HUMAN PAPILLOMAVIRUSES IN HEAD AND NECK SQUAMOUS CELL CARCINOMAS

THE RELEVANCE OF HPV16 VARIANTS

DANIELA COCHICHO TAVARES DE SOUSA

Tese para obtenção do grau de Doutor em Biomedicina

Doutoramento em associação entre:

Faculdade de Ciências Médicas | NOVA Medical School - Universidade NOVA de Lisboa

Universidade de Aveiro

Janeiro, 2023

HUMAN PAPILLOMAVIRUSES IN HEAD AND NECK SQUAMOUS CELL CARCINOMAS: THE RELEVANCE OF HPV16 VARIANTS

Daniela Cochicho Tavares de Sousa

Supervisor: Ana Félix MD PhD
Faculdade de Ciências Médicas | NOVA Medical School da Universidade NOVA de Lisboa,
Serviço de Anatomia Patológica do Instituto Português de Oncologia de Lisboa Francisco Gentil, (IPOLFG)

Co-Supervisor: Rui M. Gil da Costa PhD
LEPABE, Laboratory for Process Engineering, Environment, Biotechnology and Energy, Faculty of Engineering, University of Porto
ALiCE - Associate Laboratory in Chemical Engineering, Faculty of Engineering, University of Porto
Molecular Oncology and Viral Pathology Group, Research Center of IPO Porto (CI-IPOP) / RISE@CI-IPOP (Health Research Network), Portuguese Oncology Institute of Porto (IPO Porto), Porto Comprehensive Cancer Center (Porto.CCC)
Post-graduate Programme in Adult Health (PPGSAD), University Hospital (HUUFMA) and Morphology Department, Federal University of Maranhão, Brazil

TESE PARA OBTENÇÃO DO GRAU DE DOUTOR NA ESPECIALIDADE EM BIOMEDICINA

Janeiro, 2023

HOST INSTITUTIONS

FINANCIAL SUPPORT

This research was not directly financed but some experimental tasks and personal included in some articles were part of projects with funds and the cost were partially covered, and for that we briefly mention here:

- Fundação para a Ciência e a Tecnologia (FCT) para a Ciência e Tecnologia/Ministério da Ciência, Tecnologia e Ensino Superior (FCT/MCTES, Portugal) through national funds to iNOVA4Health.
  (UIDB/04462/2020 and UIDP/04462/2020) – IPO Lisboa
- GenomePT project supported by COMPETE 2020 Operational Programme for Competitiveness and Internationalisation (POCI), Lisboa Portugal Regional Operational Programme (Lisboa2020), Algarve Portugal Regional Operational Programme (CRESC Algarve2020), under the PORTUGAL 2020 Partnership Agreement, through the European Regional Development Fund (ERDF) and by Fundação para a Ciência e a Tecnologia (FCT).
  POCI-01-0145-FEDER-022184) - INSA
- FEDER funds through the *Programa Operacional Factores de Competitividade* – COMPETE and by National funds through the *Fundação para a Ciência e a Tecnologia* within the scope of the project Centre for Toxicogenomics and Human Health -ToxOmics).
  (UID/BIM/00009/2019) – INSA
- National funds through FCT/MCTES (PIDDAC) to 2SMART (NORTE-01-0145-FEDER-000054) supported by Norte Portugal Regional Operational Programme (NORTE 2020), under the PORTUGAL 2020 Partnership Agreement, through the European Regional Development Fund (ERDF).
  (LA/P/0045/2020 ALICE) (UIDB/00511/2020 and UIDP/00511/2020 LEPABE) – IPO Porto

LIST OF SCIENTIFIC PUBLICATIONS

These are the main scientific articles published by the author whose content was totally or partially used in the elaboration of the thesis:

I. **Detecção e Genotipagem do HPV em esfregaços e biópsias buco-faringeos**.
Daniela Cochicho, Luís Martins, Pedro Montalvão, Lígia Ferreira, Luís Oliveira, Rui Fino, Hugo Estibeiro, Ricardo Pacheco, Ana Hebe, Mário Cunha, Miguel Magalhães.
Revista Portuguesa de Otorrinolaringologia e Cirurgia Cervico- facial, vol.55, nº3 Setembro 2017.https://www.journalsporl.com/index.php/sporl/article/view/414

The author contributed to the study's idea and design, performed all laboratory technical procedures. Performed data collection and compilation. Directly contribute to analysing and interpreting data. Prepare all the writing, editing, and rewriting the final manuscript.

II. **Exploring the roles of HPV16 variants in head and neck squamous cell carcinoma: Current challenges and opportunities**.
Daniela Cochicho; Rui Gil da Costa; Ana Félix.
Virol J 18, 217 (2021). https://doi.org/10.1186/s12985-021-01688-9

The author contributed to the study´s concepts and study design, including the literature search using three databases (meta-analysis), the data collection, analysis, and interpretation. Also conducted the preparation of the manuscript, in its editing and reviewed entirely.

III. Distribution and clinical significance of HPV16 Variants in Head and Neck Squamous Cell Carcinomas: Data from a Portuguese cohort and systematic review.
Daniela Cochicho, Alexandra Nunes, João Paulo, Susana Esteves, Joana Mendonça, Luís Vieira, Luís Martins, Mário Cunha, Pedro Montalvão, Miguel Magalhães, Rui Gil da Costa, Ana Félix.
Pathobiology - DOI: 10.1159/000529723

The author contributed to the study´s concepts and study design,

including the systematic review of the literature using two databases. Also performed the data collection, analysis, and interpretation. Design and developed the entire NGS workflow, directly contribute to the analysing and interpreting of the molecular data. Conducted the preparation and editing the final manuscript.

IV. **Characterization of the Human papillomavirus 16 oncogenes in K14HPV16 mice: Sublineage A1 drives multi-organ carcinogenesis**.
Daniela Cochicho, Alexandra Nunes, João Paulo Gomes, Luís Martins, Mário Cunha, Beatriz Medeiros-Fonseca, Paula Oliveira, Margarida Bastos, Rui Medeiros, Joana Mendonça, Luís Vieira, Rui M. Gil da Costa, Ana Félix.
Int. J. Mol. Sci. 2022, 23, 12371. https://doi.org/10.3390/ijms232012371

The author contributed on the study's design, performed all laboratory technical procedures for DNA HPV molecular analysis. Performed data collection and compilation. Directly contribute to molecular analysing and interpreting data. Prepare all the writing, editing, and rewriting the final manuscript.

V. PIK3CA Gene Mutations in HNSCC: Systematic Review and Correlations with HPV Status and Patient Survival.
Cochicho D, Esteves S, Rito M, Silva F, Martins L, Montalvão P, Cunha M, Magalhães M, Gil da Costa RM, Félix A.
Cancers. 2022; 14(5):1286. https://doi.org/10.3390/cancers14051286

The author contributed to the study´s concepts and study design, including the literature search using three databases (meta-analysis), the data collection, analysis, and interpretation. Performed all laboratory technical optimization and procedures. Conducted the preparation of the manuscript, in its editing and reviewed entirely.

Fulfilling the Portuguese Law, the Candidate declares to have actively participated in the collection and study of all material and data included in this Thesis, having written all papers.

This Thesis and the all the work included in it have been approved by the:

- Ethics Research Committee of NOVA Medical School (nº 70/2019/CEFCM)
- Scientific board of Instituto Português de Oncologia de Lisboa and the Ethics Committee (UIC/1168)
- The animal model experiments were performed and approved by the University of Trás-os-Montes and Alto Douro Ethics Committee (Approval Number 10/2013) and by the Portuguese Veterinary Directorate (Approval Number 0421000/000/2014).

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF GRAPHICS

ACKNOWLEDGMENT

I would like to express my sincerest gratitude to my supervisors for the confidence and encouragement during this project's journey. An exceptional experience in your guidance and expertise, I am grateful for the accepted challenge and commitment to empower.

Professor Ana Felix
Professor Rui Gil da Costa

The host institution where the study was conducted, thankful for the opportunity to fulfil its mission and values for the development of scientific work in the treatment, study, and research of cancer.
Instituto Português de Oncologia de Lisboa Francisco Gentil

To the team of the Virology Laboratory IPOLFG, I would like to thank for all the technical and logistical support, and dedication to make this project possible.
Dr Luís Martins, and Dr Mário Cunha

I would like to thank all my co-authors for their ongoing support, expertise, and encouragement to the advancement of scientific work in oncogenic viruses and contributing to the enrichment on cancer diagnosis, prevention, and surveillance.
Dr Hugo Estibeiro, Dr Pedro Montalvão, Dr Miguel Magalhães, and team from Otolaryngologist Medicine Department from IPOLFG, Dr Susana Esteves from IPOLFG Scientific Research Unit, Dr Fernanda Silva, and Dr Miguel Rito from IPOLFG Pathology Department, Dr Rui Medeiros Group from Molecular Oncology and Viral Pathology from Portuguese Oncology Institute of Porto, Dr Luis Vieira, Dr João Paulo Gomes, and Dr Nuno Verdasca groups from Instituto Nacional Doutor Ricardo Jorge, and to the Group Centre for the Research and Technology of Agro-Environmental and Biological Sciences (CITAB) Inov4Agro from the University of Trás-os-Montes e Alto Douro (UTAD).

THESIS OUTLINE

This thesis comprises six chapters that reflect the different stages of the development of the proposed project:

Chapter 1 - Presents the background concepts that lead to the work undertaken in this thesis and describes its intention within the scope defined by the general and specific objectives of this work.

Chapter 2 – Contains the presentation of the research purpose of this thesis. It describes the study's hypothesis and its specific objectives.

Chapter 3 – Briefly, compiled the materials and methods provide the technical specifications used in our study, beginning with the description of the patient's cohort and the characterisation of tissue samples from the K14HPV16 animal model. Following the description of the technical methodology used and necessary for the global analysis, a particular overview of the design of NGS workflow construct that was needed to develop by us for the determination of HPV16 variants.

Chapter 4 – Divided in two main parts presents the descriptive analysis of the clinicopathological characteristics of the HNSCC HPV16 patient's cohort, in the first part, followed by a second part that compile in four original publications the revisions of the literature and the experimental work.

Chapter 5 - Presents a general discussion which intends to encompass the original findings with the published literature overview, clinical implications, and the contribution of this work to the knowledge of HPV role in squamous cell carcinoma of head and neck cancer.

Chapter 6 - Highlights the main conclusions to be drawn and identifies weaknesses and strengths that are helpful to discuss future research perspectives.

SUMÁRIO

O Carcinoma da cabeça-pescoço é considerado mundialmente o sexto cancro de maior incidência, responsável por 450.000 mortes/ano sendo previsto que a incidência/mortalidade aumentem na próxima década. Estes tumores aparecem em várias localizações do trato aero-digestivo superior, a maioria é originada no epitélio escamoso que correspondem a 90-95% de todos os carcinomas pavimento celular da região de cabeça-pescoço. Apesar da sua origem comum no epitélio escamoso, os tumores das diferentes localizações da região cabeça-pescoço apresentam elevada heterogeneidade nas características fenotípicas, etiológicas e biológicas. Os fatores de risco registados são o consumo de tabaco e álcool, descritos como as causas mais comuns para o aparecimento destes tumores, e a infeção pelo Papilomavírus Humano como factor de risco associado aos tumores que surgem na orofaringe. O projecto desenvolvido propôs-se obter informações sobre alguns aspectos específicos do carcinoma pavimento celular da região cabeça-pescoço através de uma análise molecular detalhada de uma série de casos. O desenho do estudo foi desenvolvido em tarefas distintas e estruturadas de forma a conseguir: 1- obter os dados relativos às suas variáveis clínico-patológicas, 2- identificar as diferentes variantes do HPV16 presentes e 3- determinar a frequência de mutações do gene *PIK3CA*.

A coorte incluída foi selecionada de um universo de pacientes com diagnóstico de carcinoma pavimento celular da região de cabeça-pescoço no Instituto Português de Oncologia Francisco Gentil de Lisboa IPOLFG, sendo composta maioritariamente de pacientes do sexo masculino com cancro na orofaringe, HPV-positivo e com hábitos de consumo activos. Adicionalmente foi estudada uma série de amostras de um modelo murino de carcinogénese orofaríngea induzida pelo HPV16, para identificar a variante do vírus presente neste modelo experimental. A sequenciação do genoma completo do HPV16 foi realizada por metodologia *Next Generation Sequencing*, desenvolvida a partir de *primers* anteriormente descritos, que foram adaptados para optimizar um fluxo de trabalho laboratorial que permite identificar variantes do HPV16 de tumores provenientes de múltiplos tecidos. Por último, a frequência de mutação *PIK3CA* fortemente associada aos tumores pavimento celular da região cabeça-pescoço foi calculada com base na detecção da presença de quatro mutações diferentes (E545D, E545K, E542K e H1047L) por metodologia de PCR em tempo real.

Os principais dados obtidos expressam a caracterização clínica da patologia na casuística de pacientes, descrevem as diferentes linhagens do HPV16 identificadas no total da coorte de pacientes, bem como o perfil mutacional por cada gene do HPV16. A análise filogenética mostra a linhagem A como a mais prevalente na nossa coorte.

No estudo exploratório das respectivas linhagens não foi possível estabelecer nenhuma associação de confiança com as variáveis clinico-patológicas, incluindo as variáveis de sobrevivência, devido ao reduzido número de casos incluídos na análise (n=35). Na análise comparativa do genoma dos 35 pacientes foram reportadas 243 variações nucleotídicas, 84 das quais a alteração resulta num aminoácido diferente. O perfil mutacional apresentado foi concordante com o previamente descrito, em 41% (100/243) das variações nucleotídicas foram publicadas por diferentes autores na patologia do colo e/ou na patologia região de cabeça-pescoço. As restantes SNV contribuem com novos dados em diferentes genes do genoma viral, incluindo os genes que até à data se encontram menos explorados, nomeadamente o gene E4, E5 e L2. Da análise exploratória do significado biológico o total de 23% (23/100) variantes foram descritas, a maioria já estudada no carcinoma do colo do útero.

No geral, os nossos resultados revelam existir uma maior variabilidade, precisamente nos genes E5 e L2, que não estavam tão estudados, e confirmaram que o gene mais conservado é o E7, o que está de acordo com o reportado previamente por outros autores.

Os nossos resultados no estudo da frequência de mutações do gene *PIK3CA* sustentam a elevada ocorrência destas mutações nos tumores de cabeça-pescoço. Na nossa coorte identificamos a presença de pelo menos uma mutação no gene *PIK3CA* em 39% dos pacientes incluídos e a mutação E545D foi a mais frequentemente detectada.

A análise do genoma do modelo murino que efetuámos permitiu pela primeira vez identificar qual a é linhagem do HPV16 presente neste modelo que é a sublinhagem A1 (Europeia). Este dado novo é muito interessante porque esta é a mais prevalente em doentes e que reforça a hipótese de que esta é uma linhagem com elevado potencial oncogénico na orofaringe, tornando este modelo ainda mais interessante.

Globalmente, os nossos dados corroboram os previamente reportados por outros grupos, destacando a importância da análise molecular para melhor compreender a distribuição das variantes do HPV16 e das mutações do *PIK3CA* e a sua possível influência na carcinogénese, diagnóstico e abordagem clínica no cancro de cabeça e pescoço.

PALAVRAS-CHAVE: CARCINOMA, OROFARINGE, HPV16 VARIANTES, PIK3CA, GENOMA, NGS

ABSTRACT

Head and Neck Carcinoma is considered the sixth most common cancer worldwide and is responsible for 450,000 deaths/year. These tumours appear in various locations of the upper aero-digestive tract. Most of these neoplasms originate in the squamous epithelium and correspond to 90 to 95% of all squamous cell carcinomas of the head and neck region. Despite their common origin in the squamous epithelium, tumours from different locations within the head and neck region show high heterogeneous phenotypic, etiological, and biological characteristics. The registered risk factors are tobacco and alcohol consumption, described as the most common causes for the appearance of these tumours, and Human Papillomavirus (HPV) infection as a risk factor associated with tumours that arise in the oropharynx.

The developed project aimed to obtain information about some specific aspects of squamous cell carcinoma of the head and neck region through a detailed molecular analysis of a case series. The study design was developed in distinct and structured tasks to: 1- obtain data related to clinical and pathological variables, 2- identify the different HPV16 variants present and 3- determine the frequency of *PIK3CA* gene mutations.

The cohort included in the study was selected from a universe of patients diagnosed with squamous cell carcinoma of the head and neck region at the Instituto Português de Oncologia Francisco Gentil de Lisboa IPOLFG, being composed mostly of male patients with oropharyngeal cancer, HPV-positive and with active consumption habits. Additionally, a series of samples from a murine model of oropharyngeal carcinogenesis induced by HPV16 was studied to identify the virus variant present in this experimental model. The sequencing of the complete genome of HPV16 was performed using Next Generation Sequencing methodology, developed from previously described primers, which were adapted to optimize a laboratory workflow that allows the identification of HPV16 variants from multiple tissues. Finally, the frequency of the *PIK3CA* mutation strongly associated with squamous cell carcinoma of the head and neck region was calculated based on the detection of the presence of four different mutations (E545D, E545K, E542K and H1047L) by real-time PCR methodology.

The main data obtained express the clinical characterization of the pathology in the series of patients, describe the different HPV16 strains identified in the total

cohort of patients, as well as the mutational profile for each HPV16 gene. Phylogenetic analysis shows lineage A as the most prevalent in our cohort. In the exploratory study of the respective strains, it was not possible to establish any reliable association with clinicopathological variables, including survival variables, due to the small number of cases included in the analysis (n=35). Regarding the genomic comparative analysis of the 35 patients, 243 nucleotide variations were reported, 84 of which integrate an amino acid alteration. The mutational profile agrees with that previously described; a total of 41% (100/243) single nucleotide variation (SNV) were published by other authors based on cervix pathology and/or in the pathology of the head-neck region. The remaining SNV that were not described in the systematic review contribute with new data in different genes of the viral genome, including those genes that until now are less explored, such as the E4, E5 and L2 gene. From the exploratory analysis of biological significance, a total of 23 variants were described, the majority studied in cervical carcinoma. Overall, the greatest variability was observed in the E5 and L2 genes that were not previously studied. Oppositely, the most conserved gene was E7, as previously reported by other authors.

The frequency of *PIK3CA* gene mutations found in our cohort support the high occurrence of these mutations among head and neck tumours. In our cohort, we identified the presence of at least one mutation in the *PIK3CA* gene in 39% of the patients included, and the E545D mutation was the most frequently detected. It should also be noted that the data from the analysis of the murine model genome allowed for the first time to identify the HPV16 lineage present as A1 (European) sublineage, which was the most prevalent in patients, supporting the hypothesis that this is a lineage with high oncogenic potential in the oropharynx and making this model even more interesting.

Overall, our data corroborate those previously reported by other groups, highlighting the importance of molecular analysis to better understand the distribution of HPV16 variants and *PIK3CA* mutations and their potential influence on carcinogenesis, diagnosis and clinical approach in head and neck cancer.

KEYWORDS: HNSCC, HPV16 VARIANTS, PIK3CA, WHOLE-GENOME, NGS, SNVS

ACRONYMS AND ABBREVIATIONS

AA - Asia-American HPV16 Variant
AF - African HPV16 Variant
AJCC - American Joint Committee on Cancer
AS - Asian HPV16 Variant
A260- Absorbance at 260 nm
A280 - Absorbance at 280 nm
bp - Base pair
cDNA - Complementary DNA
CDKN2A - Cyclin-dependent kinase inhibitor 2A
CT - Chemotherapy
CRT - Chemoradiotherapy
DFS - Disease-free survival
DNA - Deoxyribonucleic acid
E - European HPV16 variant
E HPV - Early protein
EU - European Union
FFPE - Formalin-fixed paraffin-embedded
H&E - Hematoxylin and eosin staining
HN – Head and Neck
HNSCC - Head and Neck Squamous cell carcinoma
HPV - Human Papillomavirus
HNSCC-HPV - Head and Neck Squamous cell carcinoma associated to Human Papillomavirus
HNSCC-nonHPV - Head and Neck Squamous cell carcinoma without Human Papillomavirus association
IARC - International Agency for Research on cancer
ICTV - International Committee on Taxonomy of Viruses
ISO - International Organization for Standardization
K14HPV16 - Early HPV16 genes under the control of the cytokeratin 14
LCR - Long control Region
MUT - Mutated
mRNA - Messenger RNA
NA - North American HPV16 Variant

NGS - Next-generation sequencing

NTC – No Target Control

O.C.T. - Compound Optimal cutting temperature compound

OD - Odds ratio

ORF - Open reading frame

OS - Overall survival

PCR - Polymerase chain reaction

PIK3CA - Phosphatidylinositol-4,5-biphosphate 3-kinase

PI3K - Phosphoinositide 3-kinases

PI3-K-AKT - Phosphatidylinositol-3-kinase-AK

PIP2 - Phosphatidylinositol-(4,5)-bisphosphate

PIP3 - Phosphatidylinositol-(3,4,5)-trisphosphate

PRB - Retinoblastoma protein

PTEN - Phosphatase and tensing homolog

PV - Papilloma virus

P16 - p16INK4a protein, cyclin-dependent kinase inhibitor 2A, or CDKN2A

P53 – Tumour suppressor protein 53

qPCR - Real-time PCR

RNA - Ribonucleic acid

RT – Radiotherapy

SR – Systematic review

TNM - Tumour, lymph nodes, metastasis

TCGA - Tumour Cancer Genome Atlas

UICC - Union for International Cancer Control's

USA - United States of America

WT - Wild type

WHO - World health organization

FOREWORD

Head and neck carcinoma is the sixth leading cancer by incidence worldwide and is responsible for more than 800,000 new cases yearly. HPV is present in approximately 35% of HNSCC, emerges in the lingual and palatine tonsils, and HPV16 is the most detected type. The epidemiological, etiological, and molecular data suggest that intra-type HPV variants are biologically distinct and may be associated with different risks in uterine cervical cancer progression. Nonetheless, HPV16 variants may also play an important role in head and neck carcinogenesis, but few have established their relevance. The genomic profile of HNSCC published by the Cancer Genome Atlas in 2015 broadened the development of mutation studies and highlighted a high frequency of changes. The identification of biomarkers to predict response to therapy will allow the selection of patients more likely to respond to new therapeutic combinations making a valuable contribution to clinical treatment strategies. Essentially, it is critical to refine the mutational profile of HPV-positive and HPV-negative within HNSCC patients. This project addresses a deep molecular analysis of a Portuguese cohort of HNSCC patients and explores its correlation with clinicopathological variables and their potential contribution to clinical practice.

CHAPTER 1

BACKGROUND

1. HUMAN PAPILLOMAVIRUSES

The Papilloma Virus (PV) a member of the *Papillomaviridae* family (De Villiers *et al.*, 2004), are ubiquitous viruses that infect the epidermis and mucosa of several mammals, including humans. PV infections are usually transient and asymptomatic, with no relevant clinical manifestations. However, they can cause benign tumours (warts, papilloma) that usually regress or establish persistent infections that can lead to tumours (zur Hausen, 1996). There are over 300 distinct types of PVs isolated. About 200 can infect humans – human papillomavirus (HPV), where 40 types infect the anogenital region (World Health Organization, 2010). Only a small group is associated with the development of neoplastic pathologies. Although the wide number of HPV types isolated and tropism differences, they all share the same genome arrangement (Bernard, Calleja-Macias and Dunn, 2006).

1.1 STRUCTURE AND REGULATION OF THE VIRAL GENOME

The virus particle is small (55nm-diameter), non-enveloped, suitable only for a protein capsid of icosahedral symmetry that encloses the viral DNA genome (Münger *et al.*, 2004). The viral DNA, with a length of 7,906 base pairs (bp), is circular, double-stranded and structured into three functional regions: Early Region, Late Region, and Long Control Region. These three regions are separated by two polyadenylation sites (pA) as shown in Figure 1 (Zheng and Baker, 2006). The genetic information is distributed among the eight coding sequences for proteins classified as early (E) and late (L) according to their temporal expression. The Early region occupies more than 50% of the virus genome and encodes six open reading frames (ORFs): E1, E2, E4, E5, E6 and E7. The late region contains the L1 and L2 ORFs that encode two structural proteins. In addition, there is a long control region (LCR) situated between the E6 and L1 genes, which does not contain ORFs but includes the origin of replication (ori) as well as multiple elements (enhancers) required for transcription (Münger *et al.*, 2004).

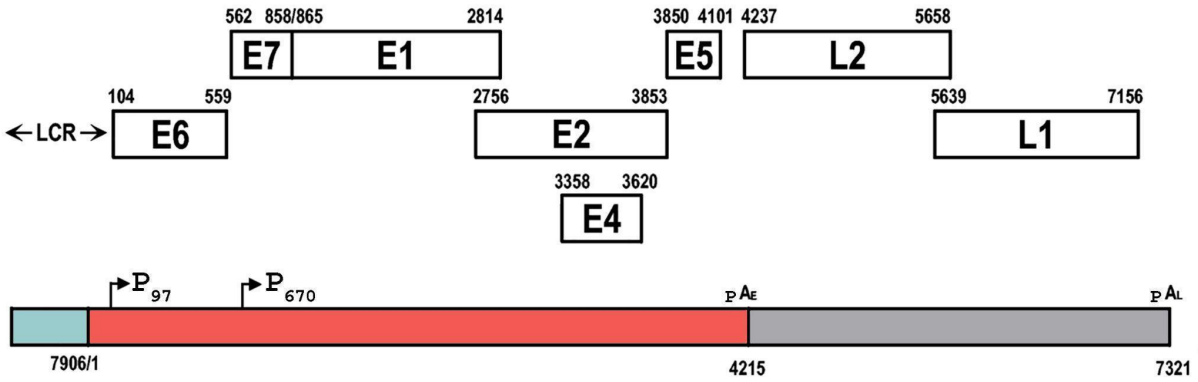


Figure 1. HPV16 complete genome representation adapted from (Zheng and Baker, 2006). The coloured line illustrates the linear form of the virus genome (≈8000 pb), presenting the three functional regions LCR (blue), Early (red) and Late (grey); the promoters P (arrows) and polyadenylation sites (pA); the early (AE) and late (AL). Above the line, white boxes illustrate the ORFs for each coding region (red and grey): the Early E6, E7, E1, E2, E4 and E5 genes; and the late genes L2 and L1. The numbers above represent the first nucleotide in the start codon and the last nucleotide position in the stop codon for each ORF.

The transcription of the genome occurs from a single strand of DNA considered to be polycistronic, with multiple introns that undergo extensive alternative RNA splicing producing multiple RNAs with several open reading frames. The transcription of the Early and Late genes occurs unidirectionally from primers and promoters. It is initiated at p97 promoter located upstream of the ORF of the E6 gene and is responsible for almost all early gene expression (Smotkin and Wettstein, 1986). Next is followed by the p670 promoter located within the ORF of the E7 gene, which is responsible for the transcription of late genes. This mechanism gives to HPV the ability to express different proteins at a given time and in the differentiation stage of the infected epithelium (Zheng and Baker, 2006). It also ensures the smooth progression of key events for the virus, such as the replication cycle, along with immune evasion.

1.2 INFECTION AND REPLICATION CYCLE

HPV accesses to the basal layer through micro abrasions in the epithelium and binds to host cells using cell surface molecules (Shafti-Keramat *et al.*, 2003; Patterson, Smith and Ozbun, 2005). The first events are induced by the interaction between the L1 structural protein and the cell surface receptor (heparin sulphate

proteoglycan) (Giroglou *et al.*, 2001). Followed by the interaction of the L2 protein with the cell surface recruiting a second receptor that allows the entrance of the virus (Kawana *et al.*, 1999), which as a non-enveloped particle, adopts mechanisms through clathrinid-mediated endocytosis (Day, Lowy and Schiller, 2003). Once inside the cell, the viral particle is transported to the nucleus, not before losing the capsid first. This process allows the uncoating of the circular genome protected in the endosome (Merle *et al.*, 1999; Nelson, Rose and Moroianu, 2002) and is mediated by the L2 protein, which can disrupt the endosome membrane, and the viral genome is released into the cytosol. Then the viral genome (8000 bp) diffuses through the cytosol, mediated thereby L2 protein, which interacts with both actin and tubulin (Yang *et al.*, 2003). Upon reaching the nucleus, the viral DNA enters through the nuclear pores.

The PVs, as DNA viruses, replicate and are assembled exclusively in the nucleus. In tissues, the replication begins in undifferentiated cells, located in the basement layer, involving host cell factors that interact with the LCR region and thus initiate the transcription of viral genes (Burd, 2003). The first proteins to be expressed are E1 and E2, necessary for viral DNA replication and to establish genomes with low copy numbers of 20-100 copies per cell (Ruesch, Stubenrauch and Laimins, 1998; Ozbun, 2002). The viral genomes replicate simultaneously with the replication of cellular DNA (You *et al.*, 2004). The infected cell, when it divides, originates new cells that remain in the basal layer, thus serving as a reservoir for the virus. Just the same these infected cells proceed with expression and replication in a tightly controlled and regulated manner by keratinocyte differentiation (zur Hausen, 1996).

The complete viral genome encodes only 8 to 10 viral proteins, which require host cell factors to regulate viral transcription and replication. The E1 and E2 proteins accompany this process, by recruiting the cellular machinery involved in DNA replication. The E1 protein is an ATP-dependent helicase which opens the replication origin, allowing viral genome replication to begin (Wilson *et al.*, 2002), while E2 interact as a transcription factor through its functions to regulate the expression of early ORF promoters (Ham *et al.*, 1991). With the epithelial differentiation, cells stop dividing and the differentiated keratinocytes do not express the crucial factors involved in DNA replication. The virus solves this problem by using viral oncoproteins E6 and E7 that act on the regulation of the cell cycle progress and blockage of apoptosis of those cells that enter the cell cycle

inappropriately (Longworth and Laimins, 2004). The E7 binds to the Rb protein (pRB), releasing the E2F transcription factor, which activates the cellular machinery necessary for DNA replication. The E6 protein indirectly regulates the expression of p53, blocking the interaction of this protein in terms of cell cycle control and subsequent apoptosis of the infected cell (zur Hausen, 1996), thus allowing the synthesis of viral DNA. The virus switches to a rolling circle DNA replication mode and amplifies its DNA to a high copy number (Burd, 2003). The remaining early genes are also expressed during this production cycle, as the E5 protein plays a cooperative role with the E7 protein in cell cycle deregulation (Conrad, Bubb and Schlegel, 1993; Faulkner Valle and Banks, 1995), while the E4 protein is the last one of the early expressed in the suprabasal layer of the epithelium. This protein interacts with the keratin filaments on the cytoplasm to collapse them (Doorbar *et al.*, 1991) and turn them permissive to the release of virions (Raj *et al.*, 2004). Capsid assembling and virion formation occur in the nucleus, essentially mediated by structural L1 and L2 proteins, which are synthesized in the cytoplasm and later translocated to the cell nucleus. The E2 protein also appears involved in virion formation, as it plays an active role in viral genome encapsulation. Finally, virions are released from differentiated keratinocytes on the mucosal surface (Florin *et al.*, 2002) (Figure 2 Replication cycle).

This overview scheme is called the productive cycle and results in benign lesions controlled by the immune response to HPV infection in immunocompetent individuals (Herrero *et al.*, 2000). An example of this is HPV uterine cervical infections, that are common in sexually active young women, but the vast majority resolve spontaneously within 1 to 2 years (Evander *et al.*, 1995; Ho *et al.*, 1998; Moscicki *et al.*, 1998). Most of these women are seroconvert, and HPV is no longer detectable (Carter *et al.*, 1995, 1996; Dillner, 1999). Although restricted to few women, an inadequate immune response can lead to viral persistence and progress to malignant lesions (Chirgwin *et al.*, 1995). In general, the replicative cycle is very long and may take months, from the infection of the basal layer cells to the appearance of lesions on the mucosal surface (Oriel, 1971; Koutsky, 1991). During this process, the virus does not produce viremia, inflammation or cell death, few or no proinflammatory cytokines are released, and there is a significant decrease in Langerhans cells recruited in the infected mucosa relative to the normal mucosa (Hawthorn *et al.*, 1988; Hubert *et al.*, 1999). This makes the presentation of viral antigens to the immune system difficult. This strategy can result in chronic and

persistent infections, as the immune system ignores its existence for a long time. The immunosuppression appears to reflect an increased risk of progression from subclinical to clinical lesions (Chirgwin *et al.*, 1995).

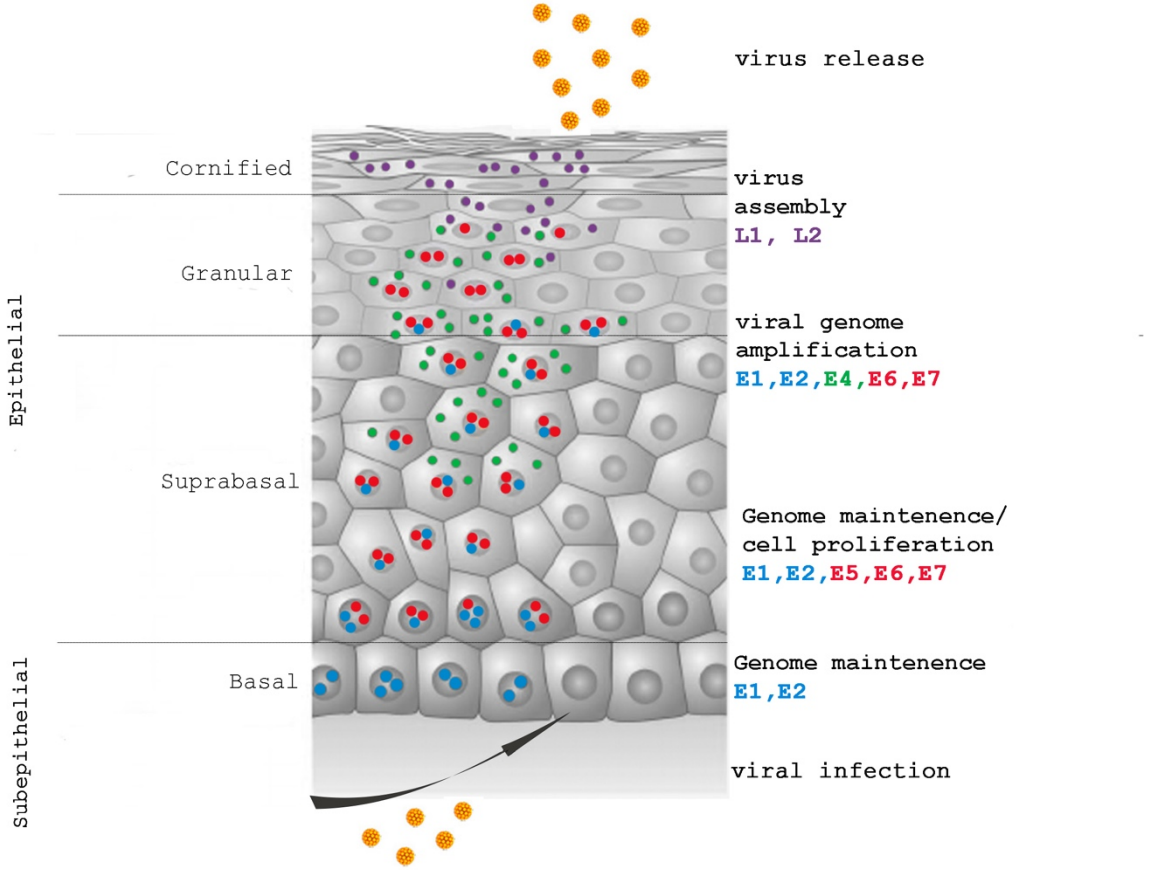


Figure 2. Representation of the HPV replication cycle. Adapted illustration (Doorbar, 2006) of the differentiated epithelium with the expressed viral markers represented by different colours. HPV particles access the basal layer, and the virus is maintained as an episome at a low copy number. The E5, E6 and E7 proteins (red) are expressed from the early promoter in the lower epithelial layers. The E4 expression (green) is triggered before E6/E7 expression ends. The E4 cells expressing with E7 are expected to contain all the early transcripts to facilitate genome amplification. Expression of the L1 and L2 virus capsid proteins (purple) completes the full genome expression and occurs in a subset of cells that express E4. Apparently, at different stages, the HPV life cycle is supported by the combination of expressed markers.

1.3 CARCINOGENESIS

In the natural history of HPV infection, tumorigenesis occurs only in very few per cent of people infected with HPV. Most are transient infections where the host's immune system resolves, within an average period of 8 to 12 months. However, in some cases, a persistent infection occurs by continuously replication of HPV in the differentiated epithelium (Burk, Chen and Van Doorslaer, 2009). One possibility to explain this, is that specific types of HPV are more prone to persistent infection than others, and this ability is genetically encoded and linked to pathogenicity (Burk, Chen and Van Doorslaer, 2009). Although persistent infections with a high-risk HPV are important and necessary but seems not able by itself a sufficient factor in the development of carcinoma. For a neoplastic transformation, it is also necessary the deregulation of viral oncogenes E6 and E7 to occur. Unlike productive infections, even in high-risk HPVs, the viral oncogenes are expressed at low levels and not enough to transform infected cells.

In conclusion, transformation and tumour progression are rare events caused by a restricted number of HPVs and occur more frequently when the virus integrates into the cell genome (zur Hausen, 1996). Apparently, the virus that appears in its episomal form (circular DNA) changes by disruption to a linear form that can integrate into the host genome and gains new abilities. In this way, it can produce changes in cellular functions favouring the replication of the viral particle (Oyervides-Muñoz *et al.*, 2018). During this integration process, some of the viral genes can be affected. Different breakpoints are described throughout the viral genome, many located within coding zones, which induces the loss of the gene in whole or part, compromising the expression of crucial proteins (Doorbar, 2006). The most common breakpoint is in the E2 gene. As this protein can act as a transcriptional repressor of early genes, by binding to viral promoters in the LCR region, its expression results in reduced transcription (zur Hausen, 1996). Therefore, E2 disruption leads to the loss of negative regulation, increasing HPV oncoproteins expression either by increasing the deregulation of their promoter or increasing the stability of their mRNA (Baker *et al.*, 1987; zur Hausen, 1996; Bechtold, Beard and Raj, 2003). Carcinoma induction is not an asset to the virus and seems to be a consequence of the virus's act in the deregulation of the cell cycle (Burd, 2003).

The trigger for the malignancy derived from the overexpression of these viral oncogenes will induce an increase in cell proliferation. The E6 interacts with p53 and induces its proteolytic degradation through the ubiquitination pathway. Specifically, by the efficient binding and consequent degradation of p53 based on the interaction with the cellular protein E6AP, which is part of the ubiquitin proteolytic complex (Zur Hausen, 2002). The sequence of events appears to be as follows: E6 binds to E6AP and recruits p53. Since E6AP has ubiquitin ligase activity, it recruits E6 (part of the ubiquitin complex), which ubiquitinates p53, and leads to its degradation. In host cells, pRB regulates the transition from the G0 phase through the G1 phase to the S phase. When hypo-phosphorylated protein it becomes inactive, and releases transcription factors such as E2F in turn to active promoters in genes necessary for DNA synthesis. The E7 oncoprotein's ability to bind to the RB protein when it is inactive (hypo-phosphorylated) causes the release of E2F (Zur Hausen, 2002). The elevated levels of E6 and E7 disestablish the chromosomes and promote an accumulation of mutations, giving cells the malignant phenotype.

Integration of viral DNA into the host genome is random, but it frequently occurs near genomically unstable zones (Liu *et al.*, 2016) at “fragile” sites that by altering the expression patterns of relevant cellular maintenance pathways and induce cellular alterations that give rise tumours (Oyervides-Muñoz *et al.*, 2018). Essentially, three possible ways are described: The first is when the integration of viral DNA takes place in host’s tumour suppressor genes, this integration into host DNA inactivates those genes and leads to uncontrolled growth (Zhao *et al.*, 2016). The second occurs at sites close to cellular oncogenes and may result in their activation. The increase in c-*myc* and c-*ras* gene expression has been related to this process. A third way can be observed by inter and intrachromosomal rearrangements between chromosomes 3 and 13 close to the site of HPV integration affecting the expression of genes resulting in increased cell proliferation (Rusan, Li and Hammerman, 2015). All those three mechanisms collectively favour cellular transformation supporting tumour growth (Androphy, Schiller and Lowy, 1985; Zur Hausen, 2002).

Transformation doesn't involve always viral oncogenes deregulation by viral E6/E7 overexpression mechanism. In fact, the viral genome also breaks in regions other than E2 (these other breaks are not related with E6/E7 overexpression, yet) and episomal form of the HPV virus instead of its integrated form can be associated

with tumoral progression (Oyervides-Muñoz *et al.*, 2018). These findings suggest that the HPV genome encloses different pathways for tumorigenesis. However, these are pathways that, although described, are not yet so well explained, and some of the elements that support the process of cell proliferation and transformation remain unknown. Nevertheless, it is relevant to discuss the genomic heterogeneity that leads to the manifestation of these tumours once their malignant transformation may attribute to specific variants of the HPV genome (Bernard, Calleja-Macias and Dunn, 2006; Burk, Chen and Van Doorslaer, 2009; Cornet *et al.*, 2012; Jackson *et al.*, 2014).

1.4 CLASSIFICATION

The different HPVs, referred to as types, share the same genomic organization and, at least, five homologous genes. The taxonomic classification of HPV types based on its nucleotide sequence proposed by E. Villiers is a phylogenetic classification (De Villiers *et al.*, 2004). Alternatively, HPV types can also be classified regarding their oncogenic potential (potential risk to induce cervical carcinoma) that is termed as epidemiological classification (Muñoz *et al.*, 2003). Molecular biology advances produced a rapid increase in the number of isolates and found few hundred new HPVs. This large number of isolates imposes the use of a improved update taxonomic classification (Van Doorslaer, 2013). The main objective of this classification is to establish the relationship between the different types and its phylogenetics' with the biological and pathogenic properties of the virus. As the L1 ORF gene is the most conserved one, it is therefore, eligible for the classification of new types. Similar HPV types are phylogenetically grouped into the same genus. Within the genus, those that most resemble each other are called species that may have distinct genomic sequences but similar biological and pathogenic properties. To consider as a new HPV genotype it implies that the whole genome clone differs from the closest phylogenetically identified genotype by more than 10% from the ORF region of the L1 gene. If the homology only diverges between 2-10%, it is classified as a subtype, and if less than 2% it defines a variant (De Villiers *et al.*, 2004). Currently, 241 genotypes have been isolated (de Villiers and Gunst, 2009; Bernard *et al.*, 2010; Van Doorslaer, 2013) and classified into 37 genera according to phylogenetic classification (Figure 3).

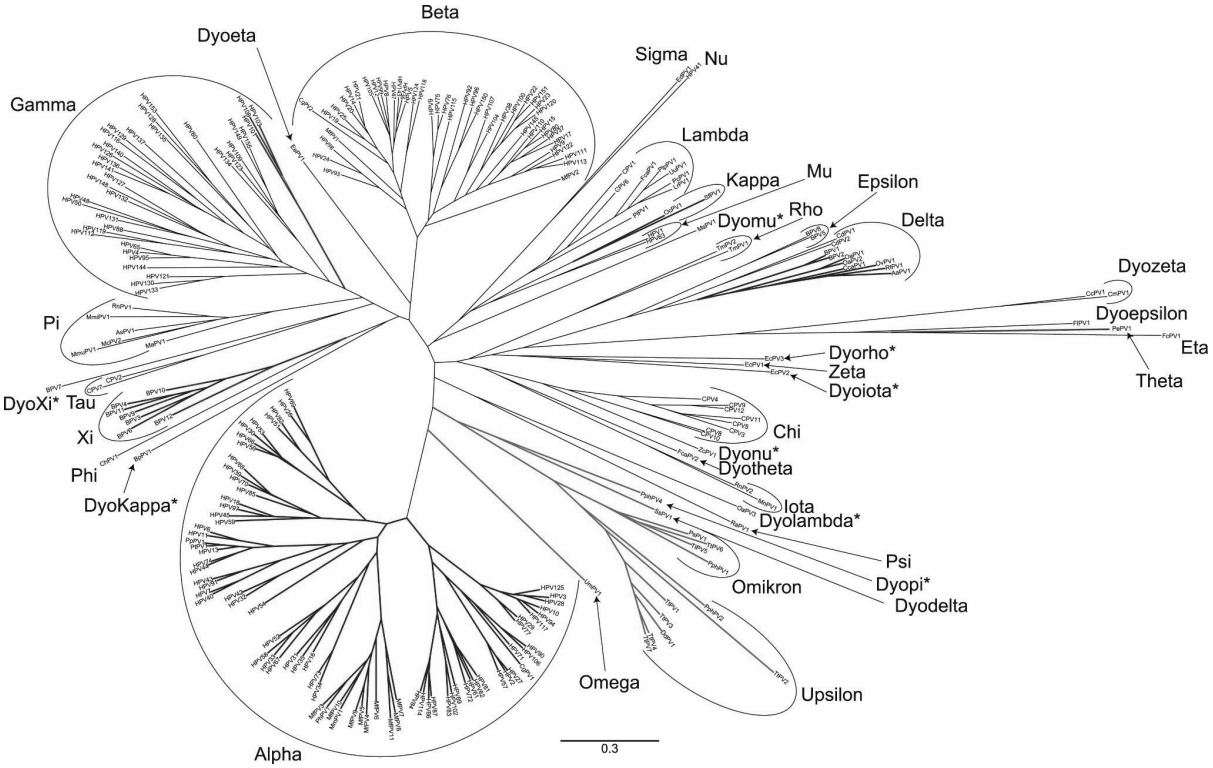


Figure 3. Phylogenetic tree of the 241 papillomavirus genotypes currently described (Van Doorslaer, 2013). The terminal number on each arm identifies a type of HPV. The larger outer semi-circles represent the genera, named according to (De Villiers *et al.*, 2004; Bernard *et al.*, 2010). Genera marked with an asterisk have been proposed to the International Committee on Taxonomy of Viruses ICTV (Van Doorslaer, 2013).

Regarding the epidemiological classification, several studies have determined the oncogenic potential of different types (Muñoz *et al.*, 2003), using the evaluation of uterine cervical carcinoma histological samples. Fifteen HPVs (16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, 68, 73 and 82) were consistently classified as high-risk, three as possible high-risk (26, 53 and 66) and twelve as low risk (6, 11, 40, 42, 43, 44, 61, 70, 72, 81 and CP6108). In general, the two classifications are concordance (Figure 4) (Muñoz *et al.*, 2003).

In the HNSCC, most HPV types belong to the high-risk group, including HPV16, HPV18, HPV39, and HPV45 (Huertas-Salgado *et al.*, 2011), with HPV16 being the most detected type, representing 90% of all cases of HNSCC-HPV (LeConte *et al.*, 2018). Significantly is a higher proportion than cervical cancer, accounting for just over 50% of cases (Burk, Harari and Chen, 2013).

		Epidemiological classification	
		HPV high-risk Type	HPV low risk Type
Phylogenetic classification	HPV high-risk Type	16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, 68, 82, 26, 53, 66	70
	HPV low risk Type	73	6, 11, 40, 42, 43, 44, 54, 61, 72, 81, CP6108

Figure 4. Comparison of Epidemiological and Phylogenetic classification. The colour diagram presents the Taxonomy by phylogenetic classification and epidemiological classification, with the distribution of the HPV Types into high- (red) and low-risk (green) groups. Above all, except (yellow) HPV 70 and HPV73, observe good agreement between the two classifications. This correlation is helpful to predict whether a new genotype may be of a high-risk sequence (Muñoz *et al.*, 2003).

1.5 HPV VARIANTS

There is also clear evidence of a wide sequence diversity within each HPV type (De Villiers *et al.*, 2004; Burk, Harari and Chen, 2013), where the genome sequencing revealed intra-type variants with genetic differences ranging between 0.5 and 1% (Ho *et al.*, 1991; Mirabello *et al.*, 2016). HPVs, being a double-stranded DNA virus, uses the efficient revision of the host's DNA polymerase for its replication, which avoids high mutation rates; therefore, changes in HPV genomes are acquired slowly, the random mutations that eventually occur within viral types generate the nucleotide polymorphisms contained in intratype variants (Gillison *et al.*, 2000; Muñoz *et al.*, 2003).

Initially, intratype variants were essentially studied in HPV16 and HPV18 and were named according to their geographical distribution depending on the origin and ethnicity of the infected population (Muñoz *et al.*, 2003). Thus, six distinct phylogenetic clusters were defined for HPV16: European (E), Asian (As), Asian-American (AA), African 1 (Af1), African 2 (Af2) (Bernard *et al.*, 1993; Ho *et al.*, 1993;

Yamada *et al.*, 1995) and North American 1 (NA1) (Yamada *et al.*, 1997; Burk, Harari and Chen, 2013; Jackson *et al.*, 2016). Posteriorly, it was proposed a more precise nomenclature, establishing lineages and sub-lineages (Burk, Harari and Chen, 2013) as shown in Figure 5. Regarding HPV16 genotype, whole-genome analysis allowed the characterization of four main variant lineages (A, B, C, D) and sixteen sublineages (A1, A2, A3, A4, B1, B2, B3, B4, C1, C2, C3, C4, D1, D2, D3, D4). According to these findings, lineage A grouped European cluster with sublineage A1, A2 and A3, and Asiatic cluster with sublineage A4. The lineage B encloses the African (Af1) cluster with sublineage B1 and the Lineage C encloses the African (Af2) cluster with appropriate sublineage (C1 to C4); Lineage D encloses the American cluster with sublineage D1, to North American cluster sublineage D2 and to Asian-American cluster sublineage D3 (Burk, Harari and Chen, 2013).

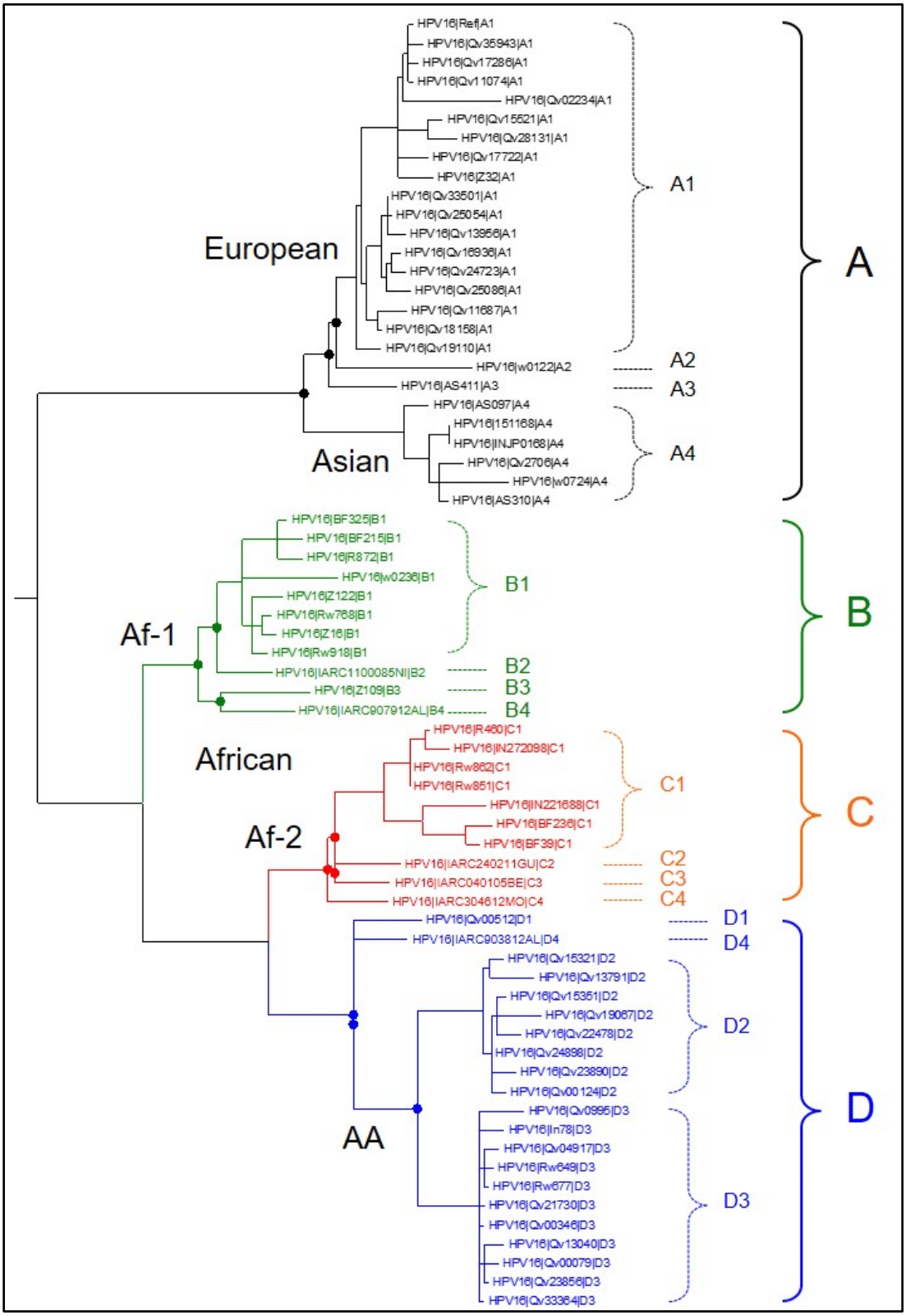


Figure 5. Phylogenetic tree from HPV16 intra-type variants. The phylogeny tree from (Burk, Harari and Chen, 2013; Mirabello *et al.*, 2018) illustrates human papillomavirus 16 (HPV16) lineages (A-D) and sublineage (A1-4, B1-4, C1-4, D1-4) relationships, highlighted on different colours, indicating the main lineage branches.

The emerging epidemiological, etiological, and molecular data suggest that the intra-type variants of HPV are biologically distinct and may be associated with different risk of cancer progression. As shown by *in vitro* studies using 3D organotypic epithelial cell cultures, the HPV16 E6 variants differ in their ability to abolish keratinocyte differentiation and to induce p53 degradation (Hoffmann *et al.*, 2004; Agrawal *et al.*, 2008; Cancer Genome Atlas Network, 2015). Therefore, it was postulated that the occurrence of nucleotide variations between intratype HPV variants reflects in their functional differences and pathogenicity (De la Cruz-Hernández *et al.*, 2005; Hochmann *et al.*, 2016; Zhao *et al.*, 2020; Hadami *et al.*, 2021). In addition, several isolates of intratype variants of HPV, previously described by different clinical studies from cohorts of carcinoma of the cervix, anal canal, oropharynx and oral cavity, contribute to a careful classification of variants of the most frequent HPV types in cancer (Bernard *et al.*, 2010; Burk, Harari and Chen, 2013; Chen *et al.*, 2015). The HPV16 T350G (L83V) variant is the most frequently found among invasive cervical cancers (Hassani *et al.*, 2015; Betiol *et al.*, 2016) and has been associated with an increased oncogenic potential than the prototype (Pakdel *et al.*, 2021), possibly by facilitating persistent viral infection (Boscolo-Rizzo *et al.*, 2009; Joseph *et al.*, 2013; Zhang *et al.*, 2015).

Following this, several studies have reported that intra-type variants of HPV16 and HPV18 alter host cells in a differently way, which translates in different risks of persistence and progression to cancer (Yamada *et al.*, 1995; Sichero *et al.*, 2007; Cornet *et al.*, 2013). Further functional studies explored differences in biological behaviour concerning the clinical evolution of patients with HPV-related cancers. Decoding the impact that these variations may be important to know the impact in the patient's prognosis and in the development of therapeutic strategies.

All these variants' impact were established for cervical cancer as the focus of most studies, compared to studies performed on other HPV-induced pathologies. In fact, in HNSCC cancer, what is known about distribution HPV16 variants resemble the knowledge acquired in cervical cancer. Indeed, authors described both share common biological features. A presenting data from a cohort study that compares genomic characteristics of HPV associated with cervical versus oropharyngeal tumours using DNA sequence analysis showed no significant differences between distribution in HPV16 variants (LeConte *et al.*, 2018) where both presenting major prevalence of the European variant.

Unlike the HPV E6 gene amplified from oropharyngeal samples reported over more nonsynonymous mutations, also for the E7 gene, no differences were found in mutation rates between the two anatomical locations (LeConte *et al.*, 2018; Zhu *et al.*, 2020). Notably, the E7 gene is conserved in both locations corroborating that the apparent restriction on E7 mutations seems to be present in the oropharynx, as well. Otherwise it is still poorly understood the potential contribution of those specific variants in cancer risk increasing, or prognosis (Gillison *et al.*, 2000) or the clinicopathological relevance of the identified variants (Hoffmann *et al.*, 2004; Badaracco *et al.*, 2007). The main reason behind this lack of conclusions in this field of knowledge may be the relatively low number of studies performed and the small size of the HNSCC HPV-associated cohorts (Hassani *et al.*, 2015). Therefore, recently this picture is starting to change, and some more studies on HPV16 variants have explored the HN region carcinomas. In 2022, a large study was published using HNSCC cases from two tertiary cancer centres. The study performed a NGS analysis of nearly 400 patients and found that specific variations have different clinical implications, namely eight HPV16 SNPs were significantly associated with worse survival (E1 gene position 1053, L2 gene positions 4410, 4539, 5050, 5254, L1 gene positions 5962, 6025, and 7173 in the upstream regulatory region) in a US population (Lang Kuhs *et al.*, 2022).

Although there is still data limitation, crucial steps have been achieved, and hopefully new studies with larger cohorts of patients and other HPV-associated tumours, leveraged due to the introduction of NGS techniques that have optimized workflows that faster and more robust sequencing results (Bernard *et al.*, 1993; Yamada *et al.*, 1995; De la Cruz-Hernández *et al.*, 2005).

2. HEAD AND NECK SQUAMOUS CELL CARCINOMA

Head and neck squamous cell carcinoma (HNSCC) are a group of tumours located in several distinct structures in the oral cavity, pharynx, and larynx subsites of the upper aerodigestive tract (Figure 6).

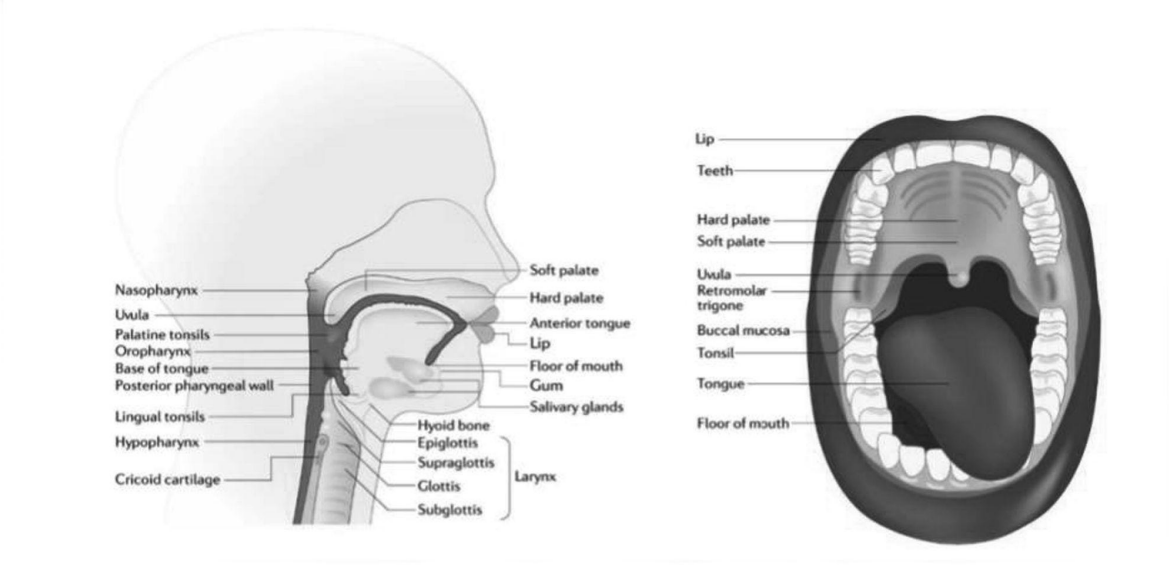


Figure 6. Illustration of the Head and Neck anatomic region adapted from (Johnson *et al.*, 2020). The cross-section (left) and front section (right) classification by each sublocation.

2.1 EPIDEMIOLOGY

Worldwide HNSCC is considered to be the sixth leading cancer regarding incidence and estimated to have an impact of ~890,000 new patients and responsible for 450,000 deaths/year (Global cancer Observatory, 2021). The incidences are rising and expected to continue to increase over the next decade (Global cancer Observatory, 2021). The age at diagnosis usually ranges between 50 and 70 years of age, and men are significantly more likely to develop the disease than women (up to a 4:1 proportion) (Bray *et al.*, 2018). Although this disease burden rate varies across countries, over the past decade, several epidemiologic studies reported higher prevalence in regions such as Southeast Asia and Australia, the majority associated with specific carcinogen substances such as chronic

consumption of alcohol and tobacco (Mork *et al.*, 2010; Chaturvedi *et al.*, 2013; Hwang *et al.*, 2015; McDermott and Bowles, 2019). Contrasting to this, the increase trending incidence, in high-income countries such as the USA and Western Europe, of HNSCC is associated with oropharyngeal HPV related cancers (Chaturvedi *et al.*, 2011; Centers for Disease Control and Prevention, 2018).

In Portugal, according to data from the National Cancer Registry, the oral cavity and pharynx are considered to be the seventh most diagnosed malignant tumour (Registo Oncológico Regional do Norte, 2010; Miranda *et al.*, 2018). Every year approximately 3,000 new cancer cases in the aerodigestive tract (Miranda *et al.*, 2018) are diagnosed, but the mortality rate from these malignant tumours is low compared to the EU average.

In conclusion, this is a relatively common human cancer, characterized by high morbidity, high mortality, and few therapeutic options, outside of standard schemes that may include surgery, chemotherapy, and radiation are available.

2.2 RISK FACTORS

The wide geographic variability in the incidence of these tumours reflects the prevalence of certain risk factors in specific regions (Johnson *et al.*, 2020). In addition, HNSCC can also occur in young patients without association with known risk factors (Rothenberg and Ellisen, 2012). In general, the most represented risk factors are the tobacco use and alcohol consumption, as well as infection with oncogenic virus such the HPV (Katabi and Lewis, 2017).

The association established between risk factors and Head and Neck subsites reported in epidemiological studies, have linked the oropharynx tumours to a previous infection with human papillomavirus (HPV) and the other HNSCC subsites, which majority are a consequence of the harmful effects of smoking and alcohol consumption and collectively referred to as HNSCC-nonHPV (Katabi and Lewis, 2017).

In these two groups (HPV and nonHPV) positive HNSCC, genetic and prognostic differences (Hwang *et al.*, 2015) reflect the critical role of the molecular profile triggered by each risk factors that contribute to its carcinogenesis (Mork *et al.*, 2010). Currently, these two groups are recognised as major carcinogenic pathways leading to HNSCC (Katabi and Lewis, 2017).

2.3 CLINICAL-PATHOLOGICAL CHARACTERIZATION

Head and neck cancer comprises very different anatomical areas and involves wide histological tissues. However, most of these tumours originate from the squamous epithelium, corresponding to 90 to 95% of all lesions (Katabi and Lewis, 2017). This cancer is characterized by its multifocal development, coupled with slow and silent growth. Despite evidence of histological progression from cellular atypia through various degrees of dysplasia, most patients are diagnosed in advanced stage disease without clinical evidence of the antecedent premalignant lesion (Nelson, Rose and Moroianu, 2002). The main clinical manifestations of the disease vary according to the site of origin of the tumour, often characterized by mucosal ulcers or tumour, dysphagia, odynophagia or otalgia, chronic cough, and neck nodes tumefaction (Johnson *et al.*, 2020).

The diagnosis is usually established based on the clinical otorhinolaryngological exam that is complemented with imaging exams to evaluate the locoregional and distant extension of the disease classified using the TNM (Tumour, Node, Metastases staging system) classification of the American Joint Committee on Cancer and the International Union for Cancer Control (Katabi and Lewis, 2017; Johnson *et al.*, 2020), and histological evaluation from tumour tissue (El-Naggar AK *et al.*, 2017). The histopathological spectrum of squamous cell carcinoma is characterized by the evaluation of cellular atypia and squamous differentiation. It defines a well-differentiated tumour, a stratified epithelium that contains mature cells organized in layers with irregular keratinization, called a “keratin pearl”. On the contrary, the poorly differentiated tumour presents mostly immature cells with nuclear pleomorphism and atypical mitoses, with a poorly defined stratification and without keratinization (Johnson *et al.*, 2020) Figure 7.

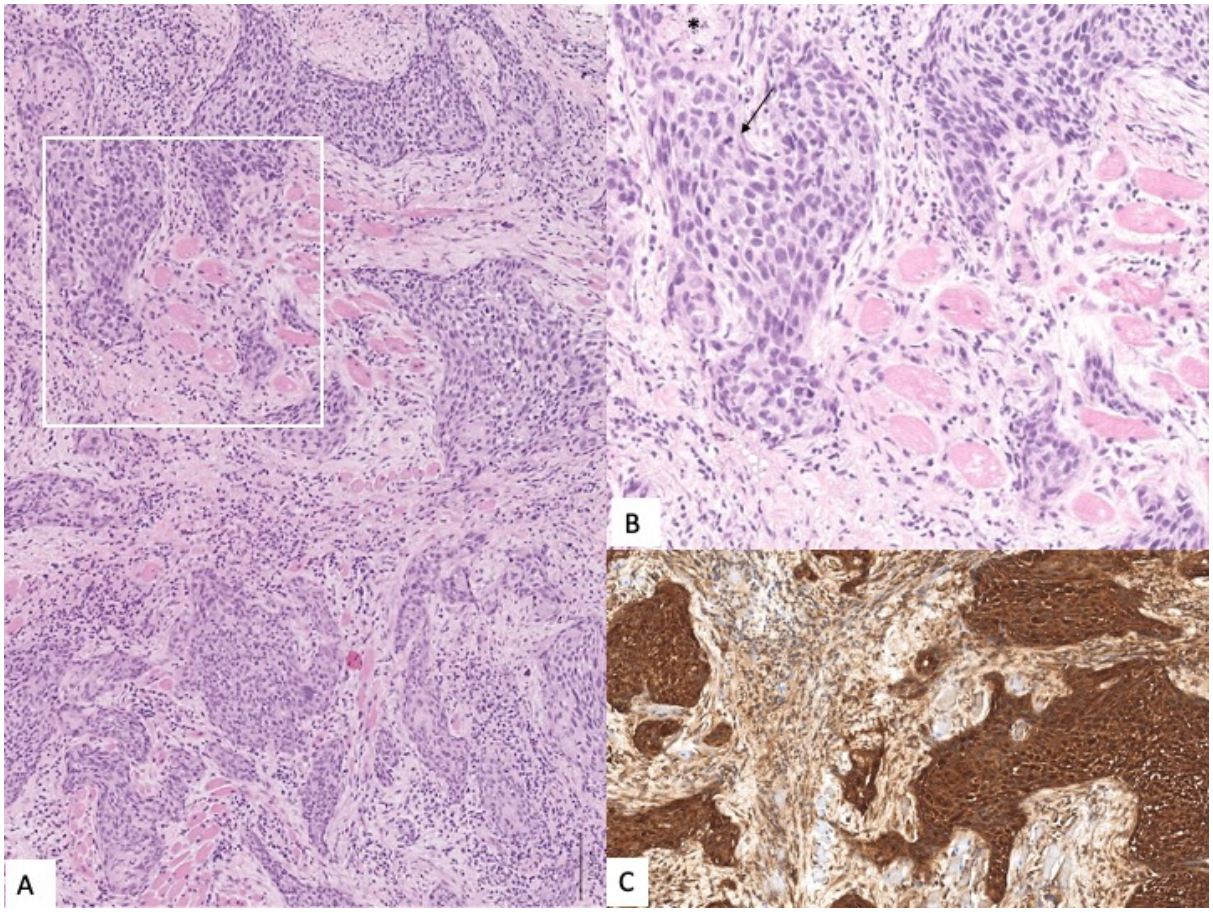


Figure 7. Squamous cell carcinoma of the oropharynx, with HPV16 DNA identified. Invasive cell nests with very irregular borders with stromal reaction (A). High power area showing poorly differentiated cells with mitotic figures (arrow) and focal necrosis (*), infiltrating striated muscle cell tissue (B); neoplastic cells show intense p16 nuclear (and also cytoplasmic) staining, a surrogate marker of HPV.

2.4 MOLECULAR CHARACTERIZATION

The HNSCC molecular analysis reveals substantial genetic instability with frequent loss or gain of chromosomal regions as noted by copy number alterations and chromosomal fusions (Cancer Genome Atlas Network, 2015). Cho and colleagues describe the comprehensive genetic profile of these tumours, characterising *TP53*, *CDKN2A*, *PIK3CA*, *HRAS*, *PTEN*, and *NOTCH1* as being changed differentially across all Head and Neck locations in contrast to *CASP8*, which is also confined within oral cavity (Loyo *et al.*, 2013; Farah, 2021). Mutations in *RAS* oncogenes are less frequent, than in other solid tumour malignancies, characterizing HNSCC as neoplasms predominately associated with loss of tumour suppressors genes (Cancer Genome Atlas Network, 2015).

Genes have been grouped into categories including cell survival and proliferation - *TP53*, *EGFR*, *PIK3CA*, and *HRAS*, cell-cycle control genes - *CDKN2A* and *CCND1*, cellular differentiation genes - *NOTCH1*, and cellular adhesion and invasion signalling - *FAT1*, have been classified (Agrawal *et al.*, 2011; Stransky *et al.*, 2011; Cancer Genome Atlas Network, 2015).

The overview of the genetic profile of HNSCC tumours (Figure 8) grouped according to Copy Number Alterations, gene mutations and gene expression profiles intent classify distinct classes of tumours based on patterns of protein expression that correlate with clinical behaviour (Rothenberg and Ellisen, 2012). The biological variability shown within HNSCC translates the division of the two main carcinogenesis pathways established for HNSCC-HPV and HNSCC-nonHPV (Farah, 2021).

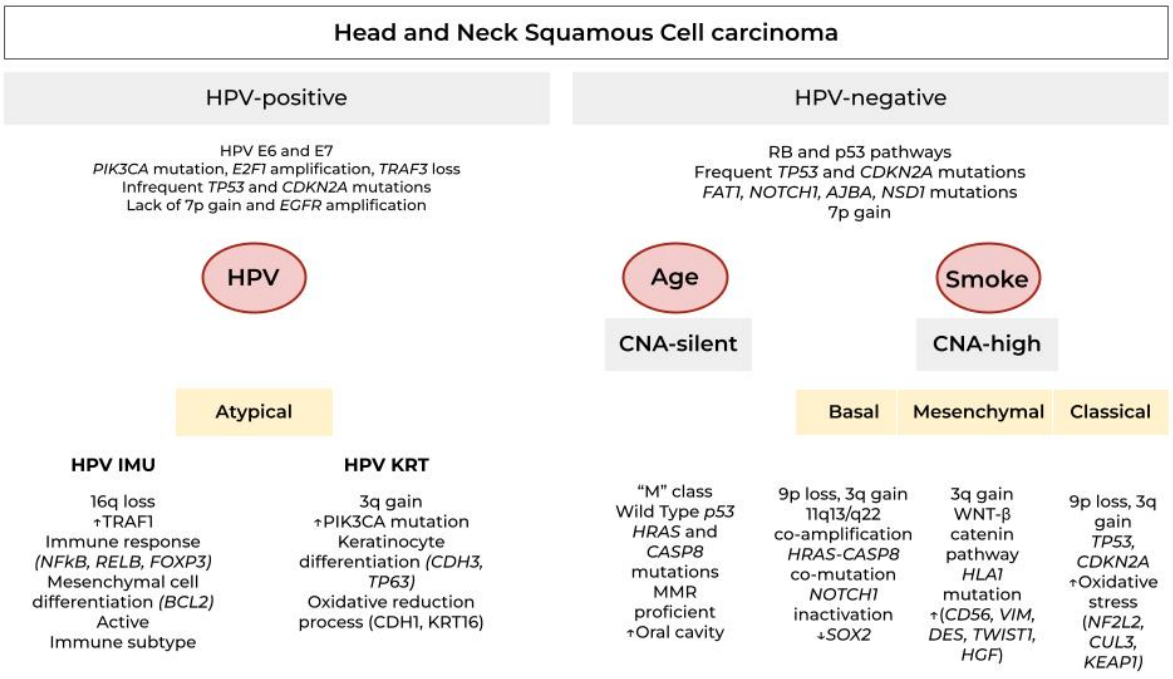


Figure 8. Molecular characterization of the HNSCC. Scheme adapted from (Farah, 2021) presenting the main profiles grouped according to copy number alteration (CNA), genetic mutations and gene expression profiles. Accordingly, the mutational analysis divided the HNSCC-nonHPV into CNA-silent and CNA-high tumours. Normally the TP53 and RB pathways are abrogated in HNSCC-nonHPV tumours but remain active in CNA-silent tumours. These CNA-silent are considered a subgroup of tumours that, although their aetiology is not well established, age apparently is the most risk factor associated and exhibiting mutations in the HRAS and CASP8 genes and often found in tumours of the oral cavity. On the other hand, HNSCC-nonHPV tumours (CNA-high), in which tobacco is the associated risk factor, exhibited changes in oncogenes such FAT1 and NOTCH1 pathway (WNT-B-catenin).

According to gene expression level the HNSCC segregate in four groups: atypical, basal, mesenchymal, and classical. The atypical subtype involves most HNSCC-HPV tumours with *PIK3CA* activating mutations and a lack of 7p amplifications (EGFR encoding region), unlike the other group that exhibit 7p gain of function. Based on RNA sequencing data, the HNSCC-HPV tumours separate into two distinct subtypes: HPV-IMU, those cases with upregulated genes activating mesenchymal and immune response, and HPV-KRT, those, enriched for keratinocyte differentiation and oxidative reduction process (Zhang *et al.*, 2016).

2.4.1 NON-HPV ASSOCIATED HNSCC

The distribution of location HNSCC-nonHPV occurs among different anatomic sites, in the context of heavy alcohol and tobacco consumption, and represents 75-85% of all HNSCCs (Rettig and D'Souza, 2015; Machiels *et al.*, 2020). Tobacco contains more than 5000 different substances, most of which have carcinogenic activity. The reactive metabolites can form DNA adducts that lead to mutations and other genetic abnormalities. This propensity depends on the balance of metabolic activation versus detoxification and DNA repair (Johnson *et al.*, 2020). Similarly, excessive alcohol consumption act synergistically with tobacco consumption. Alcohol may play as a solvent for carcinogenic elements, thus increasing the exposure of these substances in epithelial cells. It should note that the reactive metabolites from alcohol consumption also originate from DNA adducts (Balbo *et al.*, 2012). Additionally, the use of these substances causes inflammatory states in involved tissues that play a role in triggering proliferation, angiogenesis and carcinogenesis (Talamini *et al.*, 2002).

HNSCC-nonHPV appear mainly in the oral cavity (62%), followed by the larynx (26%) and oropharynx (12%), and these sites represent 87% of all HPV-negative tumours (Cancer Genome Atlas Network, 2015; Farah, 2021). Relatively to the histopathology, they are more often moderately or well-differentiated with preservation of stratification and keratinization (Johnson *et al.*, 2020).

Accordingly, to the Tumour Cancer Genome Atlas (TCGA) the two most frequent genomic alterations of HNSCC-nonHPV tumours affect *TP53* (83%) and *CDKN2A* genes (57%) (Loyo *et al.*, 2013; Johnson *et al.*, 2020). In the molecular profile (Figure

8), other frequent alterations were also described, including other tumour suppressor genes such as *NSD1, FAT1*, *NOTCH1* and *SMAD4*, genes that encode the tyrosine receptor kinases *EGFR, ERBB2* and *FGFR159*, and in the genes that encode *NRF2* (*NFE2L2*) and *KEAP1* which are regulators of oxidative stress. These alterations occur exclusively in HNSCC-nonHPV (Cancer Genome Atlas Network, 2015).

2.4.2 HPV ASSOCIATED HNSCC

The presence of HPV DNA plays a pathogenic role within HNSCC and defines a specific pathologic entity with distinct epidemiological, clinical and molecular characteristics (Chera *et al.*, 2015), when compared with HNSCC-nonHPV (Sabatini and Chiocca, 2020). Although tobacco consumption is cause of HN cancer globally, the incidence of HPV infection as a cause of oropharyngeal cancer is increasing (Hammarstedt *et al.*, 2006), as reported by the Centre's for Diseases Control and Prevention (CDC) in 2018 (Centers for Disease Control and Prevention, 2018), associating HPV infection as the underlying cause of the global rise in HNSCC incidence (Gillison *et al.*, 2008; De Bree and Leemans, 2010; Hocking *et al.*, 2011). Currently, HPV is associated with 35% of HNSCC (Taberna *et al.*, 2017) and most HNSCC-HPV associated cancer cases emerge in the lingual and palatine tonsils (El-Mofty, 2012).

The proposed pathway of HNSCC-HPV induced carcinogenesis is based on the expression of viral oncoproteins E6 and E7 on the epithelium exposed to HPV infection. These oncoproteins activated by cellular factors (Bray *et al.*, 2018) lead to the inactivation of tumour protein 53 (p53), resulting in an aberrant induction of cellular proliferation (Polanska *et al.*, 2014), and the retinoblastoma protein (pRB) to relieve repression of E2F1 resulting in activation of S phase (Nulton *et al.*, 2017) and therefore, up-regulation of p16 INK4a levels (Polanska *et al.*, 2014). The HPV oncogenes stimulate cancer formation, promoting unlimited replication and consequently gaining genomic instability and gene alterations including endogenous mutations and increased DNA damage which are associated with cancer development (Cancer Genome Atlas Network, 2015; Gillison *et al.*, 2019).

The genetic profile associated with this carcinogenesis pathway is different compared to the HNSCC-nonHPV (Figure 8). The HNSCC-HPV shows fewer

changes in chromosomal copy number, a higher frequency of E2F1 transcription factor amplifications, and the presence of deletions or inactivating mutations of *TRAF3* (Farah, 2021). Interestingly, when compared with the study of molecular profile of cervical cancer these changes were often found in HNSCC-HPV oropharyngeal cancer and rare in HPV-cervical cancer (Kreimer *et al.*, 2020).

It has been well established that HPV E6 and E7 oncogenes inactivate two tumour suppressors (p53 and pRB) besides additional alterations in crucial signalling pathways that are equally important for the transformation of oncogene-transduced cells (Gillison and Shah, 2003). In general, HPVs stimulate host cell survival and proliferation by interacting with growth factor receptor, Notch receptor, Ras and *PI3KCA* gene from four key ascending pathways, carcinogenesis occurs when activated and altered (Zhang *et al.*, 2017). Genetically, the most common mutations in HNSCC-HPV tumours include the *PIK3CA*, *KMT2D*, *PTEN*, *TRAF3*, and *FGFR3* genes, and the presence of the APOBEC mutation signature. (Gillison and Shah, 2003; Cancer Genome Atlas Network, 2015).

Despite this general genetical profile, within the HNSCC-HPV tumours exhibits several differences, and two subtypes based on gene expression were described with purported clinical and biological implications (Chung *et al.*, 2004). The immune subtype (IMU) is defined by an increased immune response, mesenchymal differentiation, and angiogenesis expression. It is also more strongly linked to epithelial-mesenchymal transition (EMT) signatures and associated with an increased expression of *BCL2*, an anti-apoptotic regulator associated with resistance to chemotherapy and radiation. The other keratinization subtype (KRT) is characterized by marked keratinization (Zhang *et al.*, 2016) and is associated with more common viral integration events. This group also shows a lower expression of the viral E2, E4, and E5 viral genes, consistent with the expression of those that are often lost after integration (Qin *et al.*, 2021). Furthermore, the copy number gain and loss partially drive the expression differences between these two subtypes (Qin *et al.*, 2021). Regarding mutation frequencies, the *PIK3CA* oncogene is the only gene showing a significant difference in the two subtypes, highly altered in KRT tumours (Qin *et al.*, 2021). Activating mutations of *PIK3CA* can increase the activity of the mTOR pathway and inhibit autophagy in HPV-positive HNSCC, leading these cells to a cancerous predisposition (Sewell *et al.*, 2014; Raudenská, Balvan and Masařík, 2021), especially those that are highly keratinized (KRT), highly stromal-like and high epithelial-to-mesenchymal transition (EMT) (Farah, 2021).

The EMT-like phenotypical change that occurs in the epithelial cells of many malignant tumours is a polarity shift that corresponds to increased invasiveness and metastatic potential (Hatakeyama *et al.*, 2014; Lefevre *et al.*, 2017; Ihler *et al.*, 2018). Indeed, a normal physiological process acquires a pathologic role in carcinogenesis, driving organ fibrosis and cancer, which are often associated with poor prognosis (Kalluri and Weinberg, 2009; Graves *et al.*, 2014; van der Heijden *et al.*, 2020). It is mediated by the tumour microenvironment and leads to a loss of cell adhesion and epithelial markers, a gain of mesenchymal markers, and an alteration of the cytoskeletal structure towards a loss of cell polarity (Thiery *et al.*, 2009; Bhaijee *et al.*, 2012; Smith, Teknos and Pan, 2013). At the molecular level, there are many observed changes related with and lead to EMT in cancer cells (Thiery *et al.*, 2009; Lefevre *et al.*, 2017). The loss of E-cadherin is seen as the hallmark for the completion of EMT in HNSCC (Thiery *et al.*, 2009; Jimenez *et al.*, 2015). E-cadherin is the main adhesion protein of epithelia and is responsible for cell attachment and epithelial polarity (Chien *et al.*, 2012). It should be noted that the loss of E-cadherin has independently been reported to be associated with poor prognosis (Fan *et al.*, 2013; Jimenez *et al.*, 2015), tumour progression (Zhang, Filho and Nör, 2012; Mitra, Mishra and Li, 2015), and metastatic spread. Expression of E-cadherin is tightly controlled through multiple signal transduction pathways and negatively regulated by transcription factors like Slug, Snail, and Twist, which in HNSCC-HPV their increased invasive potential is supposedly due to oncoproteins' HPV16 E6 and E7 able activate these transcription factors (Mirantes *et al.*, 2013; Campo, Zhang and Breuer, 2015; Lefevre *et al.*, 2017).

2.5 HNSCC CARCINOGENESIS, HPV INFECTION AND PI3K/AKT/MTOR PATHWAY

The HNSCC genomic profile published by Cancer Genome Atlas (Cancer Genome Atlas Network, 2015) has characterized more than 500 sample tumour HNSCC to date, enabling an increased knowledge of cancer genomics. Additionally, several studies in the field conclude that PI3K–AKT–mTOR oncogenic pathway is the most frequently altered in HNSCC (Cancer Genome Atlas Network, 2015; Wang *et al.*, 2017). Components of this pathway are genetically mutated in most HNSCC tumours, with frequent changes in *PIK3CA* occurring either by mutation or by

gene amplification (Lee *et al.*, 2020). In general, PI3K/AKT/mTOR signalling is active in over 90% of HNSCC, as a result of EGFR activation, *PIK3CA* mutations, *PIK3CA* amplifications, PI3K overexpression and PTEN mutation (Marquard and Jücker, 2020).

The phosphatidylinositol-3-kinase (PI3K)/Akt and the mammalian target of rapamycin (mTOR) signalling pathways are both crucial to many aspects of cell growth and survival, in physiological as well as in pathological conditions. They are interconnected and they are regarded as a single, unique pathway (Figure 9). Specifically, the PI3K/Akt pathway represents a key regulator of cell survival during a stress state (Zhang *et al.*, 2015). Since tumours normally are associated with an intrinsically stressful environment, mainly characterized by limited nutrients, oxygen supply and low pH, the role of this pathway in cancer appears to be crucial. The mammalian target of rapamycin is a serine/threonine kinase (mTOR) ubiquitously expressed in mammalian cells (Shaw and Cantley, 2006). It picks up and integrates signals initiated by nutrient intake, growth factors, and other cellular stimuli to regulate downstream signalling and protein synthesis. Through its downstream effectors, 4EBP1 and P70S6 kinase (S6K), it is involved in the initiation of ribosomal translation of mRNA into proteins necessary for cell growth, cell cycle progression, and cell metabolism.

Somatic mutations resulting in the dysregulation of downstream elements are often found in several tumour entities (Wadhwa *et al.*, 2017). The activation of the PI3K/Akt/mTOR pathway results in a profound disturbance of control of cell growth and survival, which ultimately leads to a competitive growth advantage, metastatic competence, angiogenesis, and therapy resistance. Thus, this complex pathway has been taken into consideration as one of the most attractive targets for the development of anticancer agents (Zhang *et al.*, 2017).

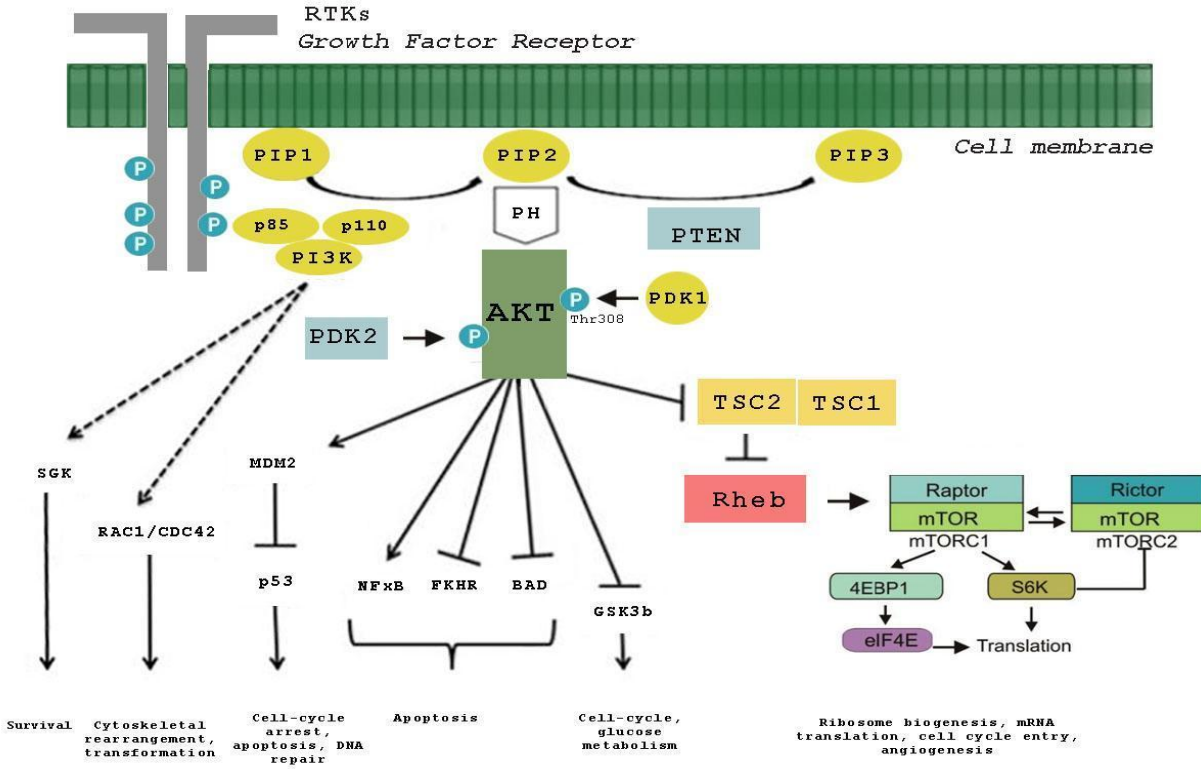


Figure 9. The PI3K/AKT/mTOR pathway. Illustration of the PI3K/AKT/mTOR pathway adapted from (Porta, Paglino and Mosca, 2014). The PI3K/AKT/mTOR pathway is activated by growth factor receptor tyrosine kinases (RTKs), which after binding to the ligand, they phosphorylate adapter proteins that interact directly with PI3K (Zhang *et al.*, 2017). The PI3K is a heterodimer containing a regulatory (p85) and a catalytic (p110) subunit. When RTKs, adapter proteins bind to the p85 regulatory subunit, the p110 catalytic subunit is released and recruited to the membrane (Zhang *et al.*, 2017). Additionally, PI3K phosphorylates the D3 hydroxyl group (Yu and Cui, 2016) and converts membrane bound PIP2 to PIP3 (Wadhwa *et al.*, 2017) and can be inactivated through dephosphorylation by PTEN (Yu and Cui, 2016). PIP3 is a second messenger that recruits' kinases with a pleckstrin homology domain (PH domain) to the membrane. There are two crucial kinases with a PH domain: phosphoinositide-dependent kinase 1 (PDK 1) and serine/threonine kinase (AKT) (Yu and Cui, 2016). The AKT contains three domains: a PH regulatory domain (N-terminal), a connecting hinge region and its kinase domain (C-terminal) and can exist under different isoforms with similar substrate specificity. The AKT recruited to the membrane by PI3K is phosphorylated by PDK1 to a specific threonine isoform that rests in its catalytic domain. This phosphorylated AKT state able converts some of its substrates, but the complete activation requires phosphorylation at serine residues in the c-terminal region induced by the mammalian target of the Rapamycin 2 complex (mTORC2). The AKT inactivation is mediated by the threonine residues dephosphorylated by protein phosphatase 2A, specifically by three different pleckstrin homology domain leucine-rich repeat protein phosphatase isoforms (Wadhwa *et al.*, 2017). The mTOR complex is a serine/threonine kinase occurring in two different complexes TORC1 and TORC2. Its activity depends on the accessibility of the active site. mTORC1, when activated, regulates via p70S6K and translation via 4E-BP1. The mTORC2 activation is mediated by the PI3K and ribosomes with the PH domain of its mSIN1 subunit (Wee and Wang, 2017).

Particularly, in HPV-positive HNSCC tumours, oncoproteins E6 and E7 interact directly in signalling pathways such PI3K/Akt/mTOR. This is thought to play a very important role in HPV-induced carcinogenesis by acting through multiple cellular and molecular events (Pim *et al.*, 2005; Contreras-Paredes *et al.*, 2009). These viral oncoproteins increase PI3K/AKT/mTOR signalling in several ways: the E7 was shown to significantly upregulate AKT activity not only through degradation of RB but also in an RB and PI3K independent manner (Figure 10). It interacts with protein phosphatase 2A (PP2A) and prevents dephosphorylation of AKT that causes significantly higher levels of phosphorylated Bcl-2 antagonist of cell death (BAD) (Marquard and Jücker, 2020). The E6 can activate Akt as well, or bind TSC2, leading to its degradation and resulting in stimulation of mTORC1.

The PI3K pathway is unique, in that all the major components of this pathway are frequently found amplified or mutated in HPV-induced cancers (Zhang *et al.*, 2017). The mutation frequency among HNSCC-HPV tumours was reported to be approximately half of that found in HNSCC-nonHPV (Stransky *et al.*, 2011). Even so, the *PIK3CA* gene remains one of the most mutated in HNSCC-HPV (Sewell *et al.*, 2014; Cancer Genome Atlas Network, 2015). Gene mutations targeting the p110α catalytic subunit were found in 56% of HPV HNSCC tumours and in 34% of HPV-negative HNSCC tumours (Cancer Genome Atlas Network, 2015).

Among the *PIK3CA* mutations observed in HNSCC, 63% occur at three specific locations encoding the p110α subunit, namely E542, E545, and H1047, known as canonical mutations (Henderson *et al.*, 2014; Miao *et al.*, 2018). Among these three activating mutations the E542 and E545 affects the helical domain of p110α, inducing PI3K hyperactivity by disrupting the regulatory activity of p85 on p110α (Miled *et al.*, 2007), and the H1047 affects the p110α kinase domain and is thought to cause a conformational change allowing easier access to the phospholipid substrate (Mandelker *et al.*, 2009).

Also, mutations targeting the p110α subunit have been associated with adverse outcomes in solid tumours although the prognostic significance in oropharyngeal SCC is still unclear. To be precise, the specific survival data did not differ between *PIK3CA* wild-type (WT) and mutated (MUT) lesions, however, the WT-PIK3CA patients had a significantly higher 3-year disease-free survival (DFS) compared to PIK3CA mutated patients. So, the prognostic value of *PIK3CA* mutations is still a matter of debate (Alexandrov and Stratton, 2014).

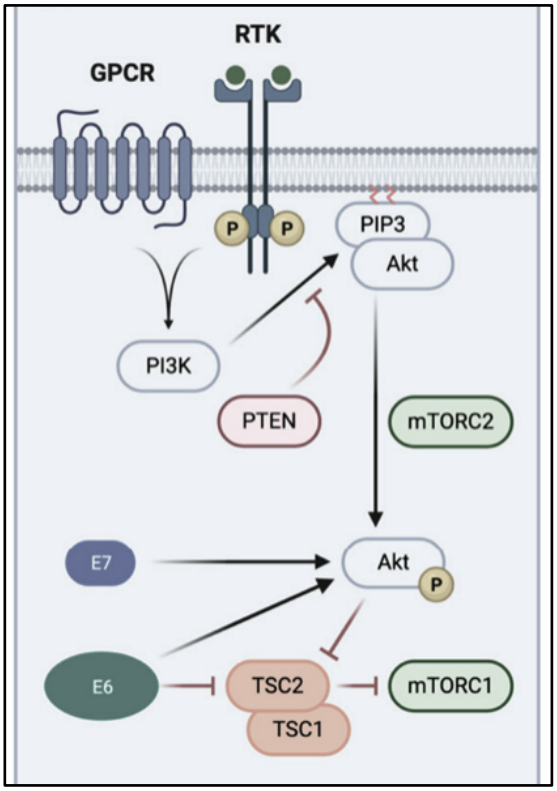


Figure 10. The PI3K pathway and HPV HNSCC. Representative scheme of the PI3K–AKT–mTOR pathway adaptation from (Medda, Duca and Chiocca, 2021). The HPV16 E7 protein binds to protein phosphatase 2A (PP2A) subunits, preventing their interaction with p-Akt and keeping it active. The E6 can activate Akt, or bind TSC2, leading to its degradation and resulting in the stimulation of mTORC1.

2.5 TREATMENT

The description of current treatment management of these patients is complex due to the variety of tumour origin sites, particularly in the definition of the surgical options, with the intention of preservation of the integrity of the surrounding anatomical structures and to ensure that the involved organs still maintain is functions. Nowadays, therapeutic decisions are made in a multidisciplinary approach, that includes several specialities and medical areas such as otorhinolaryngologists, oncologists, radiation oncologists, dentists, dieticians, psychologists, and various therapists involved in the rehabilitation area. The varying presentations defined for specific subsite profiles evolve a range of therapies including surgery, radiotherapy, chemotherapy and more recently immunotherapy (Farah, 2021). The therapeutic scheme must consider the tumour features in terms of its size, growth pattern, depth of invasion, presence and

location of invaded lymph nodes, presence of extracapsular nodal extension, and perineural and lymphatic infiltration: all these characteristics are important for the tumour staging and prognosis, which help to determine surgical margins and the need for postoperative adjuvant therapy (Machiels *et al.*, 2020). Therefore, in general, carcinomas diagnosed at an early stage of the disease (I and II) are usually treated with isolated therapy (surgery or radiotherapy). However, for carcinomas diagnosed in more advanced stages (III, IVA and IVB), combined therapy (surgery, radiotherapy and/or chemotherapy or others) is usually used.

A good therapeutic response in low stages of disease occurs and the majority of patients have a good prognosis with a 5-year survival rate of 70 to 90%, in contrast to high stage disease where the 5-year survival is between 30 to 60% (Johnson *et al.*, 2020).

All patients, at the end of active treatment phase, require regular surveillance and follow-up, as the rate of recurrence and appearance of a new primary neoplasm is frequent within tumours with the worst prognosis (Farah, 2021). Recurrences occur in 80 to 90% of cases within the first 2 to 4 years of follow-up. For that, this period of follow-up must be more frequent and thorough. Disease's recurrence has a dismal prognosis, since therapeutic options are limited by the previous treatment and tumour biological aggressiveness resulting that most of patients have a low mean survival, around 6 to 9 months.

It should be noted, that despite all treatments like surgery, radiation, and chemotherapy, approximately half of the patients will die of the disease.

The risk stratification for HNSCC by its anatomic site, stage, and histologic characteristics of the tumour, has been continuously investigated aiming to improve strategies regarding better treatment of these patients.

However, it should be noted that the inputs of the current knowledge relative to genomic alterations or gene expression profiles previously described have limited clinical utility, and management of most patients is still predominantly the "classic" (Cancer Genome Atlas Network, 2015).

In the last decades research is being actively explored on the mutational, genomic, and transcriptomic landscape and most recent available data show that HPV positivity is an indicator of better overall survival and has a substantially better prognosis after therapy compared with non-HPV+ HNSCC (Mehanna *et al.*, 2019). The initial data from HPV-positive H&N cancer cell lines have shown that cancer cells have an increased sensitivity to radiation, which may explain the enhanced

survival of HPV-positive patients (Nulton *et al.*, 2017). However, data obtained in several clinical trials provide evidence to consider de-intensified therapy for these patients (Argiris *et al.*, 2008; Smith *et al.*, 2010), but until now, data do not support any modification in contemporary treatment protocols, reducing the associated morbidity while maintaining tumour control (Hoffmann and Tribius, 2019). A large proportion of HNSCCs is still left with limited treatment options, underscoring the urgent need to identify novel clinical strategies and more importantly, biomarkers that could help to select patients for current and future therapies.

A better understanding of the mutational, genomic, and transcriptomic landscape of HNSCC has become the promise leveraged to explore new therapeutic approaches to manage this group of diseases, and it is hoped that these additional insights will further drive the correct stratification of patients enabling them to apply more specific treatments for each group of tumours (Farah, 2021).

3. K14HPV16 ANIMAL MODEL

The association of HPV with human cancer has been the subject of extensive research and considerable effort has been expended to model the diseases they cause (Doorbar, 2016). Currently, there are several useful *in vitro* and *in vivo* models that allow studying the process of cancer progression and the role of viral proteins in this process. Overall, *in vitro* models of epithelial cell differentiation can support the entire productive lifecycle of high-risk HPV types (Doorbar, 2016), and *in vivo* models are useful for studying and modulating the function of different cell populations, which serve to promote antitumor activity and reduce protumour signs (Santos *et al.*, 2017).

Despite these various efforts, the current models of HPV-associated disease have a few limitations that are linked to the fact that these viruses only replicate and complete their life cycle in humans, and cause cancers at discrete epithelial sites that are not straightforward to the model (Doorbar, 2016). Nevertheless, each model has its own potential and limitations, and adequate models should be chosen carefully (Santos *et al.*, 2017).

The K14-HPV16 *in vivo* model is valuable to explore components of the immune response against HPV-induced lesions and is promising in the development of novel, preventive and therapeutic, strategies (Santos *et al.*, 2017). This model for

HPV induced cancer relies on the cytokeratin 14 (*CK14*) gene promoter to drive the expression of all HPV16 early oncogenes and reproduces the multi-stage carcinogenic process occurring in human disease, sharing histological features with the human lesions (Du *et al.*, 2012).

K14-HPV16 transgenic mice were first developed in the 1990s using standard techniques by microinjection of B6D2/F2 embryos. A K14 expression cassette developed by E. Fuchs at the University of Chicago was prepared, which contains 2 kb of the K14 promoter/enhancer and 500 bp of a 3' flanking sequence that includes the K14 polyadenylation signal. Instead of the endogenous long control region, the viral genes are regulated by this human K14 enhancer/promoter (Arbeit, Howley and Hanahan, 1996). Three forms of HPV16 DNA of the early region were used by constructed plasmids p1203, p16Nt and p16Pt, which contain respectively a genome of wild-type HPV16 (p1203), a genome with a translation termination ligand (TTL) in ORF E1 at nucleotide 1311 (p16Nt) and a genome that contains a TTL in ORF E2 at nucleotide 2922 (p16Pt). The fragments were then excised and cloned into the BamHI site of the K14 expression plasmid to generate plasmids pK14-1203 (K14-wt), pK14-16Nt (K14-Elttl) and pK14-16Pt (K14-E2ttl). Therefore, the K14-HPV16 transgenic mice contain, within their genomes, the entire HPV16 early region (Arbeit, Howley and Hanahan, 1996), however, since the two early-region variants contain specific mutations (E1ttl and E2ttl), the functions of the E1 and E2 genes are abolished and the E6 and E7 genes remained intact in these mutants (Arbeit *et al.*, 1994). For the creation of the following transgenic lines several inbred backgrounds were carried out for several generations (minimum of five) and maintained in heterozygotes state (Arbeit *et al.*, 1994), including C57BL/6, BALB/c, and SSIN/SENCAR genetic backgrounds, however, only mice backcrossed into the FVB/n background progress to malignant squamous cell carcinomas (Coussens, Hanahan and Arbeit, 1996).

The successful targeting of HPV DNA to basal keratinocytes (Arbeit *et al.*, 1994) generated lesions in the same sites observed in cancer patients allowed a better understanding of the functions of oncoproteins in vivo (Arbeit *et al.*, 1993; Yang, Liu and Iannaccone, 1995). For instance, when female E5+/− transgenic mice were treated with oestrogen, this gene was capable of inducing cervical cancer (Smith-McCune *et al.*, 1997). But, on the other hand, the absence of the late viral coding region (L1 and L2) and the fact that gene expression is regulated by the cytokeratin 14 promoter, introduce significant differences compared to human keratinocytes

infected by HPV and constitute significant limitations of these models (Santos *et al.*, 2017).

Furthermore, this model has proved to be very profitable in several epithelial cancers, and pre-malignant lesions. Most of the early studies were predominantly performed on uterine cervix carcinomas. But later other reports have also used *in vivo* animal models in the study of other HPV-associated malignancies, such as head and neck (Jabbar *et al.*, 2010), penile (Dias *et al.*, 2022) and anus cancers (Arbeit *et al.*, 1994).

In Head and Neck tumours, Jabbar (Jabbar *et al.*, 2010) demonstrated the potential importance of studying cancer growth associated with HPV at various stages. They used a chemical carcinogen 4-nitroquinoline-1-oxide (4-NQO) to induce head and neck cancer in mice K14E6 and K14E7 transgenic models and demonstrate the E7 gene as a dominant oncogene, synergising with 4-NQO. In contrast, under the condition used in the previous studies, E6 transgenic mice treated similarly did not develop HNSCC (Strati and Lambert, 2007). These results are consistent with the dominant oncogenic properties of E7 gene in the head and neck region.

Recent data studied the distribution of oral and pharyngeal lesions in K14HPV16 transgenic mice (Mestre *et al.*, 2020). These findings provided experimental evidence that HPV16 is sufficient to drive HNSCC in the absence of chemical carcinogens and is observed a difference in their distribution. The HPV-driven lesions in the model preferentially target a clinically relevant area at the base of the tongue namely a squamocolumnar junction in the circumvallate papilla (Mestre *et al.*, 2020). Additionally, a different incidence of tongue base cancer in male and female HPV16-transgenic mice was found being these tumours more common in female mice (Neto *et al.*, 2021).

The K14HPV16 animal model can be quite useful to the study of oral carcinogenesis, However, it should be pointed out, that this model lacks a deeper molecular study, a valuable issue to explore the biological significance of the HPV16 variant based on clinical and pathological observations. The characterization of the mutational profile with the identification of the specific lineage and sublineage involved in K14HPV16 mice (phylogenetic analysis) would be useful to broaden clinicopathologic data association, in addition to improving basic and translational studies in this field.

CHAPTER 2

AIM

AIM

The main purpose of this study is to explore the hypothesis that specific HPV16 E6E7 variants influence the occurrence, the morphological and molecular phenotypes of HPV-induced HNSCC in a select population and to validate this molecular setting in a mouse model of HPV-induced HNSCC.

GENERAL HYPOTHESIS

Specific HPV16 E6 E7 variants identified in the carcinomas of the Head and Neck region may be associated with different morphological and molecular phenotypes and with clinical-pathologic settings and molecular markers of prognosis and response to therapy.

SPECIFIC OBJECTIVES

1- To identify and to study the frequency of HPV16 variants in squamous cell carcinomas of the head and neck region in a selected population, and to describe the clinical-pathological characteristics of HPV associated carcinomas, including tumour location, histological type, biological behaviour, and response to therapy.
2- To study gene mutations typically associated with HNSCC-HPV positive, as well as the expression of potential prognostic and therapeutic markers, regarding HPV variants identified in each patient sample.
3- Characterization of oral cancer in an animal model to develop a tool for experimental research:

- To identify the HPV16 variant (E6/E7) present in the K14HPV16 animal model under CK14 regulation to elucidate for a more accurate use of the animal model.

CHAPTER 3

MATERIALS AND METHODS

1. CHARACTERIZATION OF THE BIOLOGICAL SAMPLES

1.1 SQUAMOUS CELL CARCINOMA OF HEAD AND NECK SAMPLES

A detailed characterization of this material was done in the different articles that were published or submitted to publication and are presented in Chapter 4- Results.

The tumour samples are from a universe of patients diagnosed with HNSCC, treated in the Instituto Português de Oncologia de Lisboa Francisco Gentil (IPOLFG) between 2008 and 2019 and were included within the scope of the research project (ref. IPOLFG's UIC/1092), to study the prevalence of HPV in these tumour types. The patients were clinically followed up by the medical staff of the otorhinolaryngology and head and neck surgery departments of IPOLFG. The diagnosis was established by histological analysis of tumour tissue in the Pathology department. The molecular analysis and HPV DNA detection was performed in the twin tumour sample.

A total of 457 cases were selected using following selection criteria:

- Histological confirmation of squamous cell carcinoma in the Head and Neck region (HNSCC), namely the oropharynx (base of the tongue, soft palate, tonsils, wall) and oral cavity (a lip, oral mucosa, mobile tongue, floor, jaw, and hard palate)
- HPV DNA status identified
- Patients without a previous diagnosis of a malignant tumour (relapse) or synchronous diagnosis of a malignant neoplasm in another location besides the upper respiratory/digestive airway
- Only cases of patients aged over 18 years were included

1.2 TISSUE FROM THE K14HPV16 MOUSE MODEL AND WILD TYPE MICE

K14HPV16 (FVB.Cg-Tg(CK14-HPV16)wt1Dh) mice are a transgenic animal model that expresses all early HPV16 genes under the control of the cytokeratin 14 (*CK14*) gene promoter and develop spontaneously cutaneous lesions resembling those of human patients (Arbeit *et al.*, 1993, 1994). This HPV16 transgenic mouse was constructed on an FVB/n strain background and is known as K14HPV16 mouse. This mouse strain was donated by Drs Jeffrey Arbeit and Douglas Hanahan of the University of California, through the Mouse Repository of the US National Cancer Institute. This *in vivo* model is maintained and bred in bioterium of the University of Trás-os-Montes and Alto Douro according to the Portuguese (Decree-Law 1005/92 dated October the 23rd) and European (EU Directive 2010/63/EU) legislation, with 12h light/12h dark cycles, at 20–24 °C and 50 ±10% relative humidity. All the experiments were approved by the University of Trás-os-Montes and Alto Douro Ethics Committee (Approval Number 10/2013) and the Portuguese Veterinary Directorate (Approval Number 0421000/000/2014). For this study we used 4 female (30 weeks old): wild-type (n=2) and transgenic (n=2) mice, sacrificed by intraperitoneal administration of a mixture of xylazine and ketamine followed by cardiac puncture and exsanguination according to FELASA guidelines (Forbes, 2007). A total of 17 samples from the 4 animals were collected for histological and molecular analysis and prior conserved a matched-separated fixed in formalin and other snap-frozen in liquid nitrogen Optimal cutting temperature (O.C.T.) compound, respectively. The distribution of samples by animal condition is presented in Table 1. The eight samples from wild-type (WT) animals were used as controls. Overall, samples collected from tongue (n=4), skin (n=4), liver (n=4), lymph node (n=4) and oral tumour (n=1) were evaluated/analysed.

Table 1. Characterization of mouse samples collected

Animal ID	Model description	Sample tissue	Conservation
I	WT	node	Formalin | OCT
		tongue	Formalin | OCT
		liver	Formalin | OCT
		skin	Formalin | OCT
II		inguinal node	Formalin | OCT
		tongue	Formalin | OCT
		liver	Formalin | OCT
		skin	Formalin | OCT
III	MUT	node	Formalin | OCT
		tongue	Formalin | OCT
		liver	Formalin | OCT
		skin	Formalin | OCT
IV		node	Formalin | OCT
		tongue	Formalin | OCT
		liver	Formalin | OCT
		skin	Formalin | OCT
		tumour	Formalin | OCT

Legend: The animal group are designated WT wild-type FVB or MUT FVB-K14HPV16; Tissue-Tek O.C.T. Compound.

2. METHODS

2.1 HISTOLOGICAL ANALYSIS

Histopathology study for the accurate diagnosis of cancer in patients was determined from suspected tissue that has been excised by biopsy or surgical resection using the standard procedures of the Pathology Department. All samples were fixed in 10% neutral buffered formalin and paraffin embedded. Three to five µm-thick sections were stained with haematoxylin and eosin for histological classification. All cases were later reviewed by two pathologists based on their histological characteristics and were classified by histological type, (according to WHO classification), grade and stage according to the TNM system 8th edition (Amin *et al.*, 2017). Additionally, in the histological study, the immunohistochemical study of p16INk4a was performed on tumour tissue for most cases. The sample histological process was used for the mirror histological samples from transgenic mice.

2.2 HPV DNA DETECTION IN HUMAN AND MICE SAMPLES

Patient's tumours were examined for presence of genetic material (DNA) of HPV by polymerase chain reaction (PCR) method and HPV genotyping was performed to determine the specific type and stratified using the epidemiological classification (Muñoz *et al.*, 2003) for the risk of cancer development.

HPV DNA detection was performed on 457 tumour biopsies (fresh tissue) collected at the time of the diagnosis and on the 21 samples obtained from the 4 female transgenic mice. The total DNA was extracted manually using a spin silica filter column (QIAamp, a DNA Tissue Kit), or automatically using magnetic beads (Universal Kit, Seegene).

DNA integrity (quantity and purity) was evaluated by spectrophotometric reading at absorbance A260nm and A230nm. Most of the analysed samples were processed by both methodologies described above and prepared for a quantity of 100ng DNA, to proceed to amplification.

Real-time PCR (qPCR) was performed by SYBR GREEN dye with specific HPV L1 region primers to generate a 75pb amplicon (Payan *et al.*, 2007). A melting curve was performed to confirm the specificity (Tm 73 °C) and an internal amplification control (albumin gene) was performed in independent reactions, along with HPV detection.
With this assay, designed for a screening approach, samples detected for the presence of HPV DNA were identified. Only positive samples were subsequently genotyped.

2.3 HPV GENOTYPING

The identification of the HPV genotypes in the different samples was performed from the eluted DNA previously extracted. Genotype identification was processed for all samples detected with HPV. For this, two methodologies were used, which use the same amplification principles but differ in their detection method. Both commercial kits were processed according to the manufacturer's instructions. The LiRa detection technology by classical PCR (INNO-LiPA® HPV Genotyping Extra II | Fujirebio, 19355 Malvern, PA, United States) detected 24 different HPV types and qPCR multiplex (Anyplex™ II HPV28 Detection | Seegene, 20878 Gaithersburg, MD, United States) able to detect 28 HPV types. The application of these two devices intends to approach a broad spectrum of high and low-risk genotypes, allowing us to portray the presence of the different genotypes present in our cohort. From the samples in which conclusive result of the genotypes was possible, all cases of HPV16 were selected.

2.4 HPV 16 WHOLE-GENOME SEQUENCING BY NGS

Next-generation sequencing (NGS) covers a deep molecular analysis that makes it a promising tool for applications in several fields, and hopefully in clinical practice bringing improvements inputs. In this project, an NGS assay was developed to allow the identification of HPV16 variants isolated from HNSCC tumours. The design and construction of this NGS assay was developed by us aiming its implementation in a clinical laboratory, and for that reason followed the

requirements of the ISO 15189:2022 Quality and Competence requirements is an international standard that specifies quality management system requirements specific to medical laboratories and design the technical procedures in accordance with European union guidelines. A list of standardization documents relevant to NGS describes the ISO guidelines for Specimen/Sample Pre-Analysis, library preparation and NGS Data analysis (Supplementary Table 1).

2.4.1 NGS ASSAY OPTIMIZATION

The NGS workflow encompasses two fundamental pillars: laboratory performance and the bioinformatics approach. The implementation and validation require numerous steps and checkpoints regarding quality assurance procedures, and the international standards required (Jennings *et al.*, 2017; Roy *et al.*, 2018) for the correct implementation in the routine laboratory.

The considerations for optimization of the NGS workflow (Figure 11) are described in two sections: (1) Laboratory protocols and (2) Bioinformatic data analysis steps. The optimization process intends to describe the preliminary performance specification and categorize the checkpoints of quality control.

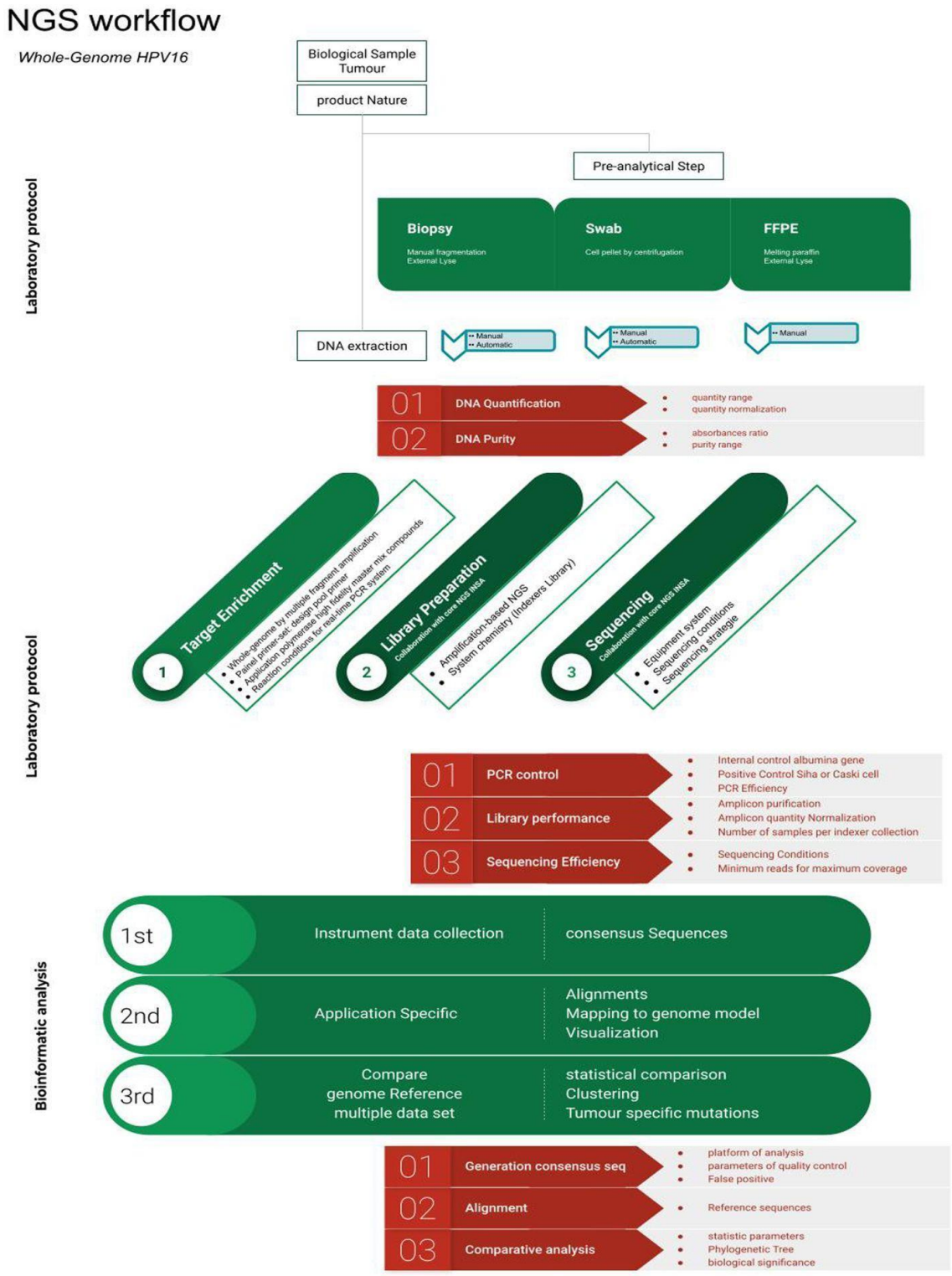


Figure 11. Design of NGS Technology workflow. Representation of the sequence of the three main stages of the process. The main developed and optimized technique (green boxes). The verification checkpoints (red boxes) and their respective analysis processing according to the respective certification standard (colour grey).

Product Nature: Sample Preparation

All biological samples used for the identification of HPV by molecular methodologies can be used for NGS. The pre-analytical sample processing is applied with the same specific procedures for the various biospecimen types (product nature) such as fresh biopsy samples conserved: liquid-based cytology, dry frozen swab samples (-80°C), or tissue derived from formalin-fixed paraffin-embedded (FFPE). In general, all samples were fresh biopsies from tumours in the Head and Neck region. The DNA extraction was performed in Automatic equipment or Manual extraction, following the manufactures specificities. The quality and suitability of DNA extracted was spectrophotometrically controlled (A260/A280).

This checkpoint allows to determine the suitability of the samples: only those with an acceptable range of quantity, purity and integrity values are validated for the next step of the NGS workflow.

Target enrichment: PCR design and optimization

Our approach was the amplification of the complete HPV16 DNA genome (~8000pb) and the primer design was based on 47 primers set previously described by Cullen et al 2015, which makes possible to cover the entire genome based on fragmented amplification producing amplicons with a size that ranged between 120-359 nucleotides. This strategy contributes to a better sensitivity of the assay, as opposed to opting for a long-fragment strategy that lowers the sensitivity values. The sensibility of the assay is an important parameter that we must overcome since the main target (viral DNA) is in a small quantity within the abundant host genomic DNA. An extensively evaluation from primer characteristics and *in silico* validation were performed. For the 103 primers that composes the 47 primers set, it was described the respective sequence, size, melt temperature, target region, and fragment size produced and evaluate secondary amplifications to other organisms with the Blast tool from PubMed (Supplement Table 2 in silico validation). For the pool primer conception was consider the *in silico* validation and intention to promote the best balance between specificity and sensibility of the assay. The best of 3 combination tested was an approach with 4 pools with 7 to 10 primer each.

(Supplement Table 2 Pool Primer Design): these were structured to amplify the viral whole-genome, avoiding the development of primer-dimer. For that, each primer was single amplified and verify within each primer pool to validate specificity. The PCR conditions were optimized for a real-time PCR system (60°C hybridization temperature) in the Applied Biosystems QuantStudio 5 Real-Time PCR System, for a reaction volume of 25uL. The mixed compounds were accurate to work with High-Fidelity DNA Polymerase, a proofreading hot-start DNA polymerase for high yield sensitivity and specificity, being over 300x more accurate than Taq-enzyme, resulting in ultra-low error rates. Validation was carried out for each pool primer with input concentration set to 1 pmol and normalised each DNA template into ~100-250ng. The reaction is performed in 3 stages for complete amplification: (1) 95°C Denaturation; following 40 cycles of segment (2) 60°C hybridization; and (3) 72°C elongation. The detection method by agarose gel was used to identify the PCR products. The 4 independent PCR products of each sample were pooled together in equal molar amounts based on the signal intensity.

Library Preparation and Sequencing

Library preparation protocols were based on Amplification-based NGS. There are several different reaction reagents available for specific protocols due to the diversity of platforms. At present, there is not a recommend protocol for any HPV whole-genome applications. Our *in-house* assay has been optimized for Illumina technology with the collaboration of the Core NGS from INSA (National Institute of Health). The optimization of the quality, quantity, and fragment length includes confirmation of expected results from positive control and validates the measurements data for the specific analytical purpose. The pooled amplicons were purified from small fragments and primers using the AMPure XP beads (Beckman Coulter) according to the manufacturer´s instructions before library preparation. The Nextera XT DNA library prep kit was used to prepare bead-based normalized dual-indexed libraries for sequencing. The validation considered fragments size to cluster efficiency and concentration of the prepared libraries since low concentration may result into a low number of sequenced reads. The library pool was sequenced on a MiSeq benchtop sequencer using a reagent kit V3 (illumina MS-102-3001). A sequencing strategy with 250-bp paired-end reads was used.

2.4.1.2 BIOINFORMATIC DATA ANALYSIS

NGS bioinformatics pipelines consists in a specific platform to analyse sequence consensus single nucleotide variant (SNV) and may be customizable based on laboratory needs. The Bioinformatics Analysis of NGS Data were developed with Core bioinformatic INSA collaboration. Primary analyses were performed on INSaFLU (https://insaflu.insa.pt/) an online platform for amplicon-based next-generation sequencing data analysis (Borges *et al.*, 2018) and used for reads' quality control, variant detection/inspection and sequence consensus generation. The genome sequence of a representative of HPV16 sublineage A1 (GenBank accession number NC001526.4) was used as a reference for mapping and SNV annotation. Regions with a depth of coverage below 10-fold were automatically masked in the INSaFLU pipeline by placing undefined bases "N" in the consensus sequence. Low-coverage regions were visually inspected using Integrative Genomics Viewer (IGV). SNVs were assumed in consensus when they displayed more than 50% intra-sample frequency. MEGA11 software (http://www.megasoftware.net) (Tamura, Nei and Kumar, 2004) was applied to calculate matrices of nucleotide distances and to perform phylogenetic reconstructions.

2.4.2 ANALYTICAL VALIDATION

The performance characteristics of this method were validated based on data measurements from Positive Control, Negative Control, and Internal Control. The cell line SiHa, positive for HPV16, was used as positive control with 100% concordance and were included in each run to evaluate PCR efficiency (HPV16 A lineage (GenBank:AF001600.1). The negative control (No Target Control) provides insight regarding chemical reagents contamination; Albumin that was also used as an internal control to distinguish true target negatives from PCR failure or inhibition.

To evaluate the assay performance: the precision of the method was established by performing positive reference control for 10 repeats and the assay performance results were analysed by agarose gel. To estimate efficiency sensibility the positive control was quantified and diluted 10x series into 4 points between 5 to 5000 copies. The assays demonstrated a wide dynamic range efficiency and can

analytically detect at 5 HPV16 copies/PCR. The specificity was accurate for each pool primer, which the specific target was analysed through fragment size in agarose gel with Biorad geldocX tool. The reproducibility was based on 10 replicates of the positive control and has demonstrated excellent reproducibility inter and intra-assay (Supplement Table 3).

2.4.3 IMPLEMENTATION AS ROUTINE PROCEDURE

For the assay implementation as a routine procedure, several actions were established based on specific requirements face to quality control metric as described in Table 2 for each section of the workflow.

Table 2. Implementation and Quality Control Metrics

NGS Workflow sections	**Requirements**	**Action**
Sample Preparation	Biological Specimen	Pre-analytical sample processing
	DNA	Monitoring quantity and purity
PCR enrichment	Reaction performance	Positive control amplification HPV16 infected cell line Siha or CaSki
	Pool Primer Efficiency	Monitoring amplicons concentration (signal intensity)
	HPV16 Whole-genome Specificity	Monitoring Fragment size (migration position)
	Target integrity	Internal control amplification albumin gene
Library Preparation and Sequencing	Library Qualification and Quantification	Monitoring number of reads
Data analysis	Bioinformatic pipeline	Standardization of analytical parameters
	Biologic significance	Data systematization into a database

2.5 DETECTION OF MUTATIONS OF THE PIK3CA GENE

Assessment of the *PIK3CA* gene mutations were performed by real-time PCR assay (Amoy Dx) for the qualitative detection of four mutations (H1047L; E542K, E545K and E545D) target in exon nine and exon twenty from the human genomic DNA extracted from tumour tissue. The assay comprises specific primers and fluorescent probes. During the nucleic acid amplification, the targeted mutant DNA is detected by FAM-labelled probes. It used an internal control system containing primers and a HEX-labelled probe for a region of genomic DNA without known mutations and polymorphisms. All the procedures of pre-analytical, analytical and data interpretation were performed according to the manufacturer's specifications.

For the PCR conditions, the preparation of the reaction components (master mix) was optimized for the eluted DNA evaluated. This amount was normalized to 15 ng of template DNA for all samples. The different compounds PIK3CA Reaction Mixes, Mixed Standard and Taq DNA polymerase were combined at different concentrations (ng/μL) for a final volume of 20μl. Therefore amplification were performed on qPCR BioRad CFX96 equipment with Cycling Parameters programmed for 3 stages defined by Temperature, Time and Cycles: Stage 1 defined by 1 cycle at 95 ℃ 5 min 1; Stage (2) defined in 15 cycles at 95 ℃ 25 s, follow decrease to 64 ℃ 20 s 15 and increase to 72 ℃ 20 s; and Stage (3) defined in 31 cycles at 93 ℃ 25 s, follow decrease 60 ℃ 35 s (Data collection of FAM and HEX/VIC) and increase to 72 ℃ 20 s. Analysis data run automatically in software to check the FAM and VIC Ct value for each sample. Based on different mutant Ct values, the detection results are divided into strong positive, weak positive or negative.

CHAPTER 4

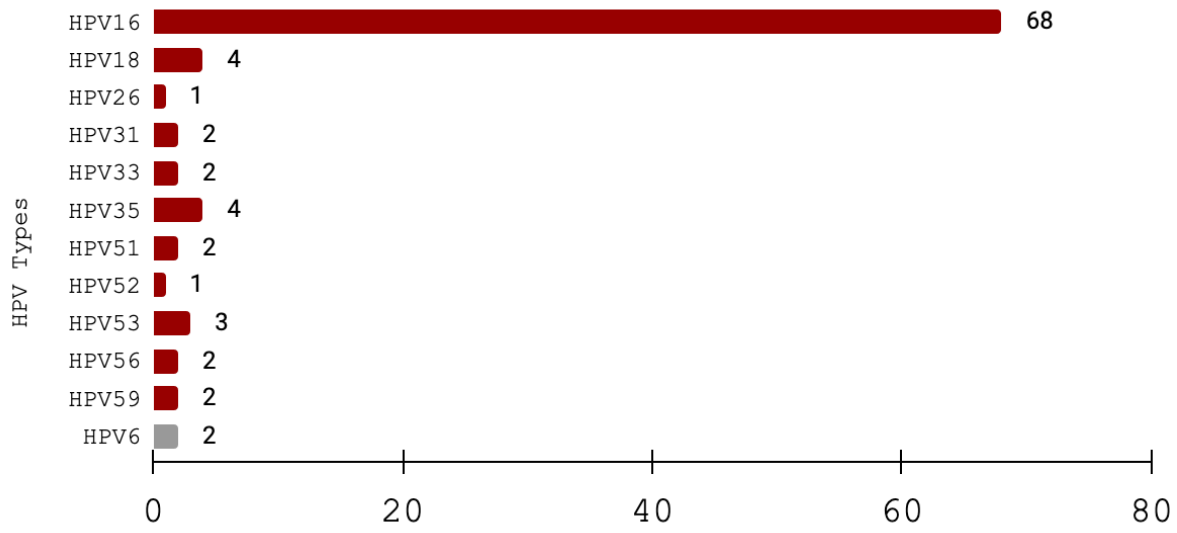
RESULTS

1. DESCRIPTIVE ANALYSIS OF THE STUDY SERIES

1.1. IDENTIFICATION OF HPV DNA IN HNSCC SAMPLES

From a total of 457 patient´s with the diagnosis of HNSCC, HPV was found in 34.4% (n=157) of the cases. The HPV genotype was achieved for 101 HNSCC-HPV despite 56 cases where genotyping was inconclusive. Single infection was present in 93 cases which the most frequent type was HPV16 (n=68), followed by HPV18 (n=4), HPV35 (n=4), HPV31 (n=2), HPV53 (n=3), HPV33 (n=2), HPV51 (n=2), HPV56 (n=2), HPV59 (n=2), HPV6 (n=2), HPV52 (n=2) and HPV26 (n=1) (Graphic 1). In 5.7% (n=8) of the cases, more than one type of HPV was detected - coinfection, where HPV16 was present with other HPV types both low-risk and/or high-risk types.

Graphic 1. Frequency HPV types of single-infection cases (n=93).

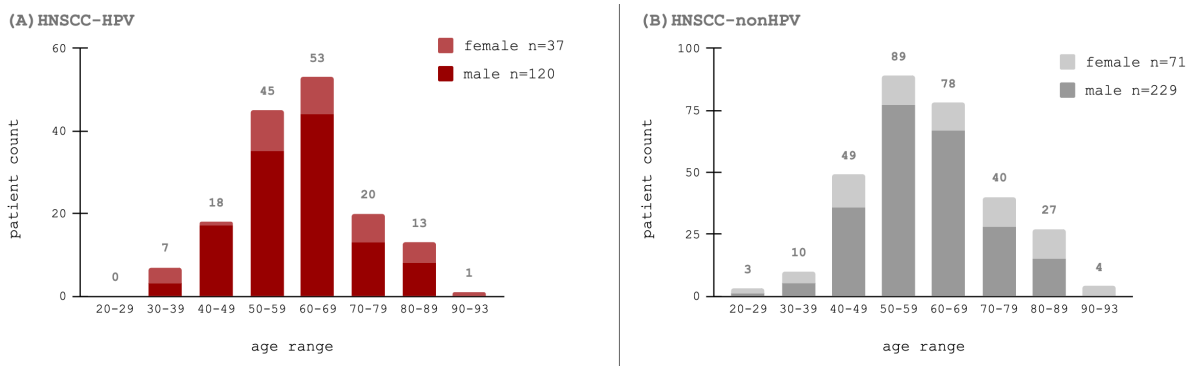


Legend: coloured red - High-Risk Types; coloured grey - Low-Risk Types

1.1.1 DEMOGRAPHIC CHARACTERISTICS

The age at diagnosis of HNSCC-HPV patients range between 36 and 91 years with a median of 61 years. Both genders are present with 76.4% (120/157) males and 23.6% (37/157) females (Graphic.2-A). In our cohort, the gender distribution is similar in both groups HNSCC-HPV and HNSCC-nonHPV with a male to female ratio of 3:1. The median age of 61 years was also recorded for HNSCC-nonHPV group, which ranged from 27 to 93 in HNSCC-nonHPV, with 76.3% (229/300) of males and 23.7% (71/300) females (Graphic.2-B).

Graphic 2. Age at diagnosis per group HNSCC-HPV (A) and HNSCC-nonHPV (B) stratified by patients' gender. The data refers to the total number of patients diagnosed with HNSCC in last decades (2008 and 2019) in IPOLFG.



1.1.2 CONSUMPTION HABITS

The consumption habits were evaluated according to self-reported in the clinical files. Data analysis was based on stratification by never drank, ex-drinker, sporadic, moderate, and severe drinker for Alcohol exposure. Tobacco consumption was defined in units as one pack/year (equal to one pack of cigarettes/day/year, with 20 cigarettes/pack) and classified as current or heavy smokers: current tobacco users included those who used tobacco ≤ 30 pack/years and heavy users were those who smoked > 30 pack/year. Patients without consumption habits were defined as not having consumed none of these substances prior to cancer diagnosis. Of the total of the 457 patients included, 41 patients had missing information in files. The analysis from the remaining 416 patients, considered two groups active consumption vs. no consumption. Most patients have an active consumption (82%;

n=341/416) with a consumption practice of both tobacco and alcohol. On the other hand, 18% (n=75/416) of the patients self-reported no consumption habits. The analysis of the distribution of HPV by consumption habits (Graphic 3) shows that a higher active consumption among HNSCC-nonHPV patients [65% (n=223/341)] compared to the consumption observed among HNSCC-HPV patients [35% (n=118/341)]. The analysis of consumption by gender reveals that the no-consumption group is mostly represented by females (Table 3).

Graphic 3. Characterization of HNSCC patients' consumption habits by HPV detection

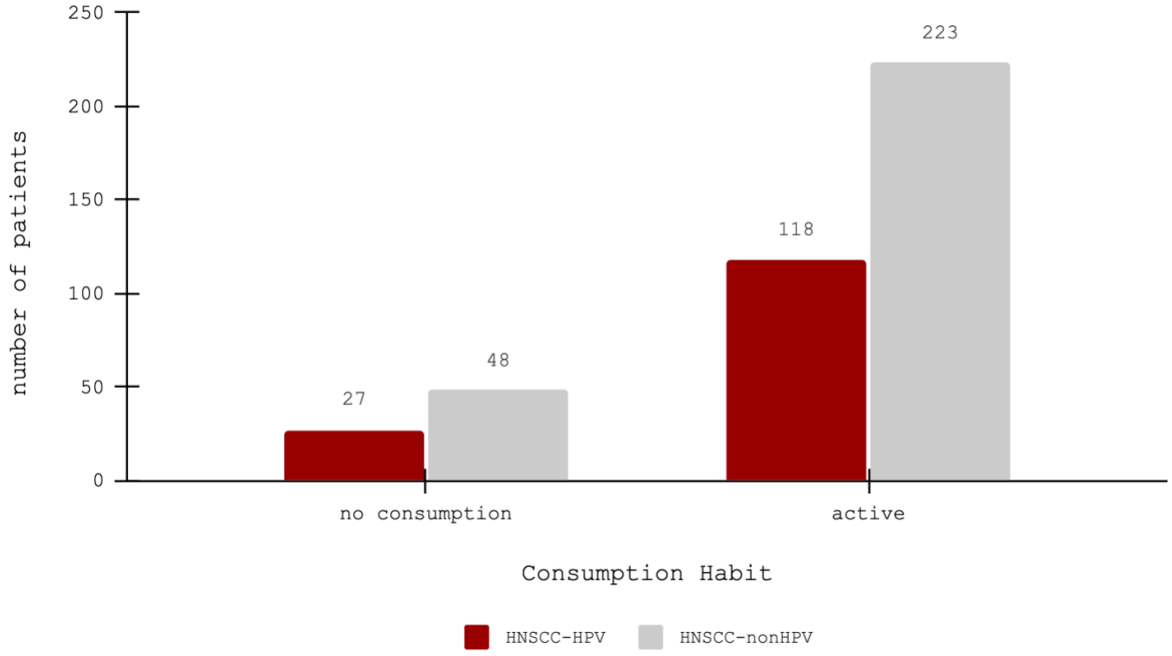


Table 3. Characterization of HNSCC patients' consumption habits by HPV and by gender

		HNSCC-HPV		HNSCC-nonHPV	
Consumption	n	male	female	male	female
no consumption	75	7	20	21	27
active	341	105	13	190	33
no data	41	8	4	18	11
Total	457	120	37	229	71

1.1.3 TUMOUR SITES

The Table 4 shows the distribution of the different tumour locations. Tumours were in different subsites of the oropharynx [45.3% (207/457)] and oral cavity [54.7% (250/457)]. The site of the tumour from HNSCC-HPV patients is more frequent in the oropharynx with 55% cases (86/157), unlike HNSCC-nonHPV patients which the tumour site is more frequent in the oral cavity with 60% cases (179/300).

Table 4. Distribution of tumour site by HPV detection

Tumour site		HNSCC-HPV		HNSCC-nonHPV	
oropharynx	207	*86*	*55%*	*121*	*40%*
base tongue	25	15	*17%*	10	*8%*
soft palate	46	10	*12%*	36	*30%*
tonsil	118	56	*65%*	62	*51%*
oropharynx wall	18	5	*6%*	13	*11%*
oral cavity	250	*71*	*45%*	*179*	*60%*
hard palate	5	2	*3%*	3	*2%*
oral mucosa	42	11	*15%*	31	*17%*
oral floor	52	13	*18%*	39	*22%*
tongue	151	45	*63%*	106	*59%*
Total	457	157		300	

1.1.4 TUMOUR HISTOLOGY

All cases included in this series were confirmed as invasive squamous cell carcinoma in the histological analysis. Tumour were diagnosed in different stages of disease at time of diagnosis (Table 5). The distribution regarding stage of disease revealed that tumours were predominately in stage II and I, followed by stage III and few cases were diagnosed at stage IV.

Table 5. Tumour stage by HPV detection

HN tumour		HNSCC-HPV		HNSCC-nonHPV	
n=457		*n=157*		*n=300*	
Tumour stage					
I	150	46	*29%*	104	*35%*
II	151	55	*35%*	96	*32%*
III	130	50	*32%*	80	*27%*
IV	7	3	*2%*	4	*1%*
no data	19	3	*2%*	16	*5%*

1.1.5 TREATMENT

This cohort of cases spans a long period of years (10) and patients were treated with different schemes, taking into consideration the clinical characteristics, tumour stage and therapeutical guidelines. Treatments used involved surgery and radiotherapy (RT) as single therapies or in association with systemic chemotherapy treatment. A treatment plan was proposed to all 457 cases, but 17 patients refused any treatment. Regarding the remaining patients, 33% (n=150) were proposed for a combination of surgery with radiotherapy, followed by 24% (n=108) with a combination of surgery with chemoradiotherapy and 17% (n=78) with a combination of chemoradiotherapy. Surgery was used as single treatment, in 12% (n=55) of the cases; single radiotherapy in 10% (n=46) and only 1% (n=3) patients were treated with single chemotherapy (CT). Most patients showed a completely local and regional tumour response except in 59 cases were persistent loco-regional disease or progression was observed (Table 6).

Table 6. Distribution of cases for therapeutic approach's

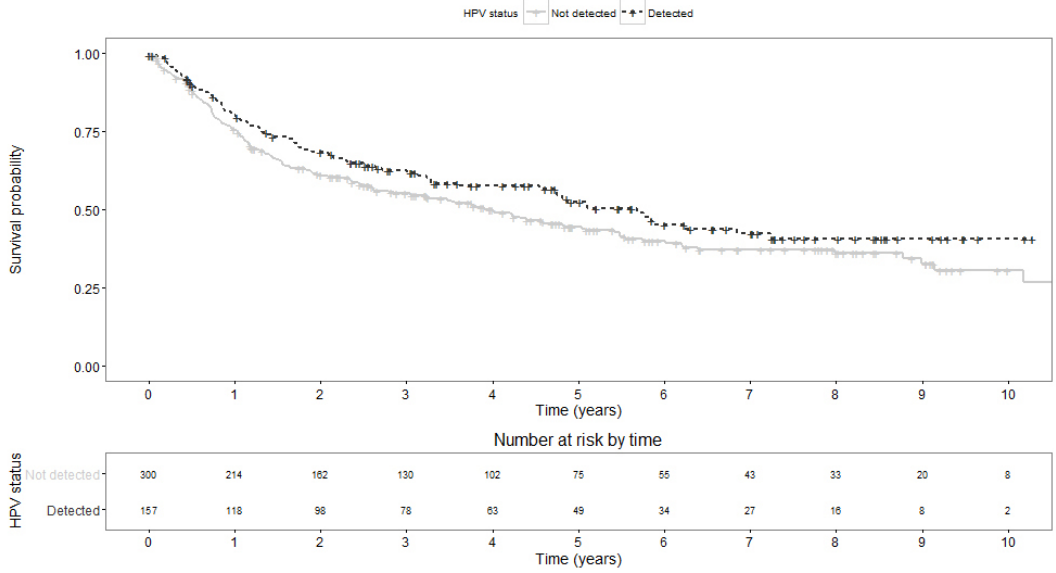
Therapy		HNSCC	
		n=457	%
Treatment Type			
surgery	SG	55	12%
radiotherapy	RT	46	10%
chemotherapy	CT	3	1%
combined	CT-RT	78	17%
combined	SG-RT	150	33%
combined	SG-QRT	108	24%
no data		17	4%
Treatment Response			
complete		333	73%
persistence		62	14%
no data		62	14%

Legend: SG surgery; RT radiotherapy; CT chemotherapy; QRT chemoradiotherapy; SG+RT surgery combined with radiotherapy; SG+QRT surgery combined with chemoradiotherapy.

1.1.6 OUTCOME: FOLLOW UP AND STATUS

The median follow-up in the whole cohort (n = 457) was 5.63 years, determined by reverse Kaplan–Meier method. A total of 238 deaths were reported during the follow-up period. The median overall survival (OS) was 4.58 years and at 3-year OS was 57%. According to the Graphic 4, no statistical differences were found in overall survival between the HNSCC-HPV patients and HNSCC-nonHPV patients, however, the median survival for HNSCC-HPV (5.53 years) was higher than for HNSCC-nonHPV (4.20 years).

Graphic 4. Overall Survival in HNSCC-nonHPV and HNSCC-HPV patients



1.2 SAMPLES FROM K14HPV16 MICE AND CONTROLS

A total of 17 samples from these 4 animals were collected for histological and molecular analysis, both from wild-type (WT, n=8 samples) and HPV16-transgenic mice (MUT, n=9 samples).

1.2.1 HPV16 DETECTION

In the total series of tissue model samples, the results of the HPV DNA status are described in Table 7. Regarding the 17 samples that were analysed from the MUT animals (n=9), HPV16 DNA was present in all tissue samples. Regarding the total 8 samples from the 2 WT mice that were analysed, HPV DNA was not detected.

Table 7. HPV DNA status in model samples

sample	Animal	sample tissue	HPV DNA
1	I WT	skin	not detected
2		liver	not detected
3		lymph node	not detected
4		tongue	not detected
5	II WT	skin	not detected
6		liver	not detected
7		lymph node	not detected
8		tongue	not detected
9	III MUT	skin	HPV16
10		liver	HPV16
11		lymph node	HPV16
12		tongue	HPV16
13	IV MUT	skin	HPV16
14		liver	HPV16
15		lymph node	HPV16
16		tongue	HPV16
17		oral tumour	HPV16

Legend: WT wild-type FVB; MUT FVB-K14HPV16

1.2.2 HISTOLOGY

Matched samples from 2 wildtype FVB/N mice and 2 K14HPV16 female mice were evaluated aiming to identify pre-malignant and malignant lesions in the tongue and document other tissues such as skin, liver, and lymph nodes, where Keratin 14 is expressed or not, that were used as controls. (Figure 12). Wildtype FVB/n animals presented normal skin, liver, and lymph nodes. HPV16-transgenic mice showed typical proliferative epithelial lesions of the skin and tongue (Figure 12)., while liver and lymph node samples showed none or mild inflammatory changes. The only oral tumour that was evaluated was confirmed to be an invasive squamous cell carcinoma of the tongue.

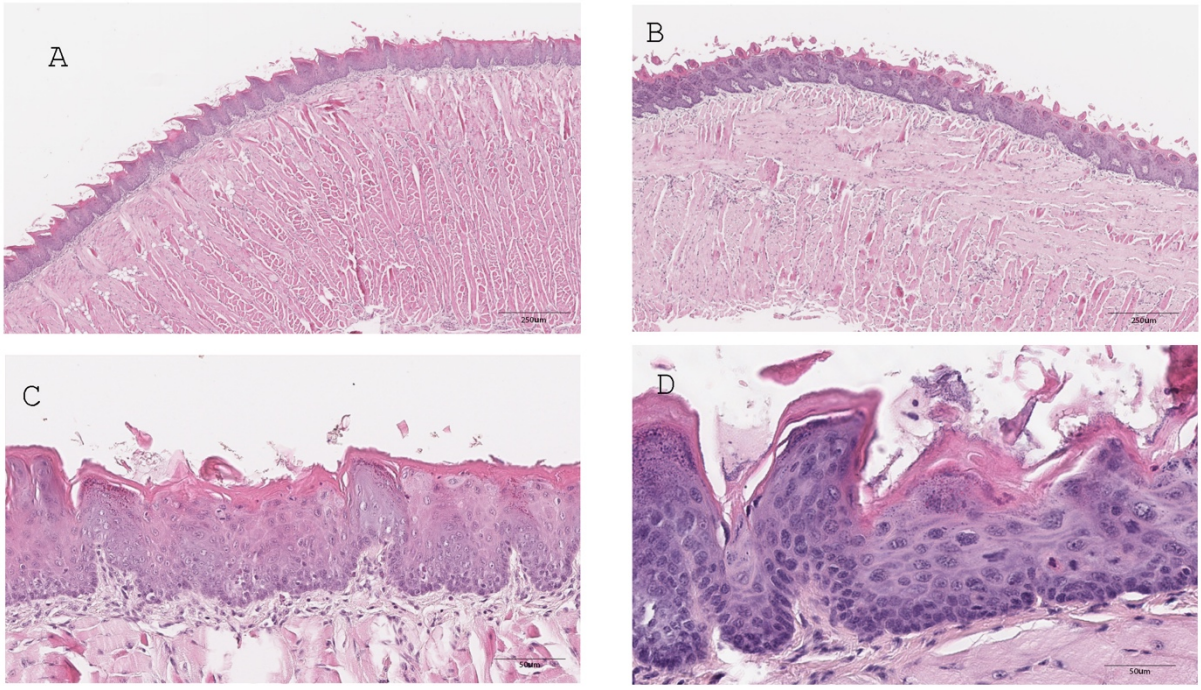


Figure 12. H&E WT and MUT tongue. H&E low power view of squamous cell epithelium of the tongue in WT mouse (A) and transgenic K14HPV16 (B). General aspect with a high-power view of the squamous epithelium of the tongue in the WT mouse (C) and in the transgenic K14HPV16 (D).

2. PUBLISHED WORK

2.1. EXPLORING THE ROLES OF HPV16 VARIANTS IN HEAD AND NECK SQUAMOUS CELL CARCINOMA: CURRENT CHALLENGES AND OPPORTUNITIES

Virology Journal. DOI: 10.1186/s12985-021-01688-9

Abstract

The incidence of squamous cell carcinomas of the head and neck (HNSCC) is consistently increasing, in association with human papillomavirus (HPV) infection, especially HPV16. HPV variants show heterogeneity in the pathogenicity of cervical cancer, but little has been established about their relevance on HNSCC. This review addresses the distribution of HPV16 variants in HNSCC and their potential contribution to clinical practice. A search was performed in PubMed using the keywords HNSCC HPV16 variants. Sixty articles were identified between 2000 and 2020 and 9 articles were selected for a systematic analysis. Clinical cohorts comprised 4 to 253 patients aged between 17 and 91 years with confirmed HPV16-positive HNSCC. Samples were collected from fresh biopsies of the tumour, oral rinse or formol fixed/paraffin-embedded tissue, from the oral cavity, oropharynx, hypopharynx, larynx and Waldeyer's tonsillar ring. HPV16 variants were identified using Sanger sequencing techniques. Seven studies addressed the HPV16 E6 gene, one studied E6 and E7, another studied L1 and one focused on the long control region. European variants represent 25-95%, Asian-American 5-57% and African 2-4% of the total isolates, suggesting a marked predominance of European strains. No correlations could be drawn with patient prognosis, partly because many studies relied on small patient cohorts. Additional studies are needed, particularly those employing next-generation sequencing techniques (NGS), which will allow faster and more accurate analysis of large numbers of samples.

Keywords: HNSCC, HPV16, HPV16 variants, next-generation sequencing

Cochicho *et al. Virol J* (2021) 18:217
https://doi.org/10.1186/s12985-021-01688-9

Virology Journal

REVIEW Open Access

Exploring the roles of HPV16 variants in head and neck squamous cell carcinoma: current challenges and opportunities

Daniela Cochicho[1,2†], Rui Gil da Costa[3,4,5†] and Ana Felix[1,6*†]

Abstract

The incidence of squamous cell carcinomas of the head and neck (HNSCC) is consistently increasing, in association with human papillomavirus (HPV) infection, especially HPV16. HPV variants show heterogeneity in the pathogenicity of cervical cancer, but little has been established about their relevance on HNSCC. This review addresses the distribution of HPV16 variants in HNSCC and their potential contribution to clinical practice. A search was performed in PubMed using the keywords HNSCC HPV16 variants. Sixty articles were identified between 2000 and 2020 and 9 articles were selected for a systematic analysis. Clinical cohorts comprised 4 to 253 patients aged between 17 and 91 years with confirmed HPV16-positive HNSCC. Samples were collected from fresh biopsies of the tumour, oral rinse or formol fixed/paraffin embedded tissue, from the oral cavity, oropharynx, hypopharynx, larynx and Waldeyer's tonsillar ring. HPV16 variants were identified using Sanger sequencing techniques. Seven studies addressed the HPV16 *E6* gene, one studied *E6* and *E7*, another studied *L1* and one focused on the long control region. European variants represent 25–95%, Asian-American 5–57% and African 2–4% of the total isolates, suggesting a marked predominance of European strains. No correlations could be drawn with patient prognosis, partly because many studies relied on small patient cohorts. Additional studies are needed, particularly those employing next generation sequencing techniques (NGS), which will allow faster and accurate analysis of large numbers of samples.

Keywords: HNSCC, HPV16, HPV16 variants, Next generation sequencing

Background

Head and neck cancer is the sixth leading cancer by incidence worldwide [1] and comprises many different pathological entities. The diagnosis is often characterized by multifocal development and presentation at an advanced stage. [2] The age at diagnosis usually ranges between 50 and 70 years of age and men are significantly more likely to develop the disease than women (up to a 4:1 proportion, depending on geographical localization) [3]. Currently, two major carcinogenic pathways leading to head and neck squamous cell carcinomas are recognized: the first is associated with risk factors like smoking and abusive alcohol consumption while the second is associated with human papillomavirus (HPV) infection [4, 5]. Over the past decade, several epidemiologic studies reported a 36.5% increase in the incidence of head and neck squamous cell carcinomas (HNSCC), particularly in high-income countries and among men < 60 years [4, 6–8]. Available data indicates that these changes specifically involve oropharyngeal cancers [9, 10]. Although tobacco consumption has decreased, the incidence of HPV-positive oropharyngeal cancers has increased [11], indicating that HPV infection is the underlying cause for the overall increase in HNSCC incidence [12–14].

*Correspondence: ana.felix@nms.unl.pt
†Daniela Cochicho, Rui Gil da Costa and Ana Felix contributed equally to this work
[1] NOVA Medical School, NOVA University of Lisbon, CEDOC, Campo Mártires da Pátria 130, 1169-056 Lisbon, Portugal
Full list of author information is available at the end of the article

Patients showing oral HPV infection are 53 times more likely to develop HNSCC [15]. Presently, HPV is present in approximately 35% of HNSCC and most HPV-positive cases emerge in lingual and palatine tonsils [16]. The presence of HPV DNA in tumour cells defines a specific pathologic entity within HNSCC with specific epidemiology, molecular characteristics, and biological behaviour [17]. HPV interferes with key signaling pathways to promote carcinogenesis via its viral oncoproteins E6 and E7, which lead to the inactivation of tumour protein 53 (p53) and the retinoblastoma protein (pRB), respectively [18]. HPV is a group of small, double-stranded DNA viruses that infect the epidermis and keratinizing mucosae. More than 100 HPV types have been identified [19, 20]. Presently, 40 different HPV types are known to infect mucosal epithelia and are categorized into low-risk and high-risk HPV types according to their epidemiologic association with cervical cancer [21]. In HNSCC, the majority of HPV types belong to this group including HPV16, HPV18, HPV39 and HPV45 [22]. HPV16 is by far, the most detected type, accounting for 90% of all HPV-positive HNSCC cases [23], a significantly greater proportion than in cervical cancer where it accounts for little over 50% of cases [24]. Sequencing of the HPV genome also revealed intra-type variants with genetic differences ranging between 0.5 and 1% [25, 26]. HPV variants arise mainly from nucleotide substitutions in some restricted positions in the genome coding region or in the noncoding region [27, 28]. The prevalence of variants for each HPV type varies significantly in different geographical areas [27]. Regarding HPV16, whole-genome analysis allowed the characterization of five distinct phylogenetic clusters named according to their original geographical distribution: European (E), Asian (As), Asian-American (AA), African 1 (Af1) and African 2 (Af2) [29–31]. Subsequently, a new branch, North American 1 (NA1) was identified [32, 33]. Currently, emerging epidemiological, etiological, and molecular data suggest that intra-type HPV variants are biologically distinct and may be associated with different risk of cervical cancer progression [28]. I*n vitro* studies using 3D organotypic epithelial cell cultures showed that HPV16 *E6* variants differ in their ability to abolish keratinocyte differentiation and to induce p53 degradation [34–36]. Those experimental results are corroborated by clinical studies on cervical cancer. The T350G (L83V) HPV16 variant is the most frequently found among invasive cervical cancers [21, 37] and has been linked to a higher oncogenic potential than the prototype [38], possibly by facilitating persistent viral infection, a critical factor for cancer development [39–41]. Despite the growing number of studies concerning HPV16 variants on cervical cancer, there are currently very few reports describing their distribution in HNSCC and nothing is established about their impact on the development, treatment response and impact on disease outcome. Nonetheless, HPV16 variants may also play an important role in head and neck carcinogenesis and studying their impact is needed due to the rising incidence of HPV-positive HNSCC [18, 23, 25, 42]. The purpose of this work is to systematically review the distribution of HPV16 variants in HNSCC and to assess the available knowledge concerning their potential contribution for HNSCC pathologic heterogeneity in published studies.

Main text

Review compilation data

A systematic review was performed on the PubMed electronic database (https://pubmed.ncbi.nlm.nih.gov) using the following keywords search criteria: HNSCC HPV16 variants. Observational studies reporting the distribution of HPV16 variants in HNSCC and published between 2000 and 2020 were included. Review articles, case reports and articles dealing with other types of cancer were excluded (Fig. 1). The selection search criteria were based on tumour localization, nature of sample and background of the variant study field. The studies were independently assessed to identify the prevalence of HPV16 variants in HNSCC and evaluate in respective cohorts for correlation with clinicopathological parameters. HNSCC

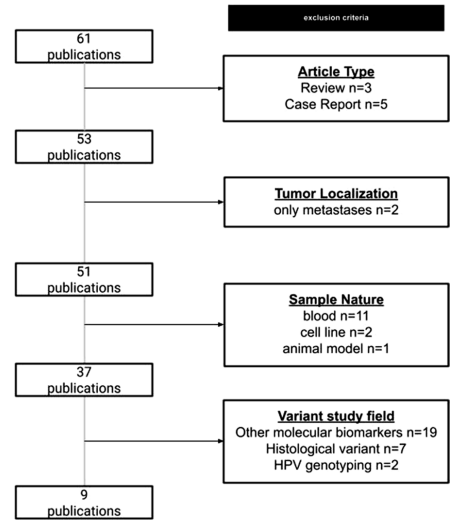


Fig. 1 Analytical flow for systematically reviewing published articles dealing with HPV16 variants in HNSCC

patients were categorized into different sites (oral cavity and oropharynx) and further into sub cohorts based on the demographic and clinical information provided. The published data were summarized using frequencies and percentages for HPV16 variants, stratified by tumour location, types of tumour samples, patient cohort size, sequencing methodology, and clinical variables such as age and gender (Fig. 2.)

General findings

From 61 records found using the search criteria HNSCC HPV16 DNA cell carcinoma variants frequency patients' cohorts, we selected 9 studies while 52 studies were excluded from further analysis, including review articles and papers which fit the exclusion criteria (Fig. 1). Selected articles report data from a total of 945 patients in the period between 2000 and 2020, distributed by 6 countries. (Table 1) Three studies were performed in Europe, three in the USA, two in Asia and Middle East (Japan and Iran) and one in South America (Brazil). HPV16 variants were evaluated in nine different clinical cohorts. All studies were different and did not contain the evaluation of the same cases. Overall, the study population consisted of patients whose age at diagnosis ranged between 17 and 91 years. The gender ratio was 1:3 or 1:4 (female to male), and all cases were histologically confirmed as HNSCC. The specimens used to evaluate HPV16 were primary tumours located in the oral cavity, oropharynx, hypopharynx, larynx and Waldeyer's tonsillar ring. Regional lymph node metastases were also included in one study. The samples used to analyze were diverse and consisted of fresh tissue, oral rinse, or formalin fixed/paraffin embedded tissue.

The use of the p16INK4a marker is to clinically detect an oncogenically active HPV infection and is considered as a surrogate marker for HPV infection [43, 44]. From the 9 studies evaluated three described the expression pattern of p16INK4a among negative HNSCC HPV(−) and HNSCC HPV16 cases. For all studies there was a significant difference in p16INK4a expression between HNSCC HPV(−) and HNSCC HPV16(+). Pakdel et al. 2020 observed the expression of p16INK4a among HPV variants and showed that a strong p16INK4a expression was present in poorly differentiated tumour tissues infected with HPV16 sublineage A2 [45].

HPV detection and genotyping

HPV was detected using PCR-based techniques employing consensus degenerate primers: MY09/11 (4 studies) MY09/11 and/or GP05/06 (2 studies), PGMY (one study) and SPF1/2 (one study); all of which are complementary to the conserved *L1* region. One study detected HPV using in situ hybridization. HPV genotyping was performed using different methods. Two studies using a commercial kit (INNO LIPA) which identifies 28 different HPV types (6, 11, 16, 18, 26, 31, 33, 35, 39, 40, 43, 44, 45, 51, 52, 53, 54, 56, 58, 59, 66, 68, 69, 70, 71, 73, 74, 82) by reverse blot hybridization. Two other studies used TaqMan PCR methods targeting the *E6* or *E6*/LCR. One study used restriction fragment length polymorphism (RFLP) analysis. Two studies performed direct Sanger sequencing of the PCR products. The HPV frequency in the different cohorts ranged from 10 to 100% of the cases, and the HPV16 distribution ranged between 67 and 100% of all HPV positive cases (Table 1). Except the study done in 2020 in

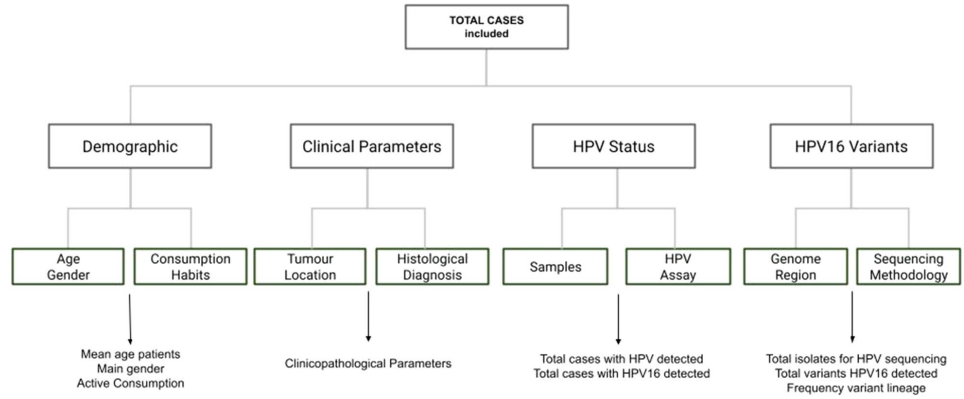


Fig. 2 Analytical variables flow for included articles dealing with HPV16 variants in HNSCC search criteria

Table 1 Identification of HPV16 variants in HNSCC in the literature (PubMed 2000–2020)

Year publication	Reference	Country	Mean age patients (yrs)	Gender (% male)	Consumption (% active)	Samples	Total cases (n)	Total cases HPV detected %(n)	Total cases HPV16 detected %(n)	Total isolates for HPV sequencing	Genome region location	Amplicon size (pb)	Sequencing
2000	[48]	USA	63	no data available	87%	Fresh tissue	253	24% (62)	90% (56)	52	E6	455	Sanger
2004	[49]	Germany	no data available	no data available	no data available	FF/PE	24	100% (24)	100% (24)	21	E6 and E7	793	Sanger
2007	[50]	Italy	63	74%	67%	Fresh tissue FF/PE	115	18% (21)	67% (14/21)	13	L1	150	Sanger
2008	[51]	USA	57	77%	76%	Oral rinse FF/PE	135	32% (44)	100% (44/44)	19	E6	609	Sanger
2009	[52]	Japan	64	86%	79%	Fresh tissue	77	10% (8)	100% (8/8)	8	E6	323	Sanger
2013	[53]	Brazil	59	100%	50%	FF/PE	4	100% (4)	100% (4/4)	4	E6	no data available	Sanger
2015	[54]	Italy	65	88%	63%	FF/PE	24	100% (10)	100% (10/10)	10	E6	no data available	Sanger
2016	[55]	USA	60	86%	No data available	FF/PE	205	18% (36)	70% (25)	21	LCR	193	Sanger
2020	[56]	Iran	56	58%	No data available	FF/PE	108	23% (25)	16% (17)	13	E6	no data available	Sanger

E6 Early gene 6; *E7* early gene 7; *LCR* long control region; *L1* late gene 1; *FF/PE* Formalin fixed Paraffin embedded

Iran, where just 16% of the HPV detected were HPV16. The most common HPV in their series was HPV16 but also followed by HPV18 and HPV11 and the cases were almost located in Oral cavity and Larynx; only 6 tonsils in all 108 cases were evaluated, explaining the low HPV frequency overall.

Identification HPV16 variants

Of the 9 studies, seven studies identified HPV16 variants by sequencing the *E*6 gene, one addressed variants in the *L*1 gene and another the long control region. In one study the *E*7 gene was also sequenced along with *E*6 (Table 1). In all studies, the identification of HPV16 variants was performed by Sanger sequencing techniques and by comparison to the reference sequence (prototype). Next generation sequencing (NGS) was not used in any of the studies evaluated. However, the reliance on conventional Sanger sequencing is likely to limit the number of cases that can be analyzed and the reliability of results. [26, 46, 47] Sanger sequencing techniques show limitations concerning the size of amplicons, which may compromise the assay sensitivity. Additionally, to overcome polymerase errors and guarantee the integrity and robustness of the results, it is necessary to perform replicates for each case or vector molecular assays [48]. Such replicates add significantly to the workload and costs and constitute a limitation to analyze large patient cohorts. In fact, only two out of nine studies included over 200 HNSCC patients and none analyzed more than 52 HPV-positive patients for HPV16 variants. Future studies aiming to study larger patient cohorts are likely to benefit from NGS techniques, which provide faster and reliable sequencing results of larger amplicons [25, 26, 46].

Prevalence of HPV16 variants

The frequency of each HPV16 variant (E, NA, AF, AS and AA) was estimated for each of the 9 studies included in the systematic review (Table 2). Gillison et al. [37] first reported data concerning the distribution of HPV16 E6 variants in HNSCC. The authors evaluated 52 HPV16-positive patients from an overall HNSCC cohort comprising 259 patients. The age at diagnosis ranged between 17 and 91 years-old (median, 63 years-old) and the majority were smokers, with or without alcohol consumption (87%). Tumours were localized in the nasopharynx (n = 2), oral cavity (n = 84), oropharynx (n = 60), hypopharynx (n = 21), larynx (n = 86). HPV-positive patients showed significantly improved disease outcomes compared with HPV-negative patients. All samples used were HNSCC fresh tumour specimens. The HPV16 variants were classified into the same phylogenetic group as the European prototype in 75% of cases, Asian in 17%, North American in 4.0% and African 1 in 4.0% of cases. Six novel variants not previously reported (E-G315T, E-G315G, E-C395G, E-A478T, E-A132T, Af1-C311, Af1-A389) were also identified. The authors remarked that the distribution of variants in HNSCC resembled that observed in cervical cancer. However, no conclusions were drawn concerning the potential contribution of specific HPV16 *E6* variants for increasing cancer risk or modifying tumour biopathology and prognosis. Four years later, Hoffmann et al. [38] identified HPV16 variants in 7 out of 21 tumour specimens of HNSCC. The authors analyzed the *E*6 and *E*7 ORFs. Altogether, the DNA samples carried HPV16 prototype European

Table 2 Frequency of HPV16 variants

Reference	Total cases HPV16 detected (*n*)	Total isolates HPV sequencing (*n*)	HPV16 variant region location	European (E)		North American	African	Asian	Asia American
				All lineages (*n*)	E-350-G (*n*)	(NA) (*n*)	(AF1|AF2) (*n*)	(AS) (*n*)	(AA) (*n*)
[48]	56	52	E6	39	6	2	2	9	0
[49]	24	21	E6 E7	15	8	0	0	0	0
[50]	14	13	L1	9	0	1	2	0	1
[51]	44	19	E6	18	4	0	0	1	0
[52]	8	8	E6	5	5	0	0	0	0
[53]	4	4	E6	2	1	0	0	0	2
[54]	10	10	E6	8	8	0	0	0	0
[55]	25	21	LCR	9	–	0	0	0	12
[56]	17	13	E6	11	–	0	0	0	2

(–) not evaluated; *AA* Asia-American HPV16 Variant; *AF* African HPV16 Variant; *AS* Asian HPV16 variant; *E* European HPV16 variant; *E6* HPV Early protein 6; *E7* HPV Early protein 7; *LCR* long control region; *L1* late gene; *NA* North American HPV16 Variant

variant (29%), the European variant T350G (38%) and 33% Euro-German variant (A131G + C712A). Again, no conclusions could be drawn concerning the clinical-pathological relevance of the HPV16 variants identified, partly because of the cohort's small size. Badaracco et al. [49], published molecular analysis data on the HPV16 *L1* ORF, from an Italian cohort (total n = 115, of which only 13 were tested for HPV variants), composed of 86 men and 29 women, with a mean age of 63.21 years old. A high percentage of patients were smokers (67%) and alcohol drinkers 43%. Tumours were localized mostly in the oral cavity (n = 60), followed by the larynx (n = 30), the oropharynx (n = 10), the tonsil area (n = 8), the hypopharynx (n = 5) and the sinus/nose (n = 2). In this study, the presence of HPV was not significantly associated with disease-free survival at 2 years. Sixty nine percent of the cases (13 cases) analyzed showed European-German variants (9 cases), 15% were African type 2 (2 cases), 8% Asian-American (1 case), and the remaining 8% (1 case) had an unclassified variant. The predominance of the European variant was unsurprising considering the Italian origin of the patients. No correlations were drawn between the presence of HPV16 variants and any epidemiological, pathological, or clinical data. Agrwal et al. [50], studied 19 HPV16 isolates from a universe of 135 HNSCC samples. The median age at diagnosis was 57 years-old, 77% of the patients were men, most patients had a history of smoking (n = 62 smokers versus 51 non-smokers) but were non-drinkers (73 non-drinkers versus 41 drinkers). All analyzed cases showed European variants and a single case carried an Asian variant. The most common European variant was E-350 T (n = 6), followed by E-350G (n = 4) and E-T131G (n = 2). Importantly, eight of the 19 isolates contained European variants with sequences unique to a single individual. Boscolo-Rizzo et al. [51] analyzed HPV E6 variants in a short (n = 8) case series. The authors showed the presence of the T350G mutation in 5 cases located in the oral cavity and oropharynx, while three tumours located in the larynx and hypopharynx contained HPV 16 prototype sequences. Joseph et al. [52] compared HPV16 variants present in four patients with bilateral tonsillar HNSCC. Two cases carried European variants while two others carried Asian-American variants. The results show that, in all 4 patients, the same HPV16 variant was present in the bilateral tumours, supporting the hypothesis that a single HPV infection, rather than independent infections with distinct agents, is responsible for those bilateral tumours. Hassani et al. [53] studied the HPV16 variants present in 10 cases of tonsillar HNSCC. The authors found that the E-350G-variant was present in 80% of cases while the European prototype was identified in the other 20%. Again, this short case series did not provide data concerning the pathobiological relevance of HPV16 variants. In the following year, Betiol et al. [54] reported the distribution of HPV16 variants in a Brazilian cohort of 21 HNSCC patients. The authors analyzed the HPV16 LCR and found that 12 (57.1%) patients carried European and 9 carried Asian-American (42.9%) variants. The authors remarked that the slight predominance of European variants accompanied observation from their normal cervical samples, suggesting that the distribution of HPV16 variants reflects the overall frequency in each studied population. The most recent study was published by Pakdel et al. [45]. Thirteen HNSCC tissue specimens tested positive for HPV16 using overlapping PCR assays and were analyzed for the presence of *E6* variants. There was a marked predominance of European variants (84.6%) followed by Asian-American (15.4%) variants.

Discussion

We reviewed the published literature on HPV16 variants in HNSCC, aiming to evaluate the biological meaning of those variants in this particular location. The clinical management of HNSCC improved greatly since the recognition of HPV-positive and HPV-negative lesions. Although tumour recurrence still occurs in 10–20% of HPV-positive HNSCC patients, the majority of these patients clearly benefit from therapeutic de-escalation [55]. The lack of adequate biomarkers to define more tailored approaches is still necessary in SCC HPV associated tumours [56]. The large majority of these tumours are associated with HPV16 and a deeper understanding of the bio pathological implications of distinct HPV16 variants would contribute to the molecular characterization of HPV-positive HNSCC and may help to define patient's subgroups that would benefit from specific therapeutic approaches. Data from cervical cancer patients showed an association between the presence of non-European HPV16 lineages, a longer viral persistence [57, 58] and an increased risk of developing high-grade cervical intraepithelial neoplasia [59]. Apparently, within individuals, HPV genomes harbour high levels of variability in HPV16 genome sequence upon normal to pre-cancer/cancer [60] revealing several fundamental discoveries and suggesting a paradigm shift from HPV16 as a single viral entity to theorize each HPV16 isolated to be a separate virus with distinct carcinogenic potential [61]. This would imply that within HPV16, the genetic variation partly predicts the risk of pre-cancer and cancer [62]. In particular, specific sublineages (A4, C, D2, and D3) have shown a significantly increased risk compared to the most common A1/A2 sublineages [26] and assured D2, for the strongest risk of cancer within glandular epithelium (adenocarcinomas) [26].

The major data from whole-sequences obtained from individual clinical specimens agrees that the genetic variation occurs more commonly on low-grade or benign HPV16 infections [61] and corrected explained In case–control analyses describing the highest amino acid changing variants HPV16 in the controls throughout the genome with cervix cancer. [62] Therefore, E7 oncogene lacks nonsynonymous (amino acid changing) variants in cervical cancers, suggesting the E7 conservation is admitted for carcinogenicity. [61] The specific conservation of the 98 amino acids of E7, that directly disrupts Rb function, was shown to be crucial for trigger carcinogenesis, owing to be commented as a highly specific target for etiologic and therapeutic research [26].

Presenting data from a cohort study that compares genomic characteristics of HPV associated with cervical versus oropharyngeal tumours using DNA sequence analysis showed no significant differences between distribution in HPV16 variants [24], both presenting major prevalence of the European variant. Instead, the HPV E6 gene amplified from oropharyngeal samples reported over more nonsynonymous mutations, but also for the E7 gene, no differences were found in mutation rates between the two anatomical locations [24, 62]. Notably, the E7 gene is conserved in both locations corroborating the recent findings on cervical cancer studies. The apparent restriction on E7 mutations seems to be present in the oropharynx, as well. Indeed, described all both share common biological features, but the important differences present in HPV-genome may explain their distinct pathophysiological mechanisms and susceptibility to treatment. Nevertheless, the invariability of E7 presents an attractive potential target for therapy at both locations. [24]

In all the studies evaluated in this analysis, the most prevalent variant was the European Variant, which belongs to the European lineage and within this lineage the E-350-G HPV16 variant was the most frequent in majority of the studies sequencing based E6 and/or E7 regions with 29% and 53% respectively, although the cohorts had or not an European origin, (cohorts were from Brazil, United States of America, Japan, Italy and Germany (Table 3).

As it was established from molecular analysis studies in cervical cancer, that the European variant T350G, was the variant frequently found in cervical intraepithelial neoplasms and cancers, and has been associated with progression to cervical cancer particularly in North European women. The detection of the T350G variant in a large proportion of HNSCC patients, therefore, indicates that this variant might also play an important role in HN carcinogenesis. Hassani et al.in 2015 supported this hypothesis, elucidating that the HPV-16 E-350G variant has a polymorphism in residue 83, a leucine for valine (L83V), probably responsible for the increased cancer risk [53].

Conclusions

In HNSCC different lineages of HPV16 variants can be identified and differ geographically. Although most reviews described only the distribution of HPV16 variants in HNSCC, LeConte in a more detailed analysis of those studies found important differences. They found the distribution of two HPV16 variant groups differ significantly in oropharyngeal cancer and cervical cancer. The European + South America (E + AS) variant groups showed a higher prevalence in the oropharyngeal samples, representing 90.2%, than in cervical carcinomas (71.4%). Moreover, the Asia-American (AA1 + AA2) variant groups were present in 22.5% of cervical cancers in contrast to 4.4% in the oropharyngeal cancers [24]. The number of cases studied does not allow us to explore the importance of those differences nor to understand the potential role determining the clinical behaviour and potential use to select treatment and prognosis. Hopefully, the new sequencing era will enrich the study of HPV and related cancers. The advances in HPV whole-genome sequencing [46] provided technically achievable large-scale longitudinal studies on HPV whole-genomic sequences and promotes an exhaustive understanding of

Table 3 Frequency HPV16 Variants by region genome location

Reference	Total number of cases	Total isolates for HPV sequencing	HPV16 variant region location	European		North American (NA)	African (AF1|AF2)	Asian (AS)	Asia American (AA)
				All lineages	E-350-G				
[50]	115	115	L1	69%	Not Detected	-	15%	-	8%
[48, 51–54, 56]	621	106	E6	78%	29%	2%	2%	9%	4%
[49]	24	21	E6E7	71%	53%	-	-	-	-
[55]	205	21	LCR	42.8%	-	-	-	-	57.2%

(–) not evaluated; *AA* Asia-American HPV16 Variant; *AF* African HPV16 Variant; *AS* Asian HPV16 Variant; *E* European HPV16 variant; *E6* HPV Early protein 6; *E7* HPV Early protein 7; *LCR* long control region; *L1* late gene; *NA* North American HPV16 variant

the viral genetic diversity within and between infected individuals and will make the link between variants and cancer risk, comprehensively [46, 61].

Abbreviations
AA: Asia-American HPV16 variant; AF: African HPV16 variant; AS: Asian HPV16 variant; E: European HPV16 variant; E6: HPV Early protein 6; E7: HPV Early protein 7; HNSCC: Head and neck squamous cell carcinoma; HPV: Human papillomavirus; NA: North American HPV16 VARIANT; NGS: Next generation sequencing.

Acknowledgements
Not applicable.

Authors' contributions
These authors contributed equally to this work. Study concepts: AF, DC, RGC. Study design: AF, DC, RGC. Data acquisition: DC, AF, RGC. Data analysis and interpretation: RGC, DC, AF. Manuscript preparation: DC. Manuscript editing: AF, DC, RGC. Manuscript review: AF, DC, RGC. All authors read and approved the final manuscript.

Funding
Not applicable.

Availability of data and materials
Not applicable.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Author details
[1]NOVA Medical School, NOVA University of Lisbon, CEDOC, Campo Mártires da Pátria 130, 1169-056 Lisbon, Portugal. [2]Virology Laboratory IPOLFG, Rua Professor Lima Bastos, 1099-023 Lisbon, Portugal. [3]LEPABE, Laboratory for Process Engineering, Environment, Biotechnology and Energy, University of Porto, Rua Dr. Roberto Frias, 4200-465 Porto, Portugal. [4]Post-graduate Programme in Adult Health (PPGSAD), University Hospital (HUUFMA) and Morphology Department, Federal University of Maranhão, Av. dos Portugueses 1966 - Vila Bacanga, São Luís, MA 65080-805, Brazil. [5]Molecular Oncology and Viral Pathology Group, Research Center of IPO Porto (CI-IPOP) / RISE@CI-IPOP (Health Research Network), Portuguese Oncology Institute of Porto (IPO Porto) / Porto Comprehensive Cancer Center (Porto.CCC), Rua Dr. António Bernardino de Almeida, 4200-072 Porto, Portugal. [6]Pathology Department IPOLFG, Rua Professor Lima Bastos, 1099-023 Lisbon, Portugal.

Received: 27 July 2021 Accepted: 28 October 2021
Published online: 08 November 2021

References

1. Ferlay J, Soerjomataram I, et al. Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. Int J Cancer. 2015. https://doi.org/10.1002/ijc.29210.
2. De Bree R, Leemans CR. Recent advances in surgery for head and neck cancer. Curr Opin Oncol. 2010. https://doi.org/10.1097/CCO.0b013e3283380009.
3. Bray F, Ferlay J, et al. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2018. https://doi.org/10.3322/caac.21492.
4. McDermott JD, Bowles DW. Epidemiology of head and neck squamous cell carcinomas: impact on staging and prevention strategies. Curr Treat Options Oncol. 2019. https://doi.org/10.1007/s11864-019-0650-5.
5. Gillison ML. Human papillomavirus-associated head and neck cancer is a distinct epidemiologic, clinical, and molecular entity. Semin Oncol. 2004. https://doi.org/10.1053/j.seminoncol.2004.09.011.
6. Chaturvedi AK, Anderson WF, et al. Worldwide trends in incidence rates for oral cavity and oropharyngeal cancers. J Clin Oncol. 2013. https://doi.org/10.1200/JCO.2013.50.3870.
7. Hwang TZ, Hsiao JR, et al. Incidence trends of human papillomavirus related head and neck cancer in Taiwan. Int J Can. 2015. https://doi.org/10.1002/ijc.29330.
8. Mork J, Moller B, et al. Time trends in pharyngeal cancer incidence in Norway 1981–2005: a subsite analysis based on a re-abstraction and recoding of registered cases. Can Causes Control. 2010. https://doi.org/10.1007/s10552-010-9567-9.
9. Zumsteg ZS, Cook-Wiens G, et al. Incidence of oropharyngeal cancer among elderly patients in the United States. JAMA Oncol. 2016. https://doi.org/10.1001/jamaoncol.2016.1804.
10. Tota JE, Best AF, et al. Evolution of the oropharynx cancer epidemic in the United States: moderation of increasing incidence in younger individuals and shift in the burden to older individuals. J Clin Oncol. 2019. https://doi.org/10.1200/JCO.19.00370.
11. Faraji F, Eisele DW, et al. Emerging insights into recurrent and metastatic human papillomavirus-related oropharyngeal squamous cell carcinoma. Laryngoscope Investig Otolaryngol. 2017. https://doi.org/10.1002/lio2.37.
12. Chaturvedi AK, Engels EA, et al. Human papillomavirus and rising oropharyngeal cancer incidence in the United States. J Clin Oncol. 2011. https://doi.org/10.1200/JCO.2011.36.4596.
13. Hammarstedt L, Lindquist D, et al. Human papillomavirus as a risk factor for the increase in incidence of tonsillar cancer. Int J Can. 2006. https://doi.org/10.1002/ijc.22177.
14. Hocking JS, Stein A, et al. Head and neck cancer in Australia between 1982 and 2005 show increasing incidence of potentially HPV-associated oropharyngeal cancers. Br J Can. 2011. https://doi.org/10.1038/sj.bjc.6606091.
15. Gillison ML, D'Souza G, et al. Distinct risk factor profiles for human papillomavirus type 6-positive and human papillomavirus type 6-negative head and neck cancers. J Natl Cancer Inst. 2008. https://doi.org/10.1093/jnci/djn025.
16. El-Mofty SK. HPV-related squamous cell carcinoma variants in the head and neck. Head Neck Pathol. 2012;6:55. https://doi.org/10.1007/s12105-012-0363-6.
17. WHO, World Health Organization. International agency for research on cancer (IARC) monographs on the evaluation of carcinogenic risks to humans. 2007;90(Human Papilloma Virus):689. https://monographs.iarc.who.int/wp-content/uploads/2018/06/mono90.pdf
18. Polanska H, Raudenska M, et al. Clinical significance of head and neck squamous cell cancer biomarkers. Oral Oncol. 2014. https://doi.org/10.1016/j.oraloncology.2013.12.008.
19. De Villiers EM, Fauquet C, et al. Classification of papillomaviruses. In: Virology. Academic Press. 2004; https://doi.org/10.1016/j.virol.2004.03.033
20. De Villiers EM, Gunst K. Characterization of seven novel human papillomavirus types isolated from cutaneous tissue, but also present in mucosal lesions. J Gen Virol. 2009. https://doi.org/10.1099/vir.0.011478-0.
21. Muñoz N, Bosch FX, et al. Epidemiologic classification of human papillomavirus types associated with cervical cancer. N Engl J Med. 2003. https://doi.org/10.1056/NEJMoa021641.
22. Kreimer AR, Clifford GM, et al. Human papillomavirus types in head and neck squamous cell carcinomas worldwide: a systemic review. Cancer Epidemiol Biomark Prev. 2005. https://doi.org/10.1158/1055-9965.EPI-04-0551.
23. Huertas-Salgado A, Martín-Gámez DC, et al. E6 molecular variants of human papillomavirus HPV type 16: an updated and unified criterion for clustering and nomenclature. Virology. 2011. https://doi.org/10.1016/j.virol.2010.10.039.
24. LeConte BA, Szaniszlo P, et al. Differences in the viral genome between HPV-positive cervical and oropharyngeal cancer. PLOS ONE. 2018. https://doi.org/10.1371/journal.pone.0203403.
25. Burk RD, Harari A, et al. Human papillomavirus genome variants. Virology. 2013. https://doi.org/10.1016/j.virol.2013.07.018.Human.

26. Mirabello L, Yeager M, et al. HPV16 sublineage associations with histology-specific cancer risk using HPV whole-genome sequences in 3200 women. J Natl Cancer Inst. 2016. https://doi.org/10.1093/jnci/djw100.
27. Sichero L, Simão Sobrinho J, et al. Oncogenic potential diverge among human papillomavirus type 16 natural variants. Virology. 2012. https://doi.org/10.1016/j.virol.2012.06.011(8).
28. Bernard HU, Calleja-Macias IE, et al. Genome variation of human papillomavirus types: phylogenetic and medical implications. Int J Cancer. 2006. https://doi.org/10.1002/ijc.21655.
29. Ho L, Chan SY, et al. Sequence variants of human papillomavirus type 16 in clinical samples permit verification and extension of epidemiological studies and construction of a phylogenetic tree. J Clin Microbiol. 1991;29:1765–72. https://doi.org/10.1128/jcm.29.9.1765-1772.1991.
30. Ho L, Chan SY, et al. The genetic drift of human papillomavirus type 16 is a means of reconstructing prehistoric viral spread and the movement of ancient human populations. J Virol. 1993;67:6413–23. https://doi.org/10.1128/jvi.67.11.6413-6423.1993.
31. Ho L, Tay SK, et al. Sequence variants of human papillomavirus type 16 from couples suggest sexual transmission with low infectivity and polyclonality in genital neoplasia. J Infect Dis. 1993. https://doi.org/10.1093/infdis/168.4.803.
32. Yamada T, Wheeler CM, et al. Human papillomavirus type 16 variant lineages in United States populations characterized by nucleotide sequence analysis of the E6, L2, and L1 coding segments. J Virol. 1995;69:7743–53.
33. Yamada T, Manos MM, et al. Human papillomavirus type 16 sequence variation in cervical cancers: a worldwide perspective. J Virol. 1997;71:2463–72. https://doi.org/10.1128/jvi.71.3.2463-2472.1997.
34. Jackson R, Rosa BA, et al. Functional variants of human papillomavirus type 16 demonstrate host genome integration and transcriptional alterations corresponding to their unique cancer epidemiology. BMC Genomics. 2016. https://doi.org/10.1186/s12864-016-3379-6.
35. Nordfors C, Sobkowiak M, et al. Human papillomavirus (HPV) 16 E6 variants in tonsillar cancer in comparison to those in cervical cancer in Stockholm, Sweden. PLOS ONE. 2012. https://doi.org/10.1371/journal.pone.0036239.
36. Stöppler MC, Ching K, et al. Natural variants of the human papillomavirus type 16 E6 protein differ in their abilities to alter keratinocyte differentiation and to induce p53 degradation. J Virol. 1996. https://doi.org/10.1128/jvi.70.10.6987-6993.1996.
37. Gillison ML, Koch WM, et al. Evidence for a causal association between human papillomavirus and a subset of head and neck cancers. J Natl Cancer Inst. 2000;92:709–20. https://doi.org/10.1093/jnci/92.9.709.
38. Hoffmann M, Lohrey C, et al. Human papillomavirus type 16 E6 and E7 genotypes in head-and-neck carcinomas. Oral Oncol. 2004. https://doi.org/10.1016/j.oraloncology.2003.10.011.
39. Xi LF, Demers GW, et al. Analysis of human papillomavirus type 16 variants indicates the establishment of persistent infection. J Infect Dis. 1995. https://doi.org/10.1093/infdis/172.3.747(3).
40. Van Belkum A, Juffermans L, et al. Genotyping human papillomavirus type 16 isolates from persistently infected promiscuous individuals and cervical neoplasia patients. J Clin Microbiol. 1995. https://doi.org/10.1128/jcm.33.11.2957-2962.1995.
41. Zhang L, Liao H, et al. Variants of human papillomavirus type 16 predispose toward persistent infection. Int J Clin Exp Pathol. 2015;87:8453.
42. The Cancer Genome Atlas Network Comprehensive genomic characterization of head and neck squamous cell carcinomas. Nature. 2015; doi: https://doi.org/10.1038/nature14129
43. Castellsagué X, Alemany L, et al. HPV involvement in head and neck cancers: comprehensive assessment of biomarkers in 3680 patients. J Natl Cancer Inst. 2016. https://doi.org/10.1093/jnci/djv403.
44. Dok R, Nuyts S. HPV positive head and neck cancers: Molecular pathogenesis and evolving treatment strategies. Cancers (Basel). 2016. https://doi.org/10.3390/cancers8040041.
45. Pakdel F, Farhadi A, et al. The frequency of high-risk human papillomavirus types, HPV16 lineages, and their relationship with p16INK4a and NF-κB expression in head and neck squamous cell carcinomas in Southwestern Iran. J Microbiol. 2020. https://doi.org/10.1007/s42770-020-00391-1.
46. Cullen M, Boland JF, et al. Deep sequencing of HPV16 genomes: A new high-throughput tool for exploring the carcinogenicity and natural history of HPV16 infection. Papillomavirus Res. 2015. https://doi.org/10.1016/j.pvr.2015.05.004.
47. Lavezzo E, Masi G, et al. Characterization of intra-type variants of oncogenic human papillomaviruses by next-generation deep sequencing of the E6/E7 region. Viruses. 2016. https://doi.org/10.3390/v8030079.
48. van der Weele P, Meijer CJLM, et al. Whole-Genome Sequencing and Variant Analysis of Hpv16 Infections. J Virol. 2017. https://doi.org/10.1128/JVI.00844-17.
49. Badaracco G, Rizzo C, et al. Molecular analyses and prognostic relevance of HPV in head and neck tumours. Oncol Rep. 2007
50. Agrawal Y, Koch WM, et al. Oral HPV infection before and after treatment for HPV16-positive and negative head and neck squamous cell carcinoma. Oral Oncol. 2008. https://doi.org/10.1158/1078-0432.CCR-08-0498.Oral.
51. Boscolo-Rizzo P, Da Mosto MC, et al. HPV-16 E6 L83V variant in squamous cell carcinomas of the upper aerodigestive tract. J Cancer Res Clin Oncol. 2009. https://doi.org/10.1007/s00432-008-0490-3.
52. Joseph AW, Ogawa T, et al. Molecular etiology of second primary tumours in contralateral tonsils of human papillomavirus-associated index tonsillar carcinomas. Oral Oncol. 2013. https://doi.org/10.1016/j.oraloncology.2012.09.009.
53. Hassani S, Castillo A, Ohori J-I, Higashi M, Kurono Y, Akiba S, Koriyama C, Molecular pathogenesis of human papillomavirus type 16 in Tonsillar squamous cell carcinoma. Anticancer Res. 2015
54. Betiol JC, de Matos LL, et al. Prevalence of human papillomavirus types and variants and p16INK4a expression in head and neck squamous cells carcinomas in São Paulo, Brazil. Infectious Agents Cancer. 2016. https://doi.org/10.1186/s13027-016-0067-8.
55. Ang KK, Harris J, et al. Human papillomavirus and survival of patients with oropharyngeal cancer. N Engl J Med. 2010. https://doi.org/10.1056/nejmoa0912217.
56. Beaty BT, Moon DH, et al. PIK3CA mutation in HPV-associated OPSCC patients receiving deintensified chemoradiation. JNCI J Natl Cancer Inst. 2019;112:855–8. https://doi.org/10.1093/jnci/djz224.
57. Villa LL, Sichero L, et al. Molecular variants of human papillomavirus types 16 and 18 preferentially associated with cervical neoplasia. J Gen Virol. 2000. https://doi.org/10.1099/0022-1317-81-12-2959.
58. Schiffman M, Rodriguez AC, et al. A population-based prospective study of carcinogenic human papillomavirus (HPV) variant lineages, viral persistence, and cervical neoplasia. NIH Public Access. 2010. https://doi.org/10.1158/0008-5472.CAN-09-4179.A.
59. Sichero L, Ferreira S, et al. High grade cervical lesions are caused preferentially by non-European variants of HPVs 16 and 18. Int J Cancer. 2007. https://doi.org/10.1002/ijc.22481.
60. Hirose Y, Onuki M, et al. Within-host variations of human papillomavirus reveal APOBEC signature mutagenesis in the viral genome. J Virol. 2018. https://doi.org/10.1128/jvi.00017-18.
61. Mirabello L, Clarke MA, et al. The intersection of HPV epidemiology, genomics and mechanistic studies of HPV-mediated carcinogenesis. Viruses. 2018. https://doi.org/10.3390/v10020080.
62. Zhu B, Xiao Y, et al. Mutations in the HPV16 genome induced by APOBEC3 are associated with viral clearance. Nat Commun. 2020. https://doi.org/10.1038/s41467-020-14730-1.

Publisher's Note

2.2 THE RELEVANCE OF HPV16 VARIANTS IN HEAD AND NECK SQUAMOUS CELL CARCINOMAS

Accepted for publication in Pathobiology.

Abstract

The variants of the human papillomavirus (HPV) are believed to play an important role in the carcinogenesis process of head and neck squamous cell carcinomas (HNSCC). This study aims to establish the prevalence of HPV16 variants in a Portuguese HNSCC cohort, describe clinical-pathological characteristics and determine the association with patient survival. We retrieved 68 HNSCC patients from the Institute of Oncology in Lisboa. DNA samples were available from tumour biopsy at the time of the primary diagnosis. Targeted NGS was used to analyse whole-genome sequences and variants were established based on phylogenetics classification. The majority of the HPV16 samples clustered in lineage A and the remaining clustering within sublineages B, C and D. Comparative genome analysis revealed a total of 1356 SNV fall within the coding regions E1, E2, E4, E5, E6, E7 early genes and L1 and L2 late genes. Our integrated approach reports a deep molecular HNSCC HPV16 mutational profile characterization, intended to describe molecular characteristics of HPV16 variants and their biological significance in HN cancer.

Keywords: K14HPV16, Variant

Research Article

Distribution and clinical significance of HPV16 variants in head and neck squamous cell carcinomas: data from a Portuguese cohort and systematic review

Daniela Cochicho[a,b], Alexandra Nunes[c], Daniel Sobral[c], João P. Gomes[c], Susana Esteves[d], Joana Mendonça[e], Luis Vieira[e], Luis Martins[b], Mário Cunha[b], Pedro Montalvão[f], Miguel Magalhães[f], Rui M. Gil da Costa[#g,h,i,j] Ana Félix[#a,l*]

[a]Research Department, NOVA Medical School University, Lisbon, Portugal;
[b]Virology Laboratory from Clinical Pathology Department, IPOLFG Portuguese Oncology Institute Francisco Gentil, Lisbon, Portugal;
[c]Bioinformatics Unit, Department of Infectious Diseases, National Institute of Health, Lisbon, Portugal;
[d]Clinical Research Unit, IPOLFG Portuguese Oncology Institute Francisco Gentil, Lisbon, Portugal;
[e]Technology and Innovation Unit, Department of Human Genetics, National Institute of Health, Lisbon, Portugal;
[f]Otorhinolaryngology Department, IPOLFG Portuguese Oncology Institute Francisco Gentil, Lisbon, Portugal;
[g]LEPABE, Laboratory for Process Engineering, Environment, Biotechnology and Energy, Faculty of Engineering, University of Porto, Rua Dr Roberto Frias, 4200-465 Porto, Portugal;
[h]ALiCE - Associate Laboratory in Chemical Engineering, Faculty of Engineering, University of Porto, Rua Dr. Roberto Frias, 4200-465 Porto, Portugal;
[i]Post-graduate Programme in Adult Health (PPGSAD), University Hospital (HUUFMA) and Morphology Department, Federal University of Maranhão, 65085-580 São Luís, Brazil;
[j]Molecular Oncology and Viral Pathology Group, Research Center of IPO Porto (CI-IPOP) / RISE@CI-IPOP (Health Research Network), Portuguese Oncology Institute of Porto (IPO Porto), Porto Comprehensive Cancer Center (Porto.CCC), 4200-072 Porto, Portugal
[l]Pathology Department, IPOLFG Portuguese Oncology Institute Francisco Gentil, Lisbon, Portugal;
# Equal contribution

Short Title: Distribution and clinical significance of HPV16 variants in HNSCC

Corresponding Author: Ana Maria Félix de Campos Pinto. CEDOC Campus NOVA Medical School, University of Lisbon, Campo dos Mártires da Pátria 130, 1169-056 Lisbon, Portugal. Email: ana.felix@nms.unl.pt; Phone: (+351) 217 200 475.

Number of Tables: 2

Number of Figures: 4

Word count: 2677 Abstract and body text (This is not to include the title page, reference list or figure legends)

Keywords: HNSCC; HPV16; HPV variants; whole-genome sequencing; oropharynx

Abstract

Introduction

Genomic variants of the human papillomavirus type 16 (HPV16) are thought to play differential roles in the susceptibility to head and neck squamous cell carcinomas (HNSCC) and its biological behaviour. This study aims to establish the prevalence of HPV16 variants in an HNSCC cohort and associate them with clinical-pathological characteristics and patient survival.
Methods
We retrieved samples and clinical data from 68 HNSCC patients. DNA samples were available from tumour biopsy at the time of the primary diagnosis. Targeted next-generation sequencing (NGS) was used to obtain whole-genome sequences and variants were established based on phylogenetic classification.
Results
74.% of samples clustered in lineage A, 5.7% lineage B, 2.9% C and 17.1% D. Comparative genome analysis revealed 243 single nucleotide variations (SNVs). Of these one hundred were previously reported, according to our systematic review. No significant associations with clinical-pathological variables or patient survival were observed. The E6 amino acid variations E31G, L83V and D25E and E7 N29S, associated with cervical cancer, were not observed, except for N29S in a single patient.
Discussion/Conclusion
These results provide a comprehensive genomic map of HPV16 in HSNCC, highlighting tissue-specific characteristics which will help design tailored therapies for cancer patients.

Introduction

Head and neck carcinoma is the sixth leading cancer by incidence worldwide [1,2]. The large majority of the cases (90%) are Squamous Cell Carcinomas (HNSCC) [3,4]. Currently, two major carcinogenic pathways are recognized: the first is associated with risk factors like smoking and abusive alcohol consumption while the second is associated with human papillomavirus (HPV) infection [5,6]. Presently, HPV is present in approximately 35% of HNSCC and most HPV-positive cases emerge in lingual and palatine tonsils [3,7]. Most available data show that HPV-associated HNSCC has a substantially better prognosis after therapy compared with HPV-negative cases [8]. Thus, the presence of HPV DNA in tumour cells defines a specific pathological entity within HNSCC with specific epidemiology, molecular characteristics, and biological behaviour [9]. In HNSCC, HPV16 is by far the most commonly detected type, accounting for 90% of all HPV-positive HNSCC cases [10,11]. Our group recently demonstrated the etiological role of HPV16 in tongue base cancer using a transgenic mouse model [12]. Sequencing of the HPV genome revealed intra-type variants with genetic differences ranging between 0.5 and 1% of all sequences [13,14]. Emerging data suggest that intra-type HPV variants are biologically distinct, have a wide broad geographical distribution and are associated with different risks of cancer progression in the uterine cervix [15] and, possibly, in the head and neck region as well [16,17]. Regarding HPV16, whole-genome analysis allowed the characterization of four main variant lineages (A, B, C, D) and ten sublineages (A1, A2, A3, A4, B1, C1, D1, D2, D3, D4) [18]. These define five distinct

phylogenetic clusters named according to their original geographical distribution: European (E), Asian (As), Asian-American (AA), African 1 (Af1), African 2 (Af2) [19,20] and North American 1 (NA1) [21,22]. Despite the growing number of studies concerning HPV16 variants in cervical cancer, there are very few reports describing their distribution in HNSCC. Recently, Lang Kuhs *et al*. conducted the first large HPV16 whole-genome sequencing study of 384 oropharynx cancers in a US population which showed that specific variations have different clinical implications [23]. Nonetheless, it remains difficult to understand the impact of HPV variants on the development and treatment of HNSCC and disease outcome. Several clinical trials are being developed aiming to evaluate therapy in HPV related tumours in Head and Neck region [24–26] including therapy de-escalation, because HPV-related tumours often show a more favourable response compared with their HPV-negative counterparts [27,28]. Moreover, in recent decades, active research is exploring mutational, genomic and transcriptomic landscape [29], in tumours HPV(+), although this research those not translated yet in different clinical therapeutical strategies based on HPV16 lineages. Several studies have explored the molecular potential of HPV lineages in the definition of subgroups within HNSCC HPV(+) patients that could tailored treatment accordingly with the molecular characteristics of the respective lineage present in the tumour [30]. However, it should be noted that the inputs of current knowledge regarding genomic alterations or gene expression profiles still have limited clinical utility and in the management of these patients who still undergo on standard therapeutic approaches [31]. A better understanding of the mutational, genomic, and transcriptomic landscape of HNSCC has become the leveraged promise to acquire different approaches for the management of this group of diseases, and it is expected that these additional insights will further drive the correct stratification of patients, allowing them to apply more specific treatments to each group of tumours [32].

The purpose of this work is to evaluate the distribution of HPV16 variants in a different HNSCC cohort from a European country and to explore the available knowledge concerning their potential contribution to HNSCC pathology.

Materials and Methods

Population

Cases were selected from a universe of 458 consecutive primary HNSCC cases diagnosed in the oropharynx or oral cavity at our tertiary cancer centre between 2008 and 2019. In total, following the eligibility criteria described in supplementary Flowchart S1, sixty-eight HPV16-positive cases were included. The clinicopathological characteristics were presented in the supplementary Table S2. The universe (n=458) was distributed by HPV16-positive [only HNSCC HPV16 single-infection patients' (n=68)] and the HPV16-negative HNSCC [comprising HNSCC HPV(-) (n=300) and HNSCC with other HPV(+) (n=90), which includes all HPVs other than HPV16 (including untyped HPV data) and co-infections with more than one type of HPV]. Different demographic and clinical variables were evaluated. Tumours were characterized histologically, and staging was established based on the 8th edition by the American Joint Committee on Cancer (AJCC) [33].

HPV detection and genotyping

HPV DNA was extracted from fresh tumour biopsies using a viral DNA extraction kit. Quantitative Real-time PCR was performed using SYBR GREEN dye with specific L1 region primers to amplify a 75pb amplicon. The albumin gene was used as an internal amplification control. The HPV16 genotyping of the positive samples was processed by commercial kit assays (INNO-LiPA®; or Anyplex™ II HPV28) and amplification products were subsequently submitted to whole-genome sequencing using the next-generation sequencing (NGS) methodology.

Next Generation Sequencing

The amplicons were purified using AMPure XP beads. The DNA library preparation was used in double-index and was sequenced on a MiSeq platform using a 250 bp end-paired reads strategy. The analysis was conducted with INSaFLU bioinformatics pipeline [34], using as reference a representative of HPV16 lineages and sub-lineages (Supplementary Table S3). All generated consensus sequences were curated as follows: i) undefined bases “N” were placed in genome regions with a depth of coverage below 10; ii) all low-coverage regions were visually inspected using Integrative Genomics Viewer, and iii) Single nucleotide variants (SNVs) with >50% intra-sample frequency were assumed in consensus. By using a conservative approach, only SNVs that exhibited the following criteria were validated and used for the analysis: i) fell within reads ≥100 total depth; ii) displayed a frequency of ≥90%; iii) did not fall within primer regions; and 4) had ≤3bp poly(T/A). MEGA11 software [35] was used to generate HPV16 phylogeny using the Neighbor-Joining method [36] with the Maximum Composite Likelihood model to compute genetic distances among taxa [35], as well as to estimate the ratio of nonsynonymous (dN) to synonymous (dS) substitutions with the p-distance model for each gene. For both analyses, the pairwise-deletion option was chosen to remove all sites containing missing data for alignment gaps.

Systematic review

The mutational profile identified was compared through a systematic review of previously published studies. HPV16-positive populations reported in those articles were retrieved for analysis to accurately compare HPV16 SNVs with our own data. The inclusion criteria were defined as described in Supplementary Flow Chart S4. This review excluded all publications before 2012 since it intends to integrate mostly studies that use NGS methodologies in the determination of variants, the same technology used in this study. The search strategy included the standard biomedicine database with defined keywords, resulting in a total of 591 articles. In all, 18 articles were selected and included for analysis.

Statistical Analysis

We conducted a descriptive analysis for the clinical and demographic characterization of our analysed cohort and between HPV16 lineages, using absolute and relative frequencies for categorical variables. Overall survival was defined as the period (in years) from the date of diagnosis to the date of death from any cause, and patients alive were censored at the date of the last follow-up assessment. Overall survival was calculated for each group (HPV16 lineage A, B, C

and D) using the Kaplan-Meier method. Disease-free survival (DFS) was evaluated in the patients with a complete response after the first treatment and was defined as the time from the end of the first treatment to locoregional disease relapse or death from any cause; patients alive without disease recurrence were censored at the date of the last follow-up assessment. The DFS was calculated using the Kaplan–Meier method We report the 3-year disease-free-survival estimates with the respective 95% CI.

Results

Patient samples

Clinical-pathological characteristics of 68 HNSCC HPV16(+) patients are summarised in Supplementary Table S2. The average age at diagnosis was 63 years old. Most patients were men (71%, n=48), and the majority self-reported active tobacco and/or alcohol consumption (72%, n=49). Fifty-five cases had an oropharyngeal location. All 68 HNSCC cases were HPV16(+) single-infection. Patients were treated with different schemes. Most patients showed complete response (n=50) but cases with the persistence of disease (n=14) were also reported.

HPV16 genome coverage and quality

From the 68 samples evaluated four samples did not meet the criterion for the amount of DNA. In 29 samples, size coverage was less than 90% and all those were excluded from further analysis. For the remaining 35 samples, the mean depth of coverage was 3567-fold, ranging between 638 to 4779-fold. On average, the validated samples had 99% of the genome covered by at least 10-fold.

Phylogenetic analysis of whole genome of HPV16 in HNSCC

The phylogenetic analysis clearly shows that the majority of the HPV16 samples clustered in lineage A, whereas the remaining grouped within sublineages B, C and D (shown in Fig. 1). According to the phylogenetic tree, 26 samples were classified as European strains (Lineage A) falling within sub-lineage A1 (n=19), A2 (n=6) and one sample within A3. No A4 strains were found. Three samples were African AF-1 and AF-2 (Lineage B and C) strains represented by two samples clustered with sub-lineage B1 (AF-1) and one sample with lineage C (AF-2). No B2 strains were found. The remaining six samples were Asian-American sub-lineage D3 (AA-1) strains.

Identification of HPV16 variants in HNSCC

Analysis from sequencing data of the 35 HNSCC tumour samples revealed a total of 1194 validated alterations (Supplementary Table S5), where 84% affected coding regions. Regarding the HPV16 genome, variable sites were found 243 sites (Table 1), where L2 (21%) was the most impacted gene, followed by E1 (17%), E2 (14%) and L1 (13%). With exception of the observed for E2 and, to a lesser extent, E6, most of the variable sites found in viral genes yielded synonymous mutations. For one D3 isolate, a disruptive inframe deletion was found at nucleotide 392 of the L2 gene. Of the 84 variant sites yielding missense substitutions, more than half (44/84) were found to be either exclusive of specific lineages or shared among distinct lineages and sublineages (shown in Fig. 2). Indeed, 26 (31%) of the 84 sites simultaneously segregated two or three lineages (B/C, C/D or B/C/D), while 18

(21%) were found to exclusively separate D, C and B lineages or A2 and A3 sublineages from the remainder. Interestingly, three of the 243 variant sites were found to differentially yield missense or synonymous variants on E2, E5 and L1 genes, depending on the HPV16 lineage. For instance, the variant genomic site 6196 was observed to yield a missense substitution (c.1422G>T | Leu474Phe) in all samples from lineages B, C and D, whereas simultaneously result in a synonymous alteration (c.1422G>A | Leu474Leu) in a sample belonging to sublineage A3.

Molecular evolution analysis

Regarding our data, none of the genes reached the positive pressure cut-off, all are less than one, which implies a negative (or purifying) selection (shown in Fig. 3). Indeed, dN/dS values ranged from 0.05 (SE 0.04) for E7 to 0.76 (SE 0.28) for E2 (shown in Fig. 3). This was not surprising as E7 was the least polymorphic viral gene among the 35 HNSCC tumour samples, whereas E2 had the highest number of nonsynonymous substitutions resulting in the highest dN/dS ratio.

Systematic Review

The 18 studies included cover four continents and have focused essentially on the E6 and E7 regions. Only 2 reported SNV data in the HPV16 genome from head-neck tumour samples, all others report cervical cancer data. The size of the included cohorts ranges between 18 and 3215 HPV16(+) tumours. From comparative analysis, we found that 100 out of the 243 variants were previously reported (Supplement Table S6). For some of these amino acid substitutions, it was also possible to predict their putative biological significance (Supplement Table S7).

Clinical pathological and outcome characterization of HPV16 lineages

The Table 2 shows the clinical and pathological characteristics of the patients with sequencing results concerning HPV16 lineage. There is a clear predominance of lineage A (n=26, 74%) followed by lineage D (n=6, 17%). In our series the lineages C and B were residual with only one (3%) and two (6%) cases, respectively.

The Lineage A is mainly composed by men (n=17; 65%), aged <65yrs old at diagnosis (n=14; 54%) with active tobacco/alcohol consumption (n=17; 65%). Most primary tumours were in the oropharynx subsites (tonsil and base of the tongue) and diagnosed on stage II (n=12; 46%). In this group the patients were treated with a combination approach of chemoradiotherapy (n=12; 46%), which almost apparently shows a complete response (n=20; 77%). Our findings suggest that infection with HPV16 lineage D seems to be associated with more advanced clinical stages at diagnosis compared to other lineages. Regarding to the outcomes, the median follow-up in the 35 patients was 46 months (95% CI 38 to 63 months). A total of 10 events of death were reported (Supplement Table S8). The median overall survival (OS) was not reached and the 4-yr OS in the whole subgroup of 35 patients was 76% (95%CI 62-92%). The 2-yr OS for the lineages A and D was 83%. Lineages B and C had a shorter OS (shown in Fig. 4). In the global HPV16(+) sample that achieved full remission following first therapy (n=48) there were 12 events (relapse or death). The DFS at 3 years of follow-up was 84.3% (95% CI, 74.3-95.7%). In the DFS analysis of the HPV16 lineages subgroup (n=35), we excluded 8 patients that did not achieve full remission after first therapy (the patient with Lineage C was

excluded due to persistence). In the remaining 27 records with HPV16 lineage A (n=20), Lineage B (n=2) and Lineage D (n=5) at 3 years of follow-up there were 4 events (relapse or death). Patients with lineage A and D had overlapping DFS with 100% at 3 years (no events). The Lineage B patients with complete response (n=2) died without relapse at median time 1.37-years.

HPV16 variants identified and clinical presentation of HNSCC

A recent publication identified eight SNVs as markers of poor prognosis [23]. Our study reports four of them individually present in thirteen patients (genome position 4410 n=1; position 4539 n=9; position 5962 n=1 and position 7173 n=2). Among these, we have ten men and three women, with an average age ≥65 and tumours in stages II-III (n=11). Most cases presented a complete response, while two cases showed disease persistence. In total, four death events were registered. The remaining patients are well without evidence of disease.

Discussion

The different oncogenic potential of HPV lineages in the development of cervical cancer has been recognized [37]. Conversely, in head-and-neck pathology, the mutational profile of HPV16 is still little explored and few studies present a whole-genome analysis of HPV-associated cancers [31]. In this context, our study of 35 cases provides a valuable contribution to a better characterization of the molecular profile of HPV-driven SCC in the head-and-neck region of a European country.

The clinical-pathological characteristics of our cohort are in line with published literature. Male gender and oropharynx location are the predominant clinical-pathological aspects of this HPV16 positive squamous cell carcinoma series.

Regarding HPV16 analysis most samples clustered in lineage A (n=26), the most common lineage described in Europe. This observation agrees with previous studies [38–41] showing a preponderance of lineage A strains, as recently reviewed [17]. This is also the case with cervical cancer [11,13,18], revealing an interesting similarity between the HPV16 strains involved in carcinogenesis at these two anatomical sites.

In the present study, we analysed and compared the frequency of different HPV16 variants. The mutation profile describes 243 variable sites, of which 100 have been previously reported in cervical pathology, nine in SCC head-and-neck pathology, and 143 are possible new changes. The largest study on HNSCC HPV16 variants and SNV was very recently published and eight HPV16 SNVs were significantly associated with worse survival in a US population [23]. This was now compared with our Portuguese series, and we found four variants (one missense and 3 synonymous) in common (positions 4410, 4539, 5962 and 7173). However, our data do not allow us to draw conclusions regarding the potential of these variants to influence the tumour's biological behaviour. Altogether, 25 patients were alive at the last follow-up, and the remaining ten had death events.

Several studies have shown that in the E6 coding region of HPV16 in the European lineage the E31G, L83V and D25E transitions were significantly associated with the development of cervical carcinoma [22,38–40]. Strikingly, none of these three substitutions was found in our data, highlighting the need for a more detailed

molecular characterization of HPV-associated cancers in different anatomic locations. In this context, it is also remarkable that the E7 N29S variant, associated with a higher oncogenic risk to cervical cancer [41], was only detected in a single HNSCC case, further highlighting the molecular differences associated with cervical and head-and-neck carcinogenesis. Similarly, previous observations from the HPV16-transgenic mouse model show differences in cancer penetrance and microRNA expression profiles between different anatomical sites [42–46], this is in the same line as what is observed in HPV-driven cancers. Taken together, these clinical and experimental data suggest that the interplay between different HPV16 strains and tissue-specific host factors differentially modulates carcinogenesis in different organs.

The E7 gene was shown to be the most conserved, corroborating recent findings on the apparent restriction on E7 gene mutations in cervical cancer and in the oropharynx, sustaining the hypothesis that carcinogenicity depends on a highly conserved E7 protein [11,23]. Interestingly, the indicator of selective pressure revealed that none of the genes reached the positive pressure cut-off, and this was especially noticeable for the E7 gene, owing to purifying selection against nonsynonymous changes.

We acknowledge that one of the limitations of the present study is the low number of cases. However, we provide high-quality whole-genome data on the HNSCC HPV16 mutational profile in 35 new cases of a Portuguese cohort, adding to the available knowledge of HPV16 variants and mutation site and amino acid changes that allow the understanding of the biological significance of each variant.

In conclusion, the present data provide an in-depth genomic characterization of HPV16 present in HNSCC samples, revealing a marked predominance of lineage A strains. A total of 243 genome variable sites were identified, of which 143 are potentially new variations. Major differences were observed concerning the distribution of SNVs in our HSNCC cohort and in another HPV-associated malignancy, cervical cancer. Overall, these data contribute for a deeper understanding of the influence of these mutations on carcinogenesis, diagnostics, and clinical management.

Statements

Acknowledgement (optional)

In the Acknowledgement section, authors must include individuals and organizations that have made substantive contributions to the research or the manuscript. An exception is where funding was provided, which should be included in Funding Sources. Please refer to the Guidelines issued by the ICMJE to determine non-author contributors that should be included in the Acknowledgement section.

Statement of Ethics

All the data were anonymized, and the study was conducted in accordance with the Helsinki Declaration and approved by the Ethics Committee of the Portuguese Oncologic Institute Francisco Gentil, Lisbon (Ref. UIC/2019/1168). A written informed consent was obtained from all participants of the study.

Conflict of Interest Statement

The authors have no conflicts of interest to declare.

Funding Sources

This study was financially supported by the Virology Laboratory Pathology Department from the Portuguese Oncology Institute of Lisboa IUIC/1168, with contributions by the GenomePT project (POCI-01-0145-FEDER-022184), supported by COMPETE 2020 - Operational Programme for Competitiveness and Internationalisation (POCI), Lisboa Portugal Regional Operational Programme (Lisboa2020), Algarve Portugal Regional Operational Programme (CRESC Algarve2020), under the PORTUGAL 2020 Partnership Agreement, through the European Regional Development Fund (ERDF), and by Fundação para a Ciência e a Tecnologia (FCT).

Author Contributions

Contribution authors AF, DC and RGC for study concepts and study design; the authors AN, DC, DS, JPG, JM, LM, LV, MC, MM, and PM performed and data acquisition; the author's AN, DC, LV and JPG analyse the genomic data; the authors AF, AN, DC, LM, MC, RGC and SE were responsible for analysis and interpretation; the author DC for manuscript preparation; the authors AF, AN, DC, DS, LM, MC, RGC and SE for manuscript editing and review. Critical revision of the article for important intellectual content was performed by AF, AN, DC, JPG, LV, MC, MM, PM and RGC. All the authors reviewed and approved the final version of the manuscript.

Data Availability Statement

The authors confirm that the data supporting the findings of this study are available within the article and its supplementary materials. Further enquiries can be directed to the corresponding author.

References

1 Ferlay J, Soerjomataram I, Rajesh Dikshit, Sultan Eser, Mathers C, Rebelo M, et al. Cancer incidence and mortality - Major patterns in GLOBOCAN 2012, worldwide and Georgia. International Journal of Cancer. 2015;136:E359–86.

2 Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA: A Cancer Journal for Clinicians. 2021;71(3):209–49.

3 El-Mofty SK. HPV-Related Squamous Cell Carcinoma Variants in the Head and Neck. Head and Neck Pathology. 2012 Jul;6(SUPPL. 1):55–62.

4 De Bree R, Leemans CR. Recent advances in surgery for head and neck cancer. Current Opinion in Oncology. 2010;22(3):186–93.

5 McDermott JD, Bowles DW. Epidemiology of Head and Neck Squamous Cell Carcinomas: Impact on Staging and Prevention Strategies. Current Treatment Options in Oncology. 2019;20(5):1–13.

6 Gillison ML. Human papillomavirus-associated head and neck cancer is a distinct epidemiologic, clinical, and molecular entity. Seminars in Oncology. 2004 Dec;31(6):744–54.

7 Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: a cancer journal for clinicians. 2018 Nov;68(6):394–424.

8 Ang KK, Harris J, Wheeler R, Weber R, Rosenthal DI, Nguyen-Tân PF, et al. Human Papillomavirus and Survival of Patients with Oropharyngeal Cancer. New England Journal of Medicine. 2010 Jul;363(1):24–35.

9 IARC. Human Papillomaviruses. [cited 2023 Jan 10].Available from: https://publications.iarc.fr/Book-And-Report-Series/Iarc-Monographs-On-The-Identification-Of-Carcinogenic-Hazards-To-Humans/Human-Papillomaviruses-2007

10 Huertas-Salgado A, Martín-Gámez DC, Moreno P, Murillo R, Bravo MM, Villa L, et al. E6 molecular variants of human papillomavirus (HPV) type 16: An updated and unified criterion for clustering and nomenclature. Virology. 2011 Feb;410(1):201–15.

11 LeConte BA, Szaniszlo P, Fennewald SM, Lou DI, Qiu S, Chen NW, et al. Differences in the viral genome between HPV-positive cervical and oropharyngeal cancer. PLoS ONE. 2018 Aug;13(8):e0203403.

12 Mestre VF, Medeiros-Fonseca B, Estêvão D, Casaca F, Silva S, Félix A, et al. HPV16 is sufficient to induce squamous cell carcinoma specifically in the tongue base in transgenic mice. Journal of Pathology. 2020 May;251(1):4–11.

13 Mirabello L, Yeager M, Cullen M, Boland JF, Chen Z, Wentzensen N, et al. HPV16 Sublineage Associations with Histology-Specific Cancer Risk Using HPV Whole-Genome Sequences in 3200 Women. Journal of the National Cancer Institute. 2016 Sep;108(9):djw100.

14 Ho L, Chan SY, Burk RD, Das BC, Fujinaga K, Icenogle JP, et al. The genetic drift of human papillomavirus type 16 is a means of reconstructing prehistoric viral spread and the movement of ancient human populations. Journal of Virology. 1993 Nov;67(11):6413–23.

15 Gillison ML, Koch WM, Capone RB, Spafford M, Westra WH, Wu L, et al. Evidence for a causal association between human papillomavirus and a subset of head and neck cancers. Journal of the National Cancer Institute. 2000 May;92(9):709–20.

16 Muñoz N, Bosch FX, de Sanjosé S, Herrero R, Castellsagué X, Shah KV, et al. Epidemiologic classification of human papillomavirus types associated with cervical cancer. The New England journal of medicine. 2003 Feb;348(6):518–27.

17 Cochicho D, Gil da Costa R, Felix A. Exploring the roles of HPV16 variants in head and neck squamous cell carcinoma: current challenges and opportunities. Virol J. 2021 Nov;18:217.

18 Burk RD, Harari A, Chen Z. Human papillomavirus genome variants. Virology. 2013 Oct;445(1–2):232–43.

19 Ho L, Chan SY, Chow V, Chong T, Tay SK, Villa LL, et al. Sequence variants of human papillomavirus type 16 in clinical samples permit verification and extension of

epidemiological studies and construction of a phylogenetic tree. Journal of Clinical Microbiology. 1991 Sep;29(9):1765–72.

20 Yamada T, Wheeler CM, Halpern AL, Stewart AC, Hildesheim A, Jenison SA. Human papillomavirus type 16 variant lineages in United States populations characterized by nucleotide sequence analysis of the E6, L2, and L1 coding segments. Journal of virology. 1995 Dec;69(12):7743–53.

21 Yamada T, Manos MM, Peto J, Greer CE, Munoz N, Bosch FX, et al. Human papillomavirus type 16 sequence variation in cervical cancers: a worldwide perspective. Journal of virology. 1997 Mar;71(3):2463–72.

22 Jackson R, Rosa BA, Lameiras S, Cuninghame S, Bernard J, Floriano WB, et al. Functional variants of human papillomavirus type 16 demonstrate host genome integration and transcriptional alterations corresponding to their unique cancer epidemiology. BMC Genomics. 2016 Dec;17(1):851.

23 Lang Kuhs KA, Faden DL, Chen L, Smith DK, Pinheiro M, Wood CB, et al. Genetic variation within the human papillomavirus type 16 genome is associated with oropharyngeal cancer prognosis. Annals of Oncology. 2022 DOI: 10.1016/j.annonc.2022.03.005

24 Bteich YT, Hosri JE, Wehbi JA, Daou LR. Current landscape of clinical trials for HPV-positive head and neck squamous cell carcinoma (HNSCC). Ecancermedicalscience. 2022;16:1447.

25 Julian R, Savani M, Bauman JE. Immunotherapy Approaches in HPV-Associated Head and Neck Cancer. Cancers. 2021 Jan;13(23):5889.

26 Lechner M, Liu J, Masterson L, Fenton TR. HPV-associated oropharyngeal cancer: epidemiology, molecular biology and clinical management. Nat Rev Clin Oncol. 2022 May;19(5):306–27.

27 Chera BS, Amdur RJ, Tepper JE, Tan X, Weiss J, Grilley-Olson JE, et al. Mature results of a prospective study of deintensified chemoradiotherapy for low-risk human papillomavirus-associated oropharyngeal squamous cell carcinoma. Cancer. 2018 Jun;124(11):2347–54.

28 Beaty BT, Moon DH, Shen CJ, Amdur RJ, Weiss J, Grilley-Olson J, et al. PIK3CA Mutation in HPV-Associated OPSCC Patients Receiving Deintensified Chemoradiation. J Natl Cancer Inst. 2020 Aug;112(8):855–8.

29 Mehanna H, Robinson M, Hartley A, Kong A, Foran B, Fulton-Lieuw T, et al. 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 (London, England). 2019 Jan;393(10166):51–60.

30 Qin T, Li S, Henry LE, Liu S, Sartor MA. Molecular tumor subtypes of hpv-positive head and neck cancers: Biological characteristics and implications for clinical outcomes. Cancers. 2021;13(11):1–21.

31 Cancer Genome Atlas Network. Comprehensive genomic characterization of head and neck squamous cell carcinomas. Nature. 2015 Jan;517(7536):576–82.

32 Farah CS. Molecular landscape of head and neck cancer and implications for therapy. Annals of Translational Medicine. 2021 May;9(10):915–915.

33 Amin MB, Greene FL, Edge SB, Compton CC, Gershenwald JE, Brookland RK, et al. The Eighth Edition AJCC Cancer Staging Manual: Continuing to build a bridge from a population-based to a more “personalized” approach to cancer staging. CA Cancer J Clin. 2017 Mar;67(2):93–9.

34 Borges V, Pinheiro M, Pechirra P, Guiomar R, Gomes JP. INSaFLU: an automated open web-based bioinformatics suite “from-reads” for influenza whole-genome-sequencing-based surveillance. Genome Med. 2018 Jun;10(1):46.

35 Tamura K, Nei M, Kumar S. Prospects for inferring very large phylogenies by using the neighbor-joining method. Proceedings of the National Academy of Sciences of the United States of America. 2004;101(30):11030–5.

36 Saitou N, Nei M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Molecular biology and evolution. 1987 Jul;4(4):406–25.

37 Gillison ML, D’Souza G, Westra W, Sugar E, Xiao W, Begum S, et al. Distinct risk factor profiles for human papillomavirus type 16-positive and human papillomavirus type 16-negative head and neck cancers. J Natl Cancer Inst. 2008 Mar;100(6):407–20.

38 Hassani S, Castillo A, Ohori JI, Higashi M, Kurono Y, Akiba S, et al. Molecular Pathogenesis of Human Papillomavirus Type 16 in Tonsillar Squamous Cell Carcinoma. Anticancer research. 2015 Dec;35(12):6633–8.

39 Bernard HU, Calleja-Macias IE, Dunn ST. Genome variation of human papillomavirus types: Phylogenetic and medical implications. International Journal of Cancer. 2006 Mar;118(5):1071–6.

40 Sichero L, Villa LL. Epidemiological and functional implications of molecular variants of human papillomavirus. Braz J Med Biol Res. 2006 Jun;39(6):707–17.

41 Song YS, Kee SH, Kim JW, Park NH, Kang SB, Chang WH, et al. Major sequence variants in E7 gene of human papillomavirus type 16 from cervical cancerous and noncancerous lesions of Korean women. Gynecol Oncol. 1997 Aug;66(2):275–81.

42 Santos C, Vilanova M, Medeiros R, Gil da Costa RM. HPV-transgenic mouse models: Tools for studying the cancer-associated immune response. Virus Research. 2017 May;235:49–57.

43 Costa AC, Santos JMO, Medeiros-Fonseca B, Oliveira PA, Bastos MMSM, Brito HO, et al. Characterizing the Inflammatory Microenvironment in K14-HPV16 Transgenic Mice: Mast Cell Infiltration and MicroRNA Expression. Cancers (Basel). 2022 Apr;14(9):2216.

44 Araújo R, Santos JMO, Fernandes M, Dias F, Sousa H, Ribeiro J, et al. Expression profile of microRNA-146a along HPV-induced multistep carcinogenesis: a study in HPV16 transgenic mice. J Cancer Res Clin Oncol. 2018 Feb;144(2):241–8.

45 Paiva I, Gil da Costa RM, Ribeiro J, Sousa H, Bastos MMSM, Faustino-Rocha A, et al. MicroRNA-21 expression and susceptibility to HPV-induced carcinogenesis - role of microenvironment in K14-HPV16 mice model. Life Sci. 2015 May;128:8–14.

46 Paiva I, Gil da Costa RM, Ribeiro J, Sousa H, Bastos M, Faustino-Rocha A, et al. A role for microRNA-155 expression in microenvironment associated to HPV-induced carcinogenesis in K14-HPV16 transgenic mice. PLoS One. 2015;10(1):e0116868.

Figure Legends

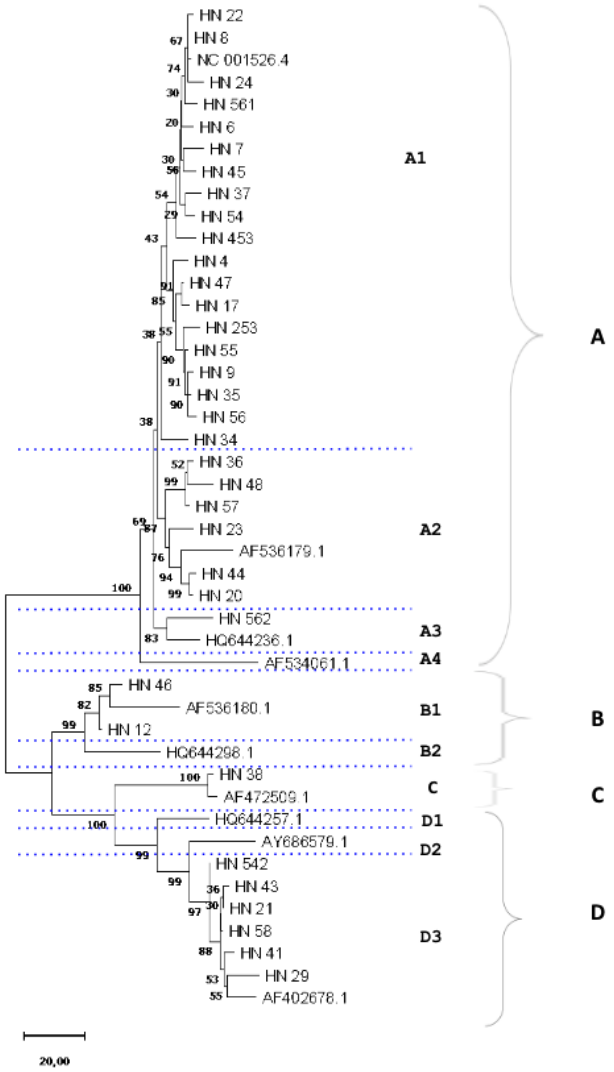


Fig.1 HPV16 phylogeny in HNSCC samples. The phylogeny tree was reconstructed using the Neighbor-Joining method [38] with the Maximum Composite Likelihood model [37]. Numbers next to the branch nodes indicate the bootstrap values (1000 replicates).

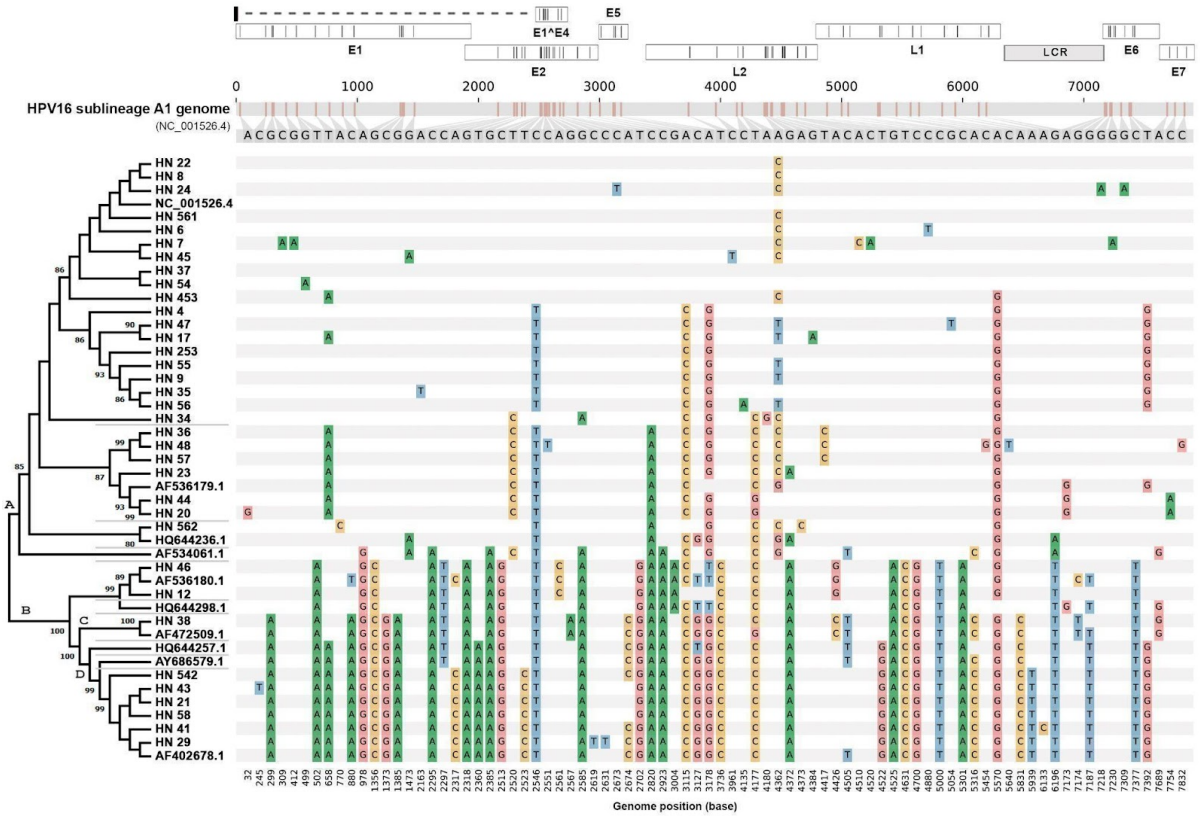


Fig.2 Non-synonymous variant sites in HPV16 genome. For each genome position of the HPV16 A1 sublineage representative (NC_001526.4), all non-synonymous mutations displayed by the 35 HNSCC tumour samples are shown. Viral coding regions are represented on the top of the figure, while on the left, the phylogenetic tree with the patient samples shown in Figure 1. Nucleotide variants are coloured-coded, with nucleotides C, G, T, and A represented by yellow, red, blue and green (coloured-coded), respectively.

Gene	dN/dS	SE
E1	0,1802	0,0606
E2	0,7565	0,2834
E5	0,2952	0,1916
L2	0,2909	0,0871
L1	0,2083	0,0804
E6	0,3359	0,2223
E7	0,0500	0,0389

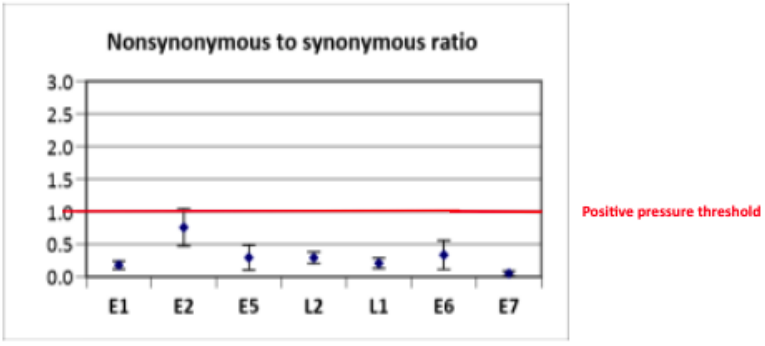


Fig.3 Nonsynonymous-to-synonymous mutation ratio for each HPV16 gene. For each gene, estimates were based on the p-distance model. Minimum and maximum values represent lower and upper limits of the standard error (SE) of the estimate (vertical bar). For better visualization, the positive pressure threshold is shown (horizontal red line).

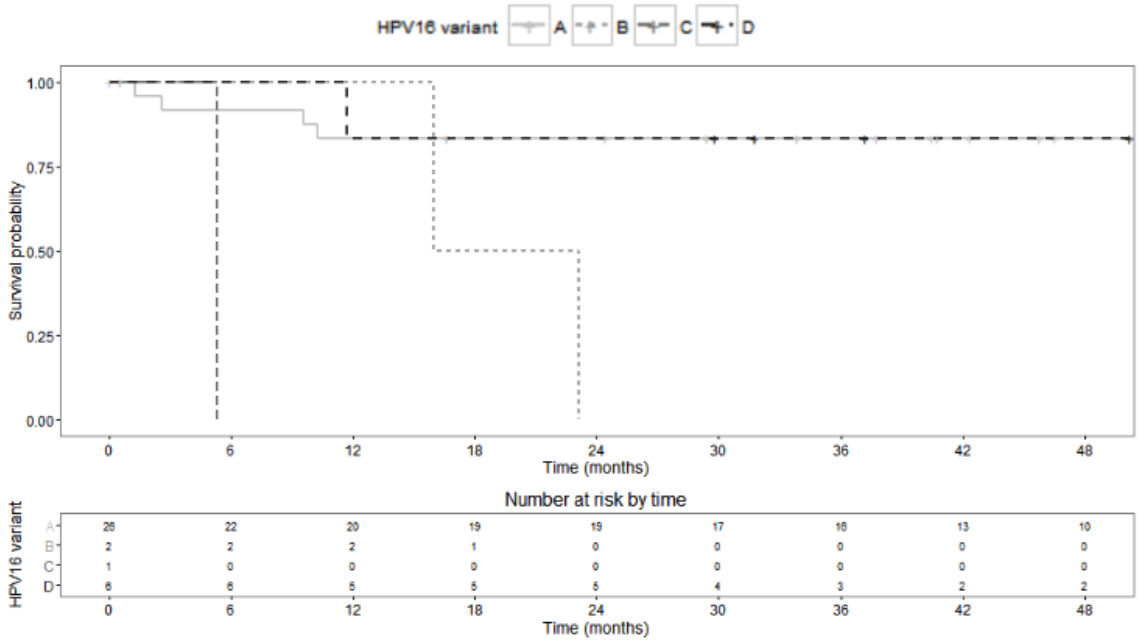


Fig.4 Kaplan–Meier plots for overall survival by each HPV lineage.

2.3 CHARACTERIZATION OF THE HPV16 E6 AND E7 ONCOGENES IN K14HPV16 MICE: SUBLINEAGE A1 DRIVES MULTI-ORGAN CARCINOGENESIS.

Int. J. Mol. Sci. DOI: https://doi.org/10.3390/ijms232012371

Abstract

The study of HPV-induced carcinogenesis relies on multiple *in vivo* mouse models, one of which relies on the cytokeratin 14 gene (*CK14*) promoter to drive the expression of all HPV early oncogenes. Our study evaluated HPV DNA from 17 samples from 4 animals, wild-type (WT, n=2) and HPV16-transgenic mice (MUT, n=2). Total DNA was extracted and detection of HPV16 was performed using a qPCR multiplex. HPV16-positive samples were subsequently whole-genome sequenced by NGS techniques. Comparative genome analysis of K14HPV16 samples revealed a total of 12 mutations to the HPV16 sublineage A1 representative strain. Most of the mutations fall within the coding regions of E1, E2, E4 and E5 early genes and of L1 and L2 late genes. The phylogenetic positioning clearly shows K14HPV16 samples clustering together in the sub-lineage A1 (NC001526.4). The knowledge of the variant is very important, and these findings will allow the rational use of this animal model to explore the role of the A1 sublineage in HPV-driven cancer.

Keywords: K14HPV16, Variant

International Journal of
Molecular Sciences

MDPI

Article

Characterization of the Human Papillomavirus 16 Oncogenes in K14HPV16 Mice: Sublineage A1 Drives Multi-Organ Carcinogenesis

Daniela Cochicho [1,2], Alexandra Nunes [3,4], João Paulo Gomes [3], Luís Martins [2], Mário Cunha [2], Beatriz Medeiros-Fonseca [5], Paula Oliveira [5], Margarida M. S. M. Bastos [6,7], Rui Medeiros [6,8,9], Joana Mendonça [10], Luis Vieira [10], Rui M. Gil da Costa [5,6,8,11,†] and Ana Felix [1,12,*,†]

1 NOVA Medical School-Universidade NOVA de Lisboa, NOVA University, 1169-056 Lisbon, Portugal
2 Virology Laboratory from Clinical Pathology Department, Portuguese Institute of Oncology Francisco Gentil (IPOLFG), 1099-023 Lisbon, Portugal
3 Genomics and Bioinformatics Unit, Department of Infectious Diseases, National Institute of Health Dr. Ricardo Jorge (INSA), 1649-016 Lisbon, Portugal
4 Faculty of Veterinary Medicine, Lusófona University, Campo Grande 376, 1749-024 Lisbon, Portugal
5 Centre for the Research and Technology of Agro-Environmental and Biological Sciences (CITAB), Inov4Agro, University of Trás-os-Montes e Alto Douro (UTAD), Quinta de Prados, 5000-801 Vila Real, Portugal
6 Laboratory for Process Engineering, Environment, Biotechnology and Energy (LEPABE), Faculty of Engineering, University of Porto, Rua Dr. Roberto Frias, 4200-465 Porto, Portugal
7 Associate Laboratory in Chemical Engineering, Faculty of Engineering (ALiCE), University of Porto, Rua Dr. Roberto Frias, 4200-465 Porto, Portugal
8 Molecular Oncology and Viral Pathology Group, Portuguese Oncology Institute of Porto (IPO Porto), Rua Dr. António Bernardino de Almeida, 4200-072 Porto, Portugal
9 Health Research Network, Research Center of Portuguese Oncology Institute of Porto (CI-IPOP/RISE@CI-IPOP), Rua Dr. António Bernardino de Almeida, 4200-072 Porto, Portugal
10 Technology and Innovation Unit, Department of Human Genetics, National Institute of Health, Av. Padre Cruz (INSA), 1649-016 Lisbon, Portugal
11 Post-Graduate Programme in Adult Health (PPGSAD), Morphology Department, University Hospital (HUUFMA), Federal University of Maranhão, São Luís 65085-580, Brazil
12 Pathology Department, Portuguese Institute of Oncology Francisco Gentil, 1099-023 Lisbon, Portugal
* Correspondence: ana.felix@nms.unl.pt
† These authors contributed equally to this work.

Abstract: The study of human papillomavirus (HPV)-induced carcinogenesis uses multiple in vivo mouse models, one of which relies on the cytokeratin 14 gene promoter to drive the expression of all HPV early oncogenes. This study aimed to determine the HPV16 variant and sublineage present in the K14HPV16 mouse model. This information can be considered of great importance to further enhance this K14HPV16 model as an essential research tool and optimize its use for basic and translational studies. Our study evaluated HPV DNA from 17 samples isolated from 4 animals, both wild-type (n = 2) and HPV16-transgenic mice (n = 2). Total DNA was extracted from tissues and the detection of HPV16 was performed using a qPCR multiplex. HPV16-positive samples were subsequently whole-genome sequenced by next-generation sequencing techniques. The phylogenetic positioning clearly shows K14HPV16 samples clustering together in the sub-lineage A1 (NC001526.4). A comparative genome analysis of K14HPV16 samples revealed three mutations to the human papillomaviruses type 16 sublineage A1 representative strain. Knowledge of the HPV 16 variant is fundamental, and these findings will allow the rational use of this animal model to explore the role of the A1 sublineage in HPV-driven cancer.

Keywords: K14HPV16; carcinogenesis; HPV16; variant; lineage

Citation: Cochicho, D.; Nunes, A.; Gomes, J.P.; Martins, L.; Cunha, M.; Medeiros-Fonseca, B.; Oliveira, P.; Bastos, M.M.S.M.; Medeiros, R.; Mendonça, J.; et al. Characterization of the Human Papillomavirus 16 Oncogenes in K14HPV16 Mice: Sublineage A1 Drives Multi-Organ Carcinogenesis. *Int. J. Mol. Sci.* **2022**, *23*, 12371. https://doi.org/10.3390/ijms232012371

Academic Editor: Alvaro Galli

Received: 18 September 2022
Accepted: 11 October 2022
Published: 15 October 2022

41. Zuna, R.E.; Moore, W.E.; Shanesmith, R.P.; Dunn, S.T.; Wang, S.S.; Schiffman, M.; Blakey, G.L.; Teel, T. Association of HPV16 E6 variants with diagnostic severity in cervical cytology samples of 354 women in a US population. *Int. J. Cancer* **2009**, *125*, 2609–2613. [CrossRef] [PubMed]
42. Hoffmann, M.; Lohrey, C.; Hunziker, A.; Kahn, T.; Schwarz, E. Human papillomavirus type 16 E6 and E7 genotypes in head-and-neck carcinomas. *Oral Oncol.* **2004**, *40*, 520–524. [CrossRef]
43. Hirose, Y.; Onuki, M.; Tenjimbayashi, Y.; Mori, S.; Ishii, Y.; Takeuchi, T.; Tasaka, N.; Satoh, T.; Morisada, T.; Iwata, T.; et al. Within-Host Variations of Human Papillomavirus Reveal APOBEC Signature Mutagenesis in the Viral Genome. *J. Virol.* **2018**, *92*, e00017-18. [CrossRef] [PubMed]
44. Jiang, M.; Baseman, J.G.; Koutsky, L.A.; Feng, Q.; Mao, C.; Kiviat, N.B.; Xi, L.F. Sequence Variation of Human Papillomavirus Type 16 and Measurement of Viral Integration by Quantitative PCR. *J. Clin. Microbiol.* **2009**, *47*, 521–526. [CrossRef] [PubMed]
45. Mirabello, L.; Yeager, M.; Yu, K.; Clifford, G.M.; Xiao, Y.; Zhu, B.; Cullen, M.; Boland, J.F.; Wentzensen, N.; Nelson, C.W.; et al. HPV16 E7 Genetic Conservation Is Critical to Carcinogenesis. *HHS Public Access* **2017**, *170*, 1164–1174. [CrossRef]
46. Wichmann, G. Gunnar Wichmann Variation of HPV Subtypes with Focus on HPV-Infection and Cancer in the Head and Neck Region. *Recent Results Cancer Res.* **2017**, *206*, 27–35. [CrossRef]
47. Park, J.S.; Shin, S.; Kim, E.-C.; Kim, J.E.; Kim, Y.B.; Oh, S.; Roh, E.Y.; Yoon, J.H. Association of human papillomavirus type 16 and its genetic variants with cervical lesion in Korea. *APMIS* **2016**, *124*, 950–957. [CrossRef] [PubMed]
48. Yang, Y.; Ren, J.; Zhang, Q. Distribution of Human Papilloma Virus Type 16 E6/E7 Gene Mutation in Cervical Precancer or Cancer- A Case Control Study in Guizhou Province. *China. J. Med. Virol.* **2016**, *88*, 345–350. [CrossRef]
49. LeConte, B.A.; Szaniszlo, P.; Fennewald, S.M.; Lou, D.I.; Qiu, S.; Chen, N.W.; Lee, J.H.; Resto, V.A. Differences in the viral genome between HPV-positive cervical and oropharyngeal cancer. *PLoS ONE* **2018**, *13*, e0203403. [CrossRef]
50. Islam, S.; Mazumder Indra, D.; Basu, M.; Roychowdhury, A.; Das, P.; Dasgupta, H.; Roy, A.; Alam, N.; Mondal, R.K.; Roychoudhury, S.; et al. Phylogenetic analysis of Human papillomavirus 16 variants isolated from Indian Breast cancer patients showed difference in genetic diversity with that of cervical cancer isolates. *Virus Res.* **2018**, *243*, 1–9. [CrossRef]
51. Kahla, S.; Hammami, S.; Kochbati, L.; Chanoufi, M.B.; Oueslati, R. HPV16 E2 variants correlated with radiotherapy treatment and biological significance in cervical cell carcinoma. *Infect. Genet. Evol.* **2018**, *65*, 238–243. [CrossRef]

17. Tamura, K.; Nei, M.; Kumar, S. Prospects for inferring very large phylogenies by using the neighbor-joining method. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 11030–11035. [CrossRef] [PubMed]
18. Saitou, N.; Nei, M. The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **1987**, *4*, 406–425. [CrossRef]
19. Giannoudis, A.; van Duin, M.; Snijders, P.J.F.; Herrington, C.S. Variation in the E2-binding domain of PHV 16 is associated with high-grade squamous intraepithelial lesions of the cervix. *Br. J. Cancer* **2001**, *84*, 1058–1063. [CrossRef]
20. Combes, J.D.; Franceschi, S. Human papillomavirus genome variants and head and neck cancers: A perspective. *Infect. Agent. Cancer* **2018**, *13*, 13. [CrossRef]
21. Burk, R.D.; Harari, A.; Chen, Z. Human papillomavirus genome variants. *Virology* **2013**, *445*, 232–243. [CrossRef] [PubMed]
22. Medeiros-Fonseca, B.; Mestre, V.F.; Estêvão, D.; Sánchez, D.F.; Cañete-Portillo, S.; Fernández-Nestosa, M.J.; Casaca, F.; Silva, S.; Brito, H.; Félix, A.; et al. HPV16 induces penile intraepithelial neoplasia and squamous cell carcinoma in transgenic mice: First mouse model for HPV-related penile cancer. *J. Pathol.* **2020**, *251*, 411–419. [CrossRef] [PubMed]
23. Gil Da Costa, R.M.; Neto, T.; Estêvão, D.; Moutinho, M.; Félix, A.; Medeiros, R.; Lopes, C.; Bastos, M.M.S.M.; Oliveira, P.A. Ptaquiloside from bracken (*Pteridium* spp.) promotes oral carcinogenesis initiated by HPV16 in transgenic mice. *Food Funct.* **2020**, *11*, 3298–3305. [CrossRef] [PubMed]
24. Jabbar, S.; Strati, K.; Shin, M.K.; Pitot, H.C.; Lambert, P.F. Human papillomavirus type 16 E6 and E7 oncoproteins act synergistically to cause head and neck cancer in mice. *Virology* **2010**, *407*, 60–67. [CrossRef]
25. Sepkovic, D.W.; Stein, J.; Carlisle, A.D.; Ksieski, H.B.; Auborn, K.; Bradlow, H.L. Diindolylmethane Inhibits Cervical Dysplasia, Alters Estrogen Metabolism and Enhances Immune Response in The K14-HPV16 Transgenic Mouse Model. *Bone* **2008**, *23*, 1–7. [CrossRef] [PubMed]
26. Galliverti, G.; Wullschleger, S.; Tichet, M.; Murugan, D.; Zangger, N.; Horton, W.; Korman, A.J.; Coussens, L.M.; Swartz, M.A.; Hanahan, D. Myeloid cells orchestrate systemic immunosuppression, impairing the efficacy of immunotherapy against HPVþ cancers. *Cancer Immunol. Res.* **2020**, *8*, 131–145. [CrossRef] [PubMed]
27. Federation of European Laboratory Animal Science Associations. *Euroguide—On the Accommodation and Care of Animals Used for Experimental and Other Scientific Purposes*; Federation of European Laboratory Animal Science Associations: Suffolk, UK, 2010; ISBN 9781853157516.
28. Wang, L.; Wang, F.; Fu, S.; Zhang, C.; Zhe, X.; Li, H.; Li, D.; Shao, R.; Pan, Z. Analysis of genetic variation in human papillomavirus type 16 E1 and E2 in women with cervical infection in Xinjiang, China. *BMC Med. Genom.* **2021**, *14*, 51. [CrossRef] [PubMed]
29. Cullen, M.; Boland, J.F.; Schiffman, M.; Zhang, X.; Wentzensen, N.; Yang, Q.; Chen, Z.; Yu, K.; Mitchell, J.; Roberson, D.; et al. Deep sequencing of HPV16 genomes: A new high-throughput tool for exploring the carcinogenicity and natural history of HPV16 infection. *Papillomavirus Res.* **2015**, *1*, 3–11. [CrossRef] [PubMed]
30. Borges, V.; Pinheiro, M.; Pechirra, P.; Guiomar, R.; Gomes, J.P. INSaFLU: An automated open web-based bioinformatics suite "from-reads" for influenza whole-genome-sequencing-based surveillance. *Genome Med.* **2018**, *10*, 46. [CrossRef] [PubMed]
31. Pérez-Quintanilla, M.; Méndez-Martínez, R.; Vázquez-Vega, S.; Espinosa-Romero, R.; Sotelo-Regil, R.; Pérez-Montiel, M.D.; Ramos-Alamillo, U.; de Jesús Cabrera-López, T.; Barquet-Muñoz, S.A.; Pérez-Plascencia, C.; et al. High prevalence of human papillomavirus and European variants of HPV 16 infecting concomitantly to cervix and oral cavity in HIV positive women. *PLoS ONE* **2020**, *15*, e0227900. [CrossRef]
32. Hirose, Y.; Yamaguchi-Naka, M.; Onuki, M.; Tenjimbayashi, Y.; Tasaka, N.; Satoh, T.; Tanaka, K.; Iwata, T.; Sekizawa, A.; Matsumoto, K.; et al. High Levels of Within-Host Variations of Human Papillomavirus 16 E1/E2 Genes in Invasive Cervical Cancer. *Front. Microbiol.* **2020**, *11*, 596334. [CrossRef] [PubMed]
33. Yamada, T.; Wheeler, C.M.; Halpern, A.L.; Stewart, A.C.; Hildesheim, A.; Jenison, S.A. Human papillomavirus type 16 variant lineages in United States populations characterized by nucleotide sequence analysis of the E6, L2, and L1 coding segments. *J. Virol.* **1995**, *69*, 7743–7753. [CrossRef] [PubMed]
34. Van der Weele, P.; King, A.J.; Meijer, C.J.L.M.L.M.; Steenbergen, R.D.M.M. HPV16 variant analysis in primary and recurrent CIN2/3 lesions demonstrates presence of the same consensus variant. *Papillomavirus Res.* **2019**, *7*, 168–172. [CrossRef]
35. Feng, Y.; Li, Y.; Zhang, Y.; Zheng, H.; Li, G. Association of genetic variants with tumor HPV16 status and survival in squamous cell carcinoma of the oropharynx. *Oral Oncol.* **2016**, *56*, 78–83. [CrossRef] [PubMed]
36. Xi, J.; Chen, J.; Xu, M.; Yang, H.; Luo, J.; Pan, Y.; Wang, X.; Qiu, L.; Yang, J.; Sun, Q. Genetic variability and functional implication of the long control region in HPV-16 variants in Southwest China. *PLoS ONE* **2017**, *12*, e0182388. [CrossRef] [PubMed]
37. Alfaro, A.; Juárez-Torres, E.; Medina-Martínez, I.; Mateos-Guerrero, N.; Bautista-Huerta, M.; Román-Bassaure, E.; Villegas-Sepúlveda, N.; Berumen, J. Different association of human papillomavirus 16 variants with early and late presentation of cervical cancer. *PLoS ONE* **2016**, *11*, e0169315. [CrossRef]
38. Ortiz-Ortiz, J.; Alarcón-Romero, L.; Jiménez-López, M.; Garzón-Barrientos, V.; Calleja-Macías, I.; Barrera-Saldaña, H.; Leyva-Vázquez, M.; Illades-Aguiar, B. Association of human papillomavirus 16 E6 variants with cervical carcinoma and precursor lesions in women from Southern Mexico. *Virol. J.* **2015**, *12*, 29. [CrossRef] [PubMed]
39. Zhang, L.; Liao, H.; Yang, B.; Geffre, C.P.; Zhang, A.; Zhou, A.; Cao, H.; Wang, J.; Zhang, Z.; Zheng, W. Variants of human papillomavirus type 16 predispose toward persistent infection. *Int. J. Clin. Exp. Pathol.* **2015**, *8*, 8453–8459.
40. Freitas, L.B.; Chen, Z.; Muqui, E.F.; Boldrini, N.A.T.; Miranda, A.E.; Spano, L.C.; Burk, R.D. Human Papillomavirus 16 Non-European Variants Are Preferentially Associated with High-Grade Cervical Lesions. *PLoS ONE* **2014**, *9*, e100746. [CrossRef]

gramme (CRESC Algarve2020), under the PORTUGAL 2020 Partnership Agreement, through the European Regional Development Fund (ERDF), and by Fundação para a Ciência e a Tecnologia (FCT). This work was also supported by Fundos FEDER through the Programa Operacional Factores de Competitividade—COMPETE and by Fundos Nacionais through the Fundação para a Ciência e a Tecnologia within the scope of the project UID/BIM/00009/2019 (Centre for Toxicogenomics and Human Health-ToxOmics); and from LA/P/0045/2020 (ALiCE), UIDB/00511/2020 and UIDP/00511/2020 (LEPABE), funded by national funds through FCT/MCTES (PIDDAC); 2SMART (NORTE-01-0145-FEDER-000054), supported by Norte Portugal Regional Operational Programme (NORTE 2020), under the PORTUGAL 2020 Partnership Agreement, through the European Regional Development Fund (ERDF).

Institutional Review Board Statement: The animal model experiments were approved by the University of Trás-os-Montes and Alto Douro Ethics Committee (Approval Number 10/2013) and the Portuguese Veterinary Directorate (Approval Number 0421000/000/2014).

Informed Consent Statement: Not applicable.

Data Availability Statement: The authors confirm that the data supporting the findings of this study are available within the article and its Supplementary materials.

Conflicts of Interest: The authors declare that there is no conflict of interest.

References

1. De Martel, C.; Plummer, M.; Vignat, J.; Franceschi, S. Worldwide burden of cancer attributable to HPV by site, country and HPV type. *Int. J. Cancer* **2017**, *141*, 664–670. [CrossRef] [PubMed]
2. Bernard, H.-U.; Burk, R.D.; Chen, Z.; Koenraad, D.v.; Al, E. Classification of Papillomaviruses (PVs) Based on 189 PV Types and Proposal of Taxonomic Amendments. *Virology* **2012**, *401*, 70–79. [CrossRef] [PubMed]
3. De Villiers, E.M.; Fauquet, C.; Broker, T.R.; Bernard, H.U.; Zur Hausen, H. Classification of papillomaviruses. *Virology* **2004**, *324*, 17–27. [CrossRef] [PubMed]
4. De Villiers, E.M.; Gunst, K. Characterization of seven novel human papillomavirus types isolated from cutaneous tissue, but also present in mucosal lesions. *J. Gen. Virol.* **2009**, *90 Pt 8*, 1999–2004. [CrossRef] [PubMed]
5. Villa, L.L.; Sichero, L.; Rahal, P.; Caballero, O.; Ferenczy, A.; Rohan, T.; Franco, E.L. Molecular variants of human papillomavirus types 16 and 18 preferentially associated with cervical neoplasia. *J. Gen. Virol.* **2000**, *81*, 2959–2968. [CrossRef] [PubMed]
6. Du, J.; Nordfors, C.; Näsman, A.; Sobkowiak, M.; Romanitan, M.; Dalianis, T.; Ramqvist, T. Human papillomavirus (HPV) 16 E6 variants in tonsillar cancer in comparison to those in cervical cancer in Stockholm, Sweden. *PLoS ONE* **2012**, *7*, e36239. [CrossRef]
7. Chan, P.K.S.; Zhang, C.; Park, J.S.; Smith-McCune, K.K.; Palefsky, J.M.; Giovannelli, L.; Coutlée, F.; Hibbitts, S.; Konno, R.; Settheetham-Ishida, W.; et al. Geographical distribution and oncogenic risk association of human papillomavirus type 58 E6 and E7 sequence variations. *Int. J. Cancer* **2013**, *132*, 2528–2536. [CrossRef]
8. Cornet, I.; Gheit, T.; Iannacone, M.R.; Vignat, J.; Sylla, B.S.; del Mistro, A.; Franceschi, S.; Tommasino, M.; Clifford, G.M. HPV16 genetic variation and the development of cervical cancer worldwide. *Br. J. Cancer* **2013**, *108*, 240–244. [CrossRef]
9. Cochicho, D.; Gil da Costa, R.; Felix, A. Exploring the roles of HPV16 variants in head and neck squamous cell carcinoma: Current challenges and opportunities. *Virol. J.* **2021**, *18*, 217. [CrossRef] [PubMed]
10. Arizmendi-Izazaga, A.; Navarro-Tito, N.; Jiménez-Wences, H.; Mendoza-Catalán, M.A.; Martínez-Carrillo, D.N.; Zacapala-Gómez, A.E.; Olea-Flores, M.; Dircio-Maldonado, R.; Torres-Rojas, F.I.; Soto-Flores, D.G.; et al. Metabolic reprogramming in cancer: Role of hpv 16 variants. *Pathogens* **2021**, *10*, 347. [CrossRef] [PubMed]
11. Santos, C.; Vilanova, M.; Medeiros, R.; Gil da Costa, R.M. HPV-transgenic mouse models: Tools for studying the cancer-associated immune response. *Virus Res.* **2017**, *235*, 49–57. [CrossRef] [PubMed]
12. Arbeit, J.M.; Münger, K.; Howley, P.M.; Hanahan, D. Progressive squamous epithelial neoplasia in K14-human papillomavirus type 16 transgenic mice. *J. Virol.* **1994**, *68*, 4358–4368. [CrossRef] [PubMed]
13. Arbeit, J.M.; Howley, P.M.; Hanahan, D. Chronic estrogen-induced cervical and vaginal squamous carcinogenesis in human papillomavirus type 16 transgenic mice. *Proc. Natl. Acad. Sci. USA* **1996**, *93*, 2930–2935. [CrossRef]
14. Moutinho, M.S.S.; Aragão, S.; Carmo, D.; Casaca, F.; Silva, S.; Ribeiro, J.; Sousa, H.; Pires, I.; Queiroga, F.; Colaço, B.; et al. Curcumin and Rutin Down-regulate Cyclooxygenase-2 and Reduce Tumor-associated Inflammation in HPV16-Transgenic Mice. *Anticancer Res.* **2018**, *38*, 1461–1466. [CrossRef] [PubMed]
15. Mestre, V.F.; Medeiros-Fonseca, B.; Estêvão, D.; Casaca, F.; Silva, S.; Félix, A.; Silva, F.; Colaço, B.; Seixas, F.; Bastos, M.M.S.M.; et al. HPV16 is sufficient to induce squamous cell carcinoma specifically in the tongue base in transgenic mice. *J. Pathol.* **2020**, *251*, 4–11. [CrossRef] [PubMed]
16. Santos, C.; Neto, T.; Ferreirinha, P.; Sousa, H.; Ribeiro, J.; Bastos, M.M.S.M.; Faustino-Rocha, A.I.; Oliveira, P.A.; Medeiros, R.; Vilanova, M.; et al. Celecoxib promotes degranulation of CD8+T cells in HPV-induced lesions of K14-HPV16 transgenic mice. *Life Sci.* **2016**, *157*, 67–73. [CrossRef]

CFX96 device (Bio-Rad). No-target controls were used for each reaction. HPV16-positive samples were subsequently submitted to whole-genome sequencing using next-generation sequencing (NGS) techniques.

4.5. HPV16 Whole-Genome Sequencing

The complete 8 kb genomes were amplified by PCR, under 4 pools of primers developed as previously described [29], using a Q5 Hot Start High-Fidelity 2X Master Mix DNA Polymerase (New England Biolabs). PCR products were evaluated for size and specificity by agarose gel analysis. The four independent PCR products of each sample were pooled together in approximately equal molar amounts based on the intensity of bands after electrophoresis. The pooled amplicons were purified from small fragments and primers using the AMPure XP beads (Beckman Coulter) according to the manufacturer's instructions before library preparation. The Nextera XT DNA library prep kit (Illumina) was used to prepare bead-based normalized dual-indexed libraries for sequencing. The library pool was sequenced on a MiSeq benchtop sequencer (Illumina) using a reagent kit V3. A sequencing strategy with 250-bp paired-end reads was used. The INSaFLU (https://insaflu.insa.pt/) (accessed on 15 July 2022), an online platform for amplicon-based next-generation sequencing data analysis [30], was used for the reads' quality control, variant detection/inspection, and sequence consensus generation. The genome sequence of a representative of HPV16 sublineage A1 (GenBank accession number NC001526.4) was used as a reference for mapping and single-nucleotide variant (SNV) annotation. Regions with a depth of coverage below 10-fold were automatically masked in the INSaFLU pipeline by placing undefined bases "N" in the consensus sequence. Low-coverage regions were visually inspected using Integrative Genomics Viewer (IGV). SNVs were assumed in consensus when they displayed more than 50% intra-sample frequency. MEGA11 software (http://www.megasoftware.net) (accessed on 4 September 2022) [17] was applied to calculate the matrices of nucleotide distances and to perform phylogenetic reconstructions. Genome sequences of representative strains from other HPV16 lineages and sublineages (Supplementary material Table S2) were also included in the analyses.

4.6. Variation Sites and their Association with Previous Studies

The SNV data were cross-referenced using the online dbSNP-PUBMED database (https://www.ncbi.nlm.nih.gov/snp/; accessed 28 January 2022) and also compared with previous HPV16 variant published data [31–51] that describe variation sites within whole-genome coding regions (Supplementary Table S3). Included articles were selected from a previously described systematic review [9] and all data from the articles that describe the mutations identified in the head and neck were analyzed; additionally, those identified in the pathology of the cervix were also included.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/ijms232012371/s1.

Author Contributions: Conceptualization, D.C., R.M.G.d.C. and A.F.; Data curation, D.C., A.N., J.P.G., L.M., M.C., B.M.-F., P.O., M.M.S.M.B., R.M., J.M. and L.V.; Formal analysis, D.C., A.N., J.P.G., R.M.G.d.C. and A.F.; Methodology, D.C., A.N., J.P.G., L.M., M.C., B.M.-F., P.O., M.M.S.M.B., R.M., J.M., L.V. and R.M.G.d.C.; Supervision, R.M.G.d.C. and A.F.; Validation, D.C., A.N., J.P.G., L.M., M.C., R.M.G.d.C. and A.F.; Visualization, L.M., M.C., B.M.-F., P.O., M.M.S.M.B., R.M. and J.M.; Writing—original draft, D.C.; Writing—review and editing, D.C., A.N., L.M., M.C., B.M.-F., L.V., R.M.G.d.C. and A.F. All authors have read and agreed to the published version of the manuscript.

Funding: This work is a result of Fundação para a Ciência e a Tecnologia (FCT); and also funded by Fundação para a Ciência e Tecnologia/Ministério da Ciência, Tecnologia e Ensino Superior (FCT/MCTES, Portugal) through national funds to iNOVA4Health (UIDB/04462/2020 and UIDP/04462/2020); from the GenomePT project (POCI-01-0145-FEDER-022184), supported by COMPETE 2020—Operational Programme for Competitiveness and Internationalisation (POCI), Lisboa Portugal Regional Operational Programme (Lisboa2020), Algarve Portugal Regional Operational Pro-

From the comparative analysis of the genome, a total of three variants in the early E1 region were validated, two of which had a missense effect. However, little can be concluded about the clinical significance of these variants in our series. In fact, as previously reported for the K14HPV16 model, the function of E1 and E2 gene products does not influence in terms of severity or extent of progression of the hyperplastic/dysplastic phenotype [12] described in a spectrum of pathologies such the skin, mucosa, and anal squamous epithelium.

The knowledge of HPV variants is very important as it is recognized that minor changes in the HPV sequence affect its carcinogenic potential [18]. Our study identified that K14HPV16 FVB/n mice harbor the HPV16 variant sublineage A1, the most common variant associated with cervical and oropharyngeal neoplasia in cancer patients. This result will allow the rational use of this animal model to explore the role of the A1 sublineage in HPV-driven cancer. Experimental results from the K14HPV16 mouse model suggest that the HPV16 A1 sublineage is associated with a high incidence of tongue base cancer in this model, while the development of cervical and penile cancers requires the action of chemical co-factors, leading us to speculate that different organ sites/microenvironments can modulate the oncogenic potential of this sublineage. However, additional studies are required to test this hypothesis.

4. Material and Methods

4.1. Animals

HPV16-transgenic mice in the FVB/n strain background—FVB.Cg-Tg(KRT14-HPV16) wt1Dh strain known as K14HPV16 mice—were generously donated by Drs. Jeffrey Arbeit and Douglas Hanahan of the University of California through the Mouse Repository of the US National Cancer Institute. Generation of K14HPV16 mice has been previously reported [12]. These animals express all early HPV16 genes under the control of the cytokeratin 14 (KRT14) gene promoter and develop cervical and cutaneous lesions resembling those of cancer patients [12,13]. The experiments were approved by the University of Trás-os-Montes and Alto Douro Ethics Committee (Approval Number 10/2013) and the Portuguese Veterinary Directorate (Approval Number 0421000/000/2014). Mice were maintained and bred according to the Portuguese (Decree-Law 1005/92 dated October the 23rd) and European (EU Directive 2010/63/EU) legislation, with 12 h light/12 h dark cycles, at 20–24 °C, and 50 ± 10% relative humidity. At 30 weeks old, 2 HPV16-transgenic and 2 matched wild-type female mice were sacrificed by intraperitoneal administration of a mixture of xylazine and ketamine, followed by cardiac puncture exsanguination, according to FELASA guidelines [27].

4.2. Samples

A total of 17 samples from 4 animals were studied, both wild-type (WT, $n = 2$) and HPV16-transgenic mice (MUT, $n = 2$). The distribution of samples by the animal condition is presented in Supplementary material Table S1. Overall, the samples collected from tongue ($n = 4$), skin ($n = 4$), liver ($n = 4$), lymph node ($n = 4$), and oral tumor ($n = 1$) were evaluated/analyzed. The samples were snap-frozen in liquid nitrogen and stored at −80 °C. Matched samples were fixed in formalin and processed for histological evaluation.

4.3. Histological Processing

The samples for histological analysis were fixed in 10% neutral buffered formalin and paraffin-embedded. Five micrometer-thick sections were stained with hematoxylin and eosin for histological classification, as previously described [13,14,21,28].

4.4. HPV16 Genotyping

Total DNA was extracted using the universal acid nucleic extraction kit (Seegene). The detection of HPV16 was performed using a qPCR multiplex (Anyplex™ II HPV28 Detection, Seegene) assay able to detect 28 HPV types. Amplification was performed in a

exhibiting differential intermediate frequencies across the nine K14HPV16 samples (from 12.3% for K14HPV16_15 to 22.1% for K14HPV16_11).

3. Discussion

HPV16 variants show a differential prevalence in diverse geographical locations and are associated with differential oncogenic potential [5–8]. Caucasian women infected with the A1/A2 variant are at a higher risk of CIN3+ compared to women of other genetic backgrounds [19]. This is also the most common variant found in head and neck oral cancers in Caucasians and cervical carcinomas [20], suggesting a heightened oncogenic potential compared with other HPV16 variants.

For the first time, the present study determined the variant lineage present in the K14HPV16 transgenic mouse model, based on the relatedness of reference sequences including ten HPV16 A, B, C, and D variant lineages. The K14HPV16 mouse model was shown to carry the HPV16 A1 sublineage. Despite their isolation from distinct tissues, the K14HPV16 samples were found to be highly genetically related among them, only differing by a mean of 1.0 $\pm$ 0.8 nucleotides and showing a mean distance of 3.1 $\pm$ 1.7 nucleotides from the phylogenetically closest sublineage A1 representative strain. On the other hand, K14HPV16 samples exhibited a mean distance to other sublineages that ranged from 22.0 $\pm$ 4.5 (sublineage A3) to 112.0 $\pm$ 10.3 (sublineage D2). Considering that variant HPV sublineages are empirically defined as exhibiting genome differences in the 0.5–1.0% range [21], respectively, these results point to a clonal origin of K14HPV16 samples, in agreement with the congenic nature of this mouse strain [12].

Having determined the variant lineage and sublineage present in K14HPV16 mice, it is now possible to discuss its association with the various kinds of neoplastic lesions observed in this mouse strain. K14HPV16 mice have been used to mimic the development of cervical cancer in the 1990s [13]. In this study, the HPV16 oncogenes were necessary but insufficient to induce cervical cancer and chronic estrogen supplementation was required for carcinogenesis. Based on the present results, it is interesting to speculate that, although the A1 sublineage is clinically associated with a heightened risk of cervical cancer, it may require hormonal co-factors for efficient carcinogenesis. Recently, our group employed K14HPV16 mice for producing the first mouse model of HPV-related penile cancer [22]. Again, the A1 lineage was necessary but insufficient to induce invasive squamous cell carcinoma and a tobacco-related co-carcinogen was needed. However, the HPV16 A1 lineage was able to induce a range of penile intraepithelial lesions in this model without additional co-factors. Interestingly, K14HPV16 mice develop squamous cell carcinomas specifically located at the tongue base, without the need for any chemical or hormonal co-carcinogens—although tumors were more frequent in female mice [15]. Oropharyngeal cancer incidence could be increased in this model by exposure to ptaquiloside, a bracken fern carcinogen, suggesting a synergistic effect [23]. Based on these observations, we speculate that the oncogenic potential of the HPV16 A1 variant lineage in this mouse model is dependent on the anatomic site: cervical and penile carcinogenesis require the presence of co-factors while oropharyngeal cancer may be solely induced by the viral oncogenes, although the predominant incidence in female mice suggests a role of the higher physiological state of estrogen. Further studies are warranted to examine the interplay between different HPV16 lineages and hormonal co-factors. Remarkably, another study using mice carrying only the HPV16 E6 and E7 transgenes without variant assignment found that oropharyngeal carcinogenesis depended on a chemical co-carcinogen [24].

The K14HPV16 mouse model was also used to test new therapies and cancer preventive strategies by our group and others [14,16,25,26]. In general, the development of lesions in this animal model is strongly associated with modulation of the immune response within the tumor microenvironment, and combined immune therapy approaches are required to prevent lesion development [26]. Now, it will be possible to assess the specific impact of those approaches on the HPV16 A1 sublineage, and it would be desirable to develop animal models that represented other HPV16 strains as well.

samples were not included in the phylogenetic analysis. Nevertheless, they share the same major mutations as the remaining K14HPV16 samples against the HPV16 sublineage A1 representative strain (see details below). The phylogenetic analysis (Figure 3) clearly shows K14HPV16 samples clustering together in the sub-lineage A1 (NC001526.4).

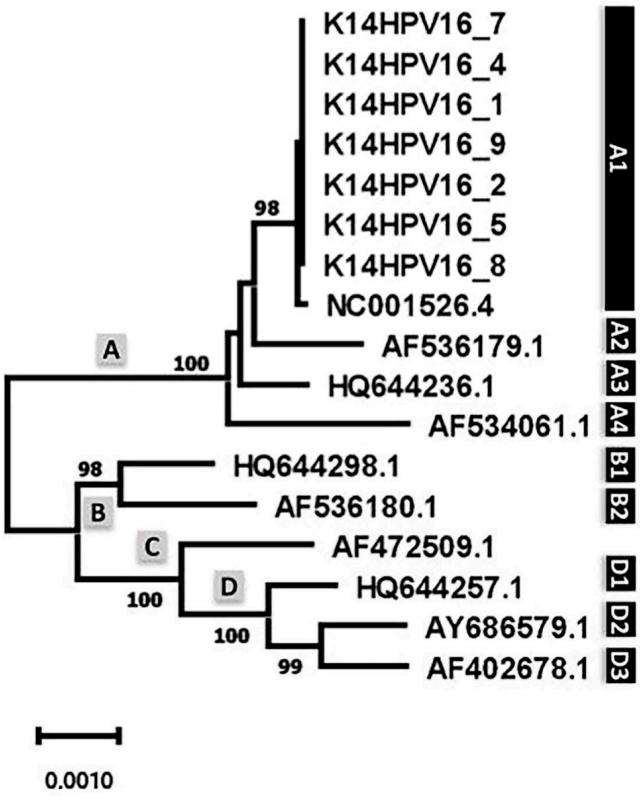


Figure 3. Phylogenetic positioning of K14HPV16 samples within the major HPV16 lineages (A–D). Legend: the phylogenetic tree was generated using the neighbor-joining method [18] with the maximum composite likelihood model [17] and depicts the genetic relationships of the obtained K14HPV16 genome consensus sequences to each representative genome. Numbers next to the branch nodes indicate the bootstrap values (1000 replicates). HPV16 lineages are shown in grey boxes next to the main tree branches, while sublineages (A1–A4, B1, B2, and D1–D3) are illustrated in black boxes.

2.4. HPV16 Sequencing

Comparative genome analysis of K14HPV16 samples revealed a total of three mutations to the HPV16 sublineage A1 representative strain. All mutations fall within the coding region of the E1 early gene, with more than half yielding amino acid changes. Two of these (330A > G | Leu110Leu and 978A > G | Ile326Met) have become evolutionarily fixed (frequency of 100%) based on the observed deep coverage supporting each alteration. Indeed, 100% of the reads mapping each position (for pos 330: ranging from 1580x for sample 6 to 15456x for sample 8; and for pos. 978; ranging from 636x for sample 6 to 11951x for sample 8) support the two variants found in all K14HPV16 samples. The last mutation is a non-synonymous minor intra-patient single nucleotide variant (166A > T | Asn56Tyr)

2. Results

2.1. Sample Characterization

Histological samples from wild-type (WT) animals showed normal histology and HPV16-transgenic (MUT) mice samples showed typical proliferative epithelial lesions of the skin and tongue, while the liver and lymph node samples from WT and HPV16-trangenic mice showed mild inflammatory changes or none at all. The oral tumor found in the HPV16-trangenic mice was identified as a squamous cell carcinoma, with minimal invasion of the tongue (Figure 2).

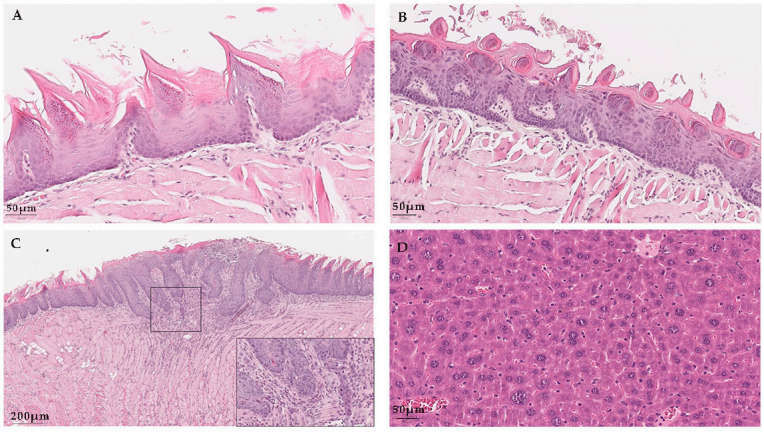


Figure 2. Histological findings of WT and MUT tongue and oral cancer. Legend: Histological samples from wild-type and transgenic K14HPV16 mice stained with H&E. (**A**) low power view of normal squamous cell epithelium of the tongue from a wild-type mouse; (**B**) low power view of mild dysplastic squamous cell epithelium with enlarged hyperchromatic nuclei in the upper layers of the epithelium from a transgenic K14HPV16 mouse and (**C**) low power view of a focally invasive squamous cell carcinoma with thickened rete ridges and dysplastic cells with hyperchromatic nuclei, irregular basement membrane and focally disrupted associated with inflammatory cells in the stroma from a transgenic K14HPV16 mouse; the inset shows the invasive front. (**D**) low power view of the liver of one transgenic K14HPV16 mouse showing normal structure and cytology.

2.2. HPV16 Genome Coverage and Quality

Among the nine HPV16-positive samples, the mean depth of coverage was 2080-fold, ranging between 870-fold for the node sample and 3726-fold for the tumor sample. On average, samples had 75% of the HPV16 genome covered, with 95% of it covered by at least 10-fold. The only exception was a ~1900 bp region comprising most of the L1 gene, the upstream regulatory region (URR) and the beginning of the E6 gene for which no amplification signal was observed for all samples. This is in agreement with the model design [12] which encompasses the entire HPV16 early coding region from bp 97 to 6152, although the gap between the L1 and E6 region (~1945 pb) compromises the hybridization efficiency of specific primers within the primer pool for that region.

2.3. K14HPV16 Lineage Classification

HPV16 variants have been classified into four major lineages (A–D) and sublineages (A1–A4, B1, B2, and D1–D3), based on their genome sequence diversity [17]. To determine the phylogenetic positioning of the HPV16 transgenes in K14HPV16 samples, the respective consensus sequences were aligned against the genomes of lineage/sublineage representative sequences (Supplementary Table S1). For K14HPV16-3 and K14HPV16-6, the generation of a consensus sequence failed (% of genome covered <70%), and those two

1. Introduction

Papillomaviruses are species-specific double-stranded DNA viruses of 8000 base pairs (bp) length (Figure 1) that have preferential tropism for epithelial cells. Infection with human papillomavirus (HPV) is the most common sexually transmissible infection and induces a range of benign (e.g., condylomas) and malignant lesions, such as cervical cancer and other anogenital squamous cell carcinomas and a growing subset of oropharyngeal squamous cell carcinomas [1]. HPV can be classified into genotypes defined by a greater than 10.0% variation in their L1 gene sequence [2]. Currently, over 200 HPV types are recognized and grouped as high-risk (e.g., HPV16 and HPV18, associated with malignant cervical lesions) or low-risk (HPV6 and HPV11, associated with benign lesions) [2–4]. Accumulating data indicate that HPV intra-type variants, lineages (defined by a 1.0–10.0% genomic variation), and sub-lineages (0.5–1.0% variation) may differ in their carcinogenic potential [5–8], as recently reviewed [9,10]. The study of HPV-induced carcinogenesis uses multiple in vivo mouse models [11], one of which relies on the Krt14 (cytokeratin 14) gene promoter to drive the expression of all HPV16 early oncogenes and specifically target basal keratinocytes (known as K14HPV16 mice) [12]. This widely used model was first developed in the 1990s and has since been used to model cervical cancer [13] and other HPV16-associated malignancies [14–16]. This animal model has also proved useful to elucidate the immune-modulatory mechanisms involved in HPV16-induced cancers and to test potential new therapies. However, the specific variant and sub-lineage involved in K14HPV16 mice remain unknown [12]. Determining which variant and sublineage are present in this animal model would help test their potential associations with the development of cancer at specific locations. Additionally, this information would help researchers. This study aimed to determine the HPV16 variant and sublineage present in the K14HPV16 mouse model, further characterizing this important research tool and potentiating its use for basic and translational studies.

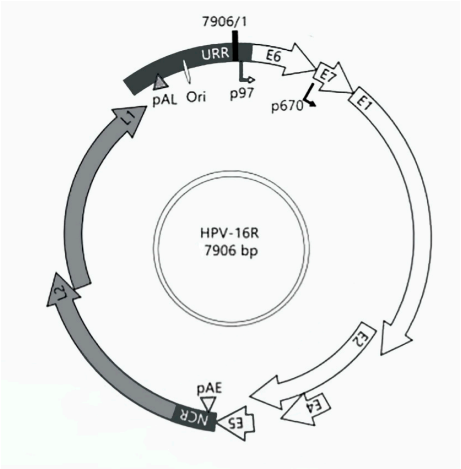


Figure 1. Schematic of the HPV-16R genome. The genome organization of 7906 base pairs (bp) arranged in a circular form represents the three functional regions, early, late, and long control (LCR); and a non-coding region (NCR). These are separated by two polyadenylation sites, described as early (pAE) and late (pAL). In total, the genome encodes eight open reading frames (ORFs), of which E1, E2, E4, E5, E6, and E7 are from the early region; and L1 and L2 are from the late region. The LCR, located between the E6 and L1 gene, is a regulatory region that includes the origin of replication (ori) and the p97 promoter.

2.4 PIK3CA GENE MUTATIONS IN HNSCC: SYSTEMATIC REVIEW AND CORRELATIONS WITH HPV STATUS AND PATIENT SURVIVAL

Cancers. DOI: 10.3390/cancers14051286

Abstract

PIK3CA mutations are believed to contribute to the pathogenesis of human papillomavirus (HPV)-associated head and neck squamous cell carcinomas (HNSCC). This study aims to establish the frequency of *PIK3CA* mutations in a Portuguese HNSCC cohort and to determine their association with HPV status and patient survival. A meta-analysis of scientific literature also done revealed widely different mutation rates in cohorts from different world regions and a trend towards improved prognosis among patients with *PIK3CA* mutations. DNA samples were available from 95 patients diagnosed with HNSCC at the Portuguese Institute of Oncology in Lisbon, between 2010 and 2019. HPV status was established based on viral DNA detected using real-time PCR. Evaluation of *PIK3CA* gene mutations was performed by real-time PCR for 4 mutations (H1047L; E542K, E545K, E545D). 37 cases were found to harbour *PIK3CA* mutations (39%) with the E545D mutation (73%) more frequent detected. There were no significant associations between mutational status and HPV status (74% WT and 68% MUT were HPV (+); p=0.489) or overall survival (OS) (3-yr OS: WT 54% and MUT 65%; p= 0.090). HPV status was the only factor significantly associated with both OS and disease-free survival (DFS), with HNSCC-HPV patients having consistently better outcomes (3-yr OS: HPV (+) 65% and HPV (-) 36%; p=0.007; DFS HPV (+) 83% and HPV (-) 43%; p=0.001). There was a statistically significant interaction effect between HPV status and *PIK3CA* mutation regarding DFS (Interaction test: p=0.026). In HNSCC-HPV patients, PIK3CA wild type is associated with a significant 4.64 times increase in the hazard of recurrence or death (HR=4.64; 95% CI 1.02-20.99; p=0.047). Overall, *PIK3CA* gene mutations are present in many patients and may help define patient subsets who can benefit from therapies targeting the PI3K pathway. The systematic assessment of *PIK3CA* gene mutations in HNSCC patients will require further methodological standardization.

Keywords: HNSCC; HPV; p16 INK4a; PIK3CA

cancers

MDPI

Article

PIK3CA Gene Mutations in HNSCC: Systematic Review and Correlations with HPV Status and Patient Survival

Daniela Cochicho [1,2], Susana Esteves [3], Miguel Rito [4], Fernanda Silva [1], Luís Martins [2], Pedro Montalvão [5], Mário Cunha [2], Miguel Magalhães [5], Rui M. Gil da Costa [6,7,8,9,†] and Ana Felix [1,4,*,†]

1 NOVA Medical School, NOVA University of Lisbon, 1099-085 Lisbon, Portugal; dcochicho@ipolisboa.min-saude.pt (D.C.); fernanda.silva@nms.unl.pt (F.S.)
2 Virology Laboratory from Clinical Pathology Department, IPOLFG, 1099-023 Lisbon, Portugal; lmartins@ipolisboa.min-saude.pt (L.M.); mcunha@ipolisboa.min-saude.pt (M.C.)
3 Clinical Research Unit, IPOLFG, 1099-023 Lisbon, Portugal; sesteves@ipolisboa.min-saude.pt
4 Pathology Department, IPOLFG, 1099-023 Lisbon, Portugal; mrito@ipolisboa.min-saude.pt
5 Otorhinolaryngology Department, IPOLFG, 1099-023 Lisbon, Portugal; pmontalvao@ipolisboa.min-saude.pt (P.M.); mmagalhaes@ipolisboa.min-saude.pt (M.M.)
6 Post-Graduate Programme in Adult Health (PPGSAD), Morphology Department, University Hospital (HUUFMA), Federal University of Maranhão, São Luís 65080-805, Brazil; rmcosta@fe.up.pt
7 LEPABE, Laboratory for Process Engineering, Environment, Biotechnology and Energy, Faculty of Engineering, University of Porto, Rua Roberto Frias, 4200-465 Porto, Portugal
8 Centre for the Research and Technology of Agro-Environmental and Biological Sciences (CITAB), Inov4Agro, University of Trás-os-Montes e Alto Douro (UTAD), Quinta de Prados, 5000-801 Vila Real, Portugal
9 Molecular Oncology and Viral Pathology Group, Research Center of IPO Porto (CI-IPOP)/RISE@CI-IPOP (Health Research Network), Portuguese Oncology Institute of Porto (IPO Porto)/Porto Comprehensive Cancer Center (Porto.CCC), 4200-162 Porto, Portugal
* Correspondence: afelix@ipolisboa.min-saude.pt
† These authors contributed equally to this work.

Simple Summary: Mutations of the PIK3CA gene are thought to contribute to the development of head and neck squamous cell carcinomas (HNSCC), especially those associated with human papillomavirus infection. Furthermore, these mutations may help identify patients who can benefit from specific targeted therapies. This study presents a systematic review of the PIK3CA mutations profile in HNSCC. The results are compared with a cohort of Portuguese patients to study the possible associations with HPV status and patient survival. The Portuguese cohort harboured PIK3CA mutations in 39% of patients, and there were no significant associations with the HPV status or with the OS. In this original case series, there was a statistically significant interaction effect between HPV status and PIK3CA mutation regarding disease-free survival. In HPV-positive patients, the PIK3CA wild-type is associated with a significant 4.64 times increase in the hazard of recurrence or death. Additional studies are needed to clarify the implications of PIK3CA mutations for patient prognosis.

Abstract: PIK3CA mutations are believed to contribute to the pathogenesis of human papillomavirus (HPV)-associated head and neck squamous cell carcinomas (HNSCC). This study aims to establish the frequency of PIK3CA mutations in a Portuguese HNSCC cohort and to determine their association with the HPV status and patient survival. A meta-analysis of scientific literature also revealed widely different mutation rates in cohorts from different world regions and a trend towards improved prognosis among patients with PIK3CA mutations. DNA samples were available from 95 patients diagnosed with HNSCC at the Portuguese Institute of Oncology in Lisbon between 2010 and 2019. HPV status was established based on viral DNA detected using real-time PCR. The evaluation of PIK3CA gene mutations was performed by real-time PCR for four mutations (H1047L; E542K, E545K, and E545D). Thirty-seven cases were found to harbour PIK3CA mutations (39%), with the E545D mutation (73%) more frequently detected. There were no significant associations between the mutational status and HPV status (74% WT and 68% MUT were HPV (+); $p = 0.489$) or overall survival

check for updates

Citation: Cochicho, D.; Esteves, S.; Rito, M.; Silva, F.; Martins, L.; Montalvão, P.; Cunha, M.; Magalhães, M.; Gil da Costa, R.M.; Felix, A. PIK3CA Gene Mutations in HNSCC: Systematic Review and Correlations with HPV Status and Patient Survival. *Cancers* **2022**, *14*, 1286. https://doi.org/10.3390/cancers14051286

Academic Editor: David Wong

Received: 5 February 2022
Accepted: 25 February 2022
Published: 2 March 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

44. Shigaki, H.; Baba, Y.; Watanabe, M.; Murata, A.; Ishimoto, T.; Iwatsuki, M.; Iwagami, S.; Nosho, K.; Baba, H. PIK3CA mutation is associated with a favorable prognosis among patients with curatively resected esophageal squamous cell carcinoma. *Clin. Cancer Res.* **2013**, *19*, 2451–2459. [CrossRef] [PubMed]
45. Zięba, S.; Chechlińska, M.; Kowalik, A.; Kowalewska, M. Genes, pathways and vulvar carcinoma-New insights from next-generation sequencing studies. *Gynecol. Oncol.* **2020**, *158*, 498–506. [CrossRef] [PubMed]
46. Qiu, W.; Tong, G.X.; Manolidis, S.; Close, L.G.; Assaad, A.M.; Su, G.H. Novel Mutant-Enriched Sequencing Identified High Frequency of PIK3CA Mutations in Pharyngeal Cancer. *Int. J. Cancer* **2008**, *122*, 1189–1194. [CrossRef] [PubMed]
47. R Core Team. *A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2021. Available online: http://www.R-project.org/ (accessed on 19 November 2021).

21. Crook, T.; Morgenstern, J.P.; Crawford, L.; Banks, L. Continued expression of HPV-16 E7 protein is required for maintenance of the transformed phenotype of cells co-transformed by HPV-16 plus EJ-ras. *EMBO J.* **1989**, *8*, 513–519. [CrossRef] [PubMed]
22. Veeraraghavalu, K.; Subbaiah, V.K.; Srivastava, S.; Chakrabarti, O.; Syal, R.; Krishna, S. Complementation of human papillomavirus type 16 E6 and E7 by Jagged1-specific Notch1-phosphatidylinositol 3-kinase signalling involves pleiotropic oncogenic functions independent of CBF1;Su(H);Lag-1 activation. *J. Virol.* **2005**, *79*, 7889–7898. [CrossRef]
23. Spangle, J.M.; Munger, K. The HPV16 E6 Oncoprotein Causes Prolonged Receptor Protein Tyrosine Kinase Signaling and Enhances Internalization of Phosphorylated Receptor Species. *PLoS Pathog.* **2013**, *9*, e1003237. [CrossRef] [PubMed]
24. Morgan, E.L.; Macdonald, A. Autocrine STAT3 activation in HPV positive cervical cancer through a virus-driven Rac1-NFκB-IL-6 signalling axis. *PLoS Pathog.* **2019**, *15*, e1007835. [CrossRef]
25. Menges, C.W.; Baglia, L.A.; Lapoint, R.; McCance, D.J. Human papillomavirus type 16 E7 up-regulates AKT activity through the retinoblastoma protein. *Cancer Res.* **2006**, *66*, 5555–5559. [CrossRef]
26. Lui, V.W.Y.; Hedberg, M.L.; Li, H.; Vangara, B.S.; Pendleton, K.; Zeng, Y.; Lu, Y.; Zhang, Q.; Du, Y.; Gilbert, B.R.; et al. Frequent mutation of the PI3K pathway in head and neck cancer defines predictive biomarkers. *Cancer Discov.* **2013**, *3*, 761–769. [CrossRef]
27. Henderson, S.; Chakravarthy, A.; Su, X.; Boshoff, C.; Fenton, T.R. APOBEC-mediated cytosine deamination links PIK3CA helical domain mutations to human papillomavirus-driven tumor development. *Cell Rep.* **2014**, *7*, 1833–1834. [CrossRef]
28. Miao, D.; Margolis, C.A.; Vokes, N.I.; Liu, D.; Taylor-Weiner, A.; Wankowicz, S.M.; Adeegbe, D.; Keliher, D.; Schilling, B.; Tracy, A.; et al. Genomic correlates of response to immune checkpoint blockade in microsatellite-stable solid tumors. *Nat. Genet.* **2018**, *50*, 1271–1281. [CrossRef]
29. Stransky, N.; Egloff, A.M.; Tward, A.D.; Kostic, A.D.; Cibulskis, K.; Sivachenko, A.; Kryukov, G.V.; Lawrence, M.S.; Sougnez, C.; McKenna, A.; et al. The mutational landscape of head and neck squamous cell carcinoma. *Science* **2011**, *333*, 1157–1160. [CrossRef]
30. Sewell, A.; Brown, B.; Biktasova, A.; Mills, G.B.; Lu, Y.; Tyson, D.R.; Issaeva, N.; Yarbrough, W.G. Reverse-phase protein array profiling of oropharyngeal cancer and significance of PIK3CA mutations in HPV-associated head and neck cancer. *Clin. Cancer Res.* **2014**, *20*, 2300–2311. [CrossRef] [PubMed]
31. Beaty, B.T.; Moon, D.H.; Shen, C.J.; Amdur, R.J.; Weiss, J.; Grilley-Olson, J.; Patel, S.; Zanation, A.; Hackman, T.G.; Thorp, B.; et al. PIK3CA Mutation in HPV-Associated OPSCC Patients Receiving Deintensified Chemoradiation. *J. Natl. Cancer Inst.* **2020**, *112*, 855–858. [CrossRef] [PubMed]
32. Mazumdar, T.; Byers, L.A.; Ng, P.K.; Mills, G.B.; Peng, S.; Diao, L.; Fan, Y.H.; Stemke-Hale, K.; Heymach, J.V.; Myers, J.N.; et al. A comprehensive evaluation of biomarkers predictive of response to PI3K inhibitors and of resistance mechanisms in head and neck squamous cell carcinoma. *Mol. Cancer Ther.* **2014**, *13*, 2738–2750. [CrossRef] [PubMed]
33. Holzhauser, S.; Wild, N.; Zupancic, M.; Ursu, R.G.; Bersani, C.; Näsman, A.; Kostopoulou, O.N.; Dalianis, T. Targeted Therapy with PI3K and FGFR Inhibitors on Human Papillomavirus Positive and Negative Tonsillar and Base of Tongue Cancer Lines with and Without Corresponding Mutations. *Front. Oncol.* **2021**, *11*, 640490. [CrossRef] [PubMed]
34. Young, N.R.; Liu, J.; Pierce, C.; Wei, T.F.; Grushko, T.; Olopade, O.I.; Liu, W.; Shen, C.; Seiwert, T.Y.; Cohen, E.E.W. Molecular phenotype predicts sensitivity of squamous cell carcinoma of the head and neck to epidermal growth factor receptor inhibition. *Mol. Oncol.* **2013**, *7*, 359–368. [CrossRef] [PubMed]
35. Michmerhuizen, N.L.; Leonard, E.; Kulkarni, A.; Brenner, J.C. Differential compensation mechanisms define resistance to PI3K inhibitors in PIK3CA amplified HNSCC. *Otorhinolaryngol. Head Neck Surg.* **2016**, *1*, 44–50. [CrossRef] [PubMed]
36. Nichols, A.C.; Palma, D.A.; Chow, W.; Tan, S.; Rajakumar, C.; Giananthony Rizzo, G.; Kevin Fung, K.; Keith Kwan, K.; Brett Wehrli, E.W.; Koropatnick, J.; et al. High frequency of activating PIK3CA mutations in human papillomavirus-positive oropharyngeal cancer. *JAMA Otolaryngol. Head Neck Surg.* **2013**, *139*, 617–622. [CrossRef] [PubMed]
37. Samuels, Y.; Wang, Z.; Bardelli, A.; Silliman, N.; Ptak, J.; Szabo, S.; Yan, H.; Gazdar, A.; Powell, S.M.; Riggins, G.J.; et al. High frequency of mutations of the PIK3CA gene in human cancers. *Science* **2004**, *304*, 554. [CrossRef]
38. Miled, N.; Yan, Y.; Hon, W.C.; Perisic, O.; Zvelebil, M.; Inbar, Y.; Schneidman-Duhovny, D.; Wolfson, H.J.; Backer, J.M.; Williams, R.L. Mechanism of two classes of cancer mutations in the phosphoinositide 3-kinase catalytic subunit. *Science* **2007**, *317*, 239–242. [CrossRef] [PubMed]
39. Mandelker, D.; Gabelli, S.B.; Schmidt-Kittler, O.; Zhu, J.; Cheong, I.; Huang, C.H.; Kinzler, K.W.; Vogelstein, B.; Amzel, L.M. A frequent kinase domain mutation that changes the interaction between PI3Kα and the membrane. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 16996–17001. [CrossRef] [PubMed]
40. LeConte, B.A.; Szaniszlo, P.; Fennewald, S.M.; Lou, D.I.; Qiu, S.; Chen, N.W.; Lee, J.H.; Resto, V.A. Differences in the viral genome between HPV-positive cervical and oropharyngeal cancer. *PLoS ONE* **2018**, *13*, e0203403. [CrossRef]
41. Samuels, Y.; Diaz, L.A., Jr.; Schmidt-Kittler, O.; Cummins, J.M.; Delong, L.; Cheong, I.; Rago, C.; Huso, D.L.; Lengauer, C.; Kinzler, K.W.; et al. Mutant PIK3CA promotes cell growth and invasion of human cancer cells. *Cancer Cell* **2005**, *7*, 561–573. [CrossRef]
42. Ogino, S.; Nosho, K.; Kirkner, G.J.; Shima, K.; Irahara, N.; Kure, S.; Chan, A.T.; Engelman, J.A.; Kraft, P.; Cantley, L.C.; et al. PIK3CA mutation is associated with poor prognosis among patients with curatively resected colon cancer. *J. Clin. Oncol.* **2009**, *27*, 1477–1484. [CrossRef]
43. Kalinsky, K.; Jacks, L.M.; Heguy, A.; Patil, S.; Drobnjak, M.; Bhanot, U.K.; Hedvat, C.V.; Traina, T.A.; Solit, D.; Gerald, W.; et al. PIK3CA mutation associates with improved outcome in breast cancer. *Clin. Cancer Res.* **2009**, *15*, 5049–5059. [CrossRef] [PubMed]

Institutional Review Board Statement: The study was conducted according to the guidelines of the Declaration of Helsinki and approved by the Institutional Ethics Committee of IPOLFG (protocol code UIC/1168, approved on 29-10-2019) and Ethics Committee of NMS|FCM-UNL (CEFCM) (protocol code 70/2019/CEFCM, approved on 23-01-2020).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The data are not publicly available due to privacy.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Ferlay, J.; Soerjomataram, I.; Dikshit, R.; Eser, S.; Mathers, C.; Rebelo, M.; Parkin, D.M.; Forman, D.; Bray, F. Cancer incidence and mortality-Major patterns in GLOBOCAN 2012, worldwide and Georgia. *Int. J. Cancer* **2015**, *136*, E359–E386. [CrossRef] [PubMed]
2. Bray, F.; Ferlay, J.; Soerjomataram, I.; Siegel, R.L.; Torre, L.A.; Jemal, A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* **2018**, *68*, 394–424. [CrossRef] [PubMed]
3. El-Mofty, S.K. HPV-related squamous cell carcinoma variants in the head and neck. *Head Neck Pathol.* **2012**, *6*, S55–S62. [CrossRef] [PubMed]
4. Koroulakis, A.; Agarwal, M. *Laryngeal Cancer*; StatPearls [Internet]; StatPearls Publishing: Treasure Island, FL, USA, 2021. [PubMed]
5. Patmore, H.S.; Ashman, J.N.E.; Stafford, N.D.; Berrieman, H.K.; MacDonald, A.; Greenman, J.; Cawkwell, L. Genetic analysis of head and neck squamous cell carcinoma using comparative genomic hybridisation identifies specific aberrations associated with laryngeal origin. *Cancer Lett.* **2007**, *258*, 55–62. [CrossRef] [PubMed]
6. Zanoni, D.K.; Patel, S.G.; Shah, J.P. Changes in the 8th Edition of the American Joint Committee on Cancer (AJCC) Staging of Head and Neck Cancer: Rationale and Implications. *Curr. Oncol. Rep.* **2019**, *21*, 52. [CrossRef] [PubMed]
7. Cancer Genome Atlas Network. Comprehensive genomic characterization of head and neck squamous cell carcinomas. *Nature* **2015**, *517*, 576–582. [CrossRef]
8. 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.; et al. Evidence for a causal association between human papillomavirus and a subset of head and neck cancers. *J. Natl. Cancer Inst.* **2000**, *92*, 709–720. [CrossRef]
9. Centres for Disease Control and Prevention. Cancers associated with human papillomavirus, United States—2011–2015. Available online: www.cdc.gov/cancer/hpv/pdf/USCS-DataBrief-No4-August2018-508.pdf (accessed on 10 July 2019).
10. 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.; et al. Human Papillomavirus and Survival of Patients with Oropharyngeal Cancer. *N. Engl. J. Med.* **2010**, *363*, 24–35. [CrossRef]
11. 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.; et al. Radiotherapy plus cetuximab or cisplatin in human papillomavirus-positive oropharyngeal cancer (NRG Oncology RTOG 1016): A randomised, multicentre, non-inferiority trial. *Lancet* **2019**, *393*, 40–50. [CrossRef]
12. Chera, B.S.; Amdur, R.J.; Tepper, J.; Qaqish, B.; Green, R.; Aumer, S.L.; Hayes, N.; Weiss, J.; Grilley-Olson, J.; Zanation, A.; et al. Phase 2 Trial of De-intensified Chemoradiation Therapy for Favorable-Risk Human Papillomavirus-Associated Oropharyngeal Squamous Cell Carcinoma. *Int. J. Radiat. Oncol. Biol. Phys.* **2015**, *93*, 976–985. [CrossRef] [PubMed]
13. Lee, M.J.; Jin, N.; Grandis, J.R.; Johnson, D.E. Alterations and molecular targeting of the GSK-3 regulator, PI3K, in head and neck cancer. *BBA Mol. Cell Res.* **2020**, *1867*, 118679. [CrossRef] [PubMed]
14. Chang, F.; Lee, J.T.; Navolanic, P.M.; Steelman, L.S.; Shelton, J.G.; Blalock, W.L.; Franklin, R.A.; McCubrey, J.A. Involvement of PI3K/Akt pathway in cell cycle progression, apoptosis, and neoplastic transformation: A target for cancer chemotherapy. *Leukemia* **2003**, *17*, 590–603. [CrossRef]
15. Xue, G.; Hemmings, B.A. PKB/akt-dependent regulation of cell motility. *J. Natl. Cancer Inst.* **2013**, *105*, 393–404. [CrossRef] [PubMed]
16. Karar, J.; Maity, A. PI3K/AKT/mTOR pathway in angiogenesis. *Front. Mol. Neurosci.* **2011**, *4*, 1–8. [CrossRef] [PubMed]
17. Van Doorslaer, K.; Burk, R.D. Evolution of human papillomavirus carcinogenicity. *Adv. Virus Res.* **2010**, *77*, 41–62. [CrossRef]
18. Zhang, S.; Wu, X.; Jiang, T.; Lu, Y.; Ma, L.; Liang, M.; Sun, X. The up-regulation of KCC1 gene expression in cervical cancer cells by IGF-II through the ERK1/2MAPK and PI3K/AKT pathways and its significance. *Eur. J. Gynaecol. Oncol.* **2009**, *30*, 29–34. [PubMed]
19. Tang, X.; Zhang, Q.; Nishitani, J.; Brown, J.; Shi, S.; Le, A.D. Overexpression of human papillomavirus type 16 oncoproteins enhances hypoxia-inducible factor 1 alpha protein accumulation and vascular endothelial growth factor expression in human cervical carcinoma cells. *Clin. Cancer Res.* **2007**, *13*, 2568–2576. [CrossRef]
20. Hyland, P.L.; McDade, S.S.; McCloskey, R.; Dickson, G.J.; Arthur, K.; McCance, D.J.; Patel, D. Evidence for alteration of EZH2, BMI1, and KDM6A and epigenetic reprogramming in human papillomavirus type 16 E6/E7-expressing keratinocytes. *J. Virol.* **2011**, *85*, 10999–11006. [CrossRef]

odds ratio across strata). We evaluated the prognostic impact of PIK3CA mutations and p16 overexpression on the overall survival (OS) and disease-free survival (DFS). Overall survival was defined as the period of time (in years) from the date of diagnosis to the date of death from any cause, patients alive were censored at the date of last follow-up assessment. Disease-free survival was evaluated in the patients with a complete response after the first treatment and was defined as the time from the end of the first treatment to disease relapse or death from any cause; patients alive without disease recurrence were censored at the date of the last follow-up assessment. We used Kaplan–Meier curves to visualize the differences in survival between subgroups defined by PIK3CA mutation, p16 overexpression, and HPV status and the log-rank test for group comparisons. Cox proportional hazards regression analysis was used to compute the hazard ratios (HRs) and 95% Confidence Intervals based on Wald statistics, with OS and DFS as the outcome variables and adjusting for age, HPV status, and drinking/smoking habits as potential confounding factors. Age and HPV status were chosen a priori as the most important confounding factors based on clinical criteria, and drinking/smoking habits were also included due to the imbalances observed between the PIK3CA MUT and WT groups in our cohort. The proportional hazards assumption was checked using statistical tests and graphical diagnosis based on Schoenfeld residuals. We also graphically assessed the functional form of the age variable in the models using Martingale residuals. As this analysis showed that the linearity assumption was not acceptable in the OS model, we adjusted the nonlinear effect of age with a smoothing spline using the "pspline" function of the R package "survival". The potential interaction between HPV status and PIK3CA mutations on survival was tested using the log-likelihood ratio test between the fitted Cox regression models with and without the interaction term. The analyses were conducted in the complete case dataset. All statistical tests were two-sided, and we considered a significance level of 5%. As this was an exploratory study, no *p*-value correction for multiple testing was done. We used the software R package version 4.1.0 (http://www.R-project.org, accessed on 1 October 2021) [47].

5. Conclusions

Overall, these results confirmed a high frequency of canonical PIK3CA mutations (substitutions H1047L and E542K, E545K, and E545D) in HNSCC, including in HPV (−) cases, suggesting that screening for PIK3CA mutations should not be restricted to HPV (+) patients. Additional studies are needed to clarify the implications of PIK3CA mutations for HNSCC patient prognosis.

Supplementary Materials: The following supporting information can be downloaded at https://www.mdpi.com/article/10.3390/cancers14051286/s1: Figure S1: Participant flow chart. Figure S2: Kaplan–Meier overall survival curves for the included and excluded groups. Figure S3: Flow chart from the meta-analysis inclusion criteria. Table S1: Systematic Review HNSCC PIK3CA qPCR. Table S2: Overall, the 17 articles selected for HNSCC PIK3CA mutation. Table S3: HNSCC PIK3CA mutational profile. Table S4: Clinical and demographic characterisation of the 295 excluded cases.

Author Contributions: Conceptualisation, A.F. and R.M.G.d.C.; Data curation: D.C., M.M., P.M., M.R., F.S., M.C. and L.M.; Formal analysis, D.C., S.E., M.R. and A.F.; Writing—original draft preparation, D.C., R.M.G.d.C. and A.F.; and Writing—review and editing, R.M.G.d.C. and A.F. All authors have read and agreed to the published version of the manuscript.

Funding: This study was financially supported by the Virology Laboratory from the Pathology Department of the Portuguese Oncology Institute of Lisboa IUIC/1168, with contributions by the Research Center of the Portuguese Oncology Institute of Porto (project no. PI86-CI-IPOP-66-2017), by Base Funding-UIDB/00511/2020 of the Laboratory for Process Engineering, Environment, Biotechnology, and Energy—LEPABE—funded by national funds through the FCT/MCTES (PIDDAC), and Project 2SMART-engineered Smart materials for Smart citizens, with reference NORTE-01-0145-FEDER-000054, supported by Norte Portugal Regional Operational Programme (NORTE 2020) under the PORTUGAL 2020 Partnership Agreement through the European Regional Development Fund (ERDF).

4.3. Clinical–Pathological Characterisation

The tumours were categorised by histological type, grade, and TNM stage. A tissue microarray (TMA) was constructed for morphological evaluation, which was performed by two pathologists for diagnostic confirmation. Tumour samples of formalin-fixed paraffin-embedded tissue from biopsies or surgical specimens were stained with haematoxylin and eosin and p16 immunohistochemistry. Tissue microarrays were made with cores 1.5 mm in diameter retrieved from different areas of the tumour. Immunohistochemistry was performed on 4-µm-thick sections for anti-human p16 (clone E6H4, Cat. Number 805–4713, Roche Tissue Diagnostics, Pleasanton 94588 CA USA, prediluted for 4 min; pre-treatment ULTRA CC1-56 min, Ventana Medical Systems, 95050 Santa Clara, CA 95050 USA) following the manufacturer's protocol, with appropriate positive and negative controls samples. Antigen detection was performed using the OptiView DAB IHC Detection Kit stained on the BenchMark ULTRA IHC/ISH Automatic staining platform (Ventana Medical Systems, USA), with diaminobenzidine as the chromogen to detect antigen expression. Tissue sections were counterstained with Mayer's haematoxylin.

4.4. Identification of HPV Infection

HPV DNA detection was performed from tumour biopsies collected in ThinPrep conservative medium. The DNA was extracted using a spin silica filter column (QIAamp, a DNA Tissue Kit). Real-time PCR (qPCR) was performed by SYBR GREEN dye with specific L1 region primers (SPF10). The internal amplification control (albumin gene) was performed in independent reactions, along with HPV detection. A melting curve was performed to confirm the specificity (Tm 73 °C) of the 75-pb amplicon. The genotyping of the positive samples was processed by 2 different commercial kits: the LiRas assay (INNO-LiPA® HPV Genotyping Extra II | Fujirebio, 19355 Malvern, PA, USA) detected 24 different HPV types and/or qPCR multiplex (Anyplex™ II HPV28 Detection | Seegene, 05548 Ogeum-ro, Songpa-gu, Seoul, Republic of Korea) able to detect 28 HPV types.

4.5. Mutation Analysis

DNA was isolated from biopsy samples of tumour tissue using QIAamp, a DNA Tissue Kit (Qiagen, Hilden, Germany). Detection of PIK3CA gene mutations (substitutions H1047L and E542K, E545K, and E545D) was conducted using commercial assays (AmoyDx® PIK3CA Kit; TaqMan® Mutation Detection Assays) based on a real-time PCR reaction. DNA isolated from PIK3CA gene mutation-positive cell lines (SW48 cell line: substitution E542K; MCF10A cell line: substitutions E545K and H1047R) was used as a positive control, and the internal control status amplified and detected a region of genomic DNA adjacent to the PIK3CA gene. Additionally, template control (NTC) was used to verify the positive contaminations. Reactions with 2.5 ng of DNA sample were run on a real-time PCR system (qPCR) using the universal mutation detection thermal cycling protocol. Amplification data were performed by qualitative analysis for the presence of each specific target according to the criteria described in the manufacturer's instructions estimating the mutation status of each sample.

4.6. Statistical Analysis

We conducted a descriptive analysis for clinical and demographic characterisation of our cohort using absolute and relative frequencies for categorical variables, mean and standard deviation for quantitative variables, or median and interquartile ranges in the case of asymmetric distribution. The prevalence of PIK3CA mutations in the overall sample was calculated as a percentage with the respective exact binomial 95% Confidence Interval. The association between PIK3CA mutation and the demographic and clinical variables was tested using Pearson's chi-square test or Fisher's exact test for categorical variables, as appropriate, and the two-sample *t*-test for age. We also tested the association between PIK3CA mutation and p16 overexpression stratified by HPV status using the Cochrane–Mantel–Haenszel exact test (the Woolf test was used to check the homogeneity of the

suggesting that different treatment strategies need to be considered based on the molecular phenotype of a tumour [34,35]. Stratifying patients by HPV DNA status is an effective approach to understand the prognostic impact of PIK3CA mutations, which may be masked when grouping HPV-positive and HPV-positive lesions together.

The main limitation of this study was the relatively small size of our patient cohort. The present results contribute to further defining the complex HNSCC landscape by determining the frequencies of canonical PIK3CA mutations in a Portuguese cohort subdivided into HPV-positive and HPV-negative patients and by showing the impact of PIK3CA mutations over DFS when patients are stratified according to their HPV DNA status.

4. Materials and Methods

4.1. Systematic Review of the Frequency of PIK3CA Mutations in HNSCC

The systematic review was performed according to the PICO model for estimating the frequency of PIK3CA mutations in HNSCC from previously published data. Population: HNSCC. Intervention: PIK3CA mutation frequency. Comparison: HNSCC HPV (+) vs. HNSCC HPV (−). The outcome it was not defined, since the expected results were not clear, not only because of the few studies that existed in the HN region but also because of the limiting case number on cohorts published. The inclusion criteria were established for the type of study (case series and case–control studies in humans); tumour sample type (FPPE and biopsies); tumour location (HN region oral cavity, oropharynx, larynx, hypopharynx, and nasopharynx); histological diagnosis (SCC); and molecular methodology performed (qPCR). It excluded all methodology based on RNA evaluation and the exclusive detection of other targets than PIK3CA reported. The search strategy contemplated the 3 standard databases on biomedicine: PubMed, Embase, and Cochrane, accessed in October and November 2021. The keywords PIK3CA and HNSCC were applied to a total of 374 articles, including 233 articles from PubMed, 154 articles from Embase, and 9 articles from Cochrane (Supplement Figure S3). Overall, 6 articles were selected (Supplement Table S1). To further analyse more consistently and better clarify the mutational profile within the subsites HN region, our previously meta-analysis excluded 66 sequencing articles that may contribute to better translational knowledge about what type of PIK3CA mutation may occur and the oncogenic characteristics. It is generally expected that the frequency of estimated PIK3CA mutations may be different than those previously described, since the results were obtained from two different molecular approaches, such as qPCR detected at much lower compared to the higher sensitivity of the sequencing analysis [46]; additionally, the qPCR design targetable few specific regions contrasted with the whole-genome analysis allowed in sequence methods, adding a completely different perspective for the mutational landscape. For the evaluation of the HNSCC PIK3CA mutation profile, 66 sequencing articles eligible were validated following the exclusion criteria (1) the data from the RNA evaluation, (2) no data PIK3CA mutation profile, and (3) no more than 5 patient cases. We intend to define the somatic mutational profile HNSCC previously reported based on 17 articles evaluated (Supplemental Table S2).

4.2. Population

Archived paraffin-embedded primary tumour biopsy from 95 patients diagnosed and/or treated between 2010 and 2019 at the Head and Neck Surgery and the Otorhinolaryngology Departments from the Instituto Português de Oncologia (IPOLFG). Inclusion criteria were histological confirmation of primary HNSCC from the oropharynx and oral cavity and availability of materials for histological and molecular analyses. Patients with HNSCC associated with HPV infection other than HPV16 and with missing information concerning the p16 marker were excluded. All demographic, clinical–pathological, and follow-up data were retrieved from medical records.

consisted of a majority of male patients with oropharyngeal, HPV-positive cancers, this result suggests a trend towards a high frequency of PIK3CA mutations among HPV (+) oropharyngeal tumours (Supplemental Data Table S1). Furthermore, these results are in accordance with the results of our systematic review, showing that PIK3CA is more commonly mutated in HPV-associated HNSCC. Nevertheless, we could not demonstrate a significant association between the mutation frequency and HPV status, as reported by some previous studies [7,36], which may be ascribed to different patient profiles in our cohorts.

Additionally, we could not demonstrate an association between PIK3CA and overall survival in the univariate or multivariable analyses, as previously reported by other research teams. Indeed, the multivariable analysis, where all the variables of interest, including $p16^{INK4a}$ status and possible confounding factors such as age, smoking, and drinking consumption habits were taken into consideration, showed that the HPV DNA status was the only variable significantly associated with overall survival. This is in agreement with previous studies showing that HPV (−) patients have a poorer response to treatment and a worse prognosis compared with HPV (+) patients [10–12]. Additionally, this also suggests that it may be useful to first stratify patients by HPV DNA status and then study the prognostic impact of PIK3CA mutations within each group. In what regards DFS, the multivariable analysis showed a statistically significant interaction effect between HPV status and PIK3CA mutation when controlling for p16 overexpression, drinking/smoking habits, and age at diagnosis. For this reason, we presented the adjusted hazard ratios concerning the association of the PIK3CA mutation with DFS in the HPV-positive and HPV-negative groups. Importantly, in HPV-positive patients, wild-type PIK3CA was associated with a significant 4.64 times increase in the hazard of recurrence or death. This suggests that stratifying HNSCC patients on the basis of their HPV and PIK3CA status is a successful approach for predicting their prognosis more accurately and that the PIK3CA mutational status has an important prognostic impact when considered in the context of HPV. Numerous studies have associated PIK3CA mutations with a more aggressive phenotype [41,42]. However, contradictory results were obtained when evaluating different human cancers, and data showing a favourable impact on prognosis has been reported in patients with breast [43] and oesophageal carcinomas [44]. For the tumours of the HN region, the published data are still unclear, as shown in our systematic review. The present results contribute to elucidating this point, showing the importance of conjugating HPV and PIK3CA data to achieve a more closely tailored prognosis.

In the present data cohort, the most frequent mutation was E545D, followed by E545K and by E542K, all in exon 9, and one mutation in exon 20, H1047R. This is in agreement with the findings from our systematic review showing that the three most frequent mutations in previous reports targeted the same hotspots as in our cohort, namely E545K and E542K on exon 9 and H1047R, T1025T, M10431, and G1049 on exon 20. These data confirmed that exon 9 is the major mutations hotspot and that the oropharynx is the anatomic subsite most commonly affected. This is not a surprising finding, considering that HPV-positive tumours show higher mutation frequencies and are associated with this location. Taken together, these observations support the hypothesis that HPV-induced HN carcinogenesis preferentially involves PIK3CA mutations. In general, the data evaluated suggests that lesions with identical histological features may harbour different PIK3CA mutations, revealing a heterogeneous landscape among subsites of the HN region.

We concluded that PIK3CA alterations seem to be more common in HPV-positive HNSCC but were also common in HPV-negative patients, emphasizing the importance of this pathway independent of HPV infection [7,36]. This is similar to other SCC arising in other sites, such as vulvar carcinoma [45]. Therefore, determining which patients carry PIK3CA mutations is of great clinical significance, as it may help define patient subsets with specific needs and who may benefit from de-intensification treatments by promising targeted therapies against the signalling pathway involved. Specifically, this will allow the selection of patients who are more likely to respond to new therapeutic combinations [32],

Table 6. Multivariable analysis of PIK3CA mutation associated with disease-free survival, Cox regression model adjusted for p16 overexpression, alcohol/tobacco consumption habits, and age at diagnosis.

	Adjusted HR *	**95% CI**	***p*-Value**
HPV status positive			
PIK3CA			
Mutated	1		
Wild-type	4.64	1.02–20.99	0.047
HPV not detected			
PIK3CA			
Mutated	1		
Wild-type	0.38	0.06–2.24	0.285

* Hazard ratio adjusted for p16 overexpression, consumption habits, and age at diagnosis.

3. Discussion

Several studies have pointed out the high frequency of PIK3CA mutations in HNSCC, especially in cases associated with HPV [30,36], and suggested that PIK3CA mutations have predictive and prognostic values [31,37]. Among the three canonical activating mutations, E542 and E545 target the helical domain of p110α, inducing PI3K hyperactivity by disrupting the regulatory activity of p85 on p110α [38]. Third, H1047 targets the p110α kinase domain and is thought to cause conformational change, allowing easier access to the phospholipid substrate [39]. Gene mutations targeting the p110α catalytic subunit were found in 56% of HPV-positive HNSCC tumours and in 34% of HPV-negative HNSCC tumours [7]. Moreover, data from a retrospective study with 87 oropharyngeal cancer samples reported that 10 out of 16 patients had activating mutations in the helical domain (E542K—8/16 and E545K—2/16) [36]. However, preliminary results from a clinical trial that stratified HNSCC patients by PIK3CA status showed no difference in the disease control rate over two months between patients with canonical PIK3CA mutations and those with WT-PIK3CA (36.4% vs. 38.9%). Thus, a deeper understanding of the prognostic value of PIK3CA mutations is needed [40]. To address this point, the present study evaluated the frequency of mutations in the PIK3CA gene in a Portuguese series with a diagnosis of SCC in the oropharynx and oral cavity and their association with HPV status and patient survival.

We also performed a systematic review to clarify the frequency of PIK3CA mutations reported in the scientific literature. We first analysed studies that employed qPCR to detect mutations, using a similar methodology to the one we used. Then, we performed a second and broader analysis dealing with studies that employed DNA sequencing methods, which allowed us to appraise the PIK3CA mutational spectrum within the HNSCC subsites in much greater detail. The first approach was not conclusive concerning the possible associations of PIK3CA mutations with HPV status and with patient prognosis, most likely due to the very limited number of studies involved. The second approach showed a higher frequency of PIK3CA mutations among the HPV-positive cases. Overall, 45 different mutations were reported, distributed both in exon 9 and exon 20, and novel mutations were reported, showing the potential of DNA sequencing techniques to refine the PIK3CA mutational landscape. HPV (+) patients were more likely to harbour mutations at E542 and E545, while HNSCC HPV (−) tumours showed more mutations at H1047, but the data remains insufficient to establish a pattern. Importantly, 4/17 studies reported improved OS or DFS for patients with PIK3CA mutations, even though most studies did not report any association, and another reported a lack of correlation between the mutational status and OS (Supplemental Table S2).

In the present study, PIK3CA mutations were found in 39% of our cases. This is the highest value among the previously published data (Borkowska 2021, García-Escudero 2018, Pattle 2017, Theurer 2016, McBride 2014, and Suda 2012). Since our patient cohort

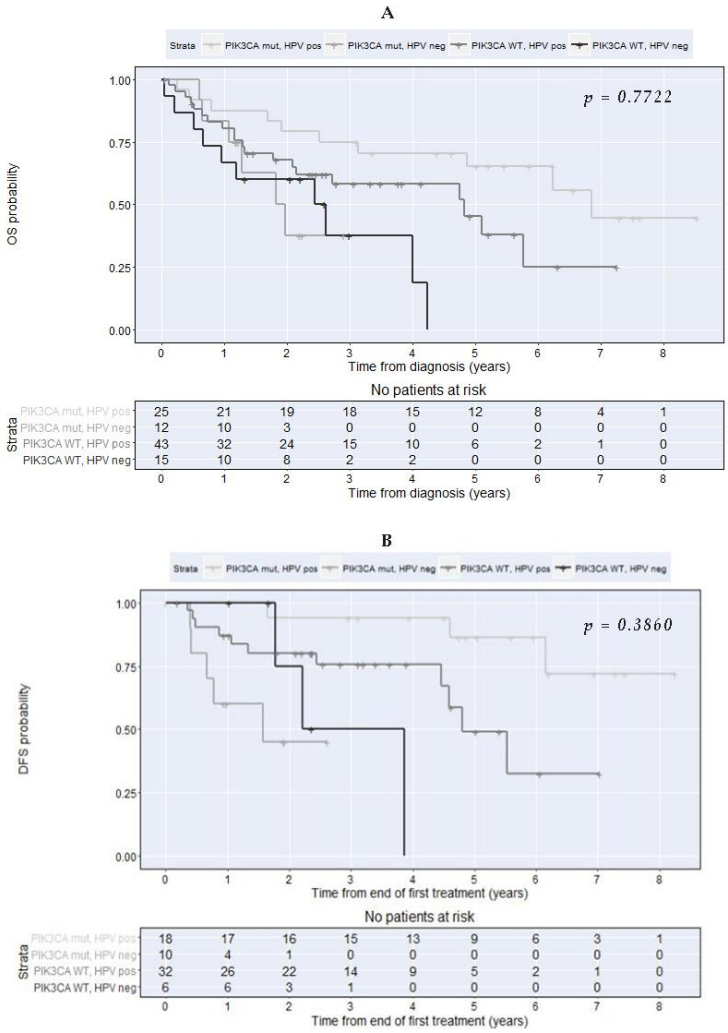


Figure 3. Kaplan–Meier plots for overall survival (**A**) and disease-free survival (**B**) by PIK3CA mutation and HPV status at diagnosis ($p = 0.3860$).

Regarding the association of PIK3CA mutation with DFS, the multivariable analysis showed a statistically significant interaction effect between HPV status and PIK3CA mutation (Interaction test: $p = 0.026$, Figure 3B) when controlling for p16 overexpression, drinking/smoking habits, and age at diagnosis. For this reason, we present the adjusted hazard ratios concerning the association of PIK3CA mutation with DFS in the groups of HPV-positive and not detected (Table 6). In patients HPV-positive, the PIK3CA wild-type was associated with a significant 4.64 times increase in the hazard of recurrence or death (HR = 4.64; 95% CI 1.02–20.99; $p = 0.047$). Conversely, no significant association could be demonstrated between the PIK3CA mutation status and disease-free survival in patients with nondetected HPV (HR = 0.38; 95% CI 0.06–2.24; $p = 0.285$).

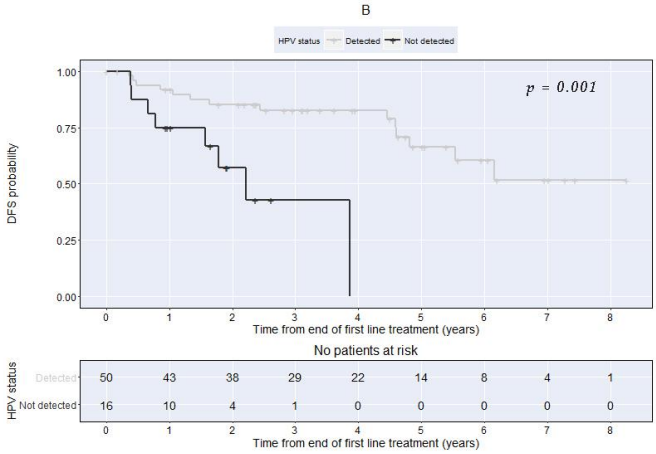


Figure 2. Kaplan–Meier curves comparing patients with detected vs. non-detected HPV at diagnosis. (**A**) Overall survival (Log-rank test chi-square = 7.3 with 1 degree of freedom, $p = 0.007$). (**B**) Disease-free survival (Log-rank test chi-square = 11 with 1 degree of freedom, $p = 0.001$).

During the multivariable analysis adjusted for p16 overexpression, the HPV status, drinking/smoking habits, and age at diagnosis, the PIK3CA wild-type was associated with a two-fold increase in the hazards of death (HR = 2.15; 95% CI 0.98–4.73); nevertheless, this was not statistically significant at a 5% significance level ($p = 0.056$, Table 5). Therefore, we could not demonstrate an independent association of the PIK3CA mutation with the overall survival, controlling for p16 overexpression, HPV status, drinking/smoking habits, and age at diagnosis (Table 5). There was no evidence of a statistically significant interaction between the PIK3CA mutational status and HPV status in what concerns the OS (Interaction test: $p = 0.7722$, Figure 3A).

Table 5. Multivariable analysis of the PIK3CA mutation and p16 status association with overall survival, Cox regression model adjusted for HPV status, alcohol/tobacco consumption habits, and age at diagnosis.

Factor	Coef (SE)	HR (95%CI)	*p*-Value
PIK3CA			
Mutated	1	1	0.056
Wild-type	0.766 (0.401)	2.15 (0.98–4.73)	
p16 overexpression			
No	1	1	0.670
Yes	−0.189 (0.439)	0.83 (0.35–1.96)	
HPV status			
Positive	1	1	0.016
Not detected	1.069 (0.443)	2.91 (1.22–6.94)	
Consumption habits			
Active	1	1	0.040
No consumption	−1.041 (0.508)	0.35 (0.13–0.96)	
Age at diagnosis *			
Pspline, linear	0.045 (0.017)	—	0.008
Pspline, nonlinear	—	—	0.016

Coef = estimated coefficient from the model; SE = standard error of coefficient; HR = Hazard ratio; 95% CI = 95% Confidence Interval. * Smoothing splines was used to adjust for age with a smoothed curve without categorisation or linearity assumption (theta = 0.973, degrees of freedom for age term = 2.66).

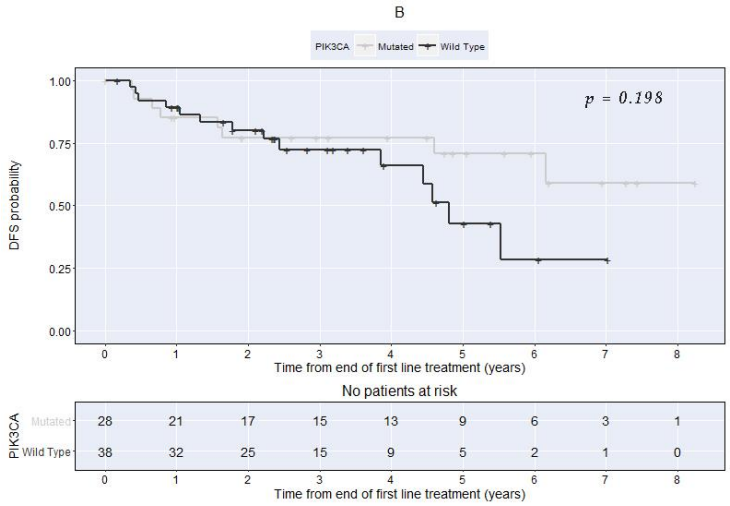


Figure 1. Kaplan–Meier curves comparing patients with PIK3CA-mutated vs. the wild-type at diagnosis (**A**). Overall survival (Log-rank test chi-square = 2.9 with 1 degree of freedom, $p = 0.090$). (**B**). Disease-free survival (Log-rank test chi-square = 1.7 with 1 degree of freedom, $p = 0.198$).

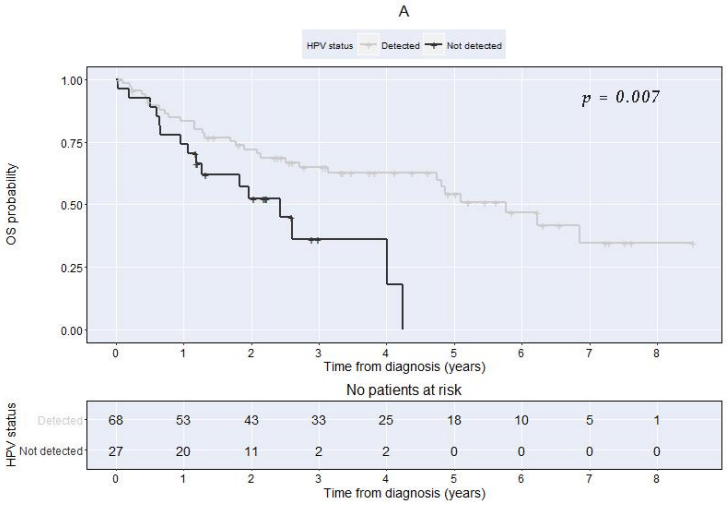


Figure 2. *Cont.*

Analysis PIK3CA, HPV Status, and p16 Immunohistochemistry

OS and DFS were determined separately for patients grouped by the PIK3CA mutational status, HPV DNA, and $p16^{INK4a}$ (Table 4 and Figures 1 and 2). During the univariable analysis, we could not demonstrate a significant association between the PIK3CA mutation status and OS or DFS (Table 4 and Figure 1A,B). The HPV status was the only factor significantly associated with both OS and DFS, with patients with positive HPV having consistently better outcomes compared with the ones with non-detected HPV (Table 4 and Figure 2A,B). Patients with p16 overexpression had significantly longer DFS than patients with no p16 overexpression, but no significant difference could be demonstrated concerning OS (Table 4).

Table 4. PIK3CA mutation, HPV, and p16 status association with the overall and disease-free survival by univariable analysis.

	Overall Survival				Disease-Free Survival			
	Median (Years)	**3-Year % (95% CI)**	**HR (95%CI)**	***p* ***	**Median, (Years)**	**3-Year % (95% CI)**	**HR (95% CI)**	***p* ***
PIK3CA								
MUT	6.2	65 (50–83)	1	0.090	NR	77 (63–95)	1	0.198
WT	4.0	54 (41–70)	1.71 (0.91–3.18)		4.8	72 (58–90)	1.80 (0.73–4.42)	
HPV status								
Positive	5.8	65 (54–78)	1	0.007	NR	83 (72–94)	1	0.001
Not detected	2.4	36 (19–69)	2.40 (1.15–4.61)		2.2	43 (21–90)	4.81 (1.74–13.29)	
p16 overexpression								
No	2.6	46 (33–64)	1	0.089	4.8	59 (44–80)	1	0.029
Yes	6.9	73 (60–90)	0.58 (0.31–1.10)		NR	91 (81–100)	0.34 (0.13–0.94)	

* Log-rank test; HR = Hazard Ratio; 95% CI = 95% Confidence Interval; NR = Not Reached; MUT = mutated; Wt = wild-type.

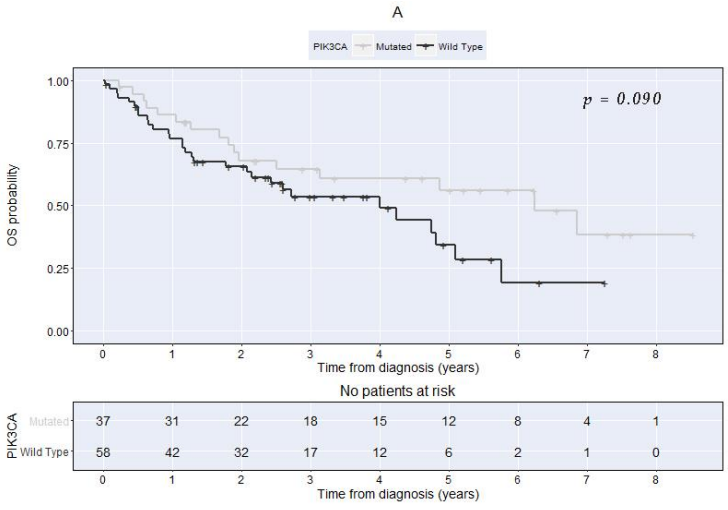


Figure 1. *Cont.*

2.2.2. Prevalence of PIK3CA Mutations

The PIK3CA mutations were present in 39% of HNSCC cases (n = 37, 95% CI: 29–49%). The prevalence of PIK3CA mutations in HNSCC HPV (+) (n = 68) and HNSCC HPV (−) (n = 27) was, respectively, 36.8% (n = 25) and 44.4% (n = 12). Next, we analysed the distribution of PIK3CA mutations in patient subgroups defined by both the HPV DNA and $p16^{INK4a}$ status, as summarised in Table 2. We could not demonstrate a statistically significant association between the presence of PIK3CA mutations and HPV DNA detection (p = 0.489, Table 1). When evaluating a possible association between PIK3CA mutation and the $p16^{INK4a}$ status without stratification by the HPV DNA status, no statistically significant association was found (p = 0.163, Table 1). Similarly, there was no evidence of an association between the $p16^{INK4a}$ status and PIK3CA mutation when considering HPV status stratification (p = 0.245).

Table 2. Frequency of PIK3CA mutations in HNSCC with and without HPV DNA and $p16^{INK4a}$ overexpression (n = 89; p16 missing data n = 6).

HNSCC		PIK3CA Gene	
		WT ***n* (%)**	**MUT** ***n* (%)**
HPV (+)	p16(+)	22 (68.8%)	10 (31.3%)
	p16(−)	17 (56.7%)	13 (43.3%)
HPV (−)	p16(+)	3 (75%)	1 (25%)
	p16(−)	12 (52.2%)	11 (47.8%)

2.2.3. Classification of PIK3CA Gene Mutations: H1047L, E542K, E545K, and E545D

HNSCC cases were screened for four different single substitutions, known as canonical PIK3CA mutations. Among the cases harbouring PIK3CA mutations (n = 37), the majority carried E545D (73%; n = 27), followed by E545K (5%; n = 2), E542K (3%; n = 1), and H1047R (3%; n = 1). We observed the occurrence of combined substitutions E545D | E542K (5.4%; n = 2), E545D | E545K (8.1%; n = 3), and H1047R | E545D (2.7%; n = 1). The distribution of PIK3CA mutations in the HPV-positive and -negative subgroups is summarised in Table 3.

Table 3. Classification of PIK3CA gene mutations (substitutions H1047L and E542K, E545K, and E545D).

HPV Status	PIK3CA Gene MUT						
	One Substitution				Two Substitutions		
	E545D % (*n*)	**E545K % (*n*)**	**E542K % (*n*)**	**H1047R % (*n*)**	**E545D | E545K % (*n*)**	**E545D | E542K % (*n*)**	**E545D | H1047R % (*n*)**
HPV (+)	76% (19)	8% (2)	4% (1)	4% (1)	8% (2)	ND	ND
HPV (−)	67% (8)	ND	ND	ND	8% (1)	17% (2)	8% (1)
Total	73% (27)	5% (2)	3% (1)	3% (1)	8% (3)	5% (2)	3% (1)

Legend: ND = not detected.

2.2.4. PIK3CA Mutations and Patient Prognosis

The median follow-up determined using the reverse Kaplan–Meier method was 4.12 years (95% CI 3.1–5.5 years). A total of 46 deaths were reported during the follow-up period. The median overall survival (OS) in the whole sample (n = 95) was 4.75 years (95% CI 2.6–6.9 years). The 3-year OS was 58% (95% CI 48–70%). Disease-free survival (DFS) was evaluated in the 66 patients with a complete response to the first-line treatment (17 patients with persistent disease and 12 with missing information concerning response to treatment were excluded from the DFS analysis). Overall, the median DFS was 6.16 years (95% CI 4.6–NA), and the 3-year DFS was 75% (64–87%).

Table 1. Clinical, pathological, and demographic characteristics and association with PIK3CA mutation.

Variable Categories	PIK3CA Evaluation WT (*n* = 58) *n* (%)	PIK3CA Evaluation MUT (*n* = 37) *n* (%)	Total Cases *n* = 95 *n* (%)	*p*-Value
Gender				
Male	45 (77.6%)	25 (67.6%)	70 (73.7%)	0.2795
Female	13 (22.4%)	12 (32.4%)	25 (26.3%)	
Age at diagnosis				
Mean (Standard Deviation)	62 (10.7)	62 (12.7)	62 (11.5)	0.9740
≥65 Years	24 (41.4%)	12 (32.4%)	36 (37.9%)	0.3807
<65 Years	34 (58.6%)	25 (67.6%)	59 (62.1%)	
Consumption habits (tobacco and/or alcohol)				
Active consumption	44 (75.9%)	22 (59.5%)	66 (69.5%)	0.0325 *
Alcohol active	18 (31.0%)	10 (27%)	28 (29.5%)	
Tobacco active	7 (12.1%)	1 (2.7%)	8 (8.4%)	
Both active	19 (32.8%)	11 (29.7%)	30 (39.6%)	
No consumption	9 (15.5%)	13 (35.1%)	22 (23.2%)	
Never	5 (8.6%)	10 (27%)	15 (15.8%)	
Nonactive	4 (6.9%)	3 (8.1%)	7 (7.4%)	
No data	5 (8.6%)	2 (5.4%)	7 (7.4%)	
Tumour Anatomic region				
Oropharynx	46 (79.3%)	29 (78.4%)	75 (78.9%)	0.9135
Oral cavity	12 (20.7%)	8 (21.6%)	20 (21.1%)	
Tumour stage				
I	4 (6.9%)	4 (10.8%)	8 (8.4%)	0.9342
II	11 (19%)	8 (21.6%)	19 (20.0%)	
III	16 (27.6%)	10 (27.0%)	26 (27.4%)	
IV	25 (43.1%)	15 (40.5%)	40 (42.1%)	
No data	2 (3.4%)	0	2 (2.1%)	
p16 IHQ				
HNSCC p16(−)	29 (50.0%)	24 (64.9%)	53 (55.8%)	0.1627
HNSCC p16(+)	25 (43.1%)	11 (29.7%)	36 (37.9%)	
No data	4 (6.9%)	2 (5.4%)	6 (6.3%)	
HPV infection				
HPV16 DNA(−)	15 (25.9%)	12 (32.4%)	27(28.4%)	0.4887
HPV16 DNA(+)	43 (74.1%)	25 (67.6%)	68 (71.6%)	
Single infection	39 (90.7%)	23 (92.0%)	62 (91.2%)	
Co infection with other HR/LR HPV	4 (9.3%)	2 (8%)	6 (8.8%)	
Primary Treatment				
Surgery	1 (1.7%)	2 (5.4%)	3 (3.2%)	0.7376
Radiotherapy (RT)	10 (17.2%)	6 (16.2%)	16 (16.8%)	
Chemotherapy (CTX)	46 (79.3%)	29 (78.4%)	75 (78.9%)	
CTX + RT	26 (56.5%)	12 (41.4%)	38 (50.7%)	
Surgery + RT	9 (19.6%)	11 (37.9%)	20 (26.7%)	
Surgery + CTX	11 (23.9%)	6 (20.7%)	17(22.7%)	
No data	1 (1.7%)	0	1 (1.1%)	
Treatment response				
Complete	38 (65.5%)	28 (75.7%)	66 (69.5%)	0.5190
Persistence	11 (19.0%)	6 (16.2%)	17 (17.9%)	
No data	9 (15.5%)	3 (8.1%)	12 (12.6%)	

Legend: No data = no information available; RT = Radiotherapy; CTX = Chemotherapy; * *p*-value calculate for active vs. no consumption group.

marker (n = 208) and all cases associated with HPV infection other than HPV16 (n = 79); the remaining 103 cases were included in the study and analysed for PIK3CA status (Figure S1). Eight cases were further excluded from the study due to technical failure in the PIK3CA mutational status evaluation. As such, our study sample comprised 95 patients. The demographic and clinical features of the excluded patients were similar to the included patients (Table S4), and both groups showed overlapping survival curves (Figure S2), which suggests the absence of selection bias.

The demographic and clinical–pathological characteristics of the 95 HNSCC patients included in the analysis are summarized in Table 1. The average age at diagnosis was 62 years old and ranged between 37 and 90 years old. Most patients were men (73.7%, n = 70) and, according to self-reported consumer habits, the majority had active tobacco and/or alcohol consumption (70%, n = 66). Stratification was performed for each consumption habit. Alcohol exposure was characterised according to qualitative data in clinical records: never drank, ex-drinker, sporadic, moderate, and severe drinkers. The classification for tobacco consumption was defined as one pack/year (equal to one pack of cigarettes/day/year, with 20 cigarettes in a pack) distributed among current and heavy smokers. Current tobacco users included those who used tobacco $\leq$ 30 pack/years. Heavy users were those who smoked > 30 pack/years. The ex-smoker group included smokers who stopped until the date of diagnosis. Cases where the pack-a-year unit information was not available were classified as unrated smokers, and the data only reported consumption in a qualitatively way. Never users of tobacco or alcohol were defined as not having consumed either of these substances prior to cancer diagnosis. The analysis considered the two groups of active consumption vs. no consumption and did not consider the missing values. In 75 cases (79%), the primary tumour was located in the oropharynx, including the palatine tonsil (n = 47), the base of the tongue (n = 12), uvula (n = 2), soft palate (n = 10), trigone retromolar (n = 1), and oropharynx wall (n = 3). In the remaining 20 cases, the primary tumour was located in the oral cavity, including the tongue body (n = 13), buccal floor (n = 4), oral mucosa (n = 2), and mandible (n = 1). HPV infection was detected in 68 HNSCC cases, including 62 cases of HPV16 single-infection and six cases of coinfection with HPV16 and other HR or LR HPV types (HPV6, HPV18, HPV53, and HPV58). Data from $p16^{INK4a}$ immunohistochemical assays were available for 89 patients. Patients were treated with different schemes involving surgery and radiotherapy (RT) as single therapies or in association with systemic antineoplastic treatment (Table 1). This treatment heterogeneity was expected, taking into consideration the clinical characteristics and tumour stage distribution observed in our cohort. Most patients showed complete response to the applied treatment (n = 66), but cases with disease persistence (n = 17) and relapse (n = 6) were also reported. In 12 cases, the treatment response could not be retrospectively evaluated (Table 1).

There were no differences between the wild-type (WT) and mutant (MUT) PIK3CA groups regarding age, primary tumour location, tumour stage at diagnosis, HPV16 infection status, and primary treatment administered (Table 1). The PIK3CA WT group had a numerically higher proportion of males (78% vs. 68%), p16 overexpression (43% vs. 30%), and complete responses to treatment (65% vs. 76%) compared to MUT PIK3CA, although none of these differences was statistically significant. There was a statistically significantly higher proportion of patients with active tobacco/alcoholic consumption habits among the WT compared to the MUT PIK3CA (76% vs. 60%) group; as such, this variable was also considered in the multivariable analysis of the prognostic impact of PIK3CA mutations.

(two cases) in exon 20). The highest reported frequency (32%) was observed in a Japanese study (Suda 2012) with a cohort of 115 cases, equally distributed across different subsites of the HN region (oral cavity $n = 31$, oropharynx $n = 25$, larynx $n = 23$, hypopharynx $n = 25$, and nasopharynx $n = 11$). Five mutations spots were evaluated in exon 9 and exon 20. The distribution of mutations by location revealed a predominance of oropharyngeal sites (oral cavity $n = 9$ (24%), oropharynx $n = 10$ (27%), larynx $n = 8$ (22%), hypopharynx $n = 9$ (24%), and nasopharynx $n = 1$ (3%) data from T Suda 2012).

In addition, we performed a second analysis to explore the HNSCC PIK3CA mutation profile in greater depth. For this second analysis, we assessed 17 selected articles comprising data from 1286 HNSCC patients published between 2006 and 2021 where mutations were detected using DNA sequencing techniques instead of PCR-based methods (Supplemental Data Table S2). The most used sequencing methodology was the classic Sanger assay ($n = 10$ studies), followed by cutting-edge NGS technology ($n = 7$). All studies evaluated tumour tissue, and the majority of samples were collected at the time of diagnostic (i.e., were treatment-naive samples). Primary tumours were located at different subsites from the HN region stratified by the oral cavity, oropharynx, larynx, hypopharynx, and nasopharynx; all data that did not specifically describe these locations were considered as not stratified. Three out of 17 studies evaluated SCC exclusively from the oral cavity, three others from the oropharynx, and one study from the hypopharynx. The remaining eight studies include data from different subsites. Overall, the most represented subsite was the oral cavity ($n = 678$ cases, reported by 11 articles), followed by the oropharynx ($n = 340$, reported by seven articles), larynx ($n = 101$, reported by five articles), hypopharynx ($n = 94$, reported by six articles), nasopharynx ($n = 40$, reported by five articles), and 32 non-stratified cases (reported by four articles). The HPV status was described in nine out of 17 studies: three studies evaluated single HPV status cohorts HPV (+) or HPV (−), and at least six studies presented both HPV (+) and HPV (−) cases. HPV (−) HNSCC was the most represented subgroup ($n = 330$ cases), followed by HNSCC HPV (+) ($n = 244$ cases). Overall, the frequency of PIK3CA mutations ranged between 3% and 36%. Only two studies reported the lowest frequency (3%) (Bruckman 2010 and Cortelazzi 2015), at least five studies presented frequencies above 20%, and three studies above 32%. Overall, 45 mutations were reported, distributed by both exon 9 and exon 20, and novel mutations were reported in a few studies. All mutations were stratified by exon, except for 30 cases reported in three studies.

The analysis based on general data for the mutational profile was supported by data reported in each of the studies. The three most frequently encountered mutations were the E545K mutation ($n = 31$ cases; data from 15 studies), followed by E542K ($n = 21$; nine studies) on exon 9 and H1047R ($n = 16$; nine studies), T1025T ($n = 16$; two studies), M10431 ($n = 8$; two studies), and G1049 ($n = 2$; three studies) on exon 20 (Supplementary Table S3). The PIK3CA-mutated cases per HN subsite were accurate; the highest number was in the oral cavity ($n = 115$, reported by seven studies), followed by oropharynx ($n = 32$, three studies), hypopharynx ($n = 10$, three studies), larynx ($n = 8$, three studies), and nasopharynx with no cases reported. We also evaluated the PIK3CA mutation (data reported from six studies) for subgroups of the HPV status. A total of 33 cases of HNSCC HPV (+) and 19 cases of HNSCC HPV (−) harboured PIK3CA mutations. In five of the 17 studies, the presence frequency of PIK3CA mutations was associated with the clinical parameters and correlated with the respective outcome (OS and DFS). Four studies estimated a better prognosis associated with the presence of the mutation (Alsofyania 2020, Lim 2019, García-Carracedo 2016, and Cohen 2011), and in one study, no correlation was found (Chau 2016).

2.2. Portuguese HNSCC Study Population

2.2.1. Clinical and Demographic Characteristics

Cases were selected from a universe of 390 consecutive primary HNSCC cases located in the oropharynx or oral cavity at our tertiary cancer centre between 2010 and 2019. We excluded all cases with HPV-negative and missing information concerning the p16INK4a

patients had significantly higher 3-year disease-free survival (DFS) compared to PIK3CA MUT patients, and a multivariable analysis for age, sex, smoking, TNM stage, and treatment showed associations between PIK3CA status and disease recurrence [31]. A greater understanding of therapies targeting the PI3K pathway has been achieved, developing new therapeutic combinations to improve the survival of HNSCC patients, but their efficacy is variable [32,33]. The identification of biomarkers to predict the response to therapy will allow the selection of patients who are more likely to respond to new therapeutic combinations [32], suggesting that different treatment strategies need to be considered based on the molecular phenotype of each tumour [34,35]. In this context, it is critical to refine the profile of PIK3CA mutations in HPV-positive and HPV-negative HNSCC patients and its association with prognosis.

To address these issues, the present study includes a systematic review of the frequency of PIK3CA mutations in HNSCC and the impact of the different detection techniques. The study also reports the prevalence of canonical PIK3CA gene mutations (substitutions for H1047L and E542K, E545K, and E545D) within an HNSCC Portuguese patient series at the time of diagnosis and conducted an exploratory analysis to test the hypotheses: (1) whether PIK3CA (canonical) mutations are associated with HPV-positive HNSCC, as evaluated by HPV DNA and p16INK4a and (2) whether PIK3CA mutations are independently associated with overall and disease-free survival.

2. Results

2.1. Systematic Review

We performed a systematic review to study the PIK3CA mutation frequency in different HNSCC cohorts (total case number range between 25 and 115 cases; total number of cases was 479) in scientific articles published from 2012 to 2021. The PIK3CA gene mutation analysis was performed by qPCR designed for specific targets in the regions known as hotspots of the PIK3CA gene exon 9 and exon 20, such E542K, E545K, E545D, H1047R, and H1047L, respectively. This analysis covers three different geographical regions: Europe (three studies), Asia (two studies), and North America (one study) (Supplemental Data Table S1).

The majority of the studies performed a mutational analysis using untreated tumour biopsy tissue, the male gender was the most represented in all studies, always representing more than half of the sample (55–93%), the mean age ranged between 63 and 65 years, and all the cohorts presented the active consumption of tobacco and/or alcohol. Almost all the studies reported a stratification by HN subsites (oral cavity, oropharynx, larynx, hypopharynx, and nasopharynx), and the majority of the cases originated from the oropharynx (five out of six studies, a total of 138 cases), followed by the oral cavity (four studies, a total of 97 cases), larynx (four studies, a total of 80 cases), hypopharynx (four studies, a total of 80 cases), and nasopharynx (three studies, a total 16 cases). All patients presented a histological diagnosis of SCC, most of them at advanced stages (III and IV). Only three of the six studies reported the HPV DNA status of the cohort. In total, 54 HNSCC HPV (+) and 219 HNSCC HPV (−) were evaluated. In three studies, the p16INK4 status was determined by immunohistochemistry with a total of 45 positive cases. The mutation frequency was estimated in a qualitative analysis of the number of cases with a mutation present. All studies reported PIK3CA gene mutations, and their frequency ranged from 8 to 32%. The lowest frequency (8%) was reported in a cohort of 113 cases (Borkowska 2021), mostly consisting of laryngeal tumours (43%) and, largely, HPV (−) lesions (77%). This low frequency of mutation detection can be related to the study's design targeting only H1047R. Four other studies presented frequencies ranging between 16% and 18% and mainly consisted of lesions from the oropharynx and oral cavity. Among these four studies, HPV (−) HNSCC cases prevailed in the two studies where the HPV status was determined (García-Escudero 2018; SB Pattle 2017). All but one of these studies evaluated at least four mutations distributed both in the exon 9 and exon 20. The highest frequency was found (taking into account only (three cases) E545D (two cases) in exon 9 and H1047R

(OS) (3-year OS: WT 54% and MUT 65%; $p = 0.090$). HPV status was the only factor significantly associated with both OS and disease-free survival (DFS), with HPV (+) patients having consistently better outcomes (3-year OS: HPV (+) 65% and HPV (−) 36%; $p = 0.007$; DFS HPV (+) 83% and HPV (−) 43%; $p = 0.001$). There was a statistically significant interaction effect between HPV status and PIK3CA mutation regarding DFS (Interaction test: $p = 0.026$). In HPV (+) patients, PIK3CA wild-type is associated with a significant 4.64 times increase in the hazard of recurrence or death (HR = 4.64; 95% CI 1.02–20.99; $p = 0.047$). Overall, PIK3CA gene mutations are present in a large number of patients and may help define patient subsets who can benefit from therapies targeting the PI3K pathway. The systematic assessment of PIK3CA gene mutations in HNSCC patients will require further methodological standardisation.

Keywords: HNSCC; HPV; p16 INK4a; PIK3CA

1. Introduction

Head and neck carcinoma (HN) is the sixth leading cancer by incidence worldwide [1], according to data published in 2018, and is responsible for more than 800,000 new cases yearly and 450,000 deaths/year worldwide [2]. It comprises many different common and rare entities, the large majority (90%) being squamous cell carcinomas (SCC) [3]. There is a wide variation in disease progression and overall patient survival at different anatomic subsites [4,5], which may be explained by its multifactorial aetiology, in which environmental factors have a strong contribution, such as smoking and alcohol consumption. Infection with human papillomavirus (HPV), mainly the high-risk (HR) type HPV16, also plays an important role in a subgroup of these tumours, particularly in oropharyngeal cancers [6]. HPV-positive and HPV-negative tumours are clinically distinct [7], and two separate carcinogenesis routes are recognized [8]. Over the past decades, the incidence of HPV-associated oropharyngeal SCC has increased, as reported by the United States Center for Disease Control (CDC) in 2018 [9]. HPV-associated HNSCC has a substantially better prognosis after therapy compared with HPV-negative cases [10]. De-intensified therapy for these patients is being actively explored to reduce the associated morbidity while maintaining tumour control [11,12].

The genomic profile of HNSCC published by The Cancer Genome Atlas (TCGA) in 2015 allowed the development of new mutation studies and highlighted a high frequency of changes in components of the phosphatidylinositol-3-kinase (PI3K) signalling pathway, pointing out PIK3CA as the most frequently altered gene [13]. Activating PIK3CA gene mutations upregulate intracellular signalling via the PI3K–protein kinase T (Akt)–mammalian target of rapamycin (mTOR) pathway, contributing to multiple hallmarks of cancer, such as resisting cell death and uncontrolled proliferation [14–16]. As described previously, HPV genome integration leads to keratinocyte immortalization and transformation by inhibiting the tumour suppressor p53 and retinoblastoma protein (pRb) but also by interacting with other pathways, including PI3K/Akt/mTOR [17–20]. Several studies have addressed the role of the PIK3CA pathway in cervical cancer and other types of HPV-associated cancers [21–25]. In the evaluation of 151 HNSCC whole-exome sequences, the PI3K pathway was found to be the most commonly altered mitogenic pathway (30.5% of tumours) compared to the JAK/STAT pathway (9.3%) and MAPK pathway (8.0%) [26]. Among the PIK3CA mutations observed in HNSCC, 63% occur at three specific locations encoding the p110α subunit, namely E542, E545, and H1047, known as canonical mutations [27,28]. The PIK3CA mutation frequency among HPV-positive tumours was reported to be approximately half of that found in HPV-negative cancers [29]. Even so, the PIK3CA gene remains one of the most mutated in HPV-associated HNSCC [7,30]. Additionally, mutations have been associated with adverse outcomes in solid tumours, but their prognostic significance in oropharyngeal SCC is still unclear. In previous studies, the specific survival data did not differ between PIK3CA wild-type (WT) and mutated (MUT) lesions. However, WT-PIK3CA

CHAPTER 5

GENERAL DISCUSSION

HNSCC and HPV type, lineage and NGS

The HNSCC associated with HPV typically arises in the oropharynx and is considered a distinct entity compared with other HN tumours named HNSCC-nonHPV. The recognition of this difference incited the development of studies in this field, to gain insights into patient risk stratification and more personalized treatments. In the early 2000s with the introduction of new molecular tools aided several research groups to explore molecular mechanism-based approaches to define genetic signatures that may play a prime role in patient stratification (Alexandrov and Stratton, 2014; Chai, Lim and Cheong, 2020).

In this context, our research project led us to explore if specific HPV16 variants could influence the occurrence, the morphology, and the molecular phenotype of HNSCC-HPV. Apparently, within each host, HPV genomes harbour high variability levels in the evolution from low-grade lesions to high-grade lesions and cancer, which may reveal several genetic signatures (Nacher *et al.*, 2020).

The growing number of studies on HPV16 variants, was mainly done in uterine cervical cancer and has exposed a very large number of different papillomavirus genomes. Moreover, the research on HPV phylogenetic lineages with clinical host specificity reports several differences within its oncogenic potential (Sichero *et al.*, 2007; Zhang *et al.*, 2015).

Our general hypothesis was that the specific morphological and molecular phenotype HPV16 variants identified in HNSCC tumour could be associated with different clinical pathologic settings. In this manner, our study was conceived to identify the frequency of HPV16 variants in squamous cell carcinomas of the head and neck region and to evaluate their correlation with clinical variables.

At the onset, we carried out a systematic review (SR) of the literature to scrutinize which sequencing methodologies were performed. We concluded that in most studies in the HN region, a limited number of samples were examined, mainly because of the labour-intensive techniques used (PCR). Sanger sequencing techniques were applied to identify HPV16 variants in all nine HN studies included in the analysis; no Next Generation Sequencing (NGS) was used, most likely due to difficulties on the access to the technique and the limitation on supporting costs to applying this new sequencing methodology. The use of conventional Sanger sequencing likely limited the number of cases analysed and introduced limitations in a whole genome analysis with respect to the size of the amplicons, which may compromise the sensitivity of the assay (Hoffmann *et al.*, 2004; Park *et al.*, 2016;

Yang, Ren and Zhang, 2016). In fact, only two of the nine studies included more than 200 patients with HNSCC, and none looked at more than 52 HPV-positive patients for HPV16 variants.

The labour time and budget required to perform large-scale HPV sequencing studies by current Sanger methods are no longer viable. The new advances on NGS techniques respond to that need for a high-throughput, cost-effective, and robust HPV sequencing process (Cullen *et al.*, 2015; Mirabello *et al.*, 2018) promoting a deeper understand of viral genetic diversity among infected individuals (Zhang *et al.*, 2015; Mirabello *et al.*, 2018).

These reasoning were one of the main reasons to adopt an NGS technology approach to determine HPV16 variants in HNSCC in our thesis.

Our objective was to contribute to the knowledge in specific aspects of the pathology of HNSCC, through a detailed molecular analysis and to validate the biological determinants found by the new sequencing tools. We were also confident that these approaches might also be important to improve knowledge of the role of HPV16 variants in other associated locations. For that, we developed an NGS workflow, designed to successfully sequence many HPV16 samples by whole genome amplification and to improve a comprehensive overview of the HPV16 mutational scenario for different HPV-associated tumours. The development of this technique in our laboratory is presented in the description of the methods in chapter 3 Assay Optimization.

Currently, there is no published consensus protocol for analysing HPV16 variants by NGS, not even in commercial kit format. Albeit, there are some descriptions of methods by various groups, with different approaches, with or without total sequencing of the HPV genome (Cullen *et al.*, 2015; Mirabello *et al.*, 2016; Mandal *et al.*, 2022). In those reports, there is also variation on optimization for different conditions of PCR, construction of libraries and sequencing equipment. Differences regarding tumour type and location are also present: most of the reports studied uterine cervical carcinomas, however some reports exist on tumours in other locations, such as head and neck (LeConte *et al.*, 2018; Lang Kuhs *et al.*, 2022) and anal tumours (Piyathilake *et al.*, 2020).

For our study, the development and design of this *in-house* method was based on a panel of primers previously described (Cullen *et al.*, 2015) for the amplification of the complete genome of the HPV virus (8000bp) from samples of uterine cervical tumours, which were adapted to our conditions. Thus, our PCR technique

integrates a different chemistry, as well as differences concerning the library construction and sequencing equipment compared with those techniques originally published. Our workflow was applied to a universe of eligible samples from a selected cohort of patients diagnosed with HNSCC and a series of tissue samples from the *in vivo* K14HPV16 mouse model.

The performance of this workflow was evaluated using our series of 77 samples, determining its sensitivity, performance, and efficiency. Specificity was ensured by validating the bioinformatics algorithm on an HPV16 positive sample used as a positive control (SiHa cell line containing HPV Lineage A). This analytical validation intended to certify the method for routine laboratory use, which until now has never been published or commercially registered.

A total of 77 samples were evaluated by this NGS workflow assay, comprising all HPV16 positive cases from the cohort of patients (n=68) and the HPV16 positive animal model tissues (n=9) samples and we obtained reads in all cases with a total genome coverage that ranged between 1-100%. The samples with low genome coverage (less than 10%) were excluded, as they may not have reached the minimum amount of amplicon, compromising the complete reading of the total genome. We realized that the minimal quantity of viral DNA (ng) obtained is crucial and turns out to be one of the steps where the procedure will have to be optimised. The sensitivity study was carried out by the analysis of cell lines infected with HPV16 (SiHa cells, containing one to two copies per cell). We processed 100 and 250 ng total DNA extracted (genomic and viral) and good results were obtained with the minimum amount of 100ng.

In the set of samples processed for sequencing, the extracted DNA was evaluated for quantity and purity. In the amplification reaction, the amount of DNA applied was normalized to 100ng for all samples, although some did not reach the minimum amount necessary: they were still processed regarding the analytical accuracy of the method.

Overall, the failure rate was 45%, and this failure rate is higher compared to that described by other studies that use the same methodology, the same panel of primers (Cullen *et al.*, 2015) and different PCR conditions designed for the pathology of the cervix. From our point of view, the main reason for this result was the low amount of viral DNA applied in our experiments, because we are detecting viral DNA that is diluted in genomic DNA, and the majority of cases carries a very low number of viral copies (Badaracco *et al.*, 2007). Sensitivity studies are our best

references in this measurement (viral copies) but even so, they have limitations, in the sense that they are carried out in cell lines, with a stable number of copies of the virus and which may be different from what is found in samples from patients positive for HPV. In our opinion, a special effort has to be done to enrich total DNA to the needed analytical requirements (amplicon quantity) for next generation sequencing efficiency. Essentially, carrying out this optimization in these samples (small number of copies) was compromised by the limited NGS resources provided (time and costs).

Experimental results

Besides generating methodological insights, a major motivation of this study was to further provide a comparative genome data analysis of the HPV16 strains identified in our HNSCC cohort, as well to determine the HPV16 lineage of the K14HPV16 model. Overall, for our comparative genome analysis, we chose a conservative approach to validate the identified variants, considering only the samples presenting a total genome coverage above 90% (horizontal reading), contrasting with some authors that had established for lower values (≈60%-80%) (Mirabello *et al.*, 2016; Lang Kuhs *et al.*, 2022). A total of 35 samples from the patient's series and 7 samples from animal model were studied. We presented the resulting data under the publication of articles, which is described in chapter 4 of the results.

Considering that this project was developed to understand the impact of HPV lineages in carcinogenesis we decided to also undertake the identification of HPV16 lineages in an animal model.

Animal models have been widely used in carcinogenesis studies, are essential to promote knowledge of tumour biology and are deeply appreciated in Medicine.

The K14HPV16 FVB/n mouse model was specifically chosen to be molecularly characterized because is an excellent model for HNSCC, as oral cancer appears spontaneously, mostly in female mice (Neto *et al.*, 2021).

This model was constructed in 1994 at the University of California (Arbeit *et al.*, 1994) and by that time, the characterization of the whole genome of HPV16 was not accessible and was not performed. In fact, it has never been done before.

Our work of sequencing different tissues from two female K14HPV16 FVB/n mice, determined that this mouse model harbours the HPV16 variant A1 sublineage and that this is consistent in all tissues analysed, as expected. Furthermore, deep

molecular characterization of that HPV16 genome reveals three variations, two of which are missense variations and are located predominant in early E1 genes. Only one of these two variations: 978A>G (Ile326Met) was found previously and described in a Chinese cohort on a uterine cervical HPV16 infection (Yao *et al.*, 2019; Wang *et al.*, 2021). In that report, nothing was established regarding its biological significance. **In our opinion, knowledge of the specific lineage in this animal model is very relevant, as it will allow researchers to explore the role of the HPV A1 subline in cancer in HNSCC or in other locations and the rational use of the animal model in therapy and other studies.**

Regarding the dataset of patients included, the main findings from comparative genome data analysis described the HPV16 variant lineages, sub lineages and SNVs identified in our HNSCC cohort. Likewise, highlights its potential importance to understand the biological behaviour of intratype variants of papillomaviruses associated with HN cancer (presented in Chapter 4.2 from Results).
In our cohort of thirty-five patients, the majority clustered into Lineage A (European). Only nine cases were distributed among Lineage B (African-1), Lineage C (African-2) and Lineage D (Asian-America). These results are similar to what is observed in uterine cervical carcinomas, sustaining that both cancers have a higher prevalence of the European lineage (LeConte *et al.*, 2018). Additionally, it is interesting to notice the convergent results from the patient cohort and the animal model, suggesting that the A1 HPV16 sublineage has a high oncogenic potential in the context of the oropharynx. The low number of non-Lineage A cases does not allow to correlate the clinical parameters and patient's survival and HPV16 lineage identification.
The limited number of HNSCC cases studied contributes to the description of lineages in HNSCC but does not allow us to explore the clinical meaning of differences. As it occurred in previous reports, no conclusions can be drawn about the potential contribution of specific HPV16 variants to increasing cancer risk or modifying biopathology and prognosis of the tumour (Gillison *et al.*, 2000) or the clinicopathological aspect of the HPV16 variants (Hoffmann *et al.*, 2004) in HNSCC.

Total Single nucleotide Variants (SNV)
In addition to the genomic comparative analysis of the 35 patients, we also did the evaluation of single nucleotide variations.

From our literature review we concluded that the study of HPV16 variants and its biological significance in HNSCC is enormously difficult because it is essentially dispersed and lacks a uniform and systematic nomenclature. Moreover, because most studies are predominantly descriptive, no inference can be obtained about their biological meaning.

In general, very few authors performed a whole-genome approach, and the majority of the studies evaluated specific regions on viral genome. The most studied regions were the L1, E6, E7 and LCR genes, lacking, in the literature, the molecular characterization of E2 gene and especially the E5 gene.

In our study we contribute with new data on the molecular profile of those genes and confirm important knowledge on other viral genes.

We report 243 variable sites in the viral genome, among them 84 yielding missense variants which integrate an amino acid alteration. The greatest variability was observed in the E5 and L2 genes, but the E2 and E6 genes show a higher proportion of missense production in relation to their respective synonyms. The viral E7 gene region, was the most conserved in our results, in agreement with previously studies done in HN tumours and uterine cervical carcinomas (Nacher *et al.*, 2020). This findings sustains the restriction of mutations in the E7 gene on both sites, oropharyngeal and tonsil, share and support the proposal this viral gene as potential therapeutic target (Strati and Lambert, 2007).

One hundred variants of the total 243 variants that were found, were previously reported. Of those variants 31 were previously shown to have a role in the carcinogenesis process or impact in the clinical context. Of these variants, the majority were studied in the context of uterine cervical carcinoma, but very recently in 2022, Lang Kuhs and co-workers (Lang Kuhs *et al.*, 2022) applied the new generation sequencing methodologies in HNSCC. They found eight HPV16 variants associated with worse prognosis in a large cohort of patients with HNSCC. In our Portuguese dataset, we identified four of those eight variants described by Kuhs *et al* (one missense and 3 synonymous) in genome positions 4410, 4539, 5962 and 7173 from L2, L1 and E6 region, respectively, that have been also previously implicated in carcinogenesis (Jabbar *et al.*, 2010). These SNV were identified in 13 patients of our HNSCC cohort, mostly diagnosed on stage II-III, represented by 10 males to 3 females and with age over 65 years. Although the low number of cases, one of which appears to be associated with a fatal outcome (4539 1167T>C). We, as others previously, can recapitulate that studies that explore HPV variant's

biological and phenotypic impact on cancer are very difficult to perform, need a very large collection of cases, require functional laboratory studies and imply a high cost.

Furthermore, in the analysis of our results, it is interesting to note that we also found discrepancies compared to what has already been described.

Remarkably, there are E6 gene modifications that are commonly found in uterine cervical cancer, namely E31G, L83V, and D25E variants (Hirose *et al.*, 2018), and were indeed found in previous HN cohort studies but did not emerge in our results. Those mutations are significantly associated with the development of cervical carcinoma (Tornesello *et al.*, 2004; Del Mistro *et al.*, 2006) and have been described in the HN region and specifically the L83V (E350G) in six studies out of nine eligible articles from the HN region in our systematic review (presented on chapter 3 Results). Interestingly, all these three substitutions in the E6 were neither present in our data, nor in Khus et al analysis (Lang Kuhs *et al.*, 2022) and others included in our systematic review (Agrawal *et al.*, 2008).

Our general data results sustain that changes (SNV) in the HPV genome do not influence the type or location of the tumour, as the same SNV are found in different anatomic tissues. This was also reported by the (LeConte *et al.*, 2018) study that compared genomic characteristics of HPV associated cervical cancer versus oropharyngeal tumours showed and did not found significant differences concerning the distribution in HPV variants (LeConte *et al.*, 2018).

PIK3CA

Finally, our last molecular analysis proposed to study the most mutated gene in HNSCC cancer, *PIK3CA*.

It is widely presumed in cancer research that the discovery and/or implication of activation of signalling pathways in cancer offers excellent new therapeutic targets, as demonstrated by the the use of BRAF or mTOR inhibitors (Wang *et al.*, 2017; Tian, Li and Zhang, 2019).

In HNSCC, *PIK3CA* mutations may contribute to the pathogenesis of HPV-associated HNSCC (Lee *et al.*, 2020). Particularly, the mutational status of the *PIK3CA* gene was eligible for further studies due to the high frequency of mutations within HNSCC, and is now actively exploited as a therapeutic target (Zhang *et al.*, 2016). *PIK3CA* mutations contribute to the uncontrolled proliferation

and survival of cancer cells and is associated with a poor patient prognosis. Determining which patients carry *PIK3CA* mutations is of great clinical importance since it can help to define subgroups of patients with specific needs and who may benefit from promising targeted therapies against the PI3K-Akt-mTOR signalling pathway.

Our study proposed a qualitive analysis of the mutations commonly associated with HNSCC tumours. The present results contribute to better defining this complex scenario, determining the frequencies of canonical *PIK3CA* mutations in a Portuguese cohort of patients with HNSCC and testing the putative associations between *PIK3CA* mutations and multiple clinicopathological parameters. In our dataset, most patients with HNSCC carrying canonical *PIK3CA* mutations had advanced disease stages (III, IV), especially among HNSCC-HPV, suggesting a negative impact on patient prognosis and corroborating previously published data (Zieba *et al.*, 2020). However, there was no significant association between mutation frequency and HPV status, and no correlation between *PIK3CA* and overall survival between univariate or multivariate correlations. **Importantly, our data evidence a high frequency of canonical *PIK3CA* mutations in HNSCC, including in HNSCC-nonHPV cases, suggesting that screening for *PIK3CA* mutations should not be restricted to the HNSCC-HPV group but can also be relevant to HNSCC-nonHPV patients, which contrasts with previous studies that indicated a higher frequency in HNSCC-HPV tumours.**

Overall, our results contribute with new data in the field of molecular analysis in oncological pathology applied to HPV16-positive oropharyngeal tumours. Currently, it is valuable to consider that tumour recurrence still occurs in 10-20% of HPV-positive HNSCC patients. Although the clinical management of HNSCC has improved since the recognition of HPV-positive and HPV-negative lesions, further therapeutic advances will require effective clinical validation of all these promising biomarkers in large-scale population studies.

Currently, in response to the clinical diagnosis, the HPV status is determined in suspicious lesions of the oropharynx by qualitative PCR methodologies that translate its presence or not. These methodologies designed to detect viral DNA only identify the genotype present. The new molecular tools developed in the present work magnify the information on this diagnosis and enrich the information on the Virus-Host dynamics, promoting better patient risk stratification and more

targeted treatments in clinical practice. Particularly, one of the strengths of this project was the development of the NGS workflow, validated against the ISO standard 15189 to certify its implementation in a clinical laboratory, thus contributing with new molecular inputs for the clinical practice among head and neck tumours.

Finally, our future perspective is to apply this workflow to clinically referred casuistic associated to HPV tumours pretending to expand, sustain and validate biomarkers to answer the clinical practice needs. Hopefully, the new sequencing era will innovate the study of HPV and related cancers.

CHAPTER 6

CONCLUDING REMARKS

Molecular analysis using next-generation sequencing methodology represents a valuable tool to consolidate genetic markers with their respective clinicopathological characteristics. The studies published in this field reinforce its application on a larger scale to improve the clinical implications of early diagnosis, prognosis, and therapeutics. The NGS methodologies implementation at the routine laboratory within the scope of HPV-associated pathology potentially increases the studies of these cohorts, provide a detailed description of mutational profiles associated with the respective clinicopathological characteristics and assesses greater robustness as a complementary of diagnosis.

A strong point of this project was the development of an NGS workflow for the identification of HPV16 variants in HNSCC tumours, optimized to analyse DNA from different types of products, and therefore also promising for its application in the context of other associated pathologies, such as the cervix and anal canal.

The analytical validation defines the characteristics of the assay based on studies the sensitivity and specificity of cell lines infect HPV16 (SiHa cell), tissue sample lesions from the K14HPV16 model and tissue tumours samples from a cohort of patients diagnosed with HNSCC.

The weaknesses are highlighted mainly by the unsatisfactory results of 33 samples to determine HPV16 variants, directly compromising the expected cohort for the phylogenetic analysis and association with the clinicopathological characteristics and survival. However, the vulnerable points identified in the global process and their resolution make it possible to refine technical flaws and contribute to increase the success of the NGS workflow.

Finally, in general, it is a technically consistent study that reveals original data from genomic analysis that directly contribute to deepening the study of the role of variants in the pathology of HNSCC and reinforcing its promissory purpose as a complementary tool of diagnosis.

In our future perspectives, we intend to expand our study, increasing the number of cases included, and invest in the improvement of bioinformatics analysis. We pretend to develop an HPV variant Database platform allowing clinical classification of the different HPV variants found. The implementation of standard methods allows this database to serve as a tool for further studies and contribute to a better translational understanding of the biologic significance of each variant.

BIBLIOGRAPHIC REFERENCES

Agrawal, N. *et al.* (2011) 'Exome sequencing of head and neck squamous cell carcinoma reveals inactivating mutations in NOTCH1', *Science (New York, N.Y.)*, 333(6046), pp. 1154–1157. Available at: https://doi.org/10.1126/science.1206923.

Agrawal, Y. *et al.* (2008) 'Oral human papillomavirus infection before and after treatment for human papillomavirus 16-positive and human papillomavirus 16-negative head and neck squamous cell carcinoma.', *Clinical cancer research : an official journal of the American Association for Cancer Research*, 14(21), pp. 7143–50. Available at: https://doi.org/10.1158/1078-0432.CCR-08-0498.

Alexandrov, L.B. and Stratton, M.R. (2014) 'Mutational signatures: The patterns of somatic mutations hidden in cancer genomes', *Current Opinion in Genetics and Development*, 24(1), pp. 52–60. Available at: https://doi.org/10.1016/j.gde.2013.11.014.

Amin, M.B. *et al.* (2017) 'The Eighth Edition AJCC Cancer Staging Manual: Continuing to build a bridge from a population-based to a more "personalized" approach to cancer staging', *CA: a cancer journal for clinicians*, 67(2), pp. 93–99. Available at: https://doi.org/10.3322/caac.21388.

Androphy, E.J., Schiller, J.T. and Lowy, D.R. (1985) 'Identification of the Protein Encoded by the E6 Transforming Gene of Bovine Papillomavirus', *Science*, 230(4724), pp. 442–445. Available at: https://doi.org/10.1126/science.2996134.

Arbeit, J.M. *et al.* (1993) 'Neuroepithelial carcinomas in mice transgenic with human papillomavirus type 16 E6/E7 ORFs', *American Journal of Pathology*, 142(4), pp. 1187–1197.

Arbeit, J.M. *et al.* (1994) 'Progressive squamous epithelial neoplasia in K14-human papillomavirus type 16 transgenic mice', *Journal of Virology*, 68(7), pp. 4358–4368. Available at: https://doi.org/10.1128/jvi.68.7.4358-4368.1994.

Arbeit, J.M., Howley, P.M. and Hanahan, D. (1996) 'Chronic estrogen-induced cervical and vaginal squamous carcinogenesis in human papillomavirus type 16 transgenic mice', *Proceedings of the National Academy of Sciences of the United States of America*, 93(7), pp. 2930–2935. Available at: https://doi.org/10.1073/pnas.93.7.2930.

Argiris, A. *et al.* (2008) 'Head and neck cancer', *The Lancet*, 371(9625), pp. 1695–1709. Available at: https://doi.org/10.1016/S0140-6736(08)60728-X.

Badaracco, G. *et al.* (2007) 'Molecular analyses and prognostic relevance of HPV in head and neck tumours', *Oncology Reports*, 17(4), pp. 931–939. Available at: https://doi.org/10.3892/or.17.4.931.

Baker, C.C. *et al.* (1987) 'Structural and transcriptional analysis of human papillomavirus type 16 sequences in cervical carcinoma cell lines', *Journal of Virology*, 61(4), pp. 962–971. Available at: https://doi.org/10.1128/jvi.61.4.962-971.1987.

Balbo, S. *et al.* (2012) 'Kinetics of DNA Adduct Formation in the Oral Cavity after Drinking Alcohol', *Cancer Epidemiology, Biomarkers & Prevention*, 21(4), pp. 601–608. Available at: https://doi.org/10.1158/1055-9965.EPI-11-1175.

Bechtold, V., Beard, P. and Raj, K. (2003) 'Human Papillomavirus Type 16 E2 Protein Has No Effect on Transcription from Episomal Viral DNA', *Journal of Virology*, 77(3), pp. 2021–2028.

Available at: https://doi.org/10.1128/JVI.77.3.2021-2028.2003.

Bernard, H.U. *et al.* (1993) 'Sequence Variants of Human Papillomavirus Type 16 from Couples Suggest Sexual Transmission with Low Infectivity and Polyclonality in Genital Neoplasia', *Journal of Infectious Diseases*, 168(4), pp. 803–809. Available at: https://doi.org/10.1093/infdis/168.4.803.

Bernard, H.-U. *et al.* (2010) 'Classification of papillomaviruses (PVs) based on 189 PV types and proposal of taxonomic amendments', *Virology*, 401(1), pp. 70–79. Available at: https://doi.org/10.1016/j.virol.2010.02.002.

Bernard, H.U., Calleja-Macias, I.E. and Dunn, S.T. (2006) 'Genome variation of human papillomavirus types: Phylogenetic and medical implications', *International Journal of Cancer*, 118(5), pp. 1071–1076. Available at: https://doi.org/10.1002/ijc.21655.

Betiol, J.C. *et al.* (2016) 'Prevalence of human papillomavirus types and variants and p16(INK4a) expression in head and neck squamous cells carcinomas in São Paulo, Brazil.', *Infectious agents and cancer*, 11(1), p. 20. Available at: https://doi.org/10.1186/s13027-016-0067-8.

Bhaijee, F. *et al.* (2012) 'Cancer stem cells in head and neck squamous cell carcinoma: a review of current knowledge and future applications', *Head & Neck*, 34(6), pp. 894–899. Available at: https://doi.org/10.1002/hed.21801.

Borges, V. *et al.* (2018) 'INSaFLU: an automated open web-based bioinformatics suite "from-reads" for influenza whole-genome-sequencing-based surveillance', *Genome Medicine*, 10(1), p. 46. Available at: https://doi.org/10.1186/s13073-018-0555-0.

Boscolo-Rizzo, P. *et al.* (2009) 'HPV-16 E6 L83V variant in squamous cell carcinomas of the upper aerodigestive tract', *Journal of Cancer Research and Clinical Oncology*, 135(4), pp. 559–566. Available at: https://doi.org/10.1007/s00432-008-0490-3.

Bray, F. *et al.* (2018) 'Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries.', *CA: a cancer journal for clinicians*, 68(6), pp. 394–424. Available at: https://doi.org/10.3322/caac.21492.

Burd, E.M. (2003) 'Human papillomavirus and cervical cancer', *Clinical Microbiology Reviews*, 16(1), pp. 1–17. Available at: https://doi.org/10.1128/CMR.16.1.1-17.2003.

Burk, R.D., Chen, Z. and Van Doorslaer, K. (2009) 'Human papillomaviruses: Genetic basis of carcinogenicity', *Public Health Genomics*, 12(5–6), pp. 281–290. Available at: https://doi.org/10.1159/000214919.

Burk, R.D., Harari, A. and Chen, Z. (2013) 'Human papillomavirus genome variants', *Virology*, 445(1–2), pp. 232–243. Available at: https://doi.org/10.1016/j.virol.2013.07.018.

Campo, L., Zhang, C. and Breuer, E.-K. (2015) 'EMT-Inducing Molecular Factors in Gynecological Cancers', *BioMed Research International*, 2015, p. 420891. Available at: https://doi.org/10.1155/2015/420891.

Cancer Genome Atlas Network (2015) 'Comprehensive genomic characterization of head and neck squamous cell carcinomas.', *Nature*, 517(7536), pp. 576–82. Available at: https://doi.org/10.1038/nature14129.

Carter, J.J. *et al.* (1995) 'Use of human papillomavirus type 6 capsids to detect antibodies in people with genital warts', *Journal of Infectious Diseases*, 172(1), pp. 11–18. Available at: https://doi.org/10.1093/infdis/172.1.11.

Carter, J.J. *et al.* (1996) 'The natural history of human papillomavirus type 16 capsid antibodies among a cohort of university women', *Journal of Infectious Diseases*, 174(5), pp. 927–936. Available at: https://doi.org/10.1093/infdis/174.5.927.

Centers for Disease Control and Prevention (2018) 'Cancers Associated with Human Papillomavirus, United States, 2011–2015', *USCS DATA BRIEF*, 4, pp. 1–2.

Chai, A.W.Y., Lim, K.P. and Cheong, S.C. (2020) 'Translational genomics and recent advances in oral squamous cell carcinoma', *Seminars in Cancer Biology*, 61(1), pp. 71–83. Available at: https://doi.org/10.1016/j.semcancer.2019.09.011.

Chaturvedi, A.K. *et al.* (2011) 'Human papillomavirus and rising oropharyngeal cancer incidence in the United States', *Journal of Clinical Oncology*, 29(32), pp. 4294–4301. Available at: https://doi.org/10.1200/JCO.2011.36.4596.

Chaturvedi, A.K. *et al.* (2013) 'Worldwide trends in incidence rates for oral cavity and oropharyngeal cancers', *Journal of Clinical Oncology*, 31(36), pp. 4550–4559. Available at: https://doi.org/10.1200/JCO.2013.50.3870.

Chen, A.A. *et al.* (2015) 'Human Papillomavirus 18 Genetic Variation and Cervical Cancer Risk Worldwide', *Journal of Virology*. Edited by L. Banks, 89(20), pp. 10680–10687. Available at: https://doi.org/10.1128/jvi.01747-15.

Chera, B.S. *et al.* (2015) 'Phase 2 trial of de-intensified chemoradiation therapy for favorable-risk human papillomavirus-associated oropharyngeal squamous cell carcinoma', *International Journal of Radiation Oncology Biology Physics*, 93(5), pp. 976–985. Available at: https://doi.org/10.1016/j.ijrobp.2015.08.033.

Chien, M.-H. *et al.* (2012) 'Effects of E-cadherin (CDH1) gene promoter polymorphisms on the risk and clinicopathologic development of oral cancer', *Head & Neck*, 34(3), pp. 405–411. Available at: https://doi.org/10.1002/hed.21746.

Chirgwin, K.D. *et al.* (1995) 'Incidence of venereal warts in human immunodeficiency virus-infected and uninfected women', *Journal of Infectious Diseases*, 172(1), pp. 235–238. Available at: https://doi.org/10.1093/infdis/172.1.235.

Chung, C.H. *et al.* (2004) 'Molecular classification of head and neck squamous cell carcinomas using patterns of gene expression', *Cancer Cell*, 5(5), pp. 489–500. Available at: https://doi.org/10.1016/s1535-6108(04)00112-6.

Conrad, M., Bubb, V.J. and Schlegel, R. (1993) 'The human papillomavirus type 6 and 16 E5 proteins are membrane-associated proteins which associate with the 16-kilodalton pore-forming protein', *Journal of Virology*, 67(10), pp. 6170–6178. Available at: https://doi.org/10.1128/jvi.67.10.6170-6178.1993.

Contreras-Paredes, A. *et al.* (2009) 'E6 variants of human papillomavirus 18 differentially modulate the protein kinase B/phosphatidylinositol 3-kinase (akt/PI3K) signaling pathway', *Virology*, 383(1), pp. 78–85. Available at: https://doi.org/10.1016/j.virol.2008.09.040.

Cornet, I. *et al.* (2012) 'Human Papillomavirus Type 16 Genetic Variants: Phylogeny and Classification Based on E6 and LCR', *Journal of Virology*, 86(12), pp. 6855–6861. Available at: https://doi.org/10.1128/jvi.00483-12.

Cornet, I. *et al.* (2013) 'HPV16 genetic variation and the development of cervical cancer worldwide', *British Journal of Cancer*, 108(1), pp. 240–244. Available at: https://doi.org/10.1038/bjc.2012.508.

Coussens, L.M., Hanahan, D. and Arbeit, J.M. (1996) 'Genetic predisposition and parameters of malignant progression in K14- HPV16 transgenic mice', *American Journal of Pathology*, 149(6), pp. 1899–1917.

Cullen, M. *et al.* (2015) 'Deep sequencing of HPV16 genomes: A new high-throughput tool for exploring the carcinogenicity and natural history of HPV16 infection', *Papillomavirus research (Amsterdam, Netherlands)*, 1, pp. 3–11. Available at: https://doi.org/10.1016/j.pvr.2015.05.004.

Day, P.M., Lowy, D.R. and Schiller, J.T. (2003) 'Papillomaviruses infect cells via a clathrin-dependent pathway', *Virology*, 307(1), pp. 1–11. Available at: https://doi.org/10.1016/S0042-6822(02)00143-5.

De Bree, R. and Leemans, C.R. (2010) 'Recent advances in surgery for head and neck cancer', *Current Opinion in Oncology*, 22(3), pp. 186–193. Available at: https://doi.org/10.1097/CCO.0b013e3283380009.

De la Cruz-Hernández, E. *et al.* (2005) 'Differential splicing of E6 within human papillomavirus type 18 variants and functional consequences', *Journal of General Virology*, 86(9), pp. 2459–2468. Available at: https://doi.org/10.1099/vir.0.80945-0.

De Villiers, E.M. *et al.* (2004) 'Classification of papillomaviruses', *Virology*, 324(1), pp. 17–27. Available at: https://doi.org/10.1016/j.virol.2004.03.033.

Del Mistro, A. *et al.* (2006) 'Human papillomavirus typing of invasive cervical cancers in Italy.', *Infectious agents and cancer*, 1(1), p. 9. Available at: https://doi.org/10.1186/1750-9378-1-9.

Dias, T.R. *et al.* (2022) 'Expression of LncRNAs in HPV-induced Carcinogenesis and Cancer Cachexia: A Study in K14-HPV16 Mice.', *Anticancer research*, 42(5), pp. 2443–2460. Available at: https://doi.org/10.21873/anticanres.15723.

Dillner, J. (1999) 'The serological response to papillomaviruses', *Seminars in Cancer Biology*, 9(6), pp. 423–430. Available at: https://doi.org/10.1006/scbi.1999.0146.

Doorbar, J. *et al.* (1991) 'Specific interaction between HPV-16 E1-E4 and cytokeratins results in collapse of the epithelial cell intermediate filament network', *Nature*. Available at: https://doi.org/10.1038/352824a0.

Doorbar, J. (2006) 'Molecular biology of human papillomavirus infection and cervical cancer', *Clinical Science*, 110(5), pp. 525–541. Available at: https://doi.org/10.1042/CS20050369.

Doorbar, J. (2016) 'Model systems of human papillomavirus-associated disease: Papillomavirus disease models', *The Journal of Pathology*, 238(2), pp. 166–179. Available at: https://doi.org/10.1002/path.4656.

Du, J. *et al.* (2012) 'Human papillomavirus (HPV) 16 E6 variants in tonsillar cancer in comparison to those in cervical cancer in Stockholm, Sweden.', *PloS one*. Edited by R. Medeiros, 7(4), p. e36239. Available at: https://doi.org/10.1371/journal.pone.0036239.

El-Mofty, S.K. (2012) 'HPV-Related Squamous Cell Carcinoma Variants in the Head and Neck', *Head and Neck Pathology*, 6(SUPPL. 1), pp. 55–62. Available at: https://doi.org/10.1007/s12105-012-0363-6.

El-Naggar AK *et al.* (2017) *WHO Classification of Head and Neck Tumours*. 4th edn. World Health Organization. Available at: https://publications.iarc.fr/Book-And-Report-

Series/Who-Classification-Of-Tumours/WHO-Classification-Of-Head-And-Neck-Tumours-2017.

Evander, M. *et al.* (1995) 'Human papillomavirus infection is transient in young women: A populationbased cohort study', *Journal of Infectious Diseases*, 171(4), pp. 1026–1030. Available at: https://doi.org/10.1093/infdis/171.4.1026.

Fan, C.-C. *et al.* (2013) 'Expression of E-cadherin, Twist, and p53 and their prognostic value in patients with oral squamous cell carcinoma', *Journal of Cancer Research and Clinical Oncology*, 139(10), pp. 1735–1744. Available at: https://doi.org/10.1007/s00432-013-1499-9.

Farah, C.S. (2021) 'Molecular landscape of head and neck cancer and implications for therapy', *Annals of Translational Medicine*, 9(10), pp. 915–915. Available at: https://doi.org/10.21037/atm-20-6264.

Faulkner Valle, G. and Banks, L. (1995) 'The human papillomavirus (HPV)-6 and HPV-16 E5 proteins co-operate with HPV-16 E7 in the transformation of primary rodent cells', *Journal of General Virology*, 76(5), pp. 1239–1245. Available at: https://doi.org/10.1099/0022-1317-76-5-1239.

Florin, L. *et al.* (2002) 'Assembly and Translocation of Papillomavirus Capsid Proteins', *Journal of Virology*, 76(19), pp. 10009–10014. Available at: https://doi.org/10.1128/JVI.76.19.10009-10014.2002.

Forbes, D. (2007) *Euroguide: on the accommodation and care of animals used for experimental and other scientific purposes; based of the revised appendix A of the European Copnvention ETS 123*. Edited by Federation of European Laboratory Animal Science Associations. London: Felasa.

Gillison, M.L. *et al.* (2000) 'Evidence for a causal association between human papillomavirus and a subset of head and neck cancers.', *Journal of the National Cancer Institute*, 92(9), pp. 709–20. Available at: https://doi.org/10.1093/jnci/92.9.709.

Gillison, M.L. *et al.* (2008) 'Distinct risk factor profiles for human papillomavirus type 16-positive and human papillomavirus type 16-negative head and neck cancers', *Journal of the National Cancer Institute*, 100(6), pp. 407–420. Available at: https://doi.org/10.1093/jnci/djn025.

Gillison, M.L. *et al.* (2019) 'Radiotherapy plus cetuximab or cisplatin in human papillomavirus-positive oropharyngeal cancer (NRG Oncology RTOG 1016): a randomised, multicentre, non-inferiority trial', *The Lancet*, 393(10166), pp. 40–50. Available at: https://doi.org/10.1016/S0140-6736(18)32779-X.

Gillison, M.L. and Shah, K.V. (2003) 'Chapter 9: Role of mucosal human papillomavirus in nongenital cancers.', *Journal of the National Cancer Institute. Monographs*, 2003(31), pp. 57–65. Available at: https://doi.org/10.1093/oxfordjournals.jncimonographs.a003484.

Giroglou, T. *et al.* (2001) 'Human Papillomavirus Infection Requires Cell Surface Heparan Sulfate', *Journal of Virology*, 75(3), pp. 1565–1570. Available at: https://doi.org/10.1128/jvi.75.3.1565-1570.2001.

Global cancer Observatory (no date) 'Global Cancer Observatory: Cancer Today.', *International Agency for Research on Cancer*. Available at: https://gco.iarc.fr/today (Accessed: 17 June 2022).

Graves, C.A. *et al.* (2014) 'The translational significance of epithelial-mesenchymal transition in head and neck cancer', *Clinical and Translational Medicine*, 3(1), p. 60.

Available at: https://doi.org/10.1186/s40169-014-0039-9.

Hadami, K. *et al.* (2021) 'Degradation of p53 by HPV16-E6 variants isolated from cervical cancer specimens of Moroccan women.', *Gene*, 791, p. 145709. Available at: https://doi.org/10.1016/j.gene.2021.145709.

Ham, J. *et al.* (1991) 'The papillomavirus E2 protein: a factor with many talents', *Trends in Biochemical Sciences*, 16(November), pp. 440–444. Available at: https://doi.org/10.1016/0968-0004(91)90172-R.

Hammarstedt, L. *et al.* (2006) 'Human papillomavirus as a risk factor for the increase in incidence of tonsillar cancer', *International Journal of Cancer*, 119(11), pp. 2620–2623. Available at: https://doi.org/10.1002/ijc.22177.

Hassani, S. *et al.* (2015) 'Molecular Pathogenesis of Human Papillomavirus Type 16 in Tonsillar Squamous Cell Carcinoma', *Anticancer research*, 35(12), pp. 6633–6638.

Hatakeyama, H. *et al.* (2014) 'Epithelial-mesenchymal transition in human papillomavirus-positive and -negative oropharyngeal squamous cell carcinoma', *Oncology Reports*, 32(6), pp. 2673–2679. Available at: https://doi.org/10.3892/or.2014.3509.

zur Hausen, H. (1996) 'Papillomavirus infections — a major cause of human cancers', *Biochimica et Biophysica Acta (BBA) - Reviews on Cancer*, 1288(2), pp. F55–F78. Available at: https://doi.org/10.1016/0304-419X(96)00020-0.

Hawthorn, R.J.S. *et al.* (1988) 'Langerhans' cells and subtypes of human papillomavirus in cervical intraepithelial neoplasia.', *BMJ*, 297(6649), pp. 643–646. Available at: https://doi.org/10.1136/bmj.297.6649.643.

van der Heijden, M. *et al.* (2020) 'Epithelial-to-mesenchymal transition is a prognostic marker for patient outcome in advanced stage HNSCC patients treated with chemoradiotherapy', *Radiotherapy and Oncology: Journal of the European Society for Therapeutic Radiology and Oncology*, 147, pp. 186–194. Available at: https://doi.org/10.1016/j.radonc.2020.05.013.

Henderson, S. *et al.* (2014) 'APOBEC-Mediated Cytosine Deamination Links PIK3CA Helical Domain Mutations to Human Papillomavirus-Driven Tumor Development', *Cell Reports*, 7(6), pp. 1833–1841. Available at: https://doi.org/10.1016/j.celrep.2014.05.012.

Herrero, R. *et al.* (2000) 'Population-Based Study of Human Papillomavirus Infection and Cervical Neoplasia in Rural Costa Rica', *JNCI: Journal of the National Cancer Institute*, 92(6), pp. 464–474. Available at: https://doi.org/10.1093/jnci/92.6.464.

Hirose, Y. *et al.* (2018) 'Within-Host Variations of Human Papillomavirus Reveal APOBEC Signature Mutagenesis in the Viral Genome', *Journal of Virology*, 92(12), pp. e00017-18. Available at: https://doi.org/10.1128/jvi.00017-18.

Ho, G.Y.F. *et al.* (1998) 'Natural History of Cervicovaginal Papillomavirus Infection in Young Women', *New England Journal of Medicine*, 338(7), pp. 423–428. Available at: https://doi.org/10.1056/nejm199802123380703.

Ho, L. *et al.* (1991) 'Sequence variants of human papillomavirus type 16 in clinical samples permit verification and extension of epidemiological studies and construction of a phylogenetic tree', *Journal of Clinical Microbiology*, 29(9), pp. 1765–1772. Available at: https://doi.org/10.1128/jcm.29.9.1765-1772.1991.

Ho, L. *et al.* (1993) 'The genetic drift of human papillomavirus type 16 is a means of

reconstructing prehistoric viral spread and the movement of ancient human populations', *Journal of Virology*, 67(11), pp. 6413–6423. Available at: https://doi.org/10.1128/jvi.67.11.6413-6423.1993.

Hochmann, J. *et al.* (2016) 'The Asian-American variant of human papillomavirus type 16 exhibits higher activation of MAPK and PI3K/AKT signaling pathways, transformation, migration and invasion of primary human keratinocytes', *Virology*, 492, pp. 145–154. Available at: https://doi.org/10.1016/j.virol.2016.02.015.

Hocking, J.S. *et al.* (2011) 'Head and neck cancer in Australia between 1982 and 2005 show increasing incidence of potentially HPV-associated oropharyngeal cancers', *British Journal of Cancer*, 104(5), pp. 886–891. Available at: https://doi.org/10.1038/sj.bjc.6606091.

Hoffmann, M. *et al.* (2004) 'Human papillomavirus type 16 E6 and E7 genotypes in head-and-neck carcinomas', *Oral Oncology*, 40(5), pp. 520–524. Available at: https://doi.org/10.1016/j.oraloncology.2003.10.011.

Hoffmann, M. and Tribius, S. (2019) 'HPV and Oropharyngeal Cancer in the Eighth Edition of the TNM Classification: Pitfalls in Practice', *Translational Oncology*, 12(8), pp. 1108–1112. Available at: https://doi.org/10.1016/j.tranon.2019.05.009.

Hubert, P. *et al.* (1999) 'Colonization of in Vitro-Formed Cervical Human Papillomavirus-Associated (Pre)Neoplastic Lesions with Dendritic Cells', *The American Journal of Pathology*, 154(3), pp. 775–784. Available at: https://doi.org/10.1016/S0002-9440(10)65324-2.

Huertas-Salgado, A. *et al.* (2011) 'E6 molecular variants of human papillomavirus (HPV) type 16: An updated and unified criterion for clustering and nomenclature', *Virology*, 410(1), pp. 201–215. Available at: https://doi.org/10.1016/j.virol.2010.10.039.

Hwang, T.Z. *et al.* (2015) 'Incidence trends of human papillomavirus-related head and neck cancer in Taiwan, 1995-2009', *International Journal of Cancer*, 137(2), pp. 395–408. Available at: https://doi.org/10.1002/ijc.29330.

Ihler, F. *et al.* (2018) 'Epithelial-Mesenchymal Transition during Metastasis of HPV-Negative Pharyngeal Squamous Cell Carcinoma', *BioMed Research International*, 2018, p. 7929104. Available at: https://doi.org/10.1155/2018/7929104.

Jabbar, S. *et al.* (2010) 'Human papillomavirus type 16 E6 and E7 oncoproteins act synergistically to cause head and neck cancer in mice.', *Virology*, 407(1), pp. 60–7. Available at: https://doi.org/10.1016/j.virol.2010.08.003.

Jackson, R. *et al.* (2014) 'Tumourigenesis driven by the human papillomavirus type 16 Asian-American E6 variant in a three-dimensional keratinocyte model', *PLoS ONE*. Edited by R.D. Burk, 9(7), p. e101540. Available at: https://doi.org/10.1371/journal.pone.0101540.

Jackson, R. *et al.* (2016) 'Functional variants of human papillomavirus type 16 demonstrate host genome integration and transcriptional alterations corresponding to their unique cancer epidemiology', *BMC Genomics*, 17(1), p. 851. Available at: https://doi.org/10.1186/s12864-016-3203-3.

Jennings, L.J. *et al.* (2017) 'Guidelines for Validation of Next-Generation Sequencing–Based Oncology Panels: A Joint Consensus Recommendation of the Association for Molecular Pathology and College of American Pathologists', *Journal of Molecular Diagnostics*, 19(3), pp. 341–365. Available at: https://doi.org/10.1016/j.jmoldx.2017.01.011.

Jimenez, L. *et al.* (2015) 'Mechanisms of Invasion in Head and Neck Cancer', *Archives of Pathology & Laboratory Medicine*, 139(11), pp. 1334–1348. Available at:

https://doi.org/10.5858/arpa.2014-0498-RA.

Johnson, D.E. *et al.* (2020) 'Head and neck squamous cell carcinoma', *Nature Reviews Disease Primers*, 6(1), p. 92. Available at: https://doi.org/10.1038/s41572-020-00224-3.

Joseph, A.W. *et al.* (2013) 'Molecular etiology of second primary tumors in contralateral tonsils of human papillomavirus-associated index tonsillar carcinomas', *Oral Oncology*, 49(3), pp. 244–248. Available at: https://doi.org/10.1016/j.oraloncology.2012.09.009.

Kalluri, R. and Weinberg, R.A. (2009) 'The basics of epithelial-mesenchymal transition', *The Journal of Clinical Investigation*, 119(6), pp. 1420–1428. Available at: https://doi.org/10.1172/JCI39104.

Katabi, N. and Lewis, J.S. (2017) 'Update from the 4th Edition of the World Health Organization Classification of Head and Neck Tumours: What Is New in the 2017 WHO Blue Book for Tumors and Tumor-Like Lesions of the Neck and Lymph Nodes.', *Head and neck pathology*, 11(1), pp. 48–54. Available at: https://doi.org/10.1007/s12105-017-0796-z.

Kawana, K. *et al.* (1999) 'Common Neutralization Epitope in Minor Capsid Protein L2 of Human Papillomavirus Types 16 and 6', *Journal of Virology*, 73(7), pp. 6188–6190. Available at: https://doi.org/10.1128/jvi.73.7.6188-6190.1999.

Koutsky, L. (1991) 'Role of epidemiology in defining events that influence transmission and natural history of anogenital papillomavirus infections', *Journal of the National Cancer Institute*, 83(14), pp. 978–979. Available at: https://doi.org/10.1093/jnci/83.14.978.

Kreimer, A.R. *et al.* (2020) 'Summary from an international cancer seminar focused on human papillomavirus (HPV)-positive oropharynx cancer, convened by scientists at IARC and NCI', *Oral Oncology*, 108(June), p. 104736. Available at: https://doi.org/10.1016/j.oraloncology.2020.104736.

Lang Kuhs, K.A. *et al.* (2022) 'Genetic variation within the human papillomavirus type 16 genome is associated with oropharyngeal cancer prognosis', *Annals of Oncology* [Preprint]. Available at: https://doi.org/10.1016/j.annonc.2022.03.005.

LeConte, B.A. *et al.* (2018) 'Differences in the viral genome between HPV-positive cervical and oropharyngeal cancer', *PLoS ONE*. Edited by M.L. Tornesello, 13(8), p. e0203403. Available at: https://doi.org/10.1371/journal.pone.0203403.

Lee, M.J. *et al.* (2020) 'Alterations and molecular targeting of the GSK-3 regulator, PI3K, in head and neck cancer.', *Biochimica et Biophysica Acta - Molecular Cell Research*, 1867(6), p. 118679. Available at: https://doi.org/10.1016/j.bbamcr.2020.118679.

Lefevre, M. *et al.* (2017) 'Epithelial to mesenchymal transition and HPV infection in squamous cell oropharyngeal carcinomas: the papillophar study', *British Journal of Cancer*, 116(3), pp. 362–369. Available at: https://doi.org/10.1038/bjc.2016.434.

Liu, F. *et al.* (2016) 'ERK signaling pathway is involved in HPV-16 E6 but not E7 oncoprotein-induced HIF-1α protein accumulation in NSCLC cells', *Oncology Research*, 23(3), pp. 109–118. Available at: https://doi.org/10.3727/096504015X14496932933610.

Longworth, M.S. and Laimins, L.A. (2004) 'Pathogenesis of Human Papillomaviruses in Differentiating Epithelia', *Microbiology and Molecular Biology Reviews*, 68(2), pp. 362–372. Available at: https://doi.org/10.1128/MMBR.68.2.362-372.2004.

Loyo, M. *et al.* (2013) 'Lessons learned from next-generation sequencing in head and neck cancer', *Head & Neck*. Edited by J.N. Myers, 35(3), pp. 454–463. Available at:

https://doi.org/10.1002/hed.23100.

Machiels, J.P. *et al.* (2020) 'Squamous cell carcinoma of the oral cavity, larynx, oropharynx and hypopharynx: EHNS-ESMO-ESTRO Clinical Practice Guidelines for diagnosis, treatment and follow-up'', *Oral Oncology*, (xxxx), p. 105042. Available at: https://doi.org/10.1016/j.oraloncology.2020.105042.

Mandal, P. *et al.* (2022) 'Predominance of genomically defined A lineage of HPV16 over D lineage in Indian patients from eastern India with squamous cell carcinoma of the cervix in association with distinct oncogenic phenotypes', *Translational Oncology*, 15(1), p. 101256. Available at: https://doi.org/10.1016/j.tranon.2021.101256.

Mandelker, D. *et al.* (2009) 'A frequent kinase domain mutation that changes the interaction between PI3Kalpha and the membrane.', *Proceedings of the National Academy of Sciences of the United States of America*, 106(40), pp. 16996–7001. Available at: https://doi.org/10.1073/pnas.0908444106.

Marquard, F.E. and Jücker, M. (2020) 'PI3K/AKT/mTOR signaling as a molecular target in head and neck cancer', *Biochemical Pharmacology*, 172, p. 113729. Available at: https://doi.org/10.1016/j.bcp.2019.113729.

McDermott, J.D. and Bowles, D.W. (2019) 'Epidemiology of Head and Neck Squamous Cell Carcinomas: Impact on Staging and Prevention Strategies', *Current Treatment Options in Oncology*, 20(5), pp. 1–13. Available at: https://doi.org/10.1007/s11864-019-0650-5.

Medda, A., Duca, D. and Chiocca, S. (2021) 'Human papillomavirus and cellular pathways: Hits and targets', *Pathogens*, 10(3), pp. 1–23. Available at: https://doi.org/10.3390/pathogens10030262.

Mehanna, H. *et al.* (2019) 'Radiotherapy plus cisplatin or cetuximab in low-risk human papillomavirus-positive oropharyngeal cancer (De-ESCALaTE HPV): an open-label randomised controlled phase 3 trial.', *Lancet (London, England)*, 393(10166), pp. 51–60. Available at: https://doi.org/10.1016/S0140-6736(18)32752-1.

Merle, E. *et al.* (1999) 'Nuclear import of HPV11 L1 capsid protein is mediated by karyopherin alpha2beta1 heterodimers.', *Journal of cellular biochemistry*, 74(4), pp. 628–37.

Mestre, V.F. *et al.* (2020) 'HPV16 is sufficient to induce squamous cell carcinoma specifically in the tongue base in transgenic mice', *Journal of Pathology*, 251(1), pp. 4–11. Available at: https://doi.org/10.1002/path.5387.

Miao, D. *et al.* (2018) 'Genomic correlates of response to immune checkpoint blockade in microsatellite-stable solid tumors', *Nature Genetics*, 50(9), pp. 1271–1281. Available at: https://doi.org/10.1038/s41588-018-0200-2.

Miled, N. *et al.* (2007) 'Mechanism of two classes of cancer mutations in the phosphoinositide 3-kinase catalytic subunit', *Science*, 317(5835), pp. 239–242. Available at: https://doi.org/10.1126/science.1135394.

Mirabello, L. *et al.* (2016) 'HPV16 Sublineage Associations with Histology-Specific Cancer Risk Using HPV Whole-Genome Sequences in 3200 Women', *Journal of the National Cancer Institute*, 108(9), p. djw100. Available at: https://doi.org/10.1093/jnci/djw100.

Mirabello, L. *et al.* (2018) 'The intersection of HPV epidemiology, genomics and mechanistic studies of HPV-mediated carcinogenesis', *Viruses*, 10(2), p. 80. Available at: https://doi.org/10.3390/v10020080.

Miranda, A. *et al.* (2018) *Registo Oncológico Nacional de Todos os Tumores na População Residente em Portugal*. Joana Ferr. Edited by RON. Instituto Português de Oncologia de Lisboa de Francisco Gentil - E. P. E.

Mirantes, C. *et al.* (2013) 'Epithelial-to-mesenchymal transition and stem cells in endometrial cancer', *Human Pathology*, 44(10), pp. 1973–1981. Available at: https://doi.org/10.1016/j.humpath.2013.04.009.

Mitra, A., Mishra, L. and Li, S. (2015) 'EMT, CTCs and CSCs in tumor relapse and drug-resistance', *Oncotarget*, 6(13), pp. 10697–10711. Available at: https://doi.org/10.18632/oncotarget.4037.

Mork, J. *et al.* (2010) 'Time trends in pharyngeal cancer incidence in Norway 1981-2005: A subsite analysis based on a reabstraction and recoding of registered cases', *Cancer Causes and Control*, 21(9), pp. 1397–1405. Available at: https://doi.org/10.1007/s10552-010-9567-9.

Moscicki, A.-B. *et al.* (1998) 'The natural history of human papillomavirus infection as measured by repeated DNA testing in adolescent and young women', *The Journal of Pediatrics*, 132(2), pp. 277–284. Available at: https://doi.org/10.1016/S0022-3476(98)70445-7.

Münger, K. *et al.* (2004) 'Mechanisms of Human Papillomavirus-Induced Oncogenesis', *Journal of Virology*, 78(21), pp. 11451–11460. Available at: https://doi.org/10.1128/jvi.78.21.11451-11460.2004.

Muñoz, N. *et al.* (2003) 'Epidemiologic classification of human papillomavirus types associated with cervical cancer.', *The New England journal of medicine*, 348(6), pp. 518–27. Available at: https://doi.org/10.1056/nejmoa021641.

Nacher, M. *et al.* (2020) 'E6/E7 Sequence Diversity of High-Risk Human Papillomaviruses in Two Geographically Isolated Populations of French Guiana', *Microorganisms*, 8(11), p. 1842. Available at: https://doi.org/10.3390/microorganisms8111842.

Nelson, L.M., Rose, R.C. and Moroianu, J. (2002) 'Nuclear import strategies of high risk HPV16 L1 major capsid protein', *Journal of Biological Chemistry*, 277(26), pp. 23958–23964. Available at: https://doi.org/10.1074/jbc.M200724200.

Neto, C.P. de O. *et al.* (2021) 'Differential incidence of tongue base cancer in male and female hpv16-transgenic mice: role of female sex hormone receptors', *Pathogens*, 10(10), pp. 1–14. Available at: https://doi.org/10.3390/pathogens10101224.

Nulton, T.J. *et al.* (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(11), pp. 17684–17699. Available at: https://doi.org/10.18632/oncotarget.15179.

Oriel, J.D. (1971) 'Natural history of genital warts.', *Sexually Transmitted Infections*, 47(1), pp. 1–13. Available at: https://doi.org/10.1136/sti.47.1.1.

Oyervides-Muñoz, M.A. *et al.* (2018) 'Understanding the HPV integration and its progression to cervical cancer', *Infection, Genetics and Evolution*, 61(February), pp. 134–144. Available at: https://doi.org/10.1016/j.meegid.2018.03.003.

Ozbun, M.A. (2002) 'Human Papillomavirus Type 31b Infection of Human Keratinocytes and the Onset of Early Transcription', *Journal of Virology*, 76(22), pp. 11291–11300. Available at: https://doi.org/10.1128/jvi.76.22.11291-11300.2002.

Pakdel, F. *et al.* (2021) 'The frequency of high-risk human papillomavirus types, HPV16 lineages, and their relationship with p16INK4a and NF-κB expression in head and neck squamous cell carcinomas in Southwestern Iran.', *Brazilian journal of microbiology : [publication of the Brazilian Society for Microbiology]*, 52(1), pp. 195–206. Available at: https://doi.org/10.1007/s42770-020-00391-1.

Park, J.S. *et al.* (2016) 'Association of human papillomavirus type 16 and its genetic variants with cervical lesion in Korea', *APMIS*, 124(11), pp. 950–957. Available at: https://doi.org/10.1111/apm.12592.

Patterson, N.A., Smith, J.L. and Ozbun, M.A. (2005) 'Human Papillomavirus Type 31b Infection of Human Keratinocytes Does Not Require Heparan Sulfate', *Journal of Virology*, 79(11), pp. 6838–6847. Available at: https://doi.org/10.1128/jvi.79.11.6838-6847.2005.

Payan, C. *et al.* (2007) 'Human Papillomavirus Quantification in Urine and Cervical Samples by Using the Mx4000 and LightCycler General Real-Time PCR Systems', *Journal of Clinical Microbiology*, 45(3), pp. 897–901. Available at: https://doi.org/10.1128/JCM.02022-06.

Pim, D. *et al.* (2005) 'Activation of the protein kinase B pathway by the HPV-16 E7 oncoprotein occurs through a mechanism involving interaction with PP2A', *Oncogene*, 24(53), pp. 7830–7838. Available at: https://doi.org/10.1038/sj.onc.1208935.

Piyathilake, C.J. *et al.* (2020) 'A rigorous exploration of anal HPV genotypes using a next-generation sequencing (NGS) approach in HIV-infected men who have sex with men at risk for developing anal cancer', *Cancer Medicine*, 9(2), pp. 807–815. Available at: https://doi.org/10.1002/cam4.2720.

Polanska, H. *et al.* (2014) 'Clinical significance of head and neck squamous cell cancer biomarkers', *Oral Oncology*, 50(3), pp. 168–177. Available at: https://doi.org/10.1016/j.oraloncology.2013.12.008.

Porta, C., Paglino, C. and Mosca, A. (2014) 'Targeting PI3K/Akt/mTOR signaling in cancer', *Frontiers in Oncology*, 4 APR(April), pp. 1–11. Available at: https://doi.org/10.3389/fonc.2014.00064.

Qin, T. *et al.* (2021) 'Molecular tumor subtypes of hpv-positive head and neck cancers: Biological characteristics and implications for clinical outcomes', *Cancers*, 13(11), pp. 1–21. Available at: https://doi.org/10.3390/cancers13112721.

Raj, K. *et al.* (2004) 'E1 ∧ E4 Protein of Human Papillomavirus Type 16 Associates with Mitochondria', *Journal of Virology*, 78(13), pp. 7199–7207. Available at: https://doi.org/10.1128/jvi.78.13.7199-7207.2004.

Raudenská, M., Balvan, J. and Masařík, M. (2021) 'Cell death in head and neck cancer pathogenesis and treatment.', *Cell death & disease*, 12(2), p. 192. Available at: https://doi.org/10.1038/s41419-021-03474-5.

Registo Oncológico Regional do Norte (2010) *Registo Oncológico Nacional 2010*. Edited by Instituto Português de Oncologia do Porto Francisco Gentil - EPE.

Rettig, E.M. and D'Souza, G. (2015) 'Epidemiology of Head and Neck Cancer', *Surgical Oncology Clinics of North America*, 24(3), pp. 379–396. Available at: https://doi.org/10.1016/j.soc.2015.03.001.

Rothenberg, S.M. and Ellisen, L.W. (2012) 'The molecular pathogenesis of head and neck squamous cell carcinoma', *Journal of Clinical Investigation*, 122(6), pp. 1951–1957. Available

at: https://doi.org/10.1172/JCI59889.

Roy, S. *et al.* (2018) 'Standards and Guidelines for Validating Next-Generation Sequencing Bioinformatics Pipelines: A Joint Recommendation of the Association for Molecular Pathology and the College of American Pathologists', *Journal of Molecular Diagnostics*, 20(1), pp. 4–27. Available at: https://doi.org/10.1016/j.jmoldx.2017.11.003.

Ruesch, M.N., Stubenrauch, F. and Laimins, L.A. (1998) 'Activation of Papillomavirus Late Gene Transcription and Genome Amplification upon Differentiation in Semisolid Medium Is Coincident with Expression of Involucrin and Transglutaminase but Not Keratin-10', *Journal of Virology*, 72(6), pp. 5016–5024. Available at: https://doi.org/10.1128/jvi.72.6.5016-5024.1998.

Rusan, M., Li, Y.Y. and Hammerman, P.S. (2015) 'Genomic landscape of human papillomavirus-associated cancers', *Clinical Cancer Research*, 21(9), pp. 2009–2019. Available at: https://doi.org/10.1158/1078-0432.CCR-14-1101.

Sabatini, M.E. and Chiocca, S. (2020) 'Human papillomavirus as a driver of head and neck cancers', *British Journal of Cancer*, 122(3), pp. 306–314. Available at: https://doi.org/10.1038/s41416-019-0602-7.

Santos, C. *et al.* (2017) 'HPV-transgenic mouse models: Tools for studying the cancer-associated immune response', *Virus Research*, 235, pp. 49–57. Available at: https://doi.org/10.1016/j.virusres.2017.04.001.

Sewell, A. *et al.* (2014) 'Reverse-phase protein array profiling of oropharyngeal cancer and significance of PIK3CA mutations in HPV-associated head and neck cancer.', *Clinical cancer research : an official journal of the American Association for Cancer Research*, 20(9), pp. 2300–11. Available at: https://doi.org/10.1158/1078-0432.CCR-13-2585.

Shafti-Keramat, S. *et al.* (2003) 'Different Heparan Sulfate Proteoglycans Serve asCellular Receptors for HumanPapillomaviruses', *Journal of Virology*, 77(24), pp. 13125–13135. Available at: https://doi.org/10.1128/jvi.77.24.13125-13135.2003.

Shaw, R.J. and Cantley, L.C. (2006) 'Ras, PI(3)K and mTOR signalling controls tumour cell growth.', *Nature*, 441(7092), pp. 424–30. Available at: https://doi.org/10.1038/nature04869.

Sichero, L. *et al.* (2007) 'High grade cervical lesions are caused preferentially by non-European variants of HPVs 16 and 18', *International Journal of Cancer*, 120(8), pp. 1763–1768. Available at: https://doi.org/10.1002/ijc.22481.

Smith, A., Teknos, T.N. and Pan, Q. (2013) 'Epithelial to mesenchymal transition in head and neck squamous cell carcinoma', *Oral Oncology*, 49(4), pp. 287–292. Available at: https://doi.org/10.1016/j.oraloncology.2012.10.009.

Smith, E.M. *et al.* (2010) 'Human papillomavirus, p16 and p53 expression associated with survival of head and neck cancer', *Infectious Agents and Cancer*, 5(1), p. 4. Available at: https://doi.org/10.1186/1750-9378-5-4.

Smith-McCune, K. *et al.* (1997) 'Cross-species comparison of angiogenesis during the premalignant stages of squamous carcinogenesis in the human cervix and K14-HPV16 transgenic mice', *Cancer Research*, 57(7), pp. 1294–1300.

Smotkin, D. and Wettstein, F.O. (1986) 'Transcription of human papillomavirus type 16 early genes in a cervical cancer and a cancer-derived cell line and identification of the E7 protein', *Proceedings of the National Academy of Sciences of the United States of America*, 83(13), pp. 4680–4684. Available at: https://doi.org/10.1073/pnas.83.13.4680.

Stransky, N. *et al.* (2011) 'The mutational landscape of head and neck squamous cell carcinoma', *Science*, 333(6046), pp. 1157–1160. Available at: https://doi.org/10.1126/science.1208130.

Strati, K. and Lambert, P.F. (2007) 'Role of Rb-Dependent and Rb-Independent Functions of Papillomavirus E7 Oncogene in Head and Neck Cancer', *Cancer Research*. Edited by A.E. Nixon, 67(24), pp. 11585–11593. Available at: https://doi.org/10.1158/0008-5472.CAN-07-3007.

Taberna, M. *et al.* (2017) 'Human papillomavirus-related oropharyngeal cancer', *Annals of Oncology*, 28(10), pp. 2386–2398. Available at: https://doi.org/10.1093/annonc/mdx304.

Talamini, R. *et al.* (2002) 'Combined effect of tobacco and alcohol on laryngeal cancer risk: a case-control study.', *Cancer causes & control : CCC*, 13(10), pp. 957–64. Available at: https://doi.org/10.1023/a:1021944123914.

Tamura, K., Nei, M. and Kumar, S. (2004) 'Prospects for inferring very large phylogenies by using the neighbor-joining method', *Proceedings of the National Academy of Sciences of the United States of America*, 101(30), pp. 11030–11035. Available at: https://doi.org/10.1073/pnas.0404206101.

Thiery, J.P. *et al.* (2009) 'Epithelial-mesenchymal transitions in development and disease', *Cell*, 139(5), pp. 871–890. Available at: https://doi.org/10.1016/j.cell.2009.11.007.

Tian, T., Li, X. and Zhang, J. (2019) 'mTOR Signaling in Cancer and mTOR Inhibitors in Solid Tumor Targeting Therapy', *International Journal of Molecular Sciences*, 20(3), p. 755. Available at: https://doi.org/10.3390/ijms20030755.

Tornesello, M.L. *et al.* (2004) 'Analysis of human papillomavirus type-16 variants in Italian women with cervical intraepithelial neoplasia and cervical cancer', *Journal of Medical Virology*, 74(1), pp. 117–126. Available at: https://doi.org/10.1002/jmv.20154.

Van Doorslaer, K. (2013) 'Evolution of the Papillomaviridae', *Virology*, 445(1–2), pp. 11–20. Available at: https://doi.org/10.1016/j.virol.2013.05.012.

de Villiers, E.M. and Gunst, K. (2009) 'Characterization of seven novel human papillomavirus types isolated from cutaneous tissue, but also present in mucosal lesions.', *The Journal of general virology*, 90(Pt 8), pp. 1999–2004. Available at: https://doi.org/10.1099/vir.0.011478-0.

Wadhwa, B. *et al.* (2017) 'Protein kinase B: emerging mechanisms of isoform-specific regulation of cellular signaling in cancer', *Anti-Cancer Drugs*, 28(6), pp. 569–580. Available at: https://doi.org/10.1097/CAD.0000000000000496.

Wang, L. *et al.* (2021) 'Analysis of genetic variation in human papillomavirus type 16 E1 and E2 in women with cervical infection in Xinjiang, China', *BMC medical genomics*, 14(1), p. 268. Available at: https://doi.org/10.1186/s12920-021-01120-9.

Wang, Z. *et al.* (2017) 'mTOR co-targeting strategies for head and neck cancer therapy', *Cancer and Metastasis Reviews*, 36(3), pp. 491–502. Available at: https://doi.org/10.1007/s10555-017-9688-7.

Wee, P. and Wang, Z. (2017) 'Epidermal growth factor receptor cell proliferation signaling pathways', *Cancers*, 9(5), pp. 1–45. Available at: https://doi.org/10.3390/cancers9050052.

Wilson, V.G. *et al.* (2002) 'Papillomavirus E1 proteins: form, function, and features.', *Virus genes*, 24(3), pp. 275–90. Available at: https://doi.org/10.1023/a:1015336817836.

World Health Organization (2010) 'Human papillomavirus laboratory manual', *WHO Press*. Available at: file:///Users/danielacochicho/Downloads/WHO_IVB_10.12_eng.pdf%3Bjsessionid=22C92FCC877957A3F66AB9C0533BEF50.pdf.

Yamada, T. *et al.* (1995) 'Human papillomavirus type 16 variant lineages in United States populations characterized by nucleotide sequence analysis of the E6, L2, and L1 coding segments.', *Journal of virology*, 69(12), pp. 7743–53. Available at: https://doi.org/10.1128/JVI.69.12.7743-7753.1995.

Yamada, T. *et al.* (1997) 'Human papillomavirus type 16 sequence variation in cervical cancers: a worldwide perspective.', *Journal of virology*, 71(3), pp. 2463–72. Available at: https://doi.org/10.1128/JVI.71.3.2463-2472.1997.

Yang, J.T., Liu, C.Z. and Iannaccone, P. (1995) 'The HPV 16 genome induces carcinomas and T-cell lymphomas in transgenic mice', *American Journal of Pathology*, 147(1), pp. 68–78.

Yang, R. *et al.* (2003) 'Interaction of L2 with β-actin directs intracellular transport of papillomavirus and infection', *Journal of Biological Chemistry*, 278(14), pp. 12546–12553. Available at: https://doi.org/10.1074/jbc.M208691200.

Yang, Y., Ren, J. and Zhang, Q. (2016) 'Distribution of human papilloma virus type 16 E6/E7 gene mutation in cervical precancer or cancer: A case control study in Guizhou Province, China', *Journal of Medical Virology*, 88(2), pp. 345–350. Available at: https://doi.org/10.1002/jmv.24333.

Yao, Y.Y. *et al.* (2019) 'Human Papillomavirus Type 16 E1 Mutations Associated with Cervical Cancer in a Han Chinese Population', *International Journal of Medical Sciences*, 16(7), pp. 1042–1049. Available at: https://doi.org/10.7150/ijms.34279.

You, J. *et al.* (2004) 'Interaction of the bovine papillomavirus E2 protein with Brd4 tethers the viral DNA to host mitotic chromosomes', *Cell*, 117(3), pp. 349–360. Available at: https://doi.org/10.1016/S0092-8674(04)00402-7.

Yu, J.S.L. and Cui, W. (2016) 'Proliferation, survival and metabolism: The role of PI3K/AKT/mTOR signalling in pluripotency and cell fate determination', *Development (Cambridge)*, 143(17), pp. 3050–3060. Available at: https://doi.org/10.1242/dev.137075.

Zhang, L. *et al.* (2015) 'Variants of human papillomavirus type 16 predispose toward persistent infection', *International Journal of Clinical and Experimental Pathology*, 8(7), pp. 8453–8459.

Zhang, Y. *et al.* (2016) 'Subtypes of HPV-Positive Head and Neck Cancers Are Associated with HPV Characteristics, Copy Number Alterations, PIK3CA Mutation, and Pathway Signatures', *Clinical Cancer Research*, 22(18), pp. 4735–4745. Available at: https://doi.org/10.1158/1078-0432.CCR-16-0323.

Zhang, Y. *et al.* (2017) 'A Pan-Cancer Proteogenomic Atlas of PI3K/AKT/mTOR Pathway Alterations', *Cancer Cell*, 31(6), pp. 820-832.e3. Available at: https://doi.org/10.1016/j.ccell.2017.04.013.

Zhang, Z., Filho, M.S. and Nör, J.E. (2012) 'The biology of head and neck cancer stem cells', *Oral Oncology*, 48(1), pp. 1–9. Available at: https://doi.org/10.1016/j.oraloncology.2011.10.004.

Zhao, J. *et al.* (2020) 'Genetic variability and functional implication of HPV16 from cervical intraepithelial neoplasia in Shanghai women', *Journal of Medical Virology*, 92(3), pp. 372–381. Available at: https://doi.org/10.1002/jmv.25618.

Zhao, J.W. *et al.* (2016) 'HPV16 integration probably contributes to cervical oncogenesis through interrupting tumor suppressor genes and inducing chromosome instability', *Journal of Experimental and Clinical Cancer Research*, 35(1), pp. 1–14. Available at: https://doi.org/10.1186/s13046-016-0454-4.

Zheng, Z.M. and Baker, C.C. (2006) 'Papillomavirus genome structure, expression, and post-transcriptional regulation', *Frontiers in Bioscience*, 11(SUPPL. 1), pp. 2286–2302. Available at: https://doi.org/10.2741/1971.

Zhu, B. *et al.* (2020) 'Mutations in the HPV16 genome induced by APOBEC3 are associated with viral clearance.', *Nature communications*, 11(1), p. 886. Available at: https://doi.org/10.1038/s41467-020-14730-1.

Zi\c eba, S. *et al.* (2020) 'Genes, pathways and vulvar carcinoma - New insights from next-generation sequencing studies', *Gynecologic Oncology*, 158(2), pp. 498–506. Available at: https://doi.org/10.1016/j.ygyno.2020.05.034.

Zur Hausen, H. (2002) 'Papillomaviruses and cancer: From basic studies to clinical application', *Nature Reviews Cancer*, 2(5), pp. 342–350. Available at: https://doi.org/10.1038/nrc798.

SUPPLEMENTARY MATERIAL

SUPPLEMENT 1 LIST OF STANDARDIZATION DOCUMENTS TO NGS

Table S1. List of standardization documents relevant to NGS in accordance with European union guidelines updated in September 2021

European Guidelines	Quality Assurance	Number	Description	Specifications	Processes
Padrões para pré-análise de espécime/amostra	EN ISO	20166-1:2018	Molecular in vitro diagnostic examinations	pre-examination	FFPE tissue - Isolated RNA
	EN ISO	20166-3:2019	Molecular in vitro diagnostic examinations	pre-examination	FFPE tissue - Isolated DNA
	EN ISO	20184-1:2018	Molecular in vitro diagnostic examinations	pre-examination	frozen tissue — Isolated RNA
	EN ISO	20184-3:2021	Molecular in vitro diagnostic examinations	pre-examination	frozen tissue — Isolated DNA
	EN ISO	20186-1:2019	Molecular in vitro diagnostic examinations	pre-examination processes	venous whole blood — Isolated cellular RNA
Library preparation and NGS-analysis	ISO/DIS	20397-1	Biotechnology-General Requirements	Massive Parallel Sequencing	Nucleic acid and library preparation
	ISO	20688-1:2020	Biotechnology-General Requirements	Nucleic acid synthesis	production and quality control of synthesized
Standards for NGS-data	ISO	20397-2:2021	Biotechnology-General Requirements	Massively parallel sequencing	Methods to evaluate the quality of sequencing data
	ISO/TR	4018:41:00	Biotechnology-General Requirements	Health informatics	Data elements and metadata pipeline
	ISO/TS	22725:40:00	Biotechnology-General Requirements	Genomics Informatics	Quality control metrics for DNA sequencing
	ISO/TR	4018:41:00	Biotechnology-General Requirements	Development of International Standards	Data Publication Preliminary Considerations and Concepts

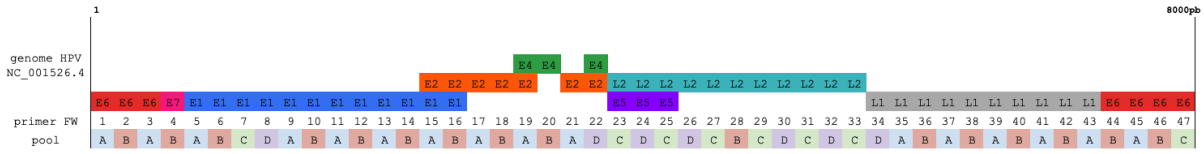
SUPPLEMENT 2 *IN SILICO* VALIDATION

Table S1. Primer's characteristics

Name	sequence 5'-3'	size (pb)	Melt Temp (°)	Region	Fragment size (pb)	Blast HPV16 001526.4	secondary amplification	
							other HPV (n)	other organisms (n)
HPV16_1_F	ACAGTTACTGCGACGTGAGC	20	57	E6	272	100%	0	0
HPV16_1_R	TGCAATCTTTGCTTTTTGTCC	21	51,7			100%		
HPV16_2_F	TCAAAAGCCACTGTGTCCTG	20	55,4	E6	252	100%	0	1
HPV16_2_R	TTCATCCTCCTCCTCTGAGC	20	55,9			100%		
HPV16_3_F	TTGCAACCAGAGACAACTGA	20	54,3	E7	248	100%	0	0
HPV16_3_R	TTCTGAGAACAGATCCCCAC	21	57,1			100%		
HPV16_4_F	GCGTACAAAGTACACACGTA	20	55,7	E7 E1	226	100%	0	20
HPV16_4_R	CTGTCATTTTGTGTGTCGTCA	21	52,9			20%		
HPV16_5_F	ACGGGATGTAATGGATGGTT	20	53,8	E1	198	100%	0	0
HPV16_5_R	TGTTGTTTTGTTTCTTGTGC	20	54,4			100%		
HPV16_6_F	TTTAACACAGCCAGAAACAGGAC	24	55,3	E1	276	100%	0	3
HPV16_6_R	CGCCCTTCTACCTGTAACATC	21	55,1			100%		
HPV16_7_F	GCTGCAAAAAGCAGATTATTTG	22	50,7	E1	260	100%	0	6
HPV16_7_R	ATTGCTGCCTTTGCATTACT	20	53,1			100%		
HPV16_8_F	TGCCAAACACCACTTACAAA	20	52,7	E1	120	100%	0	1
HPV16_8_R	CATTCCCCATCAACATGCTA	20	52,6			100%		
HPV16_8b_F	TGCCAAACACCACTTACAAA	20	52,8	E1	261	100%	0	0
HPV16_8b_R	CATTCCCCATCAACACGCTA	20	52,6			100%		
HPV16_9_F	GGTGTATTGCTGCATTTGGA	20	53,6	E1	234	100%	0	0
HPV16_9_R	TACGCAATTTTGCAACGCTCT	20	54,5			100%		
HPV16_10_F	TGTGTGTCTCCAATGTGTATGATG	24	55	E1	196	100%	0	0
HPV16_10_R	CCCATTGTACCATCGTGATAATTC	24	53,7			100%		
HPV16_11_F	ACACGTCAGAATGGATACAA	20	54,2	E1	223	100%	0	0
HPV16_11_R	CACATTGTTGTACAATCCTTT	21	52,2			100%		
HPV16_12_F	TGCACAATTGCCAGACACTA	20	54,7	E1	310	100%	0	0
HPV16_12_R	ACCTGTGTTAGCTGCACCAT	20	56,8			100%		
HPV16_13_F	AGATGTGATAGGGTAGATGATGTAGG	25	55,3	E1	320	100%	0	0
HPV16_13_R	TTGTCATCTATGTAGTTCCAACAGG	25	54,5			100%		
HPV16_14_F	GCAGATGCCAAAATAGGTATG	21	51,7			100%		
HPV16_14_R	ACTGCATTTTCCTTTTTGTC	20	51,1	E1	262	100%	0	2
HPV16_14b_R	ACTGCATTTTCCTTTTCGTC	20	51,1			100%		
HPV16_15_F	CATTTGTTGTTTTACATTTCC	22	52,2			100%		
HPV16_15_R	CATTCTAGGGCACATGTGTTT	20	53,5	E1 E2	266	100%	0	0
HPV16_15_R	CATTCTAGGGCACATTTGTTT	20	53,5			100%		
HPV16_16_F	AAAACGATGCAGACTCTTTGT	21	53,3	E1 E2	236	100%	0	0
HPV16_16_R	AGTTGTAGTTCAATTGCTTGT	21	53,5			100%		
HPV16_17_F	GTGCCAACACTGCCTGTATC	20	56,3	E2	205	100%	0	1
HPV16_17_R	TGCATATGTCTCCATCAAACTG	22	52,9			100%		
HPV16_18_F	AAACATGCATATACAGTGCAACTGC	25	55,2	E2	245	100%	0	1
HPV16_18_R	ATTACCTGACCACCCCATG	20	57,5			100%		
HPV16_19_F	TGCAGTTTAAAGATGTTGCAGA	22	52,6	E2 E4	214	100%	0	1
HPV16_19_R	CGCTGGATAGTCGTCTGTC	20	56,6			100%		
HPV16_20_F	TCTGTGTTTAGCAGCAACGA	20	56,5			100%		
HPV16_20_R	CAGTAGCGATTCGACATACTG	20	55,6	E4	217	100%	0	0
HPV16_20b_F	TCTGTGTTTAGCAGCAACGA	20	54,5			100%		
HPV16_21_F	ACCCCTGCCACACCAATAAG	20	57,5	E2	228	100%	0	0
HPV16_21_R	TATGTCCTGTCCAATGCCATG	21	55			100%		
HPV16_22_F	CCATAGTACATTTAAAAGGTGATGC	25	52,3	E4 E2	274	100%	0	0
HPV16_22_R	CGCCACGTAATGTTTCGATG	20	54,1			100%		
HPV16_23_F	TACACTTACAATATGATAGTGAATGTCAAGG	30	54,9	L2 E5	214	100%	0	0
HPV16_23_R	CGTATGTAGCACACAGCAAAACCAG	25	55,7			100%		
HPV16_24_F	GTGTGTTTTTGTGTGTCGCC	21	57,3	E5 L2	339	100%	0	0
HPV16_24_R	TGTGTGCATTGTTAAGTGATAAC	24	53,7			100%		
HPV16_25_F	GGATAACAGCGCCCTCTGC	19	58,6	E5 L2	294	100%	0	0
HPV16_25_R	AAAGTTGGGTAGCCGTGC	19	55,6			100%		
HPV16_26_F	GTTCTGCAAAACGCACAAA	19	52,4	L2	261	100%	0	1
HPV16_26_R	GGGGTCTTACAGGACTAAGT	20	56,6			100%		
HPV16_27_F	GTACAGGCGACGCACTG	18	58,8	L2	298	100%	0	2
HPV16_27_R	GGCATTATTATGTGTAGTAACAGTAGTAACAG	32	54,9			100%		
HPV16_28_F	TCAACTGATACCACACCTGCT	21	56	L2	247	100%	0	0
HPV16_28_R	CAGACCCTGCTATGCTGTG	20	57,2			100%		
HPV16_28b_R	CAGACCCCGCTATGCTGTG	20	57,2			100%		
HPV16_29_F	AGCACAAACCCTAACACAGTAACTAG	26	56,6	L2	295	100%	0	0
HPV16_29_R	TAAAATGTATTATCCACATCTATACCTTC	28	51,2			100%		
HPV16_29_F	AGCACAAATCCTAACACAGTAACTAG	26	56,6			100%		
HPV16_30_F	CACCCTGCTTTTGTAACCACTC	22	56	L2	181	100%	0	0
HPV16_30_R	GTACTGCTAGAGGTAATGCTGG	23	57,2			100%		
HPV16_31_F	GCTCCAGATCCTCACTTTTG	21	53,6			100%		
HPV16_31_R	AGTTGTTGACGCTGCATGG	20	62,3	L2	242	100%	0	2
HPV16_31b_R	AGTTGTTGACGCTGCATGTG	20	60			100%		
HPV16_32_F	GTAGAATTGCTAATAAACAAACACTACG	28	52,7	L2	358	100%	0	0
HPV16_32_R	GGAACTAATCAAGGAGTTTGTTCAG	25	55,5			100%		
HPV16_33_F	TTCCTGCAAATACAACAATTCC	22	51,7	L2	276	100%	0	0
HPV16_33_R	TACTGGGATAGGAGCTAAGTACAC	24	57,2			100%		
HPV16_34_F	CATGTTACGAAAACGACGTAAA	22	51,5	L1	295	100%	0	3
HPV16_34_R	CCAAACTTATTGGGGTCAGG	20	53,3			100%		
HPV16_35_F	CTTGCAGTTGGACATCCCTA	20	54,5	L1	219	100%	HPV18	0
HPV16_35_R	CACACCTAATGGCTACCAC	20	55,8			100%		
HPV16_36_F	GTTTGGCCCTGTGTAGGTG	19	56,1	L1	207	100%	HPV18	0
HPV16_36_R	TTCCCCTATAGCTGGTTGC	20	54,4			100%		
HPV16_37_F	AGCAAATGCAGGTGTGGATA	20	54,6	L1	261	100%	HPV18	0
HPV16_37_R	TCCAGTGGAACTTCACTTTTG	21	53			100%		
HPV16_38_F	CGGCTTTGGTGCTATGGAC	19	56,4	L1	222	100%	HPV18	0
HPV16_38_R	GCAGTAGACCCAGAGTCTTTAATG	24	55			100%		
HPV16_39_F	TGGTGAAAATGTACCAGACCA	21	53,9	L1	245	100%	HPV18	0
HPV16_39_R	GATATGGCACACACAAATGACA	22	53			100%		
HPV16_40_F	ATGGCATTTTGTTGGGGTAAC	20	53,6	L1	254	100%	HPV18	0
HPV16_40_R	ACCAAAATTCCAGTCCTCCA	20	54			100%		
HPV16_41_F	TGTGCAAAATAACCTTAACTGC	22	51,9	L1	267	100%	HPV18	0
HPV16_41_R	TGCGTCCTAAGCTAAACTGA	20	53,8			100%		
HPV16_42_F	GGAGGTACACTAGAAGATACTTATAGG	27	55,8	L1	245	100%	HPV18	0
HPV16_42_R	CAGGTGGTGGTGTAGCTTTC	22	57,9			100%		
HPV16_43_F	AAGGCCAAACCAAAATTACA	21	51,2	L1	254	100%	HPV18	0
HPV16_43_R	GCATGACACAATAGTTACACAAGC	24	54,6			100%		
HPV16_44_F	GTTTCTATGTGCTTGTATGTGCTTG	25	55,1	E6	258	100%	0	1
HPV16_44_R	CGGTTGAAGCTACAAAATGGC	21	54,9			100%		
HPV16_45_F	GCCATTTTGTAGCTTCAACCG	21	54,9	E6	285	100%	0	0
HPV16_45_R	CAAGCCAAAAATATGTGCCTAAC	23	52,9			100%		
HPV16_46_F	TAAATTACTATGCGCCAACG	20	50,7			100%		
HPV16_46_R	ACTAACCGGTTTCGGTTCAA	20	54,3	E6	302	100%	0	5
HPV16_46b_F	TAAATCACTATGCGCