have access to your medical notes to review the histology results of the tonsillar tissue that was removed. Part of this tissue will be processed in the hospital laboratory to look at the architecture of your tonsil, and part of this tissue will be tested at the University of Derby for the presence of HPV. The remaining tissue will be stored and examined for expression of viral and human genes and proteins to look at the development of an HPV infection, and to understand where an HPV infection establishes within the tonsil tissue.

We will not inform you of the results of the tests carried out because this study is to establish a new testing procedure for oral HPV infection; there is no such testing available at present.



If you wish to participate in this study, you will need to sign and return the separate consent form. You can keep this information sheet and a copy of your consent form, so that you have sufficient information about this study, who to contact if required and also a copy of your unique identification number.

Content removed for data protection purposes

Thank you for your time in reading this information sheet and also for considering participating in this study.

Summary:

Why?

- To determine the levels of oral HPV infection in adults
- Determine if men are getting more oral HPV infection and/or a worse HPV infection when compared to women
- To investigate how many adults are infected with HPV who undergo tonsillectomies
- To investigate if there are any differences in viral infection in their tonsils and saliva for these participants
- There is no benefit for participants who enter into this trial, however it would help to advance our knowledge of oral HPV infections

Who?

- Only patients undergoing a tonsillectomy at the University Hospitals of Derby and Burton, Leicester, Telford, Nottingham, or Liverpool, **aged over 18,** and able to consent, are eligible to take part in this study
- This study is completely voluntary, and participants can withdraw at any stage

What?

- We do not foresee any increased risk for any patients that choose to enter this study
- Participating individuals will complete a questionnaire relating to lifestyle choices that have been shown to affect HPV infection, including sexual activities, smoking and drinking. The questionnaire should only take about 5 minutes to complete
- Individuals will not need to attend the hospital for any additional visits due to participation in the study
- Participating individuals will give informed consent to have your tonsillar tissue tested for the presence of HPV infection

Where?

- Your tonsillectomy will take place at the Hospitals of Derby and Burton, or Leicester, or Nottingham, or Telford, or Liverpool. Testing for presence of HPV, and subsequent laboratory analyses, will be carried out at the University of Derby

How and When?

- The study will take place over a 6-year period, from 2019 to 2025

Figure 34. Patient Information Sheet for HPV clinical study (v 5.3). Eligible participants were provided with a patient information sheet prior to their scheduled tonsillectomy surgery.

Consent Form for

Investigation into rates of Human Papillomavirus (HPV) infection in the mouth

How will we use information about you?

- We will need to use information from you, and from your medical records for this research project.
- This information will include your name and contact details, as well as the unique identifier (code number) you will generate as part of this consent process.
- People will use this information to do the research or to check your records to make sure that the research is being done properly.
- People who do not need to know who you are will not be able to see your name or contact details. Your data will have a code number instead.
- We will keep all information about you safe and secure.
- Once we have finished the study, we will keep some of the data so we can check the results. We will write our reports in a way that no-one can work out that you took part in the study.

UNIQUE IDENTIFICATION NUMBER: _ _ _ _ _ _ _ _

(1st 4 digits of date of birth + last 4 digits of mobile phone number)

Statement of Informed Consent

Please **initial** to indicate you have read the statement

- I, ……(your full name) confirm that I have read the patient information sheet dated 21/05/2024 (version 5.3). ☐
- I have read the information sheet and understand that I need to donate saliva swab samples and to complete a questionnaire dated 21/05/2024 (version 5.3). ☐
- I understand that whilst the surgical team are performing my operation they will take a sample from the tissue they are operating on. This sample will then be used in the study. ☐
- I understand that my donated samples will be screened for the presence of HPV/Merkel cell polyomavirus and that I will **not** be informed of the results of this screening. ☐
- I understand that relevant sections of my medical notes and data collected during the study may be looked at by individuals from the research team (primarily the results of the tonsil tissue analysis), from regulatory authorities or from the NHS trust, where it is relevant to my taking part in this research. I give permission for these individuals to have access to my medical records. ☐

Content removed for data protection purposes

- I understand that my donated samples and questionnaire results will only be linked by my unique identification number to ensure the supplied personal information remains pseudoanonymised. ☐
- I understand that all personal data will be kept confidential, will remain pseudoanonymised at all times, and that all information will be stored securely at the research sites ☐
- I understand that any findings from the research may be disseminated through published articles and research conferences to help aid future research. ☐
- I give my full informed consent to participate in this research study. ☐

Signature of participant ----------

UNIQUE IDENTIFICATION NUMBER: _ _ _ _ _ _ _ _

(1st 4 digits of date of birth + last 4 digits of mobile phone number)

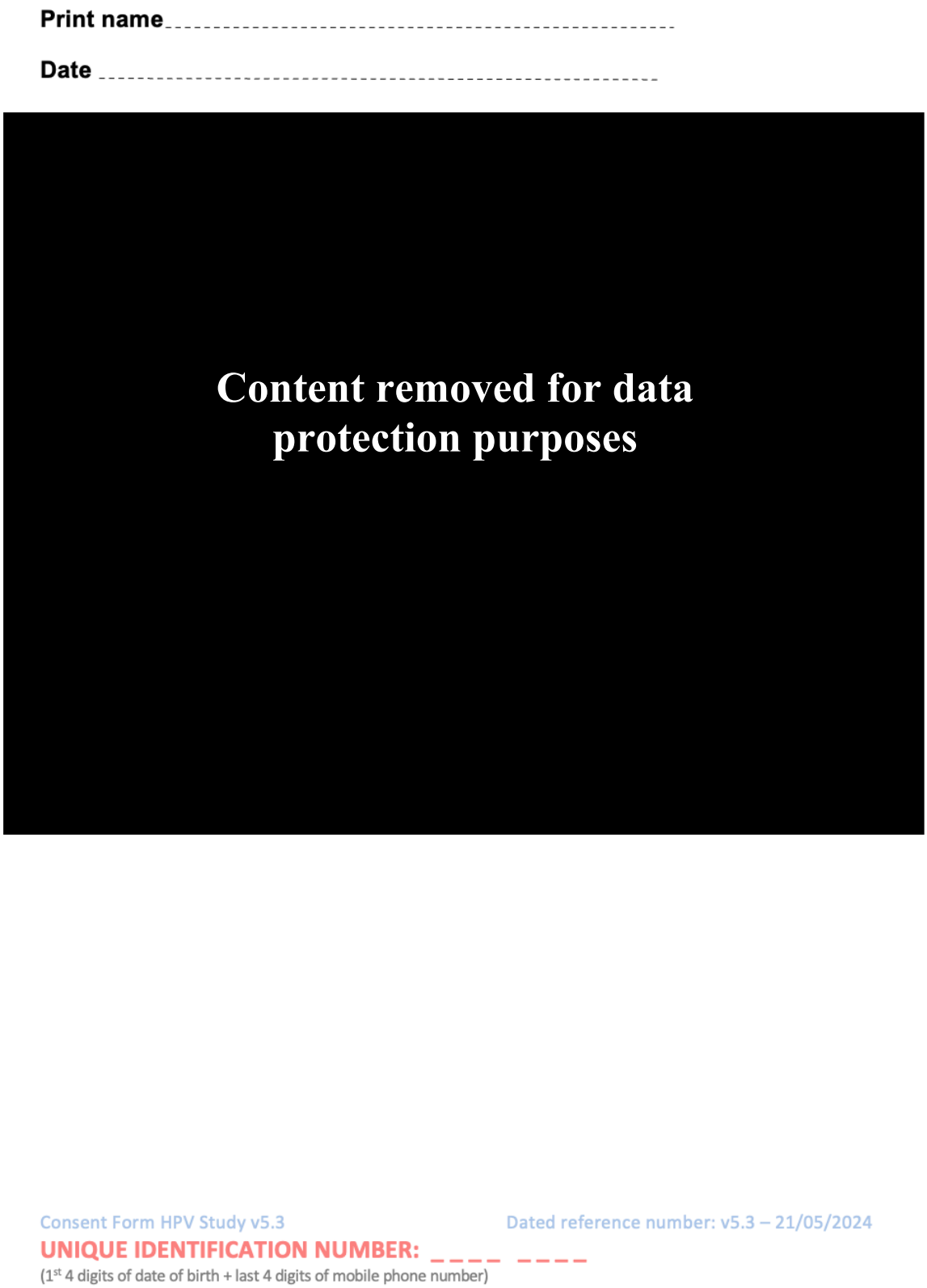


Figure 35. Consent form for HPV clinical study (v 5.3). Individuals wishing to participate in the HPV clinical study were required to sign this form to consent to the study.

UNIVERSITY OF DERBY

NHS University Hospitals of Derby and Burton NHS Foundation Trust

NHS The Shrewsbury and Telford Hospital NHS Trust

NHS University Hospitals of Leicester NHS Trust

NHS Nottingham University Hospitals NHS Trust

NHS Liverpool University Hospitals NHS Foundation Trust

I

Questionnaire for

Investigation into rates of Human Papillomavirus (HPV) infection in the mouth

This questionnaire is part of an investigation into determining the levels of HPV infection in adults. It will help the research team understand if particular lifestyle choices that people undertake can increase a person's risk of contracting this viral infection.

The questionnaire is confidential and pseudoanonymised, which means that your unique identification number is only used to ensure that your consent, each donated sample, and questionnaire can be linked together. Some of the questions are of a personal nature; please try to answer these questions in as truthful and open manner as possible.

How will we use information about you?

- We will need to use information from you, and from your medical records for this research project.
- This information will include your name and contact details, as well as the unique identifier (code number) you will generate as part of this consent process.
- People will use this information to do the research or to check your records to make sure that the research is being done properly.
- People who do not need to know who you are will not be able to see your name or contact details. Your data will have a code number instead.
- We will keep all information about you safe and secure.
- Once we have finished the study, we will keep some of the data so we can check the results. We will write our reports in a way that no-one can work out that you took part in the study.

[redacted]
Content removed for data protection purposes

UNIQUE IDENTIFICATION NUMBER: _ _ _ _ _ _ _ _

Question 1: **Please state your age in years here:** ……

Question 2: **Are you male or female?** Male ☐ Female ☐ Other ☐

Question 3: **What is your current relationship status?** *(Tick* **ALL** *that apply)*

Married and/or in long-term relationship (>1 yr) ☐

In a short-term relationship (<1yr) ☐

Single ☐

Divorced and/or Widowed ☐

Question 4: **What is your ethnic background & nationality?** …… (e.g. White British)

Question 5: **Do you currently smoke (cigarettes or otherwise) even if only occasionally?**

Yes ☐ **No** ☐

If no, please move to question 14

Question 6: **How often do you smoke cigarettes?** *(Tick* **ONE** *that applies)*

Daily	☐	Once a month	☐
3-5 times/week	☐	Infrequently (few times a year)	☐
1-2 times/week	☐	Never ***(move to question 9)***	☐
Few times a month	☐		

Question 7: **On average, how many cigarettes do you smoke in this timeframe (defined in question 6)?**

……

Question 8: **For how many years have you smoked cigarettes?**

Approx. ……Years

Question 9: **Do you smoke anything other than cigarettes?** *(e.g. cigars, pipes, cannabis, shisha and/or other)*

Yes ☐ **No** ☐

If no, please move to question 14

Question 10: **Please state what this is:** ……

Question 11: **How often do you smoke this?** *(Tick* **ONE** *that applies)*

Daily	☐	Few times a month	☐
3-5 times/week	☐	Once a month	☐
1-2 times/week	☐	Infrequently (few times a year)	☐

Question 12: **On average, how many of these do you smoke in this timeframe (defined in question 11)?**

……

Question 13: **For how many years have you smoked these items?**

Approx. ……Years

Question 14: **Have you previously smoked, or have your smoking behaviours changed from your current behaviours?**

Yes ☐ **No** ☐

If no, please move to question 23

Question 15: **How often did you smoke cigarettes?** *(Tick* **ONE** *that applies)*

Daily	☐	Once a month	☐
3-5 times/week	☐	Infrequently (few times a year)	☐
1-2 times/week	☐	Never *(please move to question 18)*	☐
Few times a month	☐		

Question 16: **On average, how many cigarettes did you smoke in this timeframe (defined in question 15)?**

……

Question 17: **For how many years did you smoke?** Approx. ……..Years

Question 18: **Did you smoke anything other than cigarettes?** *(e.g. cigars, pipes, cannabis, shisha and/or other)*

Yes ☐ **No** ☐

If no, please move to question 23

Question 19: **Please state what this is:** ……

Question 20: **How often did you smoke this?** *(Tick* **ONE** *that applies)*

Daily	☐	Few times a month	☐
3-5 times/week	☐	Once a month	☐
1-2 times/week	☐	Infrequently (few times a year)	☐

Question 21: **On average, how many of these did you smoke in this timeframe (defined in question 20)?**

……..

Question 22: **For how many years did you smoke these items?**

Approx. ……..Years

Question 23: **Do you currently drink alcohol?**

Yes ☐ **No** ☐

If no, please move to question 31

Question 24: How often do you currently drink alcohol? *(Tick* **ONE** *that applies)*

Daily ☐

3-5 times/week ☐

1-2 times/week ☐

Few times a month ☐

Once a month ☐

Few times a year ☐

Question 25: On average, how much alcohol do you currently drink at this frequency (defined in question 24)?
(E.g. A bottle of 9% red wine & 2 pints of cider)

……

Question 26: What type(s) of alcohol do you currently drink the most? (Tick **ALL** that apply)

Beer/Ale ☐ Wine ☐ Lager ☐ Neat Spirits ☐ Spirits & Mixer(s) ☐

Cider ☐ Alcopops (e.g. WKD) ☐ Other ☐ *Please state what*: ……

Question 27: Did you used to drink more heavily than you do now?

YES ☐ NO ☐

If no, please move to question 33

Question 28: How often did you drink alcohol? *(Tick* **ONE** *that applies)*

Daily	☐	Few times a month	☐
3-5 times/week	☐	Once a month	☐
1-2 times/week	☐	Few times a year	☐

Question 29: On average, how much did you use to drink at this frequency (defined in question 28)? *(E.g. A bottle of 9% red wine & 2 pints of cider)*

……..

Question 30: What type(s) of alcohol did you drink the most? *(Tick* **ALL** *that apply)*

Beer/Ale ☐ Wine ☐ Lager ☐ Neat Spirits ☐ Spirits & Mixer(s) ☐

Cider ☐ Alcopops (e.g. WKD) ☐ Other ☐ *Please state what*: ……..

Now please move to question 35

Question 31: If you are currently a non-drinker, did you used to drink alcohol?

Yes ☐ **No** ☐

If no, please move to question 35

Question 32: How often did you use to drink alcohol? *(Tick* **ONE** *that applies)*

Daily	☐	Few times a month	☐
3-5 times/week	☐	Once a month	☐
1-2 times/week	☐	Few times a year	☐

Question 33: On average, how much alcohol did you use to drink at this frequency (defined in question 32)?

(E.g. A bottle of 9% red wine & 2 pints of cider)

……..

Question 34: What type(s) of alcohol did you drink the most? *(Tick* **ALL** *that apply)*

Beer/Ale ☐ Wine ☐ Lager ☐ Neat Spirits ☐ Spirits & Mixer(s) ☐

Cider ☐ Alcopops (E.g. WKD) ☐ Other ☐ *Please state what*: ……..

The following questions are of a personal nature, but are essential for the successful completion of the aims of this investigation and as such, please answer each question truthfully. You may wish to ensure you answer these questions in an area that cannot be observed for your own privacy.

Question 35: Do you engage in open-mouth kissing (with and without tongue)**?**

Yes ☐ No ☐

Question 36: Are you? **Heterosexual** ☐ *(Tick* **ONE** *that applies)*

Homosexual ☐

Bisexual ☐

Other ☐ *Please give details: .....*

Question 37: Have you ever had sexual intercourse?

Yes ☐ No ☐

If no, please move to question 49

Question 38: Using the following scale, please state how often in the last year, have you used a condom? *(Tick* **ONE** *that applies)*

Always ☐ Mostly ☐ Sometimes ☐ Rarely ☐ Never ☐

Question 39: Please explain your answer to question 38

……

Question 40: Using the following scale, please state how many of your total number of sexual partners were one night stands? *(Tick* **ONE** *that applies)*

Always ☐ Mostly ☐ Sometimes ☐ Rarely ☐ Never ☐

Question 41: How many sexual partners have you had? *(Tick* **ONE** *that applies)*

1-5 ☐ 6-10 ☐ 11-20 ☐ 21-30 ☐ 31-40 ☐ 41-50 ☐ 50+ ☐

Question 42: **Are you currently in a monogamous relationship?**

Yes ☐ No ☐

If no, please move to question 44

Question 43: **How long have you been in this relationship for?** .....Years..... Months

Question 44: **Have you had sexual intercourse within the last year?** Yes ☐ No ☐

Question 45: **Which sexual activities have you performed/currently perform?** *(Tick* **ALL** *that apply)*

Vaginal sex ☐ Anal sex ☐ Oral sex ☐ Foreplay ☐ Masturbation ☐

Question 46: **Have you ever tested positive for a sexual transmitted infection (STI)?**

Yes ☐ No ☐

If no, please move to question 48

Question 47: **What type of STI were you diagnosed with?**

.....

Question 48: **Have you ever been diagnosed with an HPV infection?**

Yes ☐ No ☐

If no, please move to question 50

Question 49: **When were you diagnosed with an HPV infection?** .....

Question 50: **Have you been diagnosed with an HPV-related cancer and/or cancer precursor?**

Yes ☐ No ☐

If no, please move to question 52

Question 51: What sort of HPV-related cancer/cancer precursor were you diagnosed with (eg. CIN Stage II)?

……

Question 52: Have you had the HPV vaccination? Yes ☐ No ☐ Unsure ☐

If no, please move to the END

Question 53: Did you receive the HPV vaccination via:

National Health Service (NHS) ☐

Private UK healthcare ☐

Private overseas healthcare ☐

Public overseas healthcare ☐

-End of Questionnaire-

Thank You for Your Participation

Figure 36. Questionnaire for HPV clinical study (v 5.3). Consenting individuals to the HPV clinical study were invited to complete this questionnaire, that contained questions designed to understand each person's frequency in certain lifestyle characteristics, many of which are associated with an increased risk of oral HPV infection.

Appendix B: University of Derby Ethical Approval

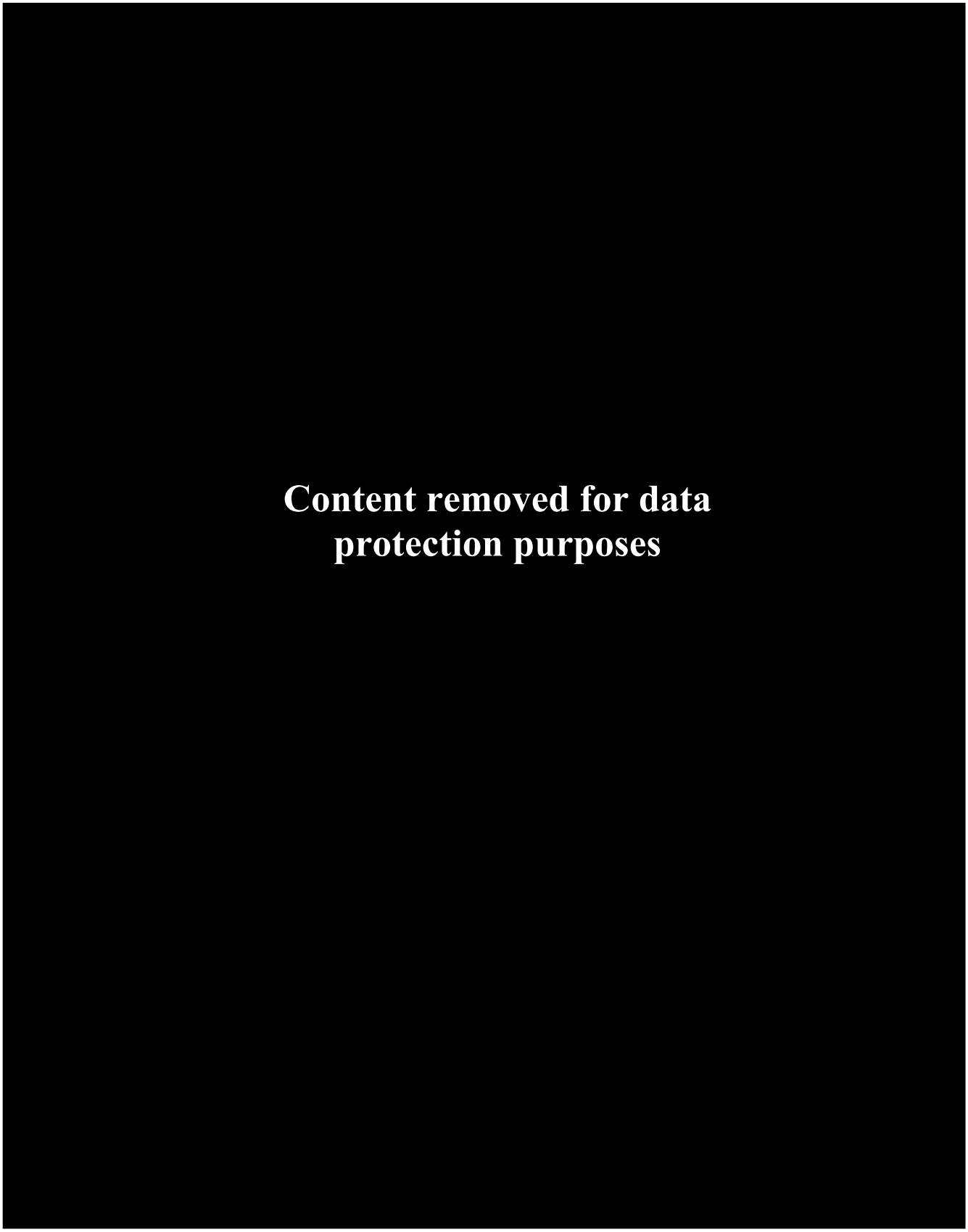


Figure 37. Ethical approval for this study from the University of Derby's College of Science and Engineering REC. Reference: ETH2021-2440

Content removed for data protection purposes

Content removed for data protection purposes

Content removed for data protection purposes

Content removed for data protection purposes

Content removed for data protection purposes

Content removed for data protection purposes

Content removed for data protection purposes

Content removed for data protection purposes

Content removed for data protection purposes

Content removed for data protection purposes

Content removed for data protection purposes

Figure 38. Ethics application for this study from the University of Derby's College of Science and Engineering REC. Reference: ETH2021-2440

Appendix C: HBRC documentation

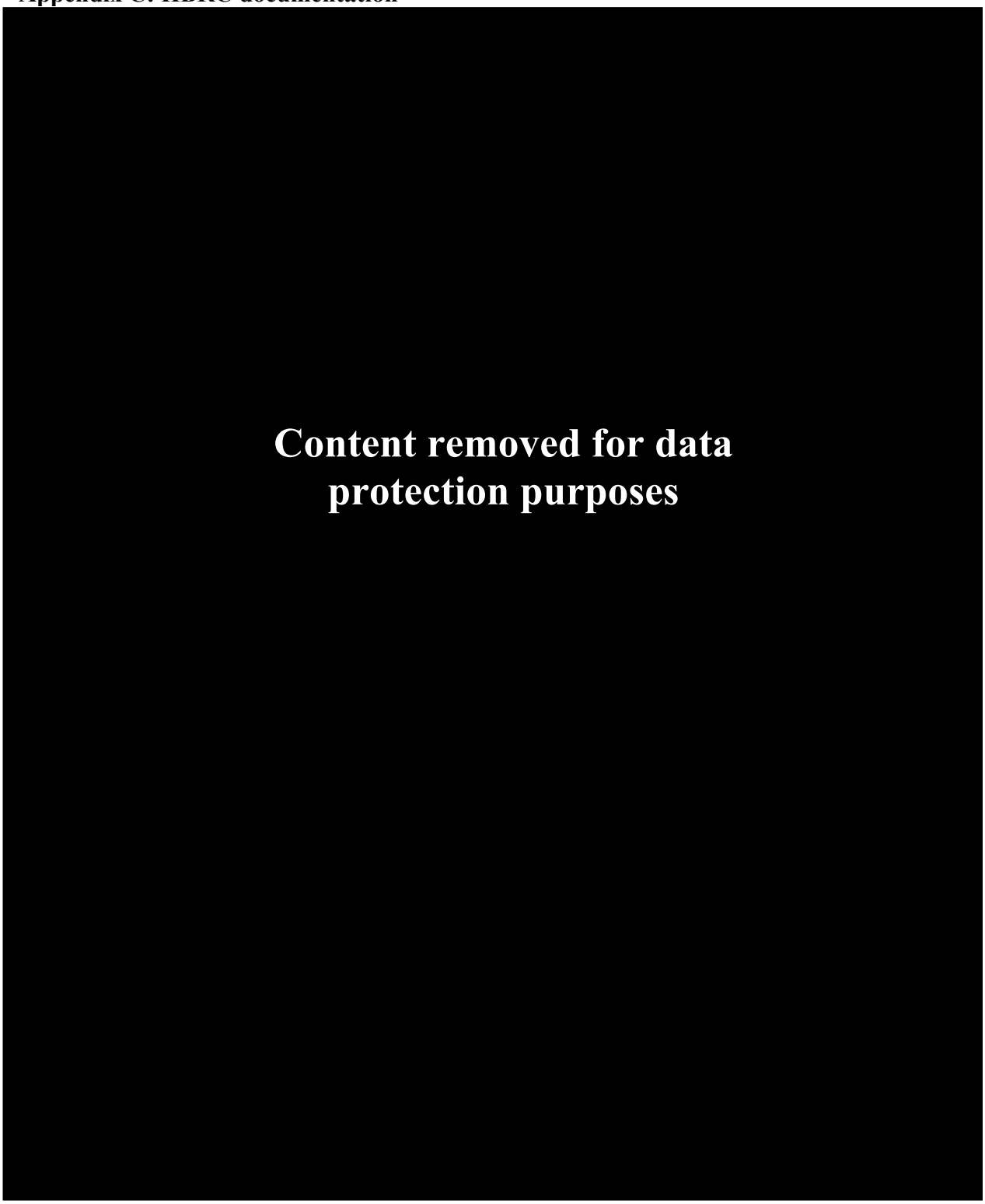


Figure 39. Confirmation of application to purchase HBRC biomaterials. Application number: 21-734. This application refers to HNBR non-cancerous palatine tonsil tissues used in this study.

Content removed for data protection purposes

Content removed for data protection purposes

Content removed for data protection purposes

Figure 40. HRA approval from the REC for the HBRC. Rec reference number: 20/NW/0001.

APPENDIX D: A modified qPCR assay to detect HPV-DNA in non-cancerous tonsil tissue.

D.0. Introduction

There are several consensus primer pairs that have been designed for the detection of multiple HPV viral types by PCR, including MY09/MY11 (Husnjak et al., 2000), GP5+/GP6+ (de Roda Husman et al., 1995), and FAP59/FAP64 (Forslund et al., 1999). They all target the late viral gene, L1, of the HPV genome. However, they differ on their expected amplicon size and slightly in their efficiency of binding to certain viral types. Importantly for this project, they can all detect and amplify a broad range of viral types. Here, we aimed to develop a reproducible PCR assay that could both accurately and stringently identify the presence of HPV within non-cancerous tonsil tissue.

Throughout the literature, these primers sets have often been used in a standard PCR assay that is followed by gel electrophoresis for product analysis and hybridisation for viral typing (Snijders et al., 1990, Qu et al., 1997, Erhart et al., 2016). This multi-step approach for HPV-DNA detection is low throughput when analysing high numbers of samples, and lacks sensitivity. Previously in our laboratory a qualitative qPCR assay that made use of the MY09/MY11 primers was developed to detect HPV within DNA extracted from oral swabs with a melt curve analysis of PCR products used to determine HPV-status (Whitton, 2020, Whitton et al., 2024). We recapitulated this principle, testing various HPV L1 primer pairs and qPCR master mixes to modify the assay to enable identification of HPV within tonsil DNA.

We have therefore developed a qPCR assay for the specific detection of HPV-DNA by L1 amplification within human tonsil tissue. The approach uses SYBR green chemistry followed by melt curve analysis to categorise HPV-positive and HPV-negative samples. This appendices chapter describes the optimisation process of this assay and the rationale for the selection of master mix, consensus primer pair, and the final PCR cycling parameters used.

D.1. Comparison of the suitability of HPV L1 consensus primer pairs for HPV detection in tonsillar gDNA

The use of consensus primers in PCR is a common method of HPV-DNA detection that we selected as the basis for our assay development as it would allow a high throughput of testing multiple gDNA samples. Moreover, this approach is cost-effective in enabling testing for multiple HPV viral types in one single assay, rather than several screens required for other methods (e.g., type-specific HPV primers). Our first aim was to develop a modified qPCR assay that could accurately define the HPV-status of the tonsil tissues of our prospective cohort to permit further pathogenesis analyses.

In the first instance, we tested primer pairs that are frequently used throughout the literature: MY09/MY11 (Husnjak et al., 2000), GP5+/GP6+ (de Roda Husman et al., 1995), and FAP59/FAP64 (Forslund et al., 1999). Primer sequences and their expected amplicon size are outlined in Table 20. HPV-18 pGEMII-WT plasmid (a kind gift from Dr Sally Roberts, University of Birmingham) and DNA extracted from HPV-16 positive SiHa cells (Friedl et al., 1970) were selected as the positive controls to be used routinely throughout this assay development. PowerTrack™ SYBR™ Green [ThermoFisher] and GoTaq® qPCR [Promega] master mixes were tested with each of the L1 consensus primer pairs. Other qPCR master mixes were trialled, but these data are not shown.

Each reaction for all the master mix and primer pair combination was set up to a total volume of 20 µL as follows: 1X qPCR master mix, 200 nM forward primer, 200 nM reverse primer, 2.5 ng/µL gDNA, and nuclease-free water to the total volume. Five gDNA samples determined as HPV-positive in previous work were used for this optimisation process. All reactions were thermocycled for an initial denaturation of 3 minutes at 95°C, then 40 cycles of 20 seconds at 95°C, 30 seconds at 53°C (MY09/MY11), 48°C (GP5+/GP6+), 50°C (FAP59/FAP64) and then an extension step at 72°C (10 seconds). A melt curve stage followed: 95°C - 15 seconds, 60°C – 1 minute increasing up to 95°C, with data collected every 0.3°C. Melt curve analysis was used to assess any PCR product generated in each reaction.

Firstly, we scrutinised the accuracy of MY09/MY11 primers with both PowerTrack™ SYBR™ Green [ThermoFisher] and GoTaq® qPCR [Promega] master mixes for HPV-DNA detection. These primers can amplify an ~450 base pair (bp) region of the HPV L1 gene

(Husnjak et al., 2000). With both master mixes, a peak with a melting temperature of ~81°C indicated the amplification of the HPV L1 gene in both HPV controls (HPV-18 plasmid and SiHa DNA) using MY09/MY11 primers. However, there were no PCR products whose melt curve aligned with HPV controls in the 5 known HPV-positive samples, indicating no L1 amplification (Figure 41). In addition, the presence of non-specific binding and large primer dimers (particularly present in the amplification products of the GoTaq® master mix [Figure 41B]) showed that at these initial cycling conditions MY09/MY11 primers were not appropriate for our samples.

The FAP59/FAP64 primer pair were assessed concurrently with each of the master mixes from ThermoFisher and Promega. FAP59/FAP64 primers target an ~480 bp region of the HPV L1 gene (Forslund et al., 1999). FAP59/FAP64 primers successfully amplified HPV-18 plasmid, indicated by a peak with a Tm of ~84°C in the melt curve plot; however, regardless of master mix used, the primers were unable to amplify L1 gene in DNA from HPV-16 positive SiHa cells (Figure 42). As a result, at these cycling parameters FAP59/FAP64 primers were considered not suitable for our HPV-DNA qPCR assay as the HPV controls were imperative for confidence that our method could define the unknown HPV-status of tonsillar gDNA samples.

Similarly, we assessed GP5+/GP6+ primers. These target a smaller region of the L1 viral gene of approximately 150 bp (de Roda Husman et al., 1995). HPV controls were amplified with GP5+/GP6+ and each master mix. The use of PowerTrack™ SYBR™ Green qPCR master mix [ThermoFisher] resulted in a different melt curve profile of HPV controls; the PCR products of the HPV-18 plasmid dissociated at a Tm of ~79°C, whereas HPV-16 positive SiHa DNA PCR products did so at a Tm of 77°C (Figure 43A), demonstrating an increased sensitivity to differentiate between the two L1 gene sequences. Conversely, the melt curves for both controls with the GoTaq® qPCR master mix both had a peak at a Tm of ~79°C (Figure 43B). HPV L1 gene was also detected in some of our known HPV-positive tonsillar gDNA samples using both master mixes; however, the combination of GP5+/GP6+ and the ThermoFisher master mix resulted in distinct product peaks with less non-specific binding than those of the GoTaq® master mix. To confirm this specificity, the PowerTrack™ SYBR™ Green master mix PCR products were analysed by gel electrophoresis (Figure 44). Here, a band at 150 bp confirmed successful amplification of our target gene in the HPV controls. Despite a ‘positive’ peak being

present in the melt curve analysis of all 5 HPV-positive gDNA samples, only 2 had a visible band at 150 bp when electrophoresed suggesting that differentiating HPV-status by melt curve analysis may allow a more sensitive detection of a low-level HPV infection.

Taken together, these data suggested that GP5+/GP6+ primers used with the PowerTrack™ SYBR™ Green qPCR master mix [ThermoFisher] were most suited to tonsillar gDNA samples, and were taken forward for further methodology optimisation.

Table 20. HPV L1 gene consensus primers. The name of each primer with their sequence and expected amplicon size.

Primer name	Sequence (5' → 3')	Amplicon size (bp)
MY09 (F)	CGT CCM ARR GGA WAC TGA TC	450
MY11 (R)	GCM CAG GGW CAT AAY AAT GG	450
GP5+ (F)	GTT TGT TAC TGT GGT AGA TAC	150
GP6+ (R)	GAA AAA TAA ACT GTA AAT CAT ATT CC	150
FAP59 (F)	TAA CWG T5G G5C AYC CWT ATT	480
FAP64 (R)	CCW ATA TCW VHC AT5 TC5 CCA TC	480

(F) = forward primer, (R) = reverse primer

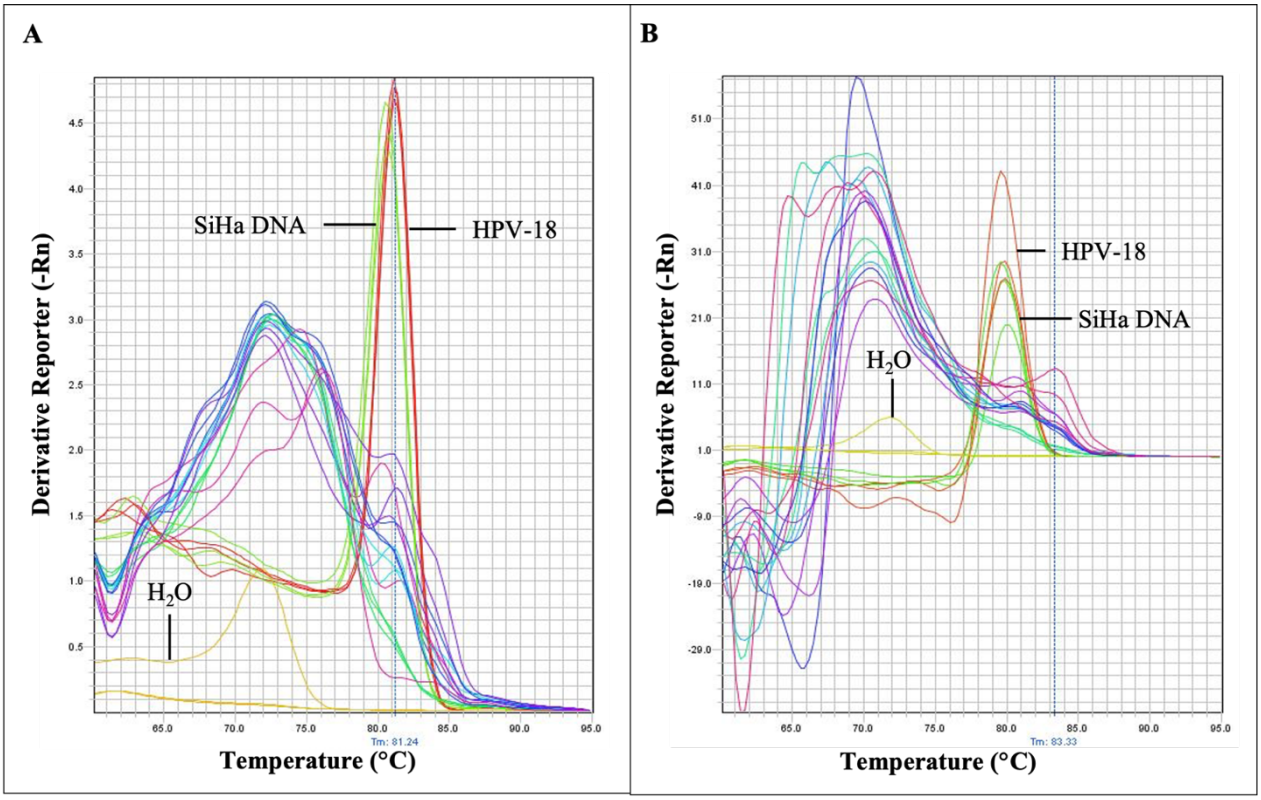


Figure 41. Melt curve analysis of products generated by qPCR using L1 consensus primers, MY09/MY11. PCR was performed using the AB® StepOnePlus™ Real-Time PCR system. Data was analysed in the corresponding software, version 2.2.2. Temperature (°C) is displayed against derivative reporter values (-Rn). The melt curves of qPCR products generated from HPV controls (HPV-18 plasmid [red] and SiHa DNA [green]), and 5 tonsil gDNA samples (believed to be HPV-positive from previous work) are displayed. **(A)** ThermoFisher PowerTrack™ SYBR™ qPCR master mix. No HPV-DNA was amplified in tonsil gDNA. **(B)** Promega GoTaq® qPCR master mix. No HPV-DNA was amplified in tonsil gDNA.

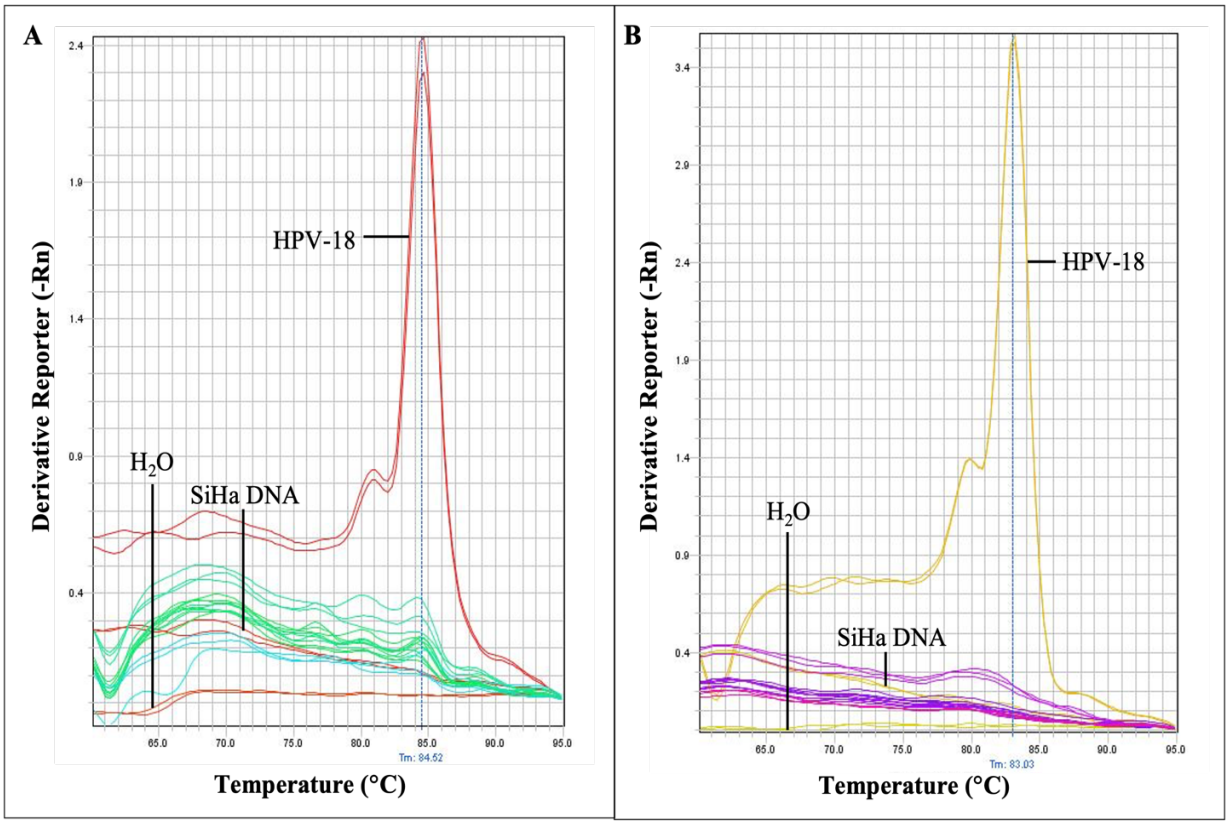


Figure 42. Melt curve analysis of products generated by qPCR using L1 consensus primers, FAP59/FAP64. PCR was performed using the AB® StepOnePlus™ Real-Time PCR system. Data was analysed in the corresponding software, version 2.2.2. Temperature (°C) is displayed against derivative report values (-Rn). The melt curves of qPCR products generated from HPV controls and 5 tonsil gDNA samples (believed to be HPV-positive from previous work) are shown. (A) ThermoFisher PowerTrack™ SYBR™ qPCR master mix. No amplification of HPV in tonsillar gDNA samples [green] or SiHa DNA [red]. (B) Promega GoTaq® qPCR mix. No amplification of HPV in tonsillar gDNA samples [purple] or HPV-16 positive SiHa DNA [yellow].

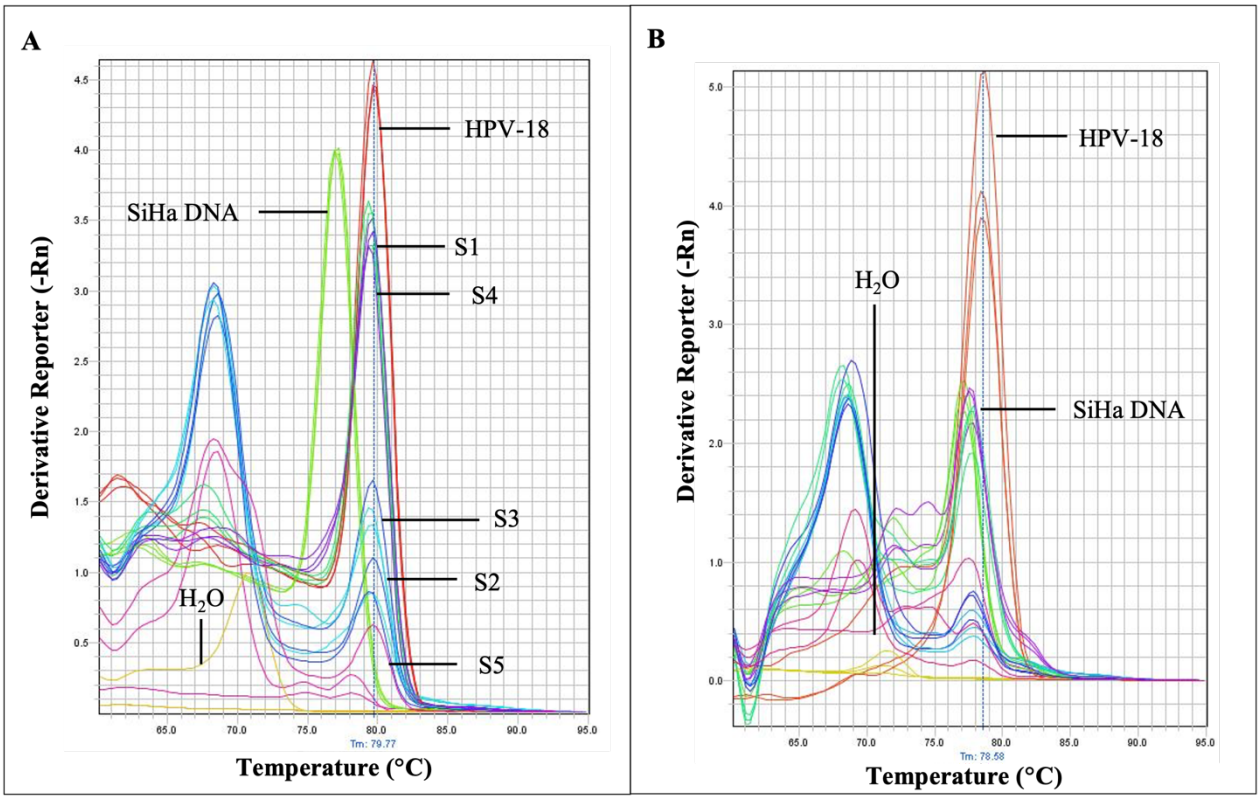


Figure 43. Melt curve analysis of products generated by qPCR using L1 consensus primers, GP5+/GP6+. PCR was performed using the AB® StepOnePlus™ Real-Time PCR system. Data was analysed in the corresponding software, version 2.2.2. Temperature (°C) is displayed against derivative report values (-Rn). The melt curves of qPCR products generated from HPV controls (HPV-18 plasmid [red] and SiHa DNA [green]), and 5 tonsil gDNA samples (believed to be HPV-positive from previous work) are shown. (A) ThermoFisher PowerTrack™ SYBR™ qPCR master mix. HPV-DNA was detected in the tonsil gDNA samples analysed. (B) Promega GoTaq® qPCR mix. Small peaks indicative of HPV L1 amplification were present in some tonsil gDNA samples.

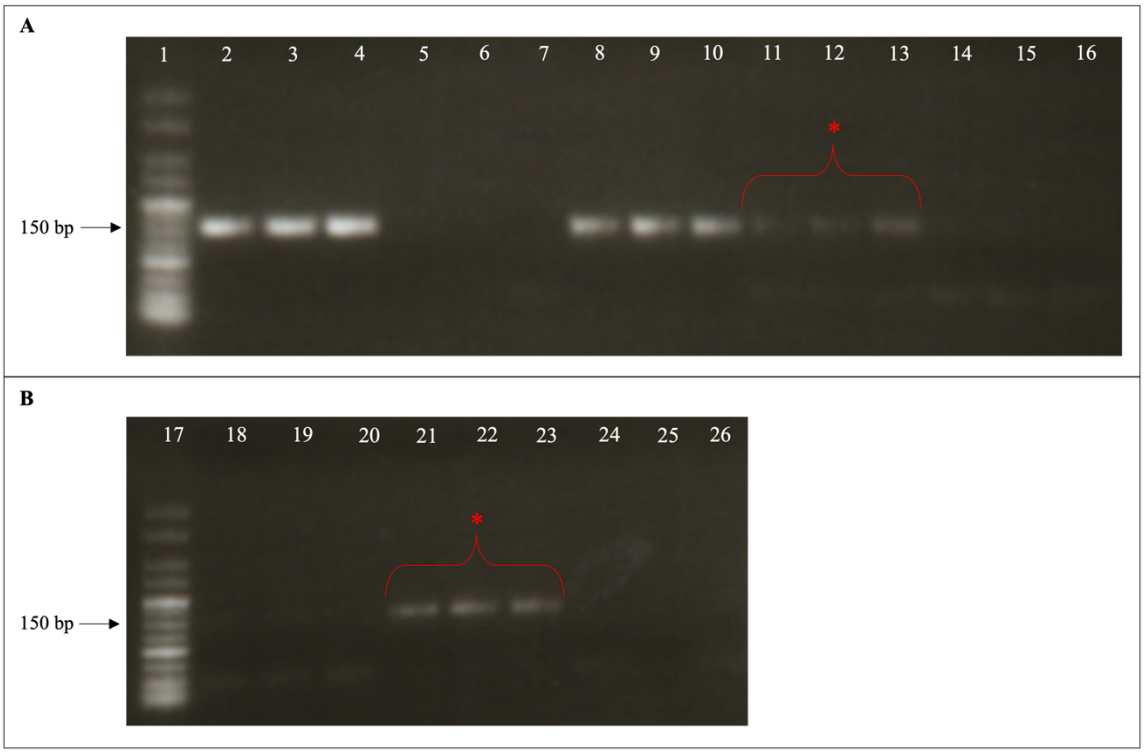


Figure 44. Analysis of qPCR products by gel electrophoresis. Amplicons produced using GP5+/GP6+ primers and ThermoFisher's PowerTrack™ SYBR™ qPCR Master Mix were electrophoresed through 2% agarose gels, against a 25 bp DNA ladder [Bioline]. A band at approximately 150 bp was used to indicate amplification of the HPV L1 gene by GP5+/GP6+ consensus primers (positive controls: HPV-18 plasmid in lanes 2-4 and SiHa DNA in lanes 8-10; negative control: lanes 5-7). A 150 bp band was present in HPV-positive tonsil DNA sample 1 (lanes 11-13) and sample 4 (21-23), indicated by an asterisk. Gel images were taken using a Geneflash Syngene Bio Imaging Gel Doc System + UV/White Transilluminator [Scientific Laboratory Supplies]

D.2. The development of an assay to detect HPV-DNA in non-cancerous tonsil tissue using GP5+/GP6+ primers and PowerTrack™ SYBR™ Green Master Mix [ThermoFisher]

After selecting a master mix (PowerTrack™ SYBR™ Green [ThermoFisher]) and an L1 consensus primer pair (GP5+/GP6+), we optimised the concentration of PCR components and cycling conditions so that the assay was stringent enough to detect a genuine HPV infection without causing false identification of HPV-positive samples by introducing non-specific binding. For this stage of the optimisation five different HPV-positive tonsillar gDNA samples were selected and used to refine the optimal gDNA and primer concentration, and annealing temperature for our sample type.

This qPCR assay is a qualitative assessment of HPV amplification, rather than a quantitative approach and was designed to be a 'yes/no' for viral DNA detection in tonsil samples. We therefore estimated the amount of PCR product produced by the level of fluorescence indicated by the derivative reporter values (-Rn) in the melt curve analysis. To assess how altering parameters such as the final gDNA concentration per reaction was affecting HPV detection by qPCR, we defined a -Rn value of >2.5 as strong positive, ≤2.5 - ≥1.5 as positive, and <1.5 - ≤0.5 as weak positive.

D.2.1. gDNA concentration optimisation

Firstly, we looked to select the most appropriate gDNA concentration for our HPV-DNA qPCR assay. Too much DNA added can 'overload' the reaction; however, we were aware that HPV-DNA in non-cancerous tissues may be present at a very low copy number and difficult to detect. To determine the optimal gDNA concentration we set up a gradient ranging from a final DNA concentration per reaction of 0.25 – 2.5 ng/µL within 20 µL reactions consisting of 1X PowerTrack™ SYBR qPCR master mix [ThermoFisher], 200 nM GP5+, 200 nM GP6+. All reactions were subjected to the following cycling conditions: an initial denaturation at 95°C (3 minutes), followed by 40 cycles of: 95°C – 20 seconds, 48°C – 30 seconds, 72°C – 10 seconds, followed by a melt curve.

Melt curve analysis of qPCR products revealed that when a final tonsil gDNA concentration of 2.0 ng/µL or 2.5 ng/µL was added per 20 µL reaction, all 5 HPV-positive samples displayed strong peaks in their melt curve plot that aligned with the HPV-18 plasmid control. The extent

of HPV amplification as indicated by -Rn value was the same in both these gDNA concentrations with 2/5 samples defined as 'strong positive' (-Rn = >2.5) and 3/5 samples as 'weak positive' (-Rn = <1.5 - ≤0.5) (Table 21). Curiously, at lower gDNA concentrations, we were frequently seeing a shift in the melt curve analysis from a peak aligning with HPV-18 plasmid to one that aligned with SiHa DNA (Figures 45A and 45C). We speculated that by altering the gDNA concentration, we were possibly detecting a dual HPV infection of more than one HPV viral type; however Sanger sequencing of these products revealed that this shift to HPV-16 was due to non-specific binding. Indeed, searching the FASTA sequences of the PCR products generated with 0.25 ng/μL gDNA in NCBI BLAST showed we were detecting a gene on human chromosome 10 (Figure 45D), but that the original alignment of melt curve peaks to HPV-18 plasmid at higher gDNA concentrations was specific amplification of HPV L1 (viral type 18) (Figure 45A-B).

We concluded that lowering the gDNA concentration was decreasing the specificity of the GP5+/GP6+ primers. Indeed, when adding 0.5 ng/μL and 0.25 ng/μL total gDNA per reaction, none of the known HPV-positive samples showed amplification of viral gene, L1. The observed shift to a non-specific binding peak, assumed to be chromosome 10, was even present within 1/5 HPV-positive samples when a final DNA concentration of 1.5 ng/μL was used (Table 21). We therefore decided to continue our assay optimisation with final gDNA concentrations of 2.0 or 2.5 ng/μL.

Table 21. Analysis of PCR amplification of known HPV-positive tonsil gDNA samples and melt curve peak alignment with HPV controls, across a DNA concentration gradient. The amount of product generated by qPCR cycling of each sample at all gDNA concentrations trialled is described as strong positive (-Rn = >2.5), positive (-Rn = ≤2.5 - ≥1.5), or weak positive (-Rn = <1.5 - ≤0.5). A green colour gradient demonstrates the level of signal detected by analysis of PCR amplification in the melt curve plot, with darker green a higher -Rn value and lighter green for a weaker signal detected. For alignment control, the cell colour is used to demonstrate whether the alignment is due to non-specific binding [red] or genuine L1 HPV gene amplification [green]. The total number (out of 5) of samples displaying a genuine HPV positive peak in their melt curve analysis is shown. At each DNA concentration, all HPV controls amplified as expected.

			DNA concentration (ng/µL)					
			2.5	2.0	1.5	1.0	0.5	0.25
Sample	TD1R3	Peak size in melt curve plot (-Rn value)	Strong Positive	Strong Positive	Strong Positive	Strong Positive	Strong Positive	Strong Positive
		Alignment with control	HPV-18	HPV-18	HPV-18 & non-specific	Non-specific	Non-specific	Non-specific
	TD1L3	Peak size in melt curve plot (-Rn value)	Strong Positive	Strong Positive	Positive	Positive	Positive	Positive
		Alignment with control	HPV-18	HPV-18	HPV-18	HPV-18	Non-specific	Non-specific
	TD1L4	Peak size in melt curve plot (-Rn value)	Weak Positive	Weak Positive	Weak Positive	Weak Positive	Strong Positive	Strong Positive
		Alignment with control	HPV-18	HPV-18	HPV-18	HPV-18	Non-specific	Non-specific
	TD1L7	Peak size in melt curve plot (-Rn value)	Weak Positive	Weak Positive	Weak Positive	Positive	Strong Positive	Strong Positive
		Alignment with control	HPV-18	HPV-18	HPV-18	Non-specific	Non-specific	Non-specific
	TD1L8	Peak size in melt curve plot (-Rn value)	Weak Positive	Weak Positive	Weak Positive	Positive	Positive	Positive
		Alignment with control	HPV-18	HPV-18	HPV-18	Non-specific	Non-specific	Non-specific
	Control	HPV-18	Positive	Positive	Positive	Positive	Positive	Positive
		SiHa DNA	Positive	Positive	Positive	Positive	Positive	Positive
		Total gDNA HPV-positive	5	5	4	2	0	0

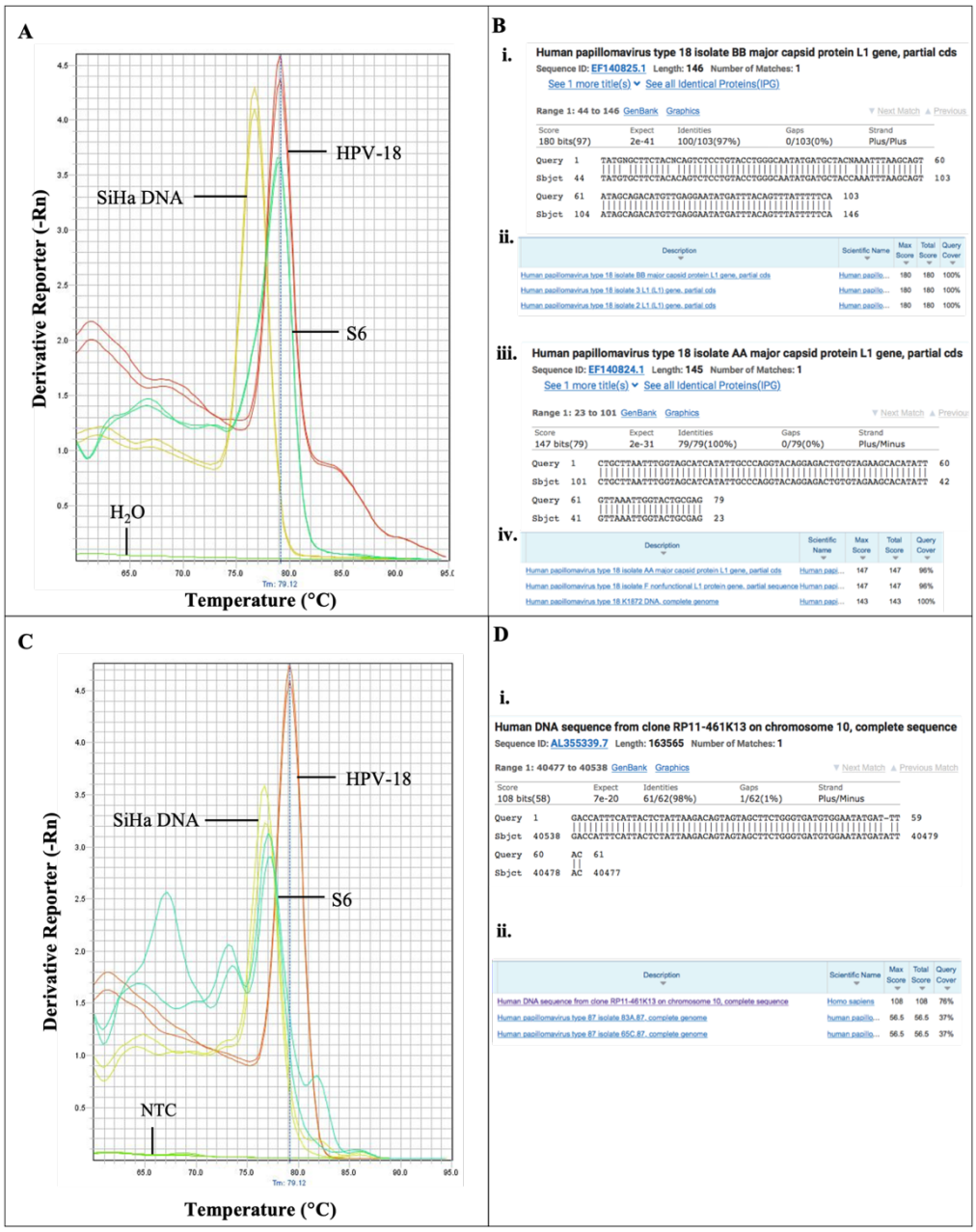


Figure 45. Shift in melt curve alignment of tonsillar gDNA amplification products at lower final gDNA concentration; melt curve analysis and corresponding Sanger sequencing data. Melt curve data was generated using the AB® StepOnePlus™ Real-Time PCR system, and data analysed in their software (v 2.2.2). Temperature (°C) is displayed against derivative reporter values (-Rn). Sanger sequencing data was searched using NCBI Blast database. (A, C): Melt curve of qPCR products generated from controls HPV-18 [red] and SiHa DNA [yellow] are displayed. A representative tonsillar gDNA sample (S6) is displayed in green. (A): 2.5 ng/µL final DNA concentration; S6 peak aligned with HPV-18 control. (B): NCBI BLAST search results of the S6 amplicon Sanger sequencing data from panel A; (i) GP5+ primer, (ii) GP6+ primer, confirming HPV-18 amplification. (C): 0.25ng/µL final DNA concentration. S6 product aligned with SiHa DNA (HPV-16 control). (D): NCBI BLAST search results of the S6 amplicon from panel C; (i) GP5+ primer, (ii) GP6+ primer; showing non-specific amplification of chromosome 10.

D.2.2. GP5+/GP6+ primer concentration optimisation

We next scrutinised the primer concentration and, using the same five HPV-positive gDNA samples, determined the ideal concentration of GP5+/GP6+ to use in our assay. A final concentration of 200 nM of each forward and reverse primers had been used in all the initial master mix and gDNA concentration optimisation steps based on the reports of GP5+/GP6+ concentration described previously, and the recommended primer concentration for use with PowerTrack™ SYBR qPCR master mix (Fuessel Haws et al., 2004, ThermoFisher Scientific, 2022a). Given that 200nM GP5+ and 200 nM GP6+ was successful in combination with the two proposed final tonsil gDNA concentrations, we selected a range of similar primer concentrations to trial. In each instance, melt curve plots were compared to results generated using the original concentration of 200 nM of each primer. To determine the optimal primer concentration with the 2 potential gDNA concentrations we set up a concentration gradient ranging from 250 nM to 100 nM of each primer, with either 2.0 ng/µL or 2.5 ng/µL gDNA in 20 µL reactions with 1X PowerTrack™ SYBR qPCR master mix [ThermoFisher]. All reactions were subject to a cycling program as described at 3.2.1, followed by a melt curve stage.

Melt curve analysis revealed that a higher concentration of each primer (250 nM rather than 200 nM) did not particularly alter the specificity of the assay to detect HPV-DNA in the five known HPV-positive tonsil gDNA samples (Table 22). At a primer concentration of 250 nM, the stringency of the assay was compromised in one instance to align with SiHa DNA, which, given this shift, we again believed to be non-specific binding (Table 22). Similarly, when considering gDNA concentration at 200 nM primer, 2.0 ng/µL gDNA appeared to be less efficient, with all tonsil gDNA samples being considered ‘weakly positive’ for HPV, compared to 2/5 ‘strongly positive’ when 2.5 ng/µL gDNA was added to each reaction. 2.5 ng/µL gDNA was therefore chosen as the optimised gDNA concentration for this HPV screen.

Interestingly, lowering the primer concentration to 150 nM or 100 nM of each GP5+ and GP6+ primer resulted in the loss of amplification of the SiHa DNA control (Figure 46). Similarly to HPV controls at 100 nM primer, the amplification of an otherwise ‘strong positive’ sample was also lost. The amplification of products visualised through melt curve peaks of our HPV controls is essential to subsequently define the HPV-status of unknown tonsillar gDNA samples, rendering a primer concentration 150 nM and below unsuitable for this assay. Taken

together, we selected a primer concentration of 200 nM rather than 250 nM of each forward and reverse GP5+/GP6+ consensus primers to be used in our HPV-DNA qPCR assay, as a result of the compromise in L1 amplification in 1/5 sample with 250 nM.

Table 22. Analysis of PCR amplification of known HPV-positive tonsil gDNA samples and melt curve peak analysis across a GP5+/GP6+ concentration gradient, with two gDNA concentrations. The amount of PCR product amplified by qPCR of each tonsil DNA sample at all primer concentrations trialled at either 2.5 or 2.0 ng/µL gDNA per reaction, is described as strong positive (-Rn = >2.5), positive (-Rn = ≤2.5 - ≥1.5), or weak positive (-Rn = <1.5 - ≤0.5). A green colour gradient demonstrates the level of signal detected by melt curve analysis, with darker green more HPV-positive and lighter green for a weaker signal detected. Reactions with no products detected are considered negative with no control alignment (indicated by N/A). For alignment with HPV control, the cell colour is used to demonstrate whether the alignment is due to non-specific binding [red] or genuine L1 HPV gene amplification [green]. The total number (out of 5) of samples displaying a genuine HPV positive peak in their melt curve analysis is shown.

			GP5+/GP6+ Primer concentration (nM)							
			250 F/250 R		200 F/200 R		150 F/150 R		100 F/100 R	
			2.0	2.5	2.0	2.5	2.0	2.5	2.0	2.5
Sample	TD1R3	Peak size in melt curve plot (-Rn value)	Strong Positive	Strong Positive	Weak Positive	Strong Positive	Negative	Negative	Negative	Negative
		Alignment with control	HPV-18	HPV-18	HPV-18	HPV-18	N/A	N/A	N/A	N/A
	TD1L3	Peak size in melt curve plot (-Rn value)	Strong Positive	Strong Positive	Weak Positive	Strong Positive	Negative	Positive	Negative	Negative
		Alignment with control	HPV-18	HPV-18	HPV-18	HPV-18	N/A	HPV-18	N/A	N/A
	TD1L4	Peak size in melt curve plot (-Rn value)	Positive	Weak Positive	Weak Positive	Weak Positive	Negative	Negative	Positive	Weak Positive
		Alignment with control	HPV-18	HPV-18	HPV-18	HPV-18	N/A	N/A	HPV-18	HPV-18
	TD1L7	Peak size in melt curve plot (-Rn value)	Strong Positive	Weak Positive	Weak Positive	Weak Positive	Strong Positive	Strong Positive	Weak Positive	Weak Positive
		Alignment with control	HPV-16	HPV-18	HPV-18	HPV-18	Non-specific*	Non-specific*	HPV-18	HPV-18
	TD1L8	Peak size in melt curve plot (-Rn value)	Strong Positive	Positive	Weak Positive	Weak Positive	Strong Positive	Strong Positive	Strong Positive	Weak Positive
		Alignment with control	Non-specific	HPV-18	HPV-18	HPV-18	Non-specific*	HPV-18	HPV-18	HPV-18
	Control	HPV-18	Positive	Positive	Positive	Positive	Positive	Positive	Positive	Positive
		SiHa DNA	Positive	Positive	Positive	Positive	Negative	Negative	Negative	Negative
		Total gDNA HPV-positive	4	5	5	5	0	2	3	3

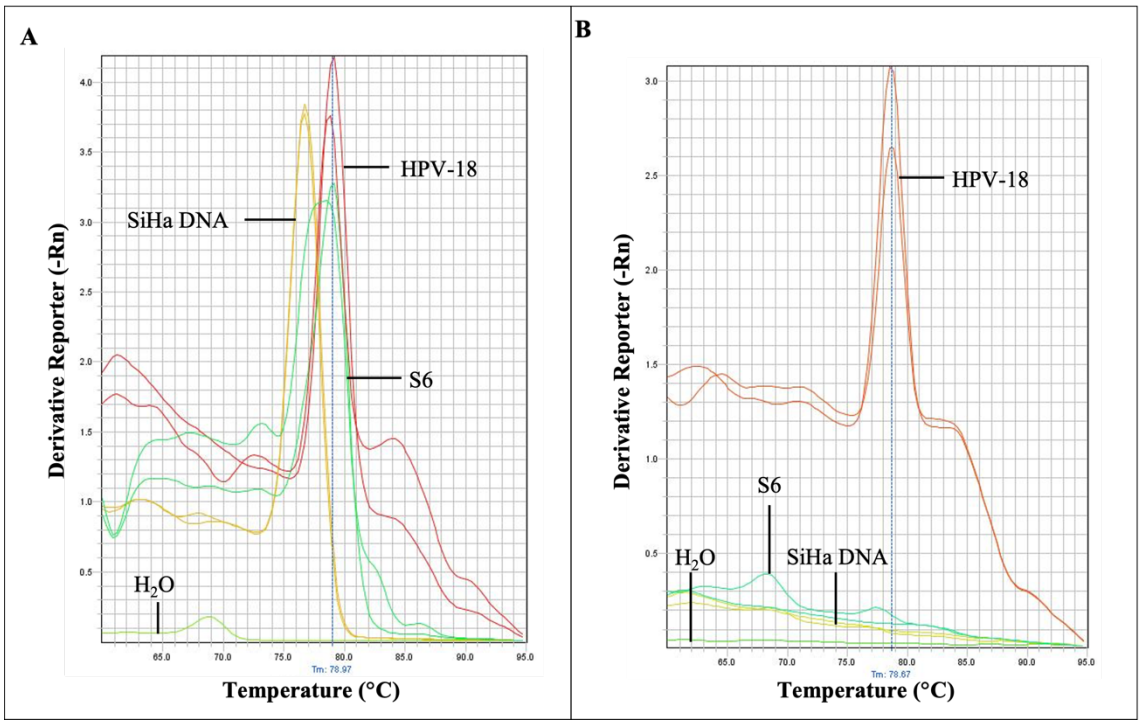


Figure 46. Loss of amplification products for SiHa DNA control with reduced GP5+/GP6+ primer concentration. Melt curve data was generated using the AB® StepOnePlus™ Real-Time PCR system, and data analysed in their software (v 2.2.2). Temperature (°C) is displayed against derivative reporter values (-Rn). Melt curve of qPCR products generated from controls HPV-18 [red] and SiHa DNA [yellow] are displayed. A representative tonsillar gDNA samples (S6; TD1R3) is displayed in turquoise. (A) 250 nM GP5+ and 250 nM GP6+ added to the final reaction volume; amplification of all samples (including HPV controls), tonsillar gDNA S6 peak aligned with HPV-18 control. (B) 100 nM GP5+ and 100 nM GP6+ added to the final reaction volume; amplification of HPV-18 only, with loss of products for SiHa DNA and S6.

D.2.3. Annealing temperature gradient

The final consideration of our HPV-DNA qPCR screen modification process was to select an appropriate annealing temperature compatible with the chosen primer pair and, suitable for the concentrations of primer and gDNA already optimised. In these earlier development stages we had used an annealing temperature of 48°C. To determine the optimal annealing temperature we examined the melt curves generated across a large annealing temperature gradient (ranging from 40-52°C). For this trial, the final composition of all reactions was: 1X PowerTrack™ SYBR qPCR master mix [ThermoFisher], 200 nM GP5+, 200 nM GP6+, 2.5 ng/μL gDNA. Nuclease free H_2O was added to a total volume of 20 μL. The PCR cycling conditions were as described above (D.2.2); however, the 30 second annealing period trialled the temperature range of 40-52°C.

Melt curve analysis of the five HPV-positive samples used in both gDNA and primer concentration optimisation showed successful amplification of HPV L1 in all samples using annealing temperatures of 52, 50, and 48°C. At any lower annealing temperatures (than 48°C) tested in our gradient analysis, we again observed a shift to a product melting peak aligning with HPV-16 positive SiHa DNA in at least some of the known HPV-positive samples. At 46°C and below, this shift was seen in all tonsillar DNA samples (Table 23). An example of this shift can be seen in tonsil gDNA sample 'S6' whose PCR products align with HPV-18 plasmid at 52°C [Figure 47A], but shift to a lower Tm and alignment with SiHa DNA control at 44°C [Figure 47C]. Sanger sequencing of PCR products revealed that, as before, alignment with HPV-18 plasmid at 52°C was true HPV-18 L1 amplification [Figure 47B], but the shift in the melt curve peak seen at lower annealing temperatures was due to a reduction in our assay specificity, and represented the amplification of a gene on human chromosome 10 instead [Figure 47D].

Table 23. Analysis of PCR amplification of known HPV-positive tonsil gDNA samples and melt curve peak alignment with HPV controls, across an annealing temperature gradient. The amount of product produced by qPCR cycling of each sample at all annealing temperatures is described as strong positive (-Rn = >2.5), positive (-Rn = ≤2.5 - ≥1.5), or weak positive (-Rn = <1.5 - ≤0.5). A green colour gradient demonstrates the level of signal detected by analysis of PCR amplification in the melt curve plot, with darker green more HPV-positive and lighter green for a weaker signal detected. For alignment control, the cell colour is used to demonstrate whether the alignment is due to non-specific binding [red] or genuine L1 HPV gene amplification [green]. The total number (out of 5) of samples displaying a genuine HPV-positive peak in their melt curve analysis is shown.

			Annealing temperature (°C)									
			52	50	48	47	46	45	44	43	42	40
Sample	TD1R3	Peak size in melt curve plot (-Rn value)	Strong Positive	Strong Positive	Strong Positive	Strong Positive	Strong Positive	Strong Positive	Strong Positive	Strong Positive	Strong Positive	Strong Positive
		Alignment with control	HPV-18	HPV-18	HPV-18	Non-specific	Non-specific	Non-specific	Non-specific	Non-specific	Non-specific	Non-specific
	TD1L3	Peak size in melt curve plot (-Rn value)	Strong Positive	Strong Positive	Strong Positive	Strong Positive	Strong Positive	Strong Positive	Strong Positive	Strong Positive	Strong Positive	Strong Positive
		Alignment with control	HPV-18	HPV-18	HPV-18	HPV-18	Non-specific	Non-specific	Non-specific	Non-specific	Non-specific	Non-specific
	TD1L4	Peak size in melt curve plot (-Rn value)	Strong Positive	Positive	Weak Positive	Weak Positive	Weak Positive	Positive	Strong Positive	Strong Positive	Strong Positive	Strong Positive
		Alignment with control	HPV-18	HPV-18	HPV-18	HPV-18	HPV-18	Non-specific	Non-specific	Non-specific	Non-specific	Non-specific
	TD1L7	Peak size in melt curve plot (-Rn value)	Weak Positive	Weak Positive	Weak Positive	Positive	Strong Positive	Strong Positive	Strong Positive	Strong Positive	Strong Positive	Strong Positive
		Alignment with control	HPV-18	HPV-18	HPV-18	Non-specific	Non-specific	Non-specific	Non-specific	Non-specific	Non-specific	Non-specific
	TD1L8	Peak size in melt curve plot (-Rn value)	Positive	Positive	Positive	Strong Positive	Strong Positive	Strong Positive	Strong Positive	Strong Positive	Strong Positive	Strong Positive
		Alignment with control	HPV-18	HPV-18	HPV-18	Non-specific	Non-specific	Non-specific	Non-specific	Non-specific	Non-specific	Non-specific
	Control	HPV-18	Positive	Positive	Positive	Positive	Positive	Positive	Positive	Positive	Positive	Positive
		SiHa DNA	Positive	Positive	Positive	Positive	Positive	Positive	Positive	Positive	Positive	Positive
		Total gDNA HPV-positive	5	5	5	2	1	0	0	0	0	0

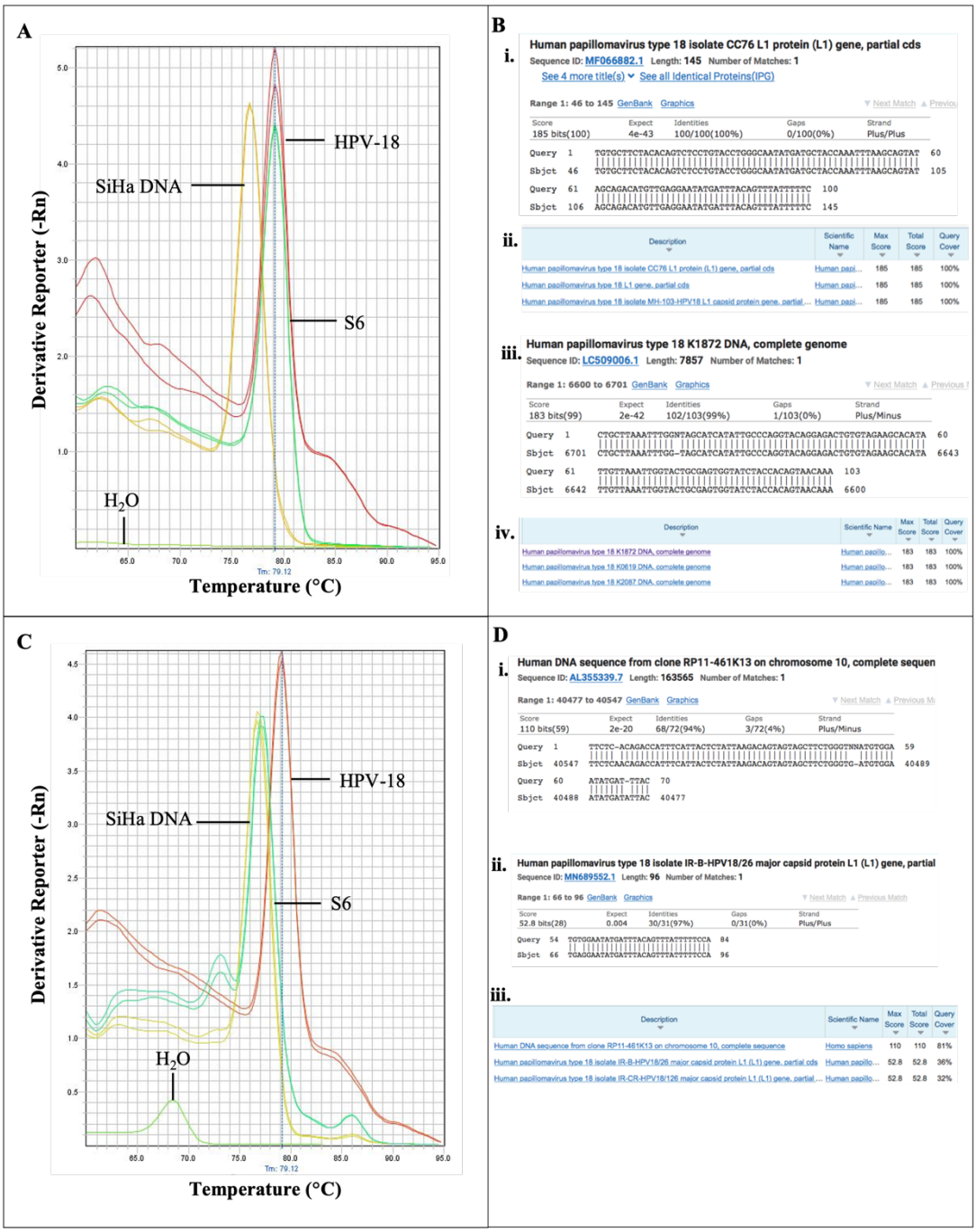


Figure 47. Shift in melt curve alignment of tonsillar gDNA amplification products at lower annealing temperature; melt curve analysis and corresponding Sanger sequencing data. Melt curve plot data was generated using the AB® StepOnePlus™ Real-Time PCR system, and data analysed in their software (v 2.2.2). Temperature (°C) is plotted against the derivative reporter value (-Rn). Sanger sequencing data from S6 (TD1R3) amplicon was searched using NCBI BLAST. Melt curves generated from qPCR products of HPV-18 plasmid, SiHa DNA, and tonsil DNA sample S6 at 52°C (A) and 44°C (C). (B): NCBI BLAST search of sequencing data generated from S6 amplicon in panel A; (i-ii) GP5+ primer, (iii-iv) GP6+ primer, both confirming HPV-18 amplification. (D): NCBI BLAST search of sequencing data generated from S6 amplicon in panel C; (i-iii): GP5+ primer; showing non-specific amplification of chromosome 10.

D.2.4. Further annealing temperature optimisation on a wider scale

As a result of these data generated in our initial annealing temperature gradient analysis, we continued the optimisation process but focussed on temperatures of ≥48°C. We next tested each potential annealing temperature on a broader number of gDNA samples that had been isolated from different human tonsils to enable a better representation of how frequently non-specific binding was being introduced to the assay. We focussed on 48°C, 50°C, 51°C, and 52°C. All individual reactions were set up and cycled as described in D.2.2., with the annealing temperature changing across a gradient.

Using a small number of tonsil gDNA samples, an annealing temperature of 48°C resulted in 100% of known HPV-positive samples (N=5) displaying a positive peak of genuine L1 amplification (section D.2.3). Conversely, melt curve analysis of amplification products generated from a larger number of tonsillar gDNA samples (N=36) at 48°C showed a peak of potential non-specific binding (products that aligned with SiHa DNA control) in 55.5% samples tested (Figure 48). Previously, this alignment of PCR products with SiHa control had been a result of a shift from products suggesting genuine HPV L1 amplification with other parameters; however, in these samples (N=36) several of those with non-specific binding peaks at 48°C appeared negative for any amplification at higher annealing temperatures. Gel electrophoresis of these potential non-specific PCR products arising from an annealing temperature of 48°C resulted in 2 bands at ~175 and ~125 bp. Representative samples were sent for Sanger sequencing, and analysis of FASTA sequences revealed these bands to represent a human gene on chromosome 3 (175 bp) and chromosome 10 (125 bp) (Figure 48 B,D). In some cases, a small region of the FASTA sequence (~30 bp) aligned with the HPV-16 or -18 genome. Interestingly, this non-specific product was also evident in some gDNA samples at a higher annealing temperature of 50°C (10/36, 27.7%; data not shown). These data suggested that 48°C and 50°C were not suitable annealing temperatures to be used for the accurate detection of HPV within tonsillar gDNA by qPCR.

We next compared the use of either 51°C or 52°C as an annealing temperature for several tonsillar gDNA samples (N=71). We had previously observed very little difference between these temperatures in our initial optimisation, but wanted to examine any difference in PCR products across a broader range of tonsil gDNA samples. To determine the optimal annealing temperature to be used routinely in our modified HPV-DNA qPCR screen, we considered: 1.

the number of samples displaying peaks indicative of potential non-specific primer binding, and 2. the quantity of PCR product generated in HPV-positive samples by the -Rn value. Overall, we found that there was more most likely non-specific binding amplifying human products introduced at 51°C with some alignment to SiHa DNA (Figure 49).

Conversely, with an annealing temperature of 52°C, there were fewer samples displaying non-specific binding, and that, generally, those samples considered to be HPV-positive displayed a stronger peak in their melt curve analysis. An example of a sample considered weak/probable positive (-Rn = 0.5) at 51°C but defined as HPV-positive at 52°C (-Rn= ~1.5) is shown in Figure 49A and Figure 49B, respectively. These PCR products were confirmed to be HPV-18 by Sanger sequencing (data not shown).

Overall, at 51°C 28.2% (N=20) appeared positive, and at 52°C 29.6% (N=21) of the same gDNA samples were considered positive. However, there were 10 samples that had a peak that aligned with HPV-16 at 51°C, but only 1 at 52°C. Based on the pattern observed throughout this optimisation process, these peaks were considered a result of non-specific binding.

To investigate this further, primer annealing temperatures of >52°C were also analysed. There was little difference in the melt curve analysis of 3 higher annealing temperatures (52-55°C) compared to that of 52°C. A peak indicative of a large primer dimer was still present within the melt curve plots of many tonsil DNA samples. Moreover, the amplification of HPV L1 was lost in some samples considered to be HPV-positive (data not shown). As a result of all these data together, we selected an annealing temperature of 52°C for our GP5+/GP6+ HPV-DNA qPCR assay.

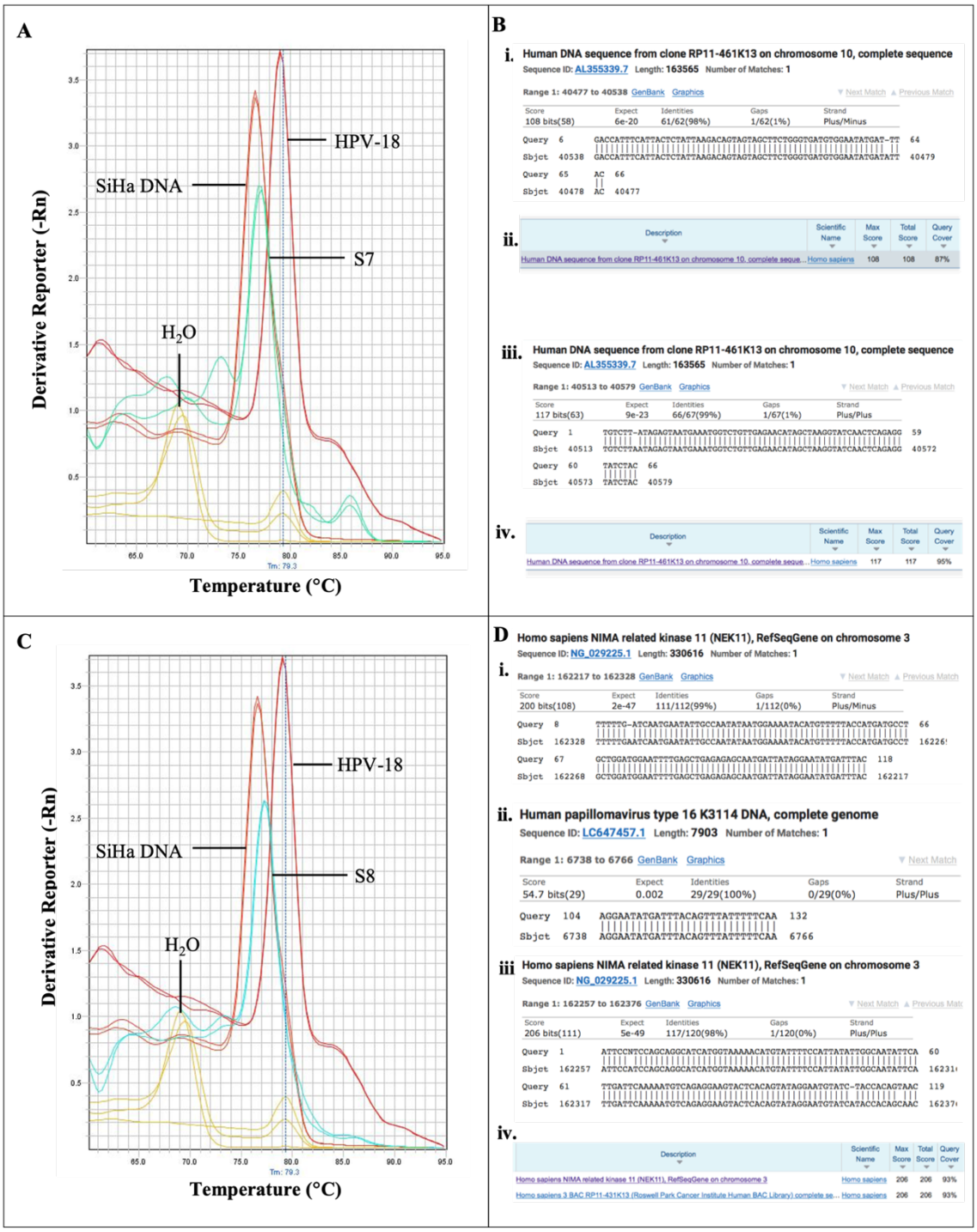


Figure 48. Non-specific amplification of GP5+/GP6+ products from tonsillar gDNA with an annealing temperature of 48°C. Melt curve data was generated using the AB® StepOnePlus Real-Time PCR system, and data analysed in the corresponding software (v 2.2.2). Temperature is plotted against derivative reporter (-Rn). Sanger sequencing data of each PCR amplicon was searched using NCBI BLAST. Melt curve plots of (A) Sample 7 (S7; TDL11) and (C) Sample 8 (S8; TDR11) are shown alongside the sequence alignment and BLAST results with (B) i-ii. GP5+ and iii-iv. GP6+. (D) The top alignments of the sequencing of PCR product in panel C with GP5+ (i-ii) and GP6+ (iii-iv). In both instances sequencing reveals non-specific amplification of human genes on chromosomes 10 and 3.

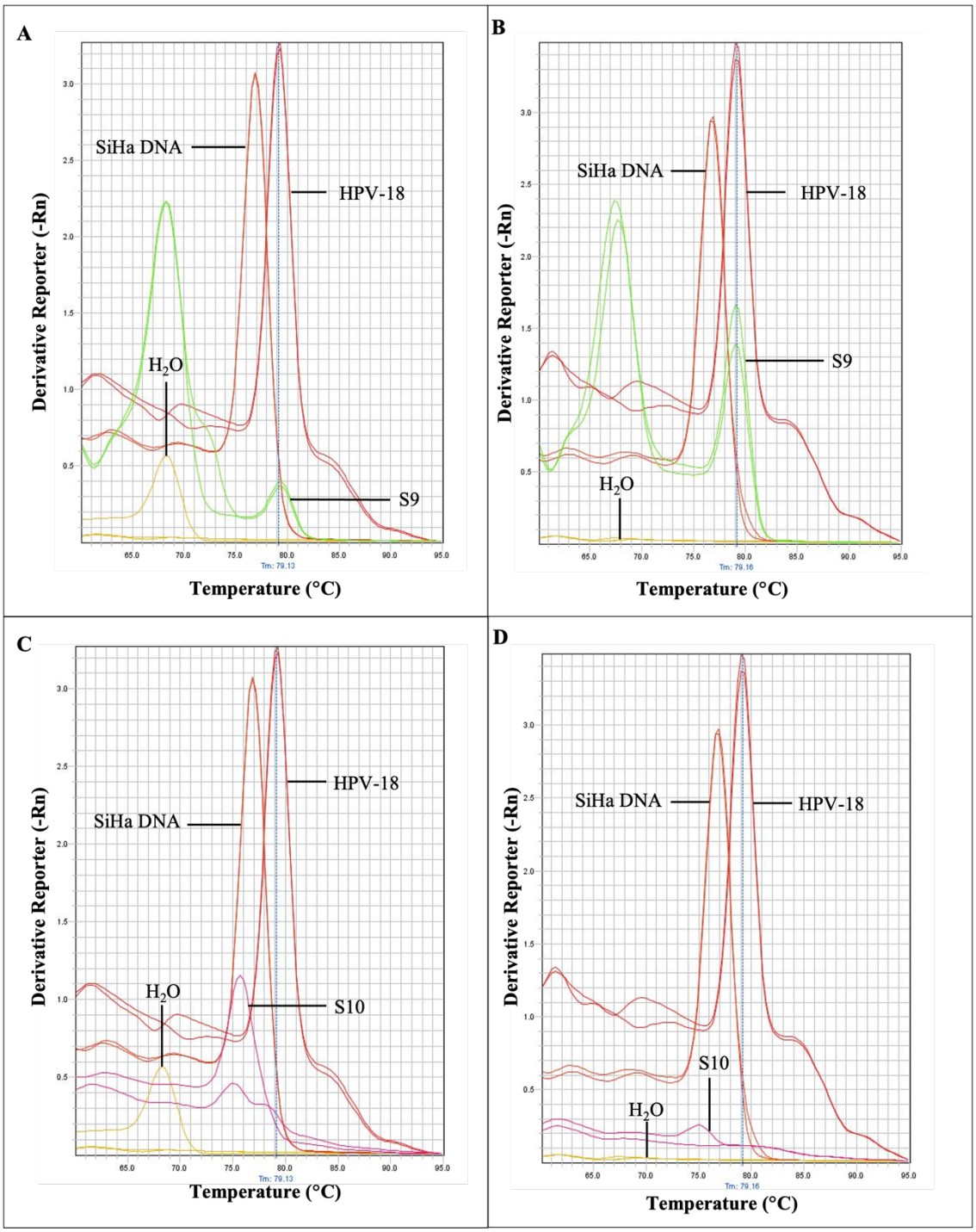


Figure 49. Generation of increased amounts of specific PCR product with limited non-specific GP5+/GP6+ primer binding at annealing temperature 52°C. Melt curve data was generated using the AB® StepOnePlus Real-Time PCR system, and data analysed in the corresponding software (v 2.2.2). Temperature is plotted against derivative reporter (-Rn). Melt curves generated from qPCR products of HPV-18 plasmid, SiHa DNA controls and tonsillar gDNA, S9 (TDR4) at an annealing temperature of (A): 51°C and (B): 52°C. Melt curve plots generated from qPCR products of HPV-18 plasmid, SiHa DNA and tonsillar gDNA, S10 (TD22L4) at an annealing temperature of (C): 51°C and (D): 52°C.

D.4.1. Alternative enzymes and PCR approaches for reliable HPV detection

Unfortunately, several gDNA samples that had been isolated from formalin-fixed tonsil tissues presented unreliable and inconsistent peaks in their melt curve analysis, and in those without these peaks, there was a general lack of HPV-DNA amplified within these samples.

To address the lack of, or inconsistent, HPV detection with formalin-fixed tonsil tissue using our modified qPCR assay, especially given the difference in the frequency of viral presence in these compared to fresh-frozen tissues, we tested different qPCR SYBR master mixes, as well as trialling a standard endpoint PCR approach for low copy DNA, as we wondered whether using different enzymes and approaches may increase the detection rate and/or reproducibility of results.

We selected 2 qPCR master mixes that made use of an intercalating dye for indication of amplification. These were: qPCRBIO SyGreen Mix [PCR Biosystems] and Luna® Universal qPCR Master Mix [New England Biolabs]. In addition, we selected 5 different master mixes to trial within a standard PCR assay. These master mixes included: Platinum II Hot-Start Green [ThermoFisher], Phire Hot-Start II [ThermoFisher], DreamTaq [ThermoFisher], VeriFi® Hot Start Mix [PCR Biosystems], and VeriFi® Proofreading Hot Start Mix [PCR Biosystems]; notably some of these are marketed for complex PCRs with limited template. We considered the use of other consensus primers, but as both MY09/MY11 target a 450 bp (Husnjak et al., 2000) and FAP59/FAP64 a 480 bp (Forslund et al., 1999) region of HPV L1 gene, we decided to continue with GP5+/GP6+ as they target a much smaller region of the L1 viral gene (150 bp) that seemed more suitable for potentially very fragmented DNA (Lenze et al., 2012). Four 10% NBF fixed (~24-96 hours) samples with potential peaks were selected for analysis with each of the master mixes (qPCR or standard). Fresh-frozen samples were analysed alongside as HPV-positive (N=2), and HPV-negative (N=2) controls, based on our modified HPV-DNA qPCR screen.

When thermocycled using two alternative qPCR master mixes (Luna® Universal qPCR Master Mix [NEB] and qPCRBIO SyGreen Mix [PCR Biosystems]) we found that irrespective of HPV-status by our consensus screen, surprisingly all samples appeared HPV-positive. There was a strong peak in the melt curve analysis that aligned with HPV-18 plasmid for all 8 tonsillar gDNA samples analysed (Figure 50B and 50C). Despite some samples showing quite late amplification by Ct value, gel electrophoresis of PCR products confirmed the presence of a

band at ~150 bp to suggest successful L1 gene amplification; however, Sanger sequencing was not able to detect reliable sequences from the amplicons (data not shown). For comparison, the same 8 tonsil gDNA samples were assayed using PowerTrack™ SYBR Green Master Mix [ThermoFisher] as normal. Indeed, melt curve analysis showed tonsil DNA samples to be HPV-positive or HPV-negative as previously defined (Figure 50A).

Next, the selected tonsillar gDNA samples were subject to PCR amplification using each of the master mixes in a standard PCR assay. Gel electrophoresis was used to examine the presence of PCR product and a band at ~150 bp suggested successful amplification of viral gene, L1. We found that HPV-18 plasmid was detectable by gel electrophoresis after thermocycling with all five master mixes, highlighted by a single asterisk in panels A-E in Figure 51. HPV-18 and HPV-16 was detectable in DNA isolated from both HeLa and SiHa cells, respectively, using only the Phire Hot-Start II [ThermoFisher] (Figure 51B) but not the other four master mixes. HPV-18 DNA from HeLa cells was detectable with VeriFi® Proofreading Hot Start master mix, but SiHa DNA was undetectable (Figure 51E) and HPV-16 was amplified from SiHa DNA using the VeriFi Hot Start mix, but there was no amplification of HeLa DNA (Figure 51D). HPV L1 was detected in tonsil gDNA samples with only two master mixes tested; we found the Phire Hot-Start II master mix the most accurate as one out of two known HPV-positive samples displayed a band at ~150 bp (Figure 51B), but there was no amplification of HPV within the formalin-fixed tissues. The VeriFi® Proofreading Hot Start Mix amplified HPV L1 gene in both known HPV-positive samples and interestingly, 2/4 formalin-fixed gDNA samples also showed HPV amplification; however, this master mix was unable to amplify HPV within the SiHa DNA control and so was not investigated further (Figure 51E).

Thus, despite these attempts, we have been unable to reliably detect HPV in formalin-fixed samples (with the exception of one sample by INNO-LiPA genotyping). Together, all these data support that idea that some of the formalin-fixed samples will be HPV-positive, but that long periods of fixation have negatively impacted the integrity of DNA.

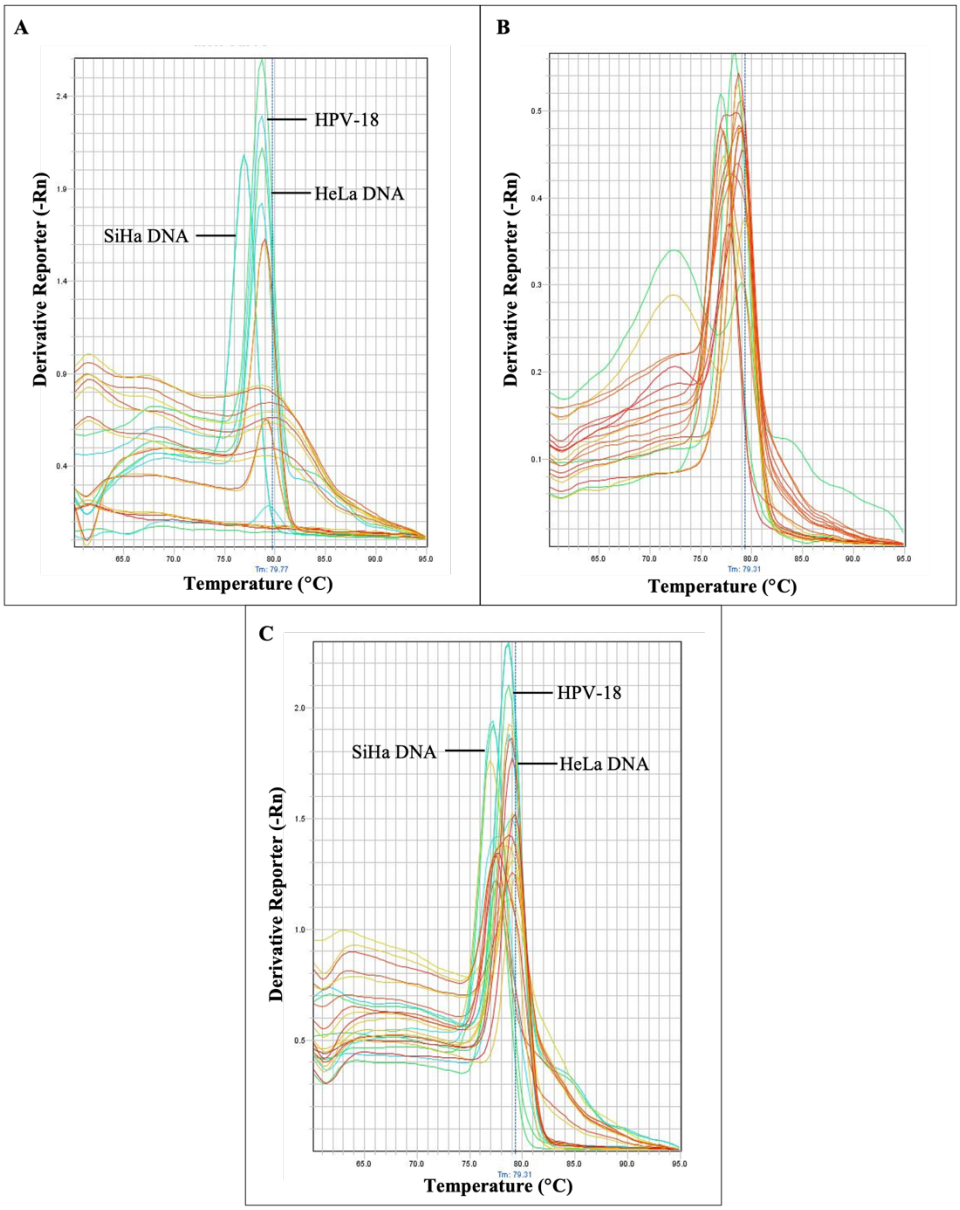


Figure 50. Melt curve analysis of products generated using GP5+/GP6+ consensus primers with different SYBR master mixes. qPCR was performed using the AB® StepOnePlus™ Real-Time PCR system. Data were analysed using their software (v 2.2.2). Temperature (°C) is displayed against derivative reporter values (-Rn). With each master mix, HPV controls (HPV-18 plasmid, SiHa DNA, and HeLa DNA [green/turquoise] and tonsillar gDNA were thermocycled. gDNA samples were selected based on both their tissue of origin (formalin-fixed or fresh-frozen) and HPV-status, as defined with our routinely used modified HPV-DNA qPCR screen. **(A)** PowerTrack™ SYBR Green qPCR master mix [ThermoFisher] showed amplification of all controls, and HPV-positive and HPV-negative gDNA samples had peaks in their melt curve analysis, as expected. **(B)** qPCRBIO SyGreen Mix [PCR Biosystems] and **(C)** Luna® Universal qPCR Master Mix [New England Biolabs] – all samples had strong peaks in their melt curve analysis that aligned with HPV controls. In panel B, HPV controls are not indicated as they are difficult to distinguish from tonsillar gDNA samples.

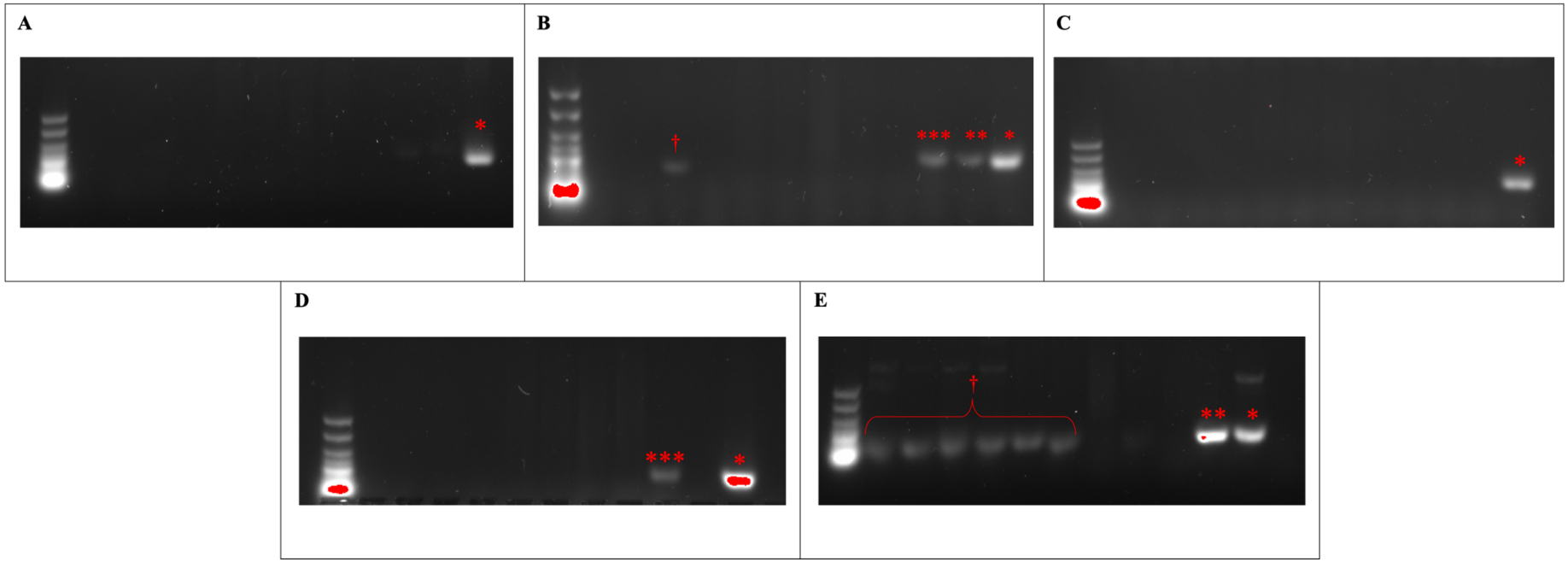


Figure 51. Analysis of PCR products generated using GP5+/GP6+ consensus primers with a variety of master mixes in a standard PCR approach by gel electrophoresis. Amplicons produced with GP5+/GP6+ primers and **(A)** Platinum™ II Hot-Start Green [ThermoFisher], **(B)** Phire Hot-Start II [ThermoFisher], **(C)** DreamTaq [ThermoFisher], **(D)** VeriFi® Hot Start Mix [PCR Biosystems] **or (E)** VeriFi® Proofreading Hot-Start Mix [PCR Biosystems] were analysed on 2% agarose gels. A band at approximately 150 bp indicated successful amplification of viral gene, L1. * = HPV-18 plasmid, ** = HeLa DNA [HPV-18 positive], *** = SiHa DNA [HPV-16 positive]. HPV-18 plasmid was successfully amplified with all master mixes. There was amplification of HPV-DNA in one the HPV-positive tonsillar gDNA samples with **(B)** Phire Hot-Start; multiple samples looked to be HPV-positive with **(E)** VeriFi® Proofreading Hot-Start PCR mixes, respectively. † = amplification of HPV-DNA in tonsillar gDNA samples.

D.3. Discussion

We based our assay optimisation on three commonly used primer pairs in HPV detection; MY09/MY11 (Husnjak et al., 2000), FAP59/FAP64 (Forslund et al., 1999), and GP5+/GP6+ (de Roda Husman et al., 1995). There are well-described protocols for the detection of HPV using all the primer pairs; traditionally, these are used as part of a standard endpoint PCR approach, with gel electrophoresis and hybridisation techniques following thermocycling to analyse PCR products and/or determine viral type (Snijders et al., 1990, Qu et al., 1997, Erhart et al., 2016). The ability to bind a broad range of viral types make consensus primers more suitable as a qualitative screen for any HPV positivity than a type-specific approach; however, the multi-stage process in standard PCR is still time inefficient, and limited for sensitivity.

Consensus primers can be used as part of a qPCR approach, rather than endpoint, with the melt curve plots used to determine PCR products in samples after cycling (Şahiner et al., 2014, Whitton et al., 2024). We built on our laboratory's previous work and scrutinised melt curve plots to distinguish between HPV-positive and HPV-negative samples of tonsillar gDNA.

The use of GP5+/GP6+ primers provided optimal results with these non-cancerous tonsil gDNA samples; however, it is of note that the annealing temperature used for our modified HPV-DNA qPCR assay (52°C) is higher than reports throughout the literature (de Roda Husman et al., 1995, Erhart et al., 2016). Nonetheless, this temperature proved to be most specific for HPV amplification with our sample type.

These data show the occurrence of non-specific binding that would appear HPV-positive by melt curve analysis with certain qPCR cycling parameters. As such, to ensure the stringency of our method, and given all that we had observed through the optimisation process, we decided that any samples with an unusual peak alignment to HPV-16 positive SiHa DNA would be repeated and/or analysed by gel electrophoresis for the presence of a 150 bp band. Finally, to determine viral type and confirm HPV-positivity, samples considered positive by consensus screening would subsequently be subjected to additional scrutiny by HPV genotyping.

Moreover, these data of the use of alternative PCR enzymes and approaches support our rigorous optimisation process of our modified qPCR assay and selection of primer pair/master mix, given that when using a standard PCR approach only one master mix (Phire Hot-Start II

[ThermoFisher]) was able to amplify the L1 gene using the GP5+/GP6+ primer pair in 1 out of 2 known HPV-positive samples.

D.3.1. Conclusion

This chapter describes an extensive optimisation process that revealed that the GP5+/GP6+ consensus primers, with PowerTrack™ SYBR™ Green qPCR master mix [ThermoFisher], and an annealing temperature of 52°C were the most suitable for accurately amplifying HPV L1 within the DNA of non-cancerous tonsil tissue. This qPCR assay was routinely used in this project for the screening of DNA from the non-cancerous palatine tonsils of our prospective cohort for HPV. These data also show that the reliable detection of HPV-DNA in tonsil gDNA isolated from formalin-fixed tissue was unlikely to be successful, irrespective of the PCR approach used.

References

DE RODA HUSMAN, A. M., WALBOOMERS, J. M., VAN DEN BRULE, A. J., MEIJER, C. J. & SNIJDERS, P. J. 1995. The use of general primers GP5 and GP6 elongated at their 3' ends with adjacent highly conserved sequences improves human papillomavirus detection by PCR. *J Gen Virol,* 76 (Pt 4)**,** 1057-62.

ERHART, S. M., RIVERO, E. R., BAZZO, M. L. & ONOFRE, A. S. 2016. Comparative evaluation of the GP5+/6+, MY09/11 and PGMY09/11 primer sets for HPV detection by PCR in oral squamous cell carcinomas. *Exp Mol Pathol,* 100**,** 13-6.

FORSLUND, O., ANTONSSON, A., NORDIN, P., STENQUIST, B. & GÖRAN HANSSON, B. 1999. A broad range of human papillomavirus types detected with a general PCR method suitable for analysis of cutaneous tumours and normal skin. *J Gen Virol,* 80 (Pt 9)**,** 2437-2443.

FRIEDL, F., KIMURA, I., OSATO, T. & ITO, Y. 1970. Studies on a new human cell line (SiHa) derived from carcinoma of uterus. I. Its establishment and morphology. *Proc Soc Exp Biol Med,* 135**,** 543-5.

FUESSEL HAWS, A. L., HE, Q., RADY, P. L., ZHANG, L., GRADY, J., HUGHES, T. K., STISSER, K., KONIG, R. & TYRING, S. K. 2004. Nested PCR with the PGMY09/11 and GP5(+)/6(+) primer sets improves detection of HPV DNA in cervical samples. *J Virol Methods,* 122**,** 87-93.

HUSNJAK, K., GRCE, M., MAGDIĆ, L. & PAVELIĆ, K. 2000. Comparison of five different polymerase chain reaction methods for detection of human papillomavirus in cervical cell specimens. *J Virol Methods,* 88**,** 125-34.

LENZE, D., MÜLLER, H.-H. & HUMMEL, M. 2012. Considerations for the use of formalin-fixed and paraffin-embedded tissue specimens for clonality analysis. *Journal of Hematopathology,* 5**,** 27-34.

QU, W., JIANG, G., CRUZ, Y., CHANG, C. J., HO, G. Y., KLEIN, R. S. & BURK, R. D. 1997. PCR detection of human papillomavirus: comparison between MY09/MY11 and GP5+/GP6+ primer systems. *J Clin Microbiol,* 35**,** 1304-10.

ŞAHINER, F., KUBAR, A., GÜMRAL, R., ARDIÇ, M., YIĞIT, N., ŞENER, K., DEDE, M. & YAPAR, M. 2014. Efficiency of MY09/11 consensus PCR in the detection of multiple HPV infections. *Diagn Microbiol Infect Dis,* 80**,** 43-9.

SNIJDERS, P. J., VAN DEN BRULE, A. J., SCHRIJNEMAKERS, H. F., SNOW, G., MEIJER, C. J. & WALBOOMERS, J. M. 1990. The use of general primers in the polymerase chain reaction permits the detection of a broad spectrum of human papillomavirus genotypes. *J Gen Virol,* 71 (Pt 1)**,** 173-81.

THERMOFISHER SCIENTIFIC. 2022. *PowerTrack™ SYBR™ Green Master Mix - USER GUIDE* [Online]. Available: https://www.thermofisher.com/document-connect/document-connect.html?url=https://assets.thermofisher.com/TFS-Assets%2FLSG%2Fmanuals%2FMAN0018825_PowerTrackSYBRGreenMasterMix_UG.pdf [Accessed 2nd July 2024].

WHITTON, A. 2020. *Oral Human Papillomavirus (HPV) Prevalence and Abundance in the UK Young Adult Population.* Doctor of Philosophy, University of Derby.

WHITTON, A. F., KNIGHT, G. L. & MARSH, E. K. 2024. Risk factors associated with oral Human Papillomavirus (HPV) prevalence within a young adult population. *BMC Public Health,* 24**,** 1485.

Appendix E: Statistically significant protein hits within non-cancerous tonsil tissue by DEA.

Table 24. All 222 significantly changed protein hits identified in HPV-positive compared to HPV-negative non-cancerous tonsil tissue. UniProt ID and protein name for each hit is shown against direction of change (i.e., increased = positive FC in HPV-positive compared to HPV-negative). The significance value is the -log10 p-value calculated in Amica. These hits were first ordered on Manhattan distance, then re-ordered as most significantly changed to least changed (by p-value).

Name	Protein Name	Change	Fold change (log2)	Significance	Manhattan distance
BID_HUMAN	BH3-interacting domain death agonist	Decreased	-3.03590199	4.886791136	7.922693126
SPG7_HUMAN	Mitochondrial inner membrane m-AAA protease component paraplegin	Decreased	-2.116411311	4.751194734	6.867606045
CTGE5_HUMAN*	Melanoma inhibitory activity protein 2	Decreased	-1.566592541	4.412491737	5.979084278
MDN1_HUMAN	Midasin	Decreased	-3.00646278	4.052495722	7.058958501999999
CATS_HUMAN	Cathepsin S	Increased	0.912799731	3.907503927	4.820303658
SPE39_HUMAN	Spermatogenesis-defective protein 39 homolog	Decreased	-1.747270527	3.859403976	5.606674503
SSBP_HUMAN	Single-stranded DNA-binding protein, mitochondrial	Increased	0.964095314	3.796662452	4.760757766
WDR5_HUMAN	WD repeat-containing protein 5	Increased	1.169944393	3.755294982	4.925239375
GPDM_HUMAN	Glycerol-3-phosphate dehydrogenase, mitochondrial	Decreased	-1.042306645	3.721212114	4.763518759
CWC22_HUMAN	Pre-mRNA-splicing factor CWC22 homolog	Decreased	-0.970379543	3.698880983	4.669260526
IPO8_HUMAN	Importin-8	Decreased	-3.668944359	3.614778149	7.283722508
VMA21_HUMAN	Vacuolar ATPase assembly integral membrane protein VMA21	Decreased	-2.984638222	3.396752752	6.381390974
TSN_HUMAN	Translin	Decreased	-1.098932342	3.290689326	4.389621668
GLCNE_HUMAN	Vacuolar ATPase assembly integral membrane protein VMA21	Decreased	-1.480600373	3.207337311	4.687937684
TMED5_HUMAN	Transmembrane emp24 domain-containing protein 5	Decreased	-1.589669238	3.113819032	4.703488269999999
NDUB6_HUMAN	NADH dehydrogenase [ubiquinone] 1 beta subcomplex subunit 6	Decreased	-2.806842813	3.101414671	5.908257484
SRSF1_HUMAN	Serine/arginine-rich splicing factor 1	Increased	0.660488899	3.068234571	3.72872347
DHRS4_HUMAN	Dehydrogenase/reductase SDR family member 4	Increased	0.641405887	3.042709811	3.684115698
PEDF_HUMAN	Pigment epithelium-derived factor	Increased	1.169837139	3.038137242	4.207974381
SCAF8_HUMAN	SR-related and CTD-associated factor 8	Decreased	-1.971910113	3.017948232	4.989858345
CHM2B_HUMAN	Charged multivesicular body protein 2b	Increased	1.023752456	2.982736212	4.006488667999999
XPO4_HUMAN	Exportin-4	Decreased	-1.842210305	2.978511348	4.820721653
RL10L_HUMAN	Ribosomal protein uL16-like	Decreased	-0.633106235	2.976472643	3.609578878
PABP4_HUMAN	Polyadenylate-binding protein 4	Decreased	-2.8129696	2.925855993	5.738825593
MED23_HUMAN	Mediator of RNA polymerase II transcription subunit 23	Decreased	-1.297772321	2.900802599	4.19857492
PTPA_HUMAN	Serine/threonine-protein phosphatase 2A activator	Decreased	-0.739585501	2.890717833	3.630303334
GOSR2_HUMAN	Golgi SNAP receptor complex member 2	Decreased	-2.794275249	2.886921755	5.681197003999999
AR6P4_HUMAN	ADP-ribosylation factor-like protein 6-interacting protein 4	Decreased	-2.18292234	2.870681414	5.053603754
MANF_HUMAN	Mesencephalic astrocyte-derived neurotrophic factor	Increased	0.972832129	2.827530763	3.800362892
KS6A6_HUMAN	Ribosomal protein S6 kinase alpha-6	Decreased	-3.030965789	2.825404999	5.856370788
KAPCA_HUMAN	cAMP-dependent protein kinase catalytic subunit alpha	Decreased	-1.311927376	2.824710029	4.136637405
K1797_HUMAN*	Focadhesin	Decreased	-2.309136582	2.823870814	5.133007396
H2AX_HUMAN	Histone H2AX	Decreased	-0.748998464	2.819681403	3.568679867
PFKAL_HUMAN	ATP-dependent 6-phosphofructokinase, liver type	Decreased	-0.923766073	2.813464892	3.737230965
RASA2_HUMAN	Ras GTPase-activating protein 2	Decreased	-1.248763234	2.794425669	4.043188903
CY1_HUMAN	Cytochrome c1, heme protein, mitochondrial	Decreased	-2.13860517	2.770281563	4.908886733
NALP2_HUMAN	NACHT, LRR and PYD domains-containing protein 2	Decreased	-1.629804215	2.764507914	4.394312129
GLMN_HUMAN	Glomulin	Decreased	-1.501850438	2.763732619	4.265583057
S10A6_HUMAN	Protein S100-A6	Increased	1.00845533	2.76294235	3.77139768
UBR1_HUMAN	E3 ubiquitin-protein ligase UBR1	Decreased	-1.821164389	2.728274298	4.549438687
SAM50_HUMAN	Sorting and assembly machinery component 50 homolog	Decreased	-3.543452731	2.72736508	6.270817811000001
SODC_HUMAN	Superoxide dismutase [Cu-Zn]	Increased	0.675070357	2.719303512	3.394373869
NAA20_HUMAN	N-alpha-acetyltransferase 20	Decreased	-0.947727154	2.711615498	3.659342652
BCL2_HUMAN	Apoptosis regulator Bcl-2	Decreased	-2.186807438	2.709566689	4.896374127
CDC42_HUMAN	Cell division control protein 42 homolog	Decreased	-1.170123417	2.692832481	3.862955898
S12A2_HUMAN	Solute carrier family 12 member 2	Decreased	-1.723222965	2.69212567	4.415348635
LMNA_HUMAN	Prelamin-A/C	Increased	0.655702274	2.676261268	3.331963542
CDIPT_HUMAN	CDP-diacylglycerol--inositol 3-phosphatidyltransferase	Decreased	-3.921494999	2.675946896	6.597441895
ALG5_HUMAN	Dolichyl-phosphate beta-glucosyltransferase	Decreased	-2.302174264	2.666044365	4.968218629
OCRL_HUMAN	Inositol polyphosphate 5-phosphatase OCRL	Decreased	-2.571635606	2.665615219	5.237250825
BAK_HUMAN	Bcl-2 homologous antagonist/killer	Decreased	-1.666995248	2.653878515	4.320873763
LYSC_HUMAN	Lysozyme C	Increased	1.59371374	2.650889421	4.244603161000001
IFIH1_HUMAN	Interferon-induced helicase C domain-containing protein 1	Decreased	-2.576827839	2.632491264	5.209319103
RAB2A_HUMAN	Ras-related protein Rab-2A	Decreased	-0.890499207	2.622151402	3.512650609

Name	Protein Name	Change	Fold change (log2)	Significance	Manhattan distance
SIRT3_HUMAN*	NAD-dependent protein deacetylase sirtuin-3, mitochondrial	Decreased	-1.088070102	2.615684048	3.70375415
URB2_HUMAN	Unhealthy ribosome biogenesis protein 2 homolog	Decreased	-1.857267859	2.603026241	4.4602941
BAAT1_HUMAN*	BRCA1-associated ATM activator 1	Decreased	-1.398931763	2.574689179	3.973620942
MOL2A_HUMAN*	MOB kinase activator 3A	Decreased	-1.262352395	2.572960262	3.835312657
AAKG1_HUMAN	5'-AMP-activated protein kinase subunit gamma-1	Decreased	-0.873525549	2.572408576	3.445934125
LACTB_HUMAN	Serine beta-lactamase-like protein LACTB, mitochondrial	Decreased	-1.28881281	2.570234481	3.859047291
SAR1A_HUMAN	Small COPII coat GTPase SAR1A	Decreased	-0.80050703	2.566906528	3.367413558
COX2_HUMAN	Cytochrome c oxidase subunit 2	Decreased	-3.676255793	2.552754188	6.229009981000001
CD123_HUMAN	Translation initiation factor eIF2 assembly protein	Decreased	-1.432204796	2.551272816	3.983477612
DOCK9_HUMAN	Dedicator of cytokinesis protein 9	Decreased	-2.418615775	2.547504387	4.966120162
COG4_HUMAN	Conserved oligomeric Golgi complex subunit 4	Decreased	-2.168822709	2.535966149	4.704788858000001
SPCS3_HUMAN	Signal peptidase complex subunit 3	Decreased	-1.299158513	2.53092119	3.830079703
ARF4_HUMAN	ADP-ribosylation factor 4	Decreased	-0.986382348	2.519581822	3.50596417
CHID1_HUMAN	Chitinase domain-containing protein 1	Decreased	-1.036962978	2.51730903	3.554272008
ATAD1_HUMAN	Outer mitochondrial transmembrane helix translocase	Decreased	-2.41760471	2.505087444	4.922692154
JAGN1_HUMAN	Protein jagunal homolog 1	Decreased	-2.345092387	2.501401666	4.846494053
OSTC_HUMAN	Oligosaccharyltransferase complex subunit OSTC	Decreased	-3.032423636	2.485157903	5.517581539
KCC2G_HUMAN	Calcium/calmodulin-dependent protein kinase type II subunit gamma	Decreased	-1.661410436	2.484478501	4.145888937
RAB4A_HUMAN	Ras-related protein Rab-4A	Decreased	-1.749511968	2.482883972	4.23239594
CNBP_HUMAN	CCHC-type zinc finger nucleic acid binding protein	Decreased	-1.487261955	2.481090637	3.968352592
ICT1_HUMAN	Large ribosomal subunit protein mL62	Decreased	-1.432175854	2.480083057	3.912258911
DAD1_HUMAN	Dolichyl-diphosphooligosaccharide--protein glycosyltransferase subunit DAD1	Decreased	-2.353275253	2.467080213	4.820355466000001
PPIF_HUMAN	Peptidyl-prolyl cis-trans isomerase F, mitochondria	Increased	0.696711574	2.458011863	3.154723437
OST48_HUMAN	Dolichyl-diphosphooligosaccharide--protein glycosyltransferase 48 kDa subunit	Decreased	-1.728621734	2.45616629	4.184788024
COX15_HUMAN	Heme A synthase COX15	Decreased	-2.00621591	2.449528999	4.455744909
NB5R1_HUMAN	NADH-cytochrome b5 reductase 1	Decreased	-1.657375154	2.437796978	4.095172132
COCA1_HUMAN	Collagen alpha-1(XII) chain	Increased	1.7387781	2.434712759	4.173490859
SURF1_HUMAN	Surfeit locus protein 1	Decreased	-2.12980356	2.432720546	4.562524106
DGLB_HUMAN	Diacylglycerol lipase-beta	Decreased	-1.527230643	2.430283472	3.957514115
GDE_HUMAN	Glycogen debranching enzyme	Decreased	-1.142614691	2.428610544	3.571225235
VPS41_HUMAN	Vacuolar protein sorting-associated protein 41 homolog	Decreased	-1.323952765	2.427695673	3.751648438
NDUA3_HUMAN	NADH dehydrogenase [ubiquinone] 1 alpha subcomplex subunit 3	Decreased	-2.373331137	2.426668684	4.799999821
BZW2_HUMAN*	eIF5-mimic protein 1	Decreased	-0.7958615	2.425060097	3.220921597
HNRCL_HUMAN*	Heterogeneous nuclear ribonucleoprotein C-like 1	Increased	0.764602042	2.419426478	3.18402852
PNCB_HUMAN	Nicotinate phosphoribosyltransferase	Decreased	-1.254767138	2.419402293	3.674169431
NDUBB_HUMAN	NADH dehydrogenase [ubiquinone] 1 beta subcomplex subunit 11, mitochondrial	Decreased	-2.192223348	2.416534948	4.608758296
2A5D_HUMAN	Serine/threonine-protein phosphatase 2A 56 kDa regulatory subunit delta isoform	Decreased	-1.4416125	2.412829939	3.854442439
IPO5_HUMAN	Importin-5	Decreased	-1.063872306	2.408962565	3.472834871
INT2_HUMAN	Integrator complex subunit 2	Decreased	-1.300653943	2.406141861	3.706795804
IMPA3_HUMAN	Golgi-resident adenosine 3',5'-bisphosphate 3'-phosphatase	Decreased	-1.662309354	2.402552959	4.064862313
QCR10_HUMAN	Cytochrome b-c1 complex subunit 10	Decreased	-1.231259447	2.40181114	3.633070587
CO7_HUMAN	Complement component C7	Increased	2.501335641	2.401682442	4.903018082999999
GFPT1_HUMAN	Glutamine--fructose-6-phosphate aminotransferase [isomerizing] 1	Decreased	-0.981394457	2.393085876	3.374480333
IF4G1_HUMAN	Eukaryotic translation initiation factor 4 gamma 1	Decreased	-0.911301832	2.3877427	3.299044532
FBXL8_HUMAN	F-box/LRR-repeat protein 8,	Decreased	-1.210967822	2.368841356	3.579809178
SC11A_HUMAN	Signal peptidase complex catalytic subunit SEC11A	Decreased	-2.052403886	2.363018011	4.415421897
XPO1_HUMAN	Exportin-1	Decreased	-1.129453782	2.359220026	3.488673808
MTL13_HUMAN*	eEF1A lysine and N-terminal methyltransferase	Decreased	-1.098327329	2.358244284	3.456571613
E41L1_HUMAN	Band 4.1-like protein 1	Decreased	-1.14224147	2.340477885	3.482719355
FINC_HUMAN	Fibronectin	Increased	1.118271013	2.331349859	3.449620872
CF211_HUMAN*	Damage-control phosphatase ARMT1	Decreased	-0.952807346	2.3302902	3.283097546
CND1_HUMAN	Condensin complex subunit 1	Decreased	-2.065892101	2.310732352	4.376624453

Name	Protein Name	Change	Fold change (log2)	Significance	Manhattan distance
RAB34_HUMAN	Ras-related protein Rab-34	Decreased	-2.113461647	2.3004879	4.413949547
MSLN_HUMAN	Mesothelin	Decreased	-1.499797741	2.288747251	3.788544992
FKB11_HUMAN	Peptidyl-prolyl cis-trans isomerase FKBP11	Decreased	-1.485095765	2.288362137	3.773457902
NU5M_HUMAN	NADH-ubiquinone oxidoreductase chain 5	Decreased	-2.072161945	2.286141093	4.358303038
MPCP_HUMAN*	Solute carrier family 25 member 3	Decreased	-3.029628546	2.28250275	5.312131296
GSTK1_HUMAN	Glutathione S-transferase kappa 1	Decreased	-0.906157927	2.27555741	3.181715337
SFXN1_HUMAN	Sideroflexin-1	Decreased	-1.148455767	2.261442035	3.409897802
PSME4_HUMAN	Proteasome activator complex subunit 4	Decreased	-1.760144488	2.259705303	4.019849791
HTRA3_HUMAN	Serine protease HTRA3	Increased	0.625998088	2.257188882	2.88318697
GCN1_HUMAN	Stalled ribosome sensor GCN1	Decreased	-1.226366168	2.25692922	3.483295388
NDUA4_HUMAN	Cytochrome c oxidase subunit NDUFA4	Decreased	-1.518737988	2.256016454	3.774754442
HNRPC_HUMAN	Heterogeneous nuclear ribonucleoproteins C1/C2	Increased	0.784207044	2.251884646	3.03609169
B2L13_HUMAN	Bcl-2-like protein 13	Decreased	-1.175125039	2.250312912	3.425437951
NDRG2_HUMAN	Protein NDRG2	Decreased	-0.628671778	2.248627425	2.877299203
HM13_HUMAN	Minor histocompatibility antigen H13	Decreased	-2.541217585	2.242123628	4.783341213
RCD1_HUMAN*	CCR4-NOT transcription complex subunit 9	Decreased	-1.340114314	2.238135131	3.578249445
6PGL_HUMAN	6-phosphogluconolactonase	Decreased	-0.71229022	2.234154635	2.946444855
AT2C1_HUMAN	Calcium-transporting ATPase type 2C member 1	Decreased	-1.111966599	2.228567747	3.340534346
IMB1_HUMAN	Importin subunit beta-1	Decreased	-1.05519428	2.222548962	3.277743242
DOCK2_HUMAN	Dedicator of cytokinesis protein 2	Decreased	-1.118189079	2.221183775	3.339372854
GOSR1_HUMAN	Golgi SNAP receptor complex member	Decreased	-1.72619987	2.220777581	3.946977451
NRM_HUMAN	Nurim	Decreased	-2.66523215	2.216671384	4.881903534
CH055_HUMAN*	Protein THEM6	Decreased	-1.992361732	2.205114439	4.197476171
PI42C_HUMAN	Phosphatidylinositol 5-phosphate 4-kinase type-2 gamma	Decreased	-0.966599464	2.204811791	3.171411255
IPO11_HUMAN	Importin-11	Decreased	-1.714240707	2.201779927	3.916020634
NDUB5_HUMAN	NADH dehydrogenase [ubiquinone] 1 beta subcomplex subunit 5, mitochondrial	Decreased	-1.900432498	2.200830307	4.101262805
PLAK_HUMAN	Junction plakoglobin	Decreased	-1.624967146	2.193957393	3.818924539
RPN2_HUMAN	Dolichyl-diphosphooligosaccharide--protein glycosyltransferase subunit 2	Decreased	-2.031133802	2.191993389	4.223127191
MGN_HUMAN	Protein mago nashi homolog	Decreased	-2.040295925	2.190280467	4.230576392
VKORL_HUMAN	Vitamin K epoxide reductase complex subunit 1-like protein 1	Decreased	-2.098162426	2.189047661	4.287210087
RTRAF_HUMAN	RNA transcription, translation and transport factor protein	Decreased	-0.693693787	2.182668168	2.876361955
K1C19_HUMAN	Keratin, type I cytoskeletal 19	Decreased	-0.693754385	2.179501535	2.87325592
XPO6_HUMAN	Exportin-6	Decreased	-1.080022772	2.175974905	3.255997677
INT10_HUMAN	Integrator complex subunit 10	Decreased	-1.230584059	2.173616439	3.404200498
SYVN1_HUMAN	E3 ubiquitin-protein ligase synoviolin	Decreased	-2.53234688	2.166010116	4.698356995999999
3BP1_HUMAN	SH3 domain-binding protein 1	Decreased	-0.677430208	2.165823218	2.843253426
K6PL_HUMAN*	ATP-dependent 6-phosphofructokinase, liver type	Decreased	-1.087901164	2.163015828	3.250916992
HDGR3_HUMAN	Hepatoma-derived growth factor-related protein 3	Increased	0.886396208	2.162678091	3.049074299
IKIP_HUMAN	Inhibitor of nuclear factor kappa-B kinase-interacting protein	Decreased	-1.723569118	2.159102604	3.882671722
RPN1_HUMAN	Dolichyl-diphosphooligosaccharide--protein glycosyltransferase subunit 1	Decreased	-1.660825176	2.155710749	3.816535925
OXA1L_HUMAN	Mitochondrial inner membrane protein OXA1L	Decreased	-1.877970712	2.152725589	4.030696301
K0090_HUMAN*	ER membrane protein complex subunit 1	Decreased	-2.733663246	2.15120225	4.884865496
FERM1_HUMAN	Fermitin family homolog 1	Decreased	-1.871633012	2.151084156	4.022717168
NU2M_HUMAN	NADH-ubiquinone oxidoreductase chain 2	Decreased	-2.173170671	2.142988397	4.316159067999999
COG6_HUMAN	Conserved oligomeric Golgi complex subunit 6	Decreased	-1.936838183	2.139408859	4.076247042
TM214_HUMAN	Transmembrane protein 214	Decreased	-1.901391629	2.139226659	4.040618288
CETN1_HUMAN	Centrin-1	Increased	1.126632462	2.132872538	3.259505
BGLR_HUMAN	Beta-glucuronidase	Increased	0.680127338	2.128962265	2.809089603
CLP1L_HUMAN	Lipid scramblase CLPTM1L	Decreased	-2.057696138	2.124152681	4.181848819
SKP1_HUMAN	S-phase kinase-associated protein 1	Decreased	-0.917222334	2.122650657	3.039872991
YMEL1_HUMAN	ATP-dependent zinc metalloprotease YME1L1	Decreased	-1.359955342	2.122016121	3.481971463
S27A3_HUMAN	Long-chain fatty acid transport protein 3	Decreased	-2.115268157	2.121642992	4.236911149
PCAT1_HUMAN	Lysophosphatidylcholine acyltransferase 1	Decreased	-1.463931682	2.120833388	3.58476507
UB2D2_HUMAN	Ubiquitin-conjugating enzyme E2 D2	Decreased	-0.78686779	2.120115076	2.906982866
ATM_HUMAN	Serine-protein kinase ATM	Decreased	-1.8351637	2.119526948	3.954690648
RAB3D_HUMAN	Ras-related protein Rab-3D	Decreased	-0.822838556	2.119193151	2.942031707
DIC_HUMAN	Mitochondrial dicarboxylate carrier	Decreased	-2.310145129	2.118934875	4.429080003999999
HACD3_HUMAN	Very-long-chain (3R)-3-hydroxyacyl-CoA dehydratase 3	Decreased	-2.030837703	2.115290767	4.14612847
PLPL6_HUMAN	Patatin-like phospholipase domain-containing protein 6	Decreased	-1.710287387	2.109724944	3.820012331

Name	Protein Name	Change	Fold change (log2)	Significance	Manhattan distance
CJ058_HUMAN*	Peroxiredoxin-like 2A	Decreased	-1.39769471	2.109138092	3.506832802
UFSP2_HUMAN	Ufm1-specific protease 2	Decreased	-2.011026291	2.107663206	4.118689497
MTCH1_HUMAN	Mitochondrial carrier homolog 1	Decreased	-2.652533199	2.106848565	4.759381764
MGST2_HUMAN	Microsomal glutathione S-transferase 2	Decreased	-2.487092615	2.106842844	4.593935459
CP20A_HUMAN	Cytochrome P450 20A1	Decreased	-1.214370632	2.100467847	3.314838479
MBOA5_HUMAN	Lysophospholipid acyltransferase 5	Decreased	-2.249942308	2.09731328	4.347255587999999
OCAD2_HUMAN	OCIA domain-containing protein 2	Decreased	-3.045504158	2.097125585	5.142629743000001
MPU1_HUMAN	Mannose-P-dolichol utilization defect 1 protein	Decreased	-2.081269555	2.096979451	4.178249006
TIM50_HUMAN	Mitochondrial import inner membrane translocase subunit TIM50	Decreased	-1.344467684	2.094416787	3.438884471
TOR2X_HUMAN	Prosalusin	Decreased	-1.84086868	2.090495907	3.931364587
KPCD_HUMAN	Protein kinase C delta type	Decreased	-0.947347937	2.090074336	3.037422273
RAB30_HUMAN	Ras-related protein Rab-30	Decreased	-1.613729477	2.088001048	3.701730525
NCLN_HUMAN	BOS complex subunit NCLN	Decreased	-2.50891108	2.086310999	4.595222079
COQ8A_HUMAN	Atypical kinase COQ8A, mitochondria	Decreased	-1.040249314	2.086231327	3.126480641
HMOX2_HUMAN	Heme oxygenase 2	Decreased	-1.114839627	2.085856685	3.200696312
TOM70_HUMAN	Mitochondrial import receptor subunit TOM70	Decreased	-1.131513113	2.081570851	3.213083964
EIF3L_HUMAN	Eukaryotic translation initiation factor 3 subunit L	Decreased	-1.01641362	2.077940763	3.094354383
ALG12_HUMAN	Dol-P-Man:Man(7)GlcNAc(2)-PP-Dol alpha-1,6-mannosyltransferase	Decreased	-1.486883888	2.077263542	3.56414743
YKT6_HUMAN	Synaptobrevin homolog YKT6	Decreased	-0.916238397	2.074908155	2.991146552
AAKG2_HUMAN	5'-AMP-activated protein kinase subunit gamma-2	Decreased	-1.160102032	2.074874458	3.23497649
MGST3_HUMAN	Glutathione S-transferase 3, mitochondrial	Decreased	-2.87572217	2.074825594	4.950547764
RAB4B_HUMAN	Ras-related protein Rab-4B	Decreased	-0.678496178	2.070086567	2.748582745
COG5_HUMAN	Fermitin family homolog 1	Decreased	-2.008665292	2.070026706	4.078691998
EP300_HUMAN	Histone acetyltransferase p300	Decreased	-0.948681172	2.066399648	3.01508082
ZW10_HUMAN	Centromere/kinetochore protein zw10 homolog	Decreased	-1.445208061	2.063878959	3.50908702
TIDC1_HUMAN	Complex I assembly factor TIMMDC1, mitochondrial	Decreased	-1.237003732	2.061625247	3.298628979
S10AD_HUMAN	Protein S100-A13	Increased	0.802789124	2.058245897	2.861035021
RRP12_HUMAN	RRP12-like protein	Decreased	-1.909200175	2.057612316	3.966812491
GHITM_HUMAN	Growth hormone-inducible transmembrane protein	Decreased	-2.615163454	2.057381981	4.672545435
NDUA9_HUMAN	NADH dehydrogenase [ubiquinone] 1 alpha subcomplex subunit 9, mitochondrial	Decreased	-1.480721116	2.056204254	3.53692537
SAR1B_HUMAN	Small COPII coat GTPase SAR1B	Decreased	-0.954928115	2.052380836	3.007308951
BPA1_HUMAN*	Dystonin	Decreased	-0.903177819	2.051998782	2.955176601
SCRIB_HUMAN	Protein scribble homolog	Decreased	-1.136675751	2.051875143	3.188550894
FDFT_HUMAN	Squalene synthase	Decreased	-1.499592048	2.05089223	3.550484278
UBE3C_HUMAN	Ubiquitin-protein ligase E3C	Decreased	-1.272937917	2.049746996	3.322684913
GCN1L_HUMAN*	Stalled ribosome sensor GCN1	Decreased	-1.1222476	2.04956159	3.17180917
CBL_HUMAN	E3 ubiquitin-protein ligase CBL	Decreased	-1.8162068	2.04748401	3.86369077
FFR_HUMAN*	Vacuolar protein sorting-associated protein 51 homolog	Decreased	-1.4212566	2.04463326	3.46588986
TMM11_HUMAN	Transmembrane protein 11, mitochondrial	Decreased	-1.6135424	2.04372791	3.6572703
AIBP_HUMAN*	NAD(P)H-hydrate epimerase	Decreased	-0.8862524	2.04192989	2.92818227
T126A_HUMAN	Transmembrane protein 126A	Decreased	-1.209822	2.03945873	3.24928072
ATP5S_HUMAN	ATP synthase subunit s, mitochondrial	Decreased	-0.9146808	2.03591681	2.95059763
BD1L1_HUMAN	Biorientation of chromosomes in cell division protein 1-like 1	Decreased	-3.9845331	2.03331651	6.01784964
DYHC1_HUMAN	Cytoplasmic dynein 1 heavy chain 1	Decreased	-1.2229074	2.03309932	3.25600667
CBX5_HUMAN	Chromobox protein homolog 5	Increased	0.72871418	2.03292153	2.76163571
TMX1_HUMAN	Thioredoxin-related transmembrane protein 1	Decreased	-2.3673753	2.02381717	4.391192419
ADT3_HUMAN	ADP/ATP translocase 3	Decreased	-1.5395422	2.02023426	3.5597765
ITPR1_HUMAN	Inositol 1,4,5-trisphosphate-gated calcium channel ITPR1	Decreased	-2.1199697	2.02019882	4.14016847
PYGB_HUMAN	Glycogen phosphorylase, brain form	Decreased	-0.8948157	2.01931326	2.91412891
NADE_HUMAN	Glutamine-dependent NAD(+) synthetase	Decreased	-1.3142285	2.01700514	3.33123362
MKRN2_HUMAN	E3 ubiquitin-protein ligase makorin-2	Decreased	-1.9798407	2.01382832	3.99366907
TM173_HUMAN*	Stimulator of interferon genes protein	Decreased	-1.8338951	2.01103842	3.84493349
BZW1_HUMAN*	eIF5-mimic protein 2	Decreased	-0.8424196	2.0098237	2.85224332
TREX1_HUMAN	Three-prime repair exonuclease 1	Decreased	-1.1333365	2.00855014	3.14188663
SYAP1_HUMAN	Synapse-associated protein 1	Decreased	-1.4747323	2.00717071	3.48190304
YIPF3_HUMAN	Protein YIPF3	Decreased	-1.3771091	2.00672592	3.38383498
RAB18_HUMAN	Ras-related protein Rab-18	Decreased	-1.0451129	2.00436985	3.04948274

* = indicates that the original UniProt ID identified by SWATH data searching in DIA-NN has now been merged with another UniProt ID. In these instances, the protein name given is of that corresponding to the merged UniProt ID. Alternatively, the protein best matched to the UniProt ID was used.

Table 25. All 394 significantly changed protein hits identified in HPV-positive male compared to HPV-positive female non-cancerous tonsil tissue. UniProt ID and protein name for each hit is shown against direction of change (i.e., increased = positive FC in HPV-positive male compared to HPV-positive female). The significance value is the -log10 p-value calculated in Amica. These hits were first ordered on Manhattan distance, then re-ordered as most significantly changed to least changed (by p-value).

Name	Protein name	Change	Fold change (log2)	Significance	Manhattan distance
PORIM_HUMAN	Porimin	Decreased	-4.451948635	5.009704775	9.46165341
LETM1_HUMAN	Mitochondrial proton/calcium exchanger protein	Decreased	-4.434403184	5.009704775	9.444107959
MRP1_HUMAN	Multidrug resistance-associated protein 1	Decreased	-3.860851735	5.009704775	8.87055651
VTI1B_HUMAN	Vesicle transport through interaction with t-SNAREs homolog 1B	Decreased	-3.615473431	5.009704775	8.625178206000001
GSTT1_HUMAN	Glutathione S-transferase theta-1	Decreased	-3.517205874	5.009704775	8.526910649000001
YLAT1_HUMAN	Y+L amino acid transporter 1	Decreased	-3.445273592	5.009704775	8.454978367
PSL2_HUMAN*	Signal peptide peptidase-like 2A	Decreased	-2.929163404	5.009704775	7.938868179
COPT1_HUMAN	High affinity copper uptake protein 1	Decreased	-2.758500074	5.009704775	7.768204849
NUMA1_HUMAN	Nuclear mitotic apparatus protein 1	Decreased	-2.026609589	5.009704775	7.036314364000001
PLS1_HUMAN	Phospholipid scramblase 1	Decreased	-4.418351975	4.828374209	9.246726184
GTR14_HUMAN	Solute carrier family 2, facilitated glucose transporter member 14	Decreased	-3.618008622	4.766739631	8.384748253
VPP1_HUMAN	V-type proton ATPase 116 kDa subunit a 1	Decreased	-2.92142475	4.703019861	7.624444610999999
GSTM4_HUMAN	Glutathione S-transferase Mu 4	Increased	4.384694194	4.467876992	8.852571185999999
IFM2_HUMAN	Interferon-induced transmembrane protein 2	Decreased	-3.818786231	4.427284789	8.246071019999999
RS8_HUMAN	Small ribosomal subunit protein eS8	Decreased	-1.708353133	4.414076791	6.122429924
RL13_HUMAN	Large ribosomal subunit protein eL13	Decreased	-1.835989248	4.411780917	6.247770164999999
CY24A_HUMAN	Cytochrome b-245 light chain	Decreased	-2.660315338	4.369599093	7.029914431
ZN512_HUMAN	Zinc finger protein 512	Decreased	-4.762086206	4.207903986	8.969990192000001
CI046_HUMAN*	Plasminogen receptor (KT)	Decreased	-3.528490308	4.198264212	7.72675452
SNG2_HUMAN	Synaptogyrin-2	Decreased	-3.394059965	4.198264212	7.592324177
EGFR_HUMAN	Epidermal growth factor receptor	Decreased	-2.95417003	4.198264212	7.152434242
RL10_HUMAN	Large ribosomal subunit protein uL16	Decreased	-2.227278437	4.198264212	6.425542649
K2C5_HUMAN	Keratin, type II cytoskeletal 5	Decreased	-2.01472138	4.198264212	6.212985592
DESP_HUMAN	Desmoplakin	Decreased	-1.696022784	4.190601341	5.886624125
VPP3_HUMAN	V-type proton ATPase 116 kDa subunit a 3	Decreased	-2.996644145	4.185246538	7.181890683000001
PTPRJ_HUMAN	Receptor-type tyrosine-protein phosphatase eta	Decreased	-2.830238395	4.15816437	6.988402765
1A68_HUMAN*	HLA class I histocompatibility antigen, A alpha chain	Decreased	-5.267222219	4.094538531	9.36176075
SCAM1_HUMAN	Secretory carrier-associated membrane protein 1	Decreased	-2.52859338	4.091463678	6.620057058
MYH9_HUMAN	Myosin-9	Decreased	-1.717686507	4.058402525	5.776089032
FLOT2_HUMAN	Flotillin-2	Decreased	-2.832661674	4.027239533	6.859901207
SYPL1_HUMAN	Synaptophysin-like protein 1	Decreased	-2.974670691	3.996897219	6.97156791
STR3N_HUMAN	STARD3 N-terminal-like protein	Decreased	-2.289226491	3.996897219	6.28612371
RL18A_HUMAN	Large ribosomal subunit protein eL20	Decreased	-2.018849707	3.995816246	6.014665953
TFR1_HUMAN	Transferrin receptor protein 1	Decreased	-2.757930895	3.883770935	6.64170183
RL37A_HUMAN	Large ribosomal subunit protein eL43	Decreased	-2.11499344	3.866946751	5.981940191
MA7D1_HUMAN	MAP7 domain-containing protein 1	Decreased	-1.64582285	3.866946751	5.512769601
NPC1_HUMAN	NPC intracellular cholesterol transporter 1	Decreased	-4.826486752	3.866560097	8.693046849
LEUK_HUMAN	Leukosialin	Decreased	-3.396634411	3.866560097	7.263194508
K2C6C_HUMAN	Keratin, type II cytoskeletal 6C	Decreased	-1.826178996	3.866560097	5.692739093
SMCA5_HUMAN	SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily A member 5	Decreased	-1.792349356	3.866560097	5.658909453
JAM1_HUMAN	Junctional adhesion molecule A	Decreased	-2.396251312	3.833211822	6.229463134
NOP56_HUMAN	Nucleolar protein 56	Decreased	-1.594471314	3.817134737	5.411606051
MPRD_HUMAN	Cation-dependent mannose-6-phosphate receptor	Decreased	-3.211033735	3.816844085	7.02787782
RL4_HUMAN	Large ribosomal subunit protein uL4	Decreased	-1.514305403	3.78377338	5.298078783
PTPRC_HUMAN	Receptor-type tyrosine-protein phosphatase C	Decreased	-2.726941127	3.770266977	6.497208104
B4GT1_HUMAN	Beta-1,4-galactosyltransferase 1	Decreased	-2.716646944	3.770266977	6.486913920999999
PHF6_HUMAN	PHD finger protein 6	Decreased	-2.492285964	3.770266977	6.262552941
RL29_HUMAN	Large ribosomal subunit protein eL29	Decreased	-2.428899957	3.770266977	6.199166934
PTCD3_HUMAN	Small ribosomal subunit protein mS39	Decreased	-2.377455176	3.769816846	6.147272021999999

Name	Protein name	Change	Fold change (log2)	Significance	Manhattan distance
GCC2_HUMAN	GRIP and coiled-coil domain-containing protein 2	Decreased	-2.676374329	3.748227984	6.424602313
K1C17_HUMAN	Keratin, type I cytoskeletal 17	Decreased	-2.020166303	3.738164003	5.758330306
MRP4_HUMAN	ATP-binding cassette sub-family C member 4	Decreased	-2.254021008	3.728634753	5.982655761
STX7_HUMAN	Syntaxin-7	Decreased	-2.435294355	3.697976607	6.133270962
RL19_HUMAN	Large ribosomal subunit protein eL19	Decreased	-4.296076916	3.684552381	7.980629297
QCR8_HUMAN	Cytochrome b-c1 complex subunit 8	Decreased	-3.174882524	3.684552381	6.859434905000001
VAMP7_HUMAN	Vesicle-associated membrane protein 7	Decreased	-3.155859805	3.684552381	6.840412186
VAMP8_HUMAN	Vesicle-associated membrane protein 8	Decreased	-3.812758462	3.611066192	7.423824654000001
AT1A1_HUMAN	Sodium/potassium-transporting ATPase subunit alpha-1	Decreased	-3.374358113	3.611066192	6.985424305
CD82_HUMAN	CD82 antigen	Decreased	-2.176990081	3.611066192	5.788056273
SATT_HUMAN	Neutral amino acid transporter A	Decreased	-2.094490453	3.611066192	5.705556645
RL17_HUMAN	Large ribosomal subunit protein uL22	Decreased	-1.60640726	3.611066192	5.217473452
SURF4_HUMAN	Surfeit locus protein 4	Decreased	-4.821255732	3.553922845	8.375178577
K2C79_HUMAN	Keratin, type II cytoskeletal 79	Decreased	-1.718781222	3.513229587	5.232010809
ITB4_HUMAN	Integrin beta-4	Decreased	-1.618620065	3.494312915	5.11293298
SC11A_HUMAN	Signal peptidase complex catalytic subunit SEC11A	Decreased	-2.554378171	3.481017801	6.035395972
CD53_HUMAN	Leukocyte surface antigen CD53	Decreased	-3.614179727	3.476582629	7.090762356
CC50A_HUMAN	Cell cycle control protein 50A	Decreased	-3.259963102	3.476582629	6.736545731
CD44_HUMAN	CD44 antigen	Decreased	-1.871862645	3.468175063	5.340037708
TBD2B_HUMAN	TBC1 domain family member 2B	Decreased	-2.044636696	3.462688508	5.507325204
RTN3_HUMAN	Reticulon-3	Decreased	-3.959447532	3.443749376	7.403196908
EMD_HUMAN	Emerin	Decreased	-1.820581562	3.433636763	5.254218325
TM9S4_HUMAN	Transmembrane 9 superfamily member 4	Decreased	-4.449175012	3.430959841	7.880134853
C109A_HUMAN*	Calcium uniporter protein, mitochondrial,	Decreased	-3.2563043	3.430959841	6.687264141
DHB11_HUMAN	Estradiol 17-beta-dehydrogenase 11	Decreased	-2.837794842	3.430959841	6.268754683
CTL2_HUMAN	Choline transporter-like protein 2	Decreased	-2.589573954	3.430959841	6.020533795
K1C16_HUMAN	Keratin, type I cytoskeletal 16	Decreased	-2.248396014	3.430959841	5.679355855
RL36L_HUMAN	Ribosomal protein eL42-like	Decreased	-2.052214752	3.430959841	5.483174592999999
VATL_HUMAN	V-type proton ATPase 16 kDa proteolipid subunit c	Decreased	-4.245991789	3.429062638	7.675054426999999
AT1B1_HUMAN	Sodium/potassium-transporting ATPase subunit beta-1	Decreased	-3.163927857	3.429062638	6.592990495
CYBR1_HUMAN	Plasma membrane ascorbate-dependent reductase CYBRD1	Decreased	-3.312273483	3.423416802	6.735690285
TM9S2_HUMAN	Transmembrane 9 superfamily member 2	Decreased	-3.791999221	3.421207267	7.213206488000001
RL15_HUMAN	Large ribosomal subunit protein eL15	Decreased	-2.145422272	3.39681664	5.542238912
SL9A1_HUMAN	Sodium/hydrogen exchanger 1	Decreased	-1.791528548	3.383757713	5.175286261
TRPV2_HUMAN	Transient receptor potential cation channel subfamily V member 2	Decreased	-2.548655623	3.380610281	5.929265904
CD63_HUMAN	CD63 antigen	Decreased	-1.991092028	3.380610281	5.371702309
RL1D1_HUMAN	Ribosomal L1 domain-containing protein 1	Decreased	-1.641775874	3.358268166	5.000044040000001
HCK_HUMAN	Tyrosine-protein kinase HCK	Decreased	-2.924197186	3.353877725	6.278074910999999
GALT1_HUMAN	Polypeptide N-acetylgalactosaminyltransferase 1	Decreased	-2.444277485	3.353877725	5.79815521
PWP2_HUMAN	Periodic tryptophan protein 2 homolog	Decreased	-2.416439456	3.353877725	5.770317180999999
TACD2_HUMAN	Tumor-associated calcium signal transducer 2	Decreased	-2.090753348	3.326635297	5.417388645
CYC_HUMAN	Cytochrome c	Decreased	-1.827118942	3.326635297	5.153754239
KI67_HUMAN	Proliferation marker protein Ki-67	Decreased	-4.351031747	3.298215353	7.6492471
DHX30_HUMAN	ATP-dependent RNA helicase DHX30	Decreased	-3.182846848	3.298215353	6.481062201
AT2B4_HUMAN	Plasma membrane calcium-transporting ATPase 4	Decreased	-2.690690577	3.277463147	5.968153724
RL28_HUMAN	Large ribosomal subunit protein eL28	Decreased	-1.618774867	3.277463147	4.896238014
PCF11_HUMAN	Pre-mRNA cleavage complex 2 protein Pcf11	Decreased	-1.753109213	3.276495072	5.029604285
RL7_HUMAN	Large ribosomal subunit protein uL30	Decreased	-1.588120676	3.276495072	4.864615748

Name	Protein name	Change	Fold change (log2)	Significance	Manhattan distance
MIPEP_HUMAN	Mitochondrial intermediate peptidase	Decreased	-2.705865706	3.25736912	5.963234826
STX4_HUMAN	Syntaxin-4	Decreased	-2.198976135	3.241542311	5.440518446
RABP2_HUMAN	Cellular retinoic acid-binding protein 2	Increased	3.321038009	3.226324702	6.547362711
RGS19_HUMAN	Regulator of G-protein signaling 19	Decreased	-3.527262062	3.225951887	6.753213949
TSN3_HUMAN	Tetraspanin-3	Decreased	-3.184976459	3.225840836	6.410817295
CO9_HUMAN	Complement component C9	Decreased	-1.950483873	3.2250825	5.175566373000001
RTN4_HUMAN	Reticulon-4	Decreased	-3.993592347	3.215768681	7.209361028
RL35A_HUMAN	Large ribosomal subunit protein eL33	Decreased	-1.619958277	3.215768681	4.835726958
RL18_HUMAN	Large ribosomal subunit protein eL18	Decreased	-1.901869473	3.209074776	5.110944249
S29A1_HUMAN	Equilibrative nucleoside transporter 1	Decreased	-2.569040481	3.202960402	5.772000883
STOM_HUMAN	Stomatin	Decreased	-2.933282973	3.186292277	6.11957525
NNTM_HUMAN	NAD(P) transhydrogenase, mitochondrial	Decreased	-1.913259575	3.186292277	5.099551852
K1C14_HUMAN	Keratin, type I cytoskeletal 14	Decreased	-1.968041463	3.180073264	5.148114726999999
RAP1B_HUMAN	Ras-related protein Rap-1b	Decreased	-1.903284849	3.168335589	5.071620438
KAPCB_HUMAN	cAMP-dependent protein kinase catalytic subunit beta	Decreased	-2.640784848	3.137754693	5.778539541000001
CD47_HUMAN	Leukocyte surface antigen CD47	Decreased	-2.897955345	3.136384536	6.034339881
CAV1_HUMAN	Caveolin-1	Decreased	-4.534222014	3.108187223	7.642409237
GNAI2_HUMAN	Guanine nucleotide-binding protein G(i) subunit alpha-2	Decreased	-1.742917605	3.095284996	4.838202601
ALG5_HUMAN	Dolichyl-phosphate beta-glucosyltransferase	Decreased	-2.723238749	3.09370645	5.816945199
PSN2_HUMAN	Presenilin-2	Decreased	-2.829628009	3.066333565	5.895961573999999
FAM3C_HUMAN	Protein FAM3C	Decreased	-2.134675841	3.050816454	5.185492295
MAIP1_HUMAN	m-AAA protease-interacting protein 1, mitochondrial	Decreased	-2.076291297	3.050816454	5.127107751
ATP68_HUMAN	ATP synthase subunit ATP5MJ, mitochondrial	Decreased	-3.557504692	3.049658029	6.607162721
1C16_HUMAN*	HLA class I histocompatibility antigen, C alpha chain	Decreased	-4.403364974	3.044652893	7.448017866999999
DPM3_HUMAN	Dolichol-phosphate mannosyltransferase subunit 3	Decreased	-2.166125405	3.044652893	5.210778297999999
AGO2_HUMAN	Protein argonaute-2	Decreased	-1.826595924	3.044652893	4.871248817
RB27A_HUMAN	Ras-related protein Rab-27A	Decreased	-1.658997944	3.044652893	4.703650837
PRAF3_HUMAN	PRA1 family protein 3	Decreased	-3.320905248	3.029181886	6.350087134
TB10B_HUMAN	TBC1 domain family member 10B	Decreased	-1.90057835	3.02534372	4.92592207
DVL1L_HUMAN*	Segment polarity protein dishevelled homolog DVL-1-like	Increased	1.904271508	2.998528933	4.902800441
RL32_HUMAN	Large ribosomal subunit protein eL32	Decreased	-1.546415016	2.988513717	4.534928733
MA1A1_HUMAN	Mannosyl-oligosaccharide 1,2-alpha-mannosidase IA	Decreased	-4.487057841	2.950315362	7.437373203
LEMD2_HUMAN	LEM domain-containing protein 2	Decreased	-3.709444524	2.950315362	6.659759886
REEP5_HUMAN	Receptor expression-enhancing protein 5	Decreased	-3.243029398	2.950315362	6.19334476
HBAZ_HUMAN	Hemoglobin subunit zeta	Decreased	-3.163660622	2.950315362	6.113975984
CTL1_HUMAN	Choline transporter-like protein 1	Decreased	-1.993006726	2.950315362	4.943322088
P4K2A_HUMAN	Phosphatidylinositol 4-kinase type 2-alpha	Decreased	-1.929000072	2.950315362	4.879315434
LR16A_HUMAN*	F-actin-uncapping protein LRRC16A	Decreased	-1.830153333	2.950315362	4.780468695
RL6_HUMAN	Large ribosomal subunit protein eL6	Decreased	-1.648866017	2.950315362	4.599181379
PODXL_HUMAN	Podocalyxin	Decreased	-2.207258781	2.944152838	5.151411619
AT1B3_HUMAN	Sodium/potassium-transporting ATPase subunit beta-3	Decreased	-4.029444231	2.940524056	6.969968287
TM9S3_HUMAN	Transmembrane 9 superfamily member 3	Decreased	-3.645482337	2.940524056	6.586006393
HBAT_HUMAN	Hemoglobin subunit theta-1	Decreased	-1.825689914	2.940524056	4.76621397
MAL2_HUMAN	Protein MAL2	Decreased	-2.91756957	2.933451579	5.851021149
RHOG_HUMAN	Rho-related GTP-binding protein RhoG	Decreased	-2.227980332	2.933451579	5.161431911
LYN_HUMAN	Tyrosine-protein kinase Lyn	Decreased	-1.606045928	2.917132086	4.523178014
RAP2B_HUMAN	Ras-related protein Rap-2b	Decreased	-2.208001262	2.916481972	5.124483233999999
VAS1_HUMAN	V-type proton ATPase subunit S1	Decreased	-2.443311593	2.905785051	5.349096643999999
SCRB2_HUMAN	Lysosome membrane protein 2	Decreased	-3.992432032	2.888490493	6.880922525
RAB8A_HUMAN	Ras-related protein Rab-8A	Decreased	-1.749478267	2.887129342	4.636607609
AAAS_HUMAN	Aladin	Decreased	-4.700045838	2.883107468	7.583153306

Name	Protein name	Change	Fold change (log2)	Significance	Manhattan distance
ALO17_HUMAN*	E3 ubiquitin-protein ligase RNF213	Decreased	-2.221370648	2.870325895	5.091696543
BYST_HUMAN	Bystin	Decreased	-2.069320063	2.867755684	4.937075747
SGPL1_HUMAN	Sphingosine-1-phosphate lyase 1	Decreased	-2.710994466	2.855604217	5.566598683
RM23_HUMAN	Large ribosomal subunit protein uL23m	Increased	2.14293694	2.855298169	4.998235108999999
ATLA2_HUMAN	Atlastin-2	Decreased	-1.850739278	2.855298169	4.706037447
KC1A_HUMAN	Casein kinase I isoform alpha	Decreased	-2.547395231	2.851040618	5.398435849
SCAM2_HUMAN	Secretory carrier-associated membrane protein 2	Decreased	-4.229252867	2.847078415	7.076331282
LCAP_HUMAN	Leucyl-cystinyl aminopeptidase	Decreased	-3.092777272	2.847078415	5.939855687
JAK1_HUMAN	Tyrosine-protein kinase JAK1	Decreased	-2.10042073	2.847078415	4.947499145
TM205_HUMAN	Transmembrane protein 205	Decreased	-3.965742982	2.843857471	6.809600453
KCC2D_HUMAN	Calcium/calmodulin-dependent protein kinase type II subunit delta	Decreased	-1.520734478	2.84287399	4.363608468000001
PLP2_HUMAN	Proteolipid protein 2	Decreased	-3.140405572	2.838762905	5.979168477
PI3R4_HUMAN	Phosphoinositide 3-kinase regulatory subunit 4	Decreased	-2.244650567	2.838762905	5.083413472
RS9_HUMAN	Small ribosomal subunit protein uS4	Decreased	-1.557148912	2.837601099	4.394750011
SYNM_HUMAN	Asparaginyl-tRNA synthetase	Decreased	-1.871081374	2.817341161	4.688422535
RER1_HUMAN	Protein RER1	Decreased	-3.335330449	2.815196746	6.150527195
YIPF3_HUMAN	Protein YIPF3	Decreased	-1.914904801	2.815196746	4.730101547
TMX1_HUMAN	Thioredoxin-related transmembrane protein 1	Decreased	-3.917532697	2.813432671	6.730965368
EFR3A_HUMAN	Protein EFR3 homolog A	Decreased	-1.59753326	2.811955058	4.409488318
CCD47_HUMAN	PAT complex subunit CCDC47	Decreased	-2.494282694	2.809684273	5.303966966999999
CALX_HUMAN	Calnexin	Decreased	-2.52782484	2.806996241	5.334821080999999
LMAN1_HUMAN	Protein ERGIC-53	Decreased	-2.540004778	2.804224568	5.344229346000001
ACACA_HUMAN	Acetyl-CoA carboxylase 1	Decreased	-2.395856153	2.804224568	5.200080721
ITB1_HUMAN	Integrin beta-1	Decreased	-2.524241683	2.799932923	5.324174606
PP4P1_HUMAN	Type 1 phosphatidylinositol 4,5-bisphosphate 4-phosphatase	Decreased	-2.00560443	2.796741861	4.802346291
RTN2_HUMAN	Reticulon-2	Decreased	-2.446372361	2.792861329	5.23923369
PLVAP_HUMAN	Plasmalemma vesicle-associated protein	Decreased	-2.210298518	2.780811444	4.991109961999999
GSLG1_HUMAN	Golgi apparatus protein 1	Decreased	-4.09492081	2.779018185	6.873938995
CD276_HUMAN	CD276 antigen	Decreased	-1.882155278	2.779018185	4.661173463
RAC1_HUMAN	Ras-related C3 botulinum toxin substrate 1	Decreased	-1.945818963	2.775637892	4.721456855
GOGA7_HUMAN	Golgin subfamily A member 7	Decreased	-1.609496511	2.771530601	4.381027112
1B38_HUMAN	HLA class I histocompatibility antigen, B alpha chain	Decreased	-1.936099333	2.765332157	4.70143149
MYOF_HUMAN	Myoferlin	Decreased	-2.792017889	2.747988957	5.540006846
LFA3_HUMAN	Lymphocyte function-associated antigen 3	Decreased	-2.40289358	2.746050729	5.148944309
NDUA4_HUMAN	Cytochrome c oxidase subunit NDUFA4	Decreased	-1.548132521	2.746050729	4.29418325
MACF1_HUMAN	Microtubule-actin cross-linking factor 1, isoforms 1/2/3/4/5	Decreased	-1.604990555	2.744095333	4.349085888
ITA2_HUMAN	Integrin alpha-2	Decreased	-3.60837105	2.724693253	6.333064303
CD9_HUMAN	CD9 antigen	Decreased	-2.230616675	2.715499108	4.946115783
RL13A_HUMAN	Large ribosomal subunit protein uL13	Decreased	-1.790049416	2.713402531	4.503451947
ABC3C_HUMAN	DNA dC->dU-editing enzyme APOBEC-3C	Decreased	-1.678879455	2.705354582	4.384234037
TOR1A_HUMAN	Torsin-1A	Decreased	-2.51779463	2.704604013	5.222398643
FA40B_HUMAN*	Striatin-interacting protein 2	Decreased	-1.642692668	2.703340074	4.346032742
CQ062_HUMAN*	Cytochrome b-245 chaperone 1	Decreased	-3.241499387	2.697507594	5.939006981
RGPD2_HUMAN	RANBP2-like and GRIP domain-containing protein 2	Decreased	-1.713182017	2.697507594	4.410689611
MYO6_HUMAN	Unconventional myosin-VI	Decreased	-1.733941028	2.696013425	4.429954453
BAP31_HUMAN	B-cell receptor-associated protein 31	Decreased	-4.463580036	2.693960969	7.157541005
ENDD1_HUMAN	Endonuclease domain-containing 1 protein	Decreased	-2.610230873	2.693960969	5.304191842
CDK2_HUMAN	Cyclin-dependent kinase 2	Decreased	-2.229112993	2.693960969	4.923073962
NDUS2_HUMAN	NADH dehydrogenase [ubiquinone] iron-sulfur protein 2, mitochondrial	Decreased	-1.824875578	2.693960969	4.518836547
WAPL_HUMAN	Wings apart-like protein homolog	Decreased	-1.509624946	2.693960969	4.203585915
SMCA2_HUMAN	Probable global transcription activator SNF2L2	Decreased	-2.377533836	2.68384947	5.061383306
DJC17_HUMAN	DnaJ homolog subfamily C member 17	Increased	2.76383171	2.659809595	5.423641305

Name	Protein name	Change	Fold change (log2)	Significance	Manhattan distance
TECR_HUMAN	Very-long-chain enoyl-CoA reductase	Decreased	-3.312975371	2.659057229	5.9720326
GNAI3_HUMAN	Guanine nucleotide-binding protein G(i) subunit alpha-3	Decreased	-1.586288082	2.65784311	4.244131192
MGST2_HUMAN	Microsomal glutathione S-transferase 2	Decreased	-3.614400577	2.654831015	6.269231592000001
THOC2_HUMAN	THO complex subunit 2	Decreased	-2.410266826	2.654831015	5.065097841
GCP3_HUMAN	Gamma-tubulin complex component 3	Decreased	-2.232911272	2.654831015	4.887742287
TP4A1_HUMAN	Protein tyrosine phosphatase type IVA 1	Decreased	-1.939109233	2.654831015	4.593940248
H4_HUMAN	Histone H4	Decreased	-1.648351691	2.654831015	4.303182706
DHRS7_HUMAN	Dehydrogenase/reductase SDR family member 7	Decreased	-3.526135841	2.654448471	6.180584312
MOT1_HUMAN	Monocarboxylate transporter 1	Decreased	-2.291700047	2.652734011	4.944434058000001
SYNE2_HUMAN	Nesprin-2	Decreased	-1.949376461	2.650461003	4.599837464
FBX7_HUMAN	F-box only protein 7	Decreased	-2.259948447	2.644512274	4.904460721
SC16A_HUMAN	Protein transport protein Sec16A	Decreased	-2.065538209	2.644512274	4.710050483
GNAI1_HUMAN	Guanine nucleotide-binding protein G(i) subunit alpha-1	Decreased	-1.599500909	2.643295108	4.242796017
CH055_HUMAN*	Protein THEM6	Decreased	-2.64086925	2.636713393	5.277582643000001
MOT4_HUMAN	Monocarboxylate transporter 4	Decreased	-2.342529172	2.636713393	4.979242565
CENPC_HUMAN	Centromere protein C	Decreased	-2.632476796	2.627867851	5.260344647
CBPD_HUMAN	Carboxypeptidase D	Decreased	-2.39769656	2.623521957	5.021218516999999
RAC2_HUMAN	Ras-related C3 botulinum toxin substrate 2	Decreased	-1.930316358	2.619708421	4.550024779
ODFP2_HUMAN	Outer dense fiber protein 2	Decreased	-2.527078253	2.614914835	5.141993088
GP1BB_HUMAN	Platelet glycoprotein Ib beta chain	Decreased	-3.919775661	2.603621042	6.523396703
AAAT_HUMAN	Neutral amino acid transporter B(0)	Decreased	-2.446832788	2.593050754	5.039883542
GOGB1_HUMAN	Golgin subfamily B member 1	Decreased	-2.546288584	2.57891194	5.125200524
CD151_HUMAN	CD151 antigen	Decreased	-3.304299694	2.563664783	5.867964477
PRAF2_HUMAN	PRA1 family protein 2	Decreased	-2.995382236	2.551832478	5.547214714000001
DIP2B_HUMAN	Disco-interacting protein 2 homolog B	Decreased	-2.596674462	2.551832478	5.148506940000001
ABD12_HUMAN	Lysophosphatidylserine lipase ABHD12	Decreased	-2.781904888	2.546448671	5.328353559
ANO6_HUMAN	Anoctamin-6	Decreased	-1.675865056	2.546448671	4.222313727
ATPK_HUMAN	ATP synthase subunit f, mitochondrial	Decreased	-3.101970091	2.539218486	5.641188577
FHOD1_HUMAN	FH1/FH2 domain-containing protein 1	Decreased	-2.225342095	2.539218486	4.764560581
SNP23_HUMAN	Synaptosomal-associated protein 23	Decreased	-2.513921205	2.53609149	5.050012694999999
NOG1_HUMAN*	GTP-binding protein 4	Decreased	-1.921111301	2.53445863	4.455569931
VKOR1_HUMAN	Vitamin K epoxide reductase complex subunit 1	Decreased	-2.802891771	2.527036981	5.329928752000001
ITPR1_HUMAN	Inositol 1,4,5-trisphosphate-gated calcium channel ITPR1	Decreased	-2.747476819	2.523095802	5.270572620999999
NAA35_HUMAN	N-alpha-acetyltransferase 35, NatC auxiliary subunit	Decreased	-1.892357242	2.523095802	4.415453043999999
XPO2_HUMAN	Exportin-2	Decreased	-1.640041762	2.523095802	4.163137563999999
GTR1_HUMAN	Solute carrier family 2, facilitated glucose transporter member 1	Decreased	-3.277049584	2.522155524	5.799205108
NCEH1_HUMAN	Neutral cholesterol ester hydrolase 1	Decreased	-2.793640676	2.522155524	5.315796199999999
CBPM_HUMAN	Carboxypeptidase M	Decreased	-1.844307124	2.522155524	4.366462648
TBB1_HUMAN	Tubulin beta-1 chain	Decreased	-2.896656987	2.49996974	5.396626727
ECE1_HUMAN	Endothelin-converting enzyme 1	Decreased	-2.960927543	2.498992253	5.459919795999999
AT2B1_HUMAN	Plasma membrane calcium-transporting ATPase 1	Decreased	-2.518641645	2.492634807	5.011276452000001
RM02_HUMAN	Large ribosomal subunit protein uL2m	Decreased	-2.095270503	2.492072786	4.587343289
H15_HUMAN	Histone H1.5	Decreased	-1.747343457	2.48287473	4.230218187
ITB5_HUMAN	Integrin beta-5	Decreased	-3.880393833	2.482335739	6.362729571999999
MUC18_HUMAN	Cell surface glycoprotein MUC18	Decreased	-1.721348184	2.47547004	4.196818223999999
LMAN2_HUMAN	Vesicular integral-membrane protein VIP36	Decreased	-2.630404737	2.4714827	5.101887437
F162A_HUMAN	Protein FAM162A	Decreased	-1.989334536	2.462779511	4.452114047
MCP_HUMAN	Membrane cofactor protein	Decreased	-1.851665499	2.45189275	4.303558249
ML12B_HUMAN	Myosin regulatory light chain 12B	Increased	1.58709026	2.451515379	4.038605639
4F2_HUMAN	Amino acid transporter heavy chain SLC3A2	Decreased	-2.449390187	2.448845569	4.898235756
R13AX_HUMAN*	Putative ribosomal protein uL13-like	Decreased	-2.019870478	2.444855834	4.464726312

Name	Protein name	Change	Fold change (log2)	Significance	Manhattan distance
RHAG_HUMAN	Ammonium transporter Rh type A	Decreased	-2.963951334	2.440352713	5.404304047
ANK1_HUMAN	Ankyrin-1	Decreased	-1.858810695	2.43750812	4.296318815
SCPDH_HUMAN*	Saccharopine dehydrogenase-like oxidoreductase	Decreased	-2.166858427	2.423617181	4.590475608
NDUA9_HUMAN	NADH dehydrogenase [ubiquinone] 1 alpha subcomplex subunit 9, mitochondrial	Decreased	-1.603743903	2.419348682	4.023092585
TOIP2_HUMAN	Torsin-1A-interacting protein 2	Decreased	-1.783633654	2.419268733	4.202902387
1C05_HUMAN*	HLA class I histocompatibility antigen, C alpha chain	Decreased	-3.183888097	2.416674811	5.600562908000001
ACSL3_HUMAN	Fatty acid CoA ligase Acsl3	Decreased	-2.537316283	2.393180584	4.930496867
RAB31_HUMAN	Ras-related protein Rab-31	Decreased	-1.50119748	2.393180584	3.894378064
MYADM_HUMAN	Myeloid-associated differentiation marker	Decreased	-3.773818255	2.384659545	6.1584778
HEAT1_HUMAN	HEAT repeat-containing protein 1	Decreased	-3.701385843	2.373742906	6.075128749
UBE4A_HUMAN	Ubiquitin conjugation factor E4 A	Decreased	-1.750701691	2.360097445	4.110799136
RHD_HUMAN	Blood group Rh(D) polypeptide	Decreased	-2.084150759	2.351164867	4.435315626
Z3H7A_HUMAN	Zinc finger CCCH domain-containing protein 7A	Decreased	-2.732319868	2.349946327	5.082266195
BASI_HUMAN	Basigin	Decreased	-1.706042729	2.349946327	4.055989056
SPTA1_HUMAN	Spectrin alpha chain, erythrocytic 1	Decreased	-1.899175739	2.34611726	4.245292999
HACD3_HUMAN	Very-long-chain (3R)-3-hydroxyacyl-CoA dehydratase 3	Decreased	-2.406559405	2.338575704	4.745135109
BI1_HUMAN	Bax inhibitor 1	Decreased	-2.485678599	2.334595577	4.820274176
RHOC_HUMAN	Rho-related GTP-binding protein RhoC	Decreased	-1.887307022	2.334595577	4.221902599
WDR36_HUMAN	WD repeat-containing protein 36	Decreased	-1.721400251	2.331380876	4.052781126999999
FAF2_HUMAN	FAS-associated factor 2	Decreased	-2.721045325	2.330569678	5.051615003
TM9S1_HUMAN	Transmembrane 9 superfamily member 1	Decreased	-2.348886265	2.330275242	4.679161507
BRI3B_HUMAN	BRI3-binding protein	Decreased	-4.579123436	2.321778536	6.900901972
GAPR1_HUMAN	Golgi-associated plant pathogenesis-related protein 1	Decreased	-1.997469291	2.317793892	4.315263183
EXOC8_HUMAN	Exocyst complex component 8	Decreased	-2.054425365	2.312409065	4.366834430000001
TMED1_HUMAN	Transmembrane emp24 domain-containing protein 1	Decreased	-2.256983027	2.310308438	4.567291465
CLAP1_HUMAN	CLIP-associating protein 1	Decreased	-1.781580676	2.309986007	4.091566683
CPT1A_HUMAN	Carnitine O-palmitoyltransferase 1, liver isoform	Decreased	-3.065424398	2.307210344	5.372634742000001
SUV3_HUMAN	ATP-dependent RNA helicase SUPV3L1, mitochondrial	Decreased	-2.918899957	2.307210344	5.226110301
SYCM_HUMAN	Probable cysteine--tRNA ligase, mitochondrial	Decreased	-2.236255623	2.307210344	4.543465966999999
ACSL1_HUMAN	Long-chain-fatty-acid--CoA ligase 1	Decreased	-1.801712623	2.307210344	4.108922967
BROX_HUMAN	BRO1 domain-containing protein BROX	Decreased	-1.701668049	2.307210344	4.008878393
KPCA_HUMAN	Protein kinase C alpha type	Decreased	-1.51709689	2.307210344	3.824307234
ELMD2_HUMAN	ELMO domain-containing protein 2	Decreased	-2.390916637	2.30501165	4.695928287
BRX1_HUMAN	Ribosome biogenesis protein BRX1 homolog	Decreased	-2.3139563	2.299532614	4.613488913999999
DYHC1_HUMAN	Cytoplasmic dynein 1 heavy chain 1	Decreased	-1.66005269	2.299532614	3.959585304
CA2D1_HUMAN	Voltage-dependent calcium channel subunit alpha-2/delta-1	Decreased	-2.786473106	2.299443639	5.085916745
UN13D_HUMAN	Protein unc-13 homolog D	Decreased	-2.073020043	2.296103869	4.369123912
RS29_HUMAN	Small ribosomal subunit protein uS14	Decreased	-1.568180559	2.292859306	3.861039865
NB5R1_HUMAN	NADH-cytochrome b5 reductase 1	Decreased	-1.590286847	2.290066805	3.880353652
ARL8B_HUMAN	ADP-ribosylation factor-like protein 8B	Decreased	-1.630899097	2.288499331	3.919398428
TMX4_HUMAN	Thioredoxin-related transmembrane protein 4	Decreased	-3.081489536	2.271617158	5.353106694
DERL1_HUMAN	Derlin-1	Decreased	-3.038074511	2.27078198	5.308856491
EST1_HUMAN	Liver carboxylesterase 1	Decreased	-1.656317399	2.266424959	3.922742358
TEX10_HUMAN	Testis-expressed protein 10	Decreased	-2.719928094	2.262295812	4.982223906
MPU1_HUMAN	Mannose-P-dolichol utilization defect 1 protein	Decreased	-2.587646715	2.262295812	4.849942527
PRKDC_HUMAN	DNA-dependent protein kinase catalytic subunit	Decreased	-1.881680601	2.262295812	4.143976413
WDR46_HUMAN	WD repeat-containing protein 46	Decreased	-1.806207124	2.262295812	4.068502936

Name	Protein name	Change	Fold change (log2)	Significance	Manhattan distance
NAA16_HUMAN	N-alpha-acetyltransferase 16, NatA auxiliary subunit	Decreased	-2.082412547	2.261255344	4.343667891
CFA20_HUMAN	Cilia- and flagella-associated protein 20	Decreased	-2.1271719	2.260705171	4.387877071
DMD_HUMAN	Dystrophin	Decreased	-2.621309224	2.260487939	4.881797163
LNP_HUMAN	Endoplasmic reticulum junction formation protein lunapark	Decreased	-1.734197738	2.260487939	3.994685677
MILK2_HUMAN	MICAL-like protein 2	Decreased	-2.988439851	2.256851148	5.245290999
ZW10_HUMAN	Centromere/kinetochore protein zw10 homolog	Decreased	-2.461314719	2.256851148	4.718165867
OSTC_HUMAN	Oligosaccharyltransferase complex subunit OSTC	Decreased	-3.698674506	2.250686903	5.949361409
CDC42_HUMAN	Cell division control protein 42 homolog	Decreased	-2.116687941	2.249258321	4.365946262
STX12_HUMAN	Syntaxin-12	Decreased	-2.457244678	2.245220184	4.702464861999999
WAHS7_HUMAN*	WASH complex subunit 7	Decreased	-2.080108289	2.245220184	4.325328473
QCR9_HUMAN	Cytochrome b-c1 complex subunit 9	Decreased	-2.088257808	2.241514743	4.329772551
AP1S1_HUMAN	AP-1 complex subunit sigma-1A	Decreased	-1.769526859	2.241514743	4.011041602000001
DDX50_HUMAN	ATP-dependent RNA helicase DDX50	Decreased	-1.709037736	2.241514743	3.950552479
ESYT1_HUMAN	Extended synaptotagmin-1	Decreased	-2.078126715	2.241170889	4.319297604
CD59_HUMAN	CD59 glycoprotein	Decreased	-2.638951036	2.238374695	4.877325731
CLCN7_HUMAN	H(+)/Cl(-) exchange transporter 7	Decreased	-1.508751074	2.238374695	3.747125769
MOB3A_HUMAN	MOB kinase activator 3A	Decreased	-2.251603736	2.237460673	4.489064409
HM13_HUMAN	Minor histocompatibility antigen H13	Decreased	-3.513053647	2.235612976	5.748666623
LIMS1_HUMAN	LIM and senescent cell antigen-like-containing domain protein 1	Decreased	-2.195897637	2.235612976	4.431510613
SPTB1_HUMAN	Spectrin beta chain, erythrocytic	Decreased	-2.153728151	2.233642559	4.387370710000001
NEMO_HUMAN	NF-kappa-B essential modulator	Decreased	-1.621993262	2.233519634	3.855512896
UN93B_HUMAN	Protein unc-93 homolog B1	Decreased	-2.077318118	2.225748071	4.303066189
SYF1_HUMAN	Pre-mRNA-splicing factor SYF1	Decreased	-2.63355646	2.224095001	4.857651461
SEM7A_HUMAN	Semaphorin-7A	Decreased	-1.570304487	2.216934532	3.787239019
ASM3B_HUMAN	Acid sphingomyelinase-like phosphodiesterase 3b	Decreased	-1.662423262	2.208870156	3.871293418
TBRG4_HUMAN*	FAST kinase domain-containing protein 4	Decreased	-2.266106924	2.198750413	4.464857337
CISD1_HUMAN	CDGSH iron-sulfur domain-containing protein 1	Decreased	-1.56910992	2.198750413	3.767860333
RAB35_HUMAN	Ras-related protein Rab-35	Decreased	-1.557646687	2.198750413	3.7563971
ADA10_HUMAN	Disintegrin and metalloproteinase domain-containing protein 10	Decreased	-1.782124787	2.188570655	3.970695442
PCAT1_HUMAN	Lysophosphatidylcholine acyltransferase 1	Decreased	-2.036831407	2.186401467	4.223232874000001
RAF1_HUMAN	RAF proto-oncogene serine/threonine-protein kinase	Decreased	-2.666964185	2.182654808	4.849618993
PREB_HUMAN*	Guanine nucleotide-exchange factor SEC12	Decreased	-2.569056463	2.181424304	4.750480767
PECA1_HUMAN	Platelet endothelial cell adhesion molecule	Decreased	-2.567588938	2.1710297	4.738618638
F120A_HUMAN	Constitutive coactivator of PPAR-gamma-like protein 1	Decreased	-1.980565878	2.168513393	4.149079271
RRP12_HUMAN	RRP12-like protein	Decreased	-2.27831927	2.165801396	4.444120666
RHOA_HUMAN	Transforming protein RhoA	Decreased	-1.90575419	2.164722236	4.070476426
SHIP2_HUMAN	Phosphatidylinositol 3,4,5-trisphosphate 5-phosphatase 2	Decreased	-1.656520656	2.164722236	3.821242892
CYB5B_HUMAN	Cytochrome b5 type B	Decreased	-2.327121987	2.156212402	4.483334388999999
PVRL2_HUMAN*	Nectin-2	Decreased	-1.882302876	2.155510212	4.037813088
FBN1_HUMAN	Fibrillin-1	Decreased	-4.898702682	2.154708122	7.053410804
MRCKB_HUMAN	Serine/threonine-protein kinase MRCK beta	Decreased	-1.979136215	2.154708122	4.133844337
DAF_HUMAN	Complement decay-accelerating factor	Decreased	-2.445162734	2.152754604	4.597917338
ZFPL1_HUMAN	Zinc finger protein-like 1	Decreased	-3.032258191	2.152466515	5.184724706
RALB_HUMAN	Ras-related protein Ral-B	Decreased	-1.882064809	2.152466515	4.034531324
LIMS2_HUMAN	LIM and senescent cell antigen-like-containing domain protein 2	Decreased	-1.612341004	2.152466515	3.764807519
PESC_HUMAN	Pescadillo homolog	Decreased	-1.584129231	2.147594245	3.731723476
CLH1_HUMAN	Clathrin heavy chain 1	Decreased	-1.55686674	2.146983143	3.703849883
HYEP_HUMAN	Epoxide hydrolase 1	Decreased	-3.313761199	2.130702202	5.444463401
AP1G2_HUMAN	AP-1 complex subunit gamma-like 2	Decreased	-1.773654761	2.129236809	3.90289157
TMM43_HUMAN	Transmembrane protein 43	Decreased	-3.439450154	2.129190202	5.568640356
NSDHL_HUMAN	Sterol-4-alpha-carboxylate 3-dehydrogenase, decarboxylating	Decreased	-2.394086894	2.123450444	4.517537338
WDR75_HUMAN	WD repeat-containing protein 75	Decreased	-1.980180221	2.113676144	4.093856365

Name	Protein name	Change	Fold change (log2)	Significance	Manhattan distance
FN3K_HUMAN	Fructosamine-3-kinase	Decreased	-2.21140056	2.113659386	4.325059946
BIG2_HUMAN	Brefeldin A-inhibited guanine nucleotide-exchange protein 2	Decreased	-2.274298501	2.110125631	4.384424131999999
IKZF1_HUMAN	DNA-binding protein Ikaros	Decreased	-1.742723616	2.105711525	3.848435141
NB5R3_HUMAN	NADH-cytochrome b5 reductase 3	Decreased	-1.657425225	2.105711525	3.76313675
RFC2_HUMAN	Replication factor C subunit 2	Decreased	-1.622745474	2.105711525	3.728456999
PLAK_HUMAN	Junction plakoglobin	Decreased	-1.504648609	2.105711525	3.610360134
PARP4_HUMAN	Protein mono-ADP-ribosyltransferase PARP4	Decreased	-1.563628529	2.103810657	3.667439186
RT05_HUMAN	Small ribosomal subunit protein uS5m	Decreased	-1.979741508	2.103079019	4.082820527
SPCS2_HUMAN	Signal peptidase complex subunit 2	Decreased	-3.627090169	2.098938283	5.726028452
RAB3A_HUMAN	Ras-related protein Rab-3A	Decreased	-4.854413765	2.0979459	6.952359665
NIBA1_HUMAN	Protein Niban 1	Decreased	-1.74752685	2.0979459	3.84547275
RRAGA_HUMAN	Ras-related GTP-binding protein A	Decreased	-1.622543236	2.0979459	3.720489136
PAPOG_HUMAN	Poly(A) polymerase gamma	Decreased	-1.733620558	2.096926299	3.830546857
ILVBL_HUMAN*	2-hydroxyacyl-CoA lyase 2	Decreased	-2.041370442	2.095402049	4.136772491
GHDC_HUMAN	GH3 domain-containing protein	Decreased	-1.563112825	2.094705223	3.657818048
TMCO1_HUMAN	Calcium load-activated calcium channel	Decreased	-2.697919233	2.092379229	4.790298462
MFS10_HUMAN	Major facilitator superfamily domain-containing protein 10	Decreased	-2.515733084	2.092379229	4.608112312999999
LCLT1_HUMAN	Lysocardiolipin acyltransferase 1	Decreased	-2.684174434	2.089681737	4.773856171
SMCA1_HUMAN	Probable global transcription activator SNF2L1	Decreased	-1.649346745	2.081251102	3.730597847
LAMB3_HUMAN	Laminin subunit beta-3	Decreased	-1.652964873	2.076587997	3.72955287
NCLN_HUMAN	BOS complex subunit NCLN	Decreased	-2.751310564	2.075188926	4.82649949
ABC3G_HUMAN	DNA dC->dU-editing enzyme APOBEC-3G	Decreased	-2.174908257	2.075144519	4.250052776
PI4KA_HUMAN	Phosphatidylinositol 4-kinase alpha	Decreased	-3.591089547	2.063910907	5.655000454
KDSR_HUMAN	3-ketodihydrosphingosine reductase	Decreased	-2.144380922	2.059688509	4.204069431
SRPRB_HUMAN	Signal recognition particle receptor subunit beta	Decreased	-2.401019284	2.05932427	4.460343554
VKORL_HUMAN	Vitamin K epoxide reductase complex subunit 1-like protein 1	Decreased	-1.948320134	2.058077127	4.006397261
IMPA3_HUMAN	Golgi-resident adenosine 3',5'-bisphosphate 3'-phosphatase	Decreased	-1.887661968	2.055530011	3.943191979
RB22A_HUMAN	Ras-related protein Rab-22A	Decreased	-2.012253695	2.048782196	4.061035891
NU205_HUMAN	Nuclear pore complex protein Nup205	Decreased	-2.158884512	2.046096421	4.204980933
CDCP1_HUMAN	CUB domain-containing protein 1	Decreased	-1.743963808	2.046096421	3.790060229
MEAK7_HUMAN	MTOR-associated protein MEAK7	Decreased	-1.707364843	2.044325139	3.751689982
FPRP_HUMAN	Prostaglandin F2 receptor negative regulator	Decreased	-1.838960591	2.034935055	3.873895646
RDH11_HUMAN	Retinol dehydrogenase 11	Decreased	-2.400723605	2.032734454	4.433458058999999
ABCD3_HUMAN	ATP-binding cassette sub-family D member 3	Decreased	-2.439706165	2.032047682	4.471753847
NCBP1_HUMAN	Nuclear cap-binding protein subunit 1	Decreased	-1.934608049	2.032047682	3.966655731
TM109_HUMAN	Voltage-gated monoatomic cation channel TMEM109	Decreased	-3.885193232	2.031514483	5.916707715
CLPT1_HUMAN	Putative lipid scramblase CLPTM1	Decreased	-2.716186546	2.031514483	4.747701029
MPZL1_HUMAN	Myelin protein zero-like protein 1	Decreased	-1.900687027	2.029974779	3.930661806
MGAT1_HUMAN	Alpha-1,3-mannosyl-glycoprotein 2-beta-N-acetylglucosaminyltransferase	Decreased	-2.2862774	2.022301036	4.308578435999999
B4GT5_HUMAN	Beta-1,4-galactosyltransferase 5	Decreased	-1.826506043	2.022301036	3.848807079
KC1AL_HUMAN	Casein kinase I isoform alpha-like	Decreased	-2.229523551	2.017193738	4.246717289
FACE1_HUMAN	CAAX prenyl protease 1 homolog	Decreased	-2.555424869	2.011574429	4.566999298
DRS7B_HUMAN	Dehydrogenase/reductase SDR family member 7B	Decreased	-2.795293391	2.008557451	4.803850842
ATP6_HUMAN	ATP synthase subunit a	Decreased	-3.385892817	2.00758417	5.393476987

* = indicates that the original UniProt ID identified by SWATH data searching in DIA-NN has now been merged with another UniProt ID. In these instances, the protein name given is of that corresponding to the merged UniProt ID. Alternatively, the protein best matched to the UniProt ID was used.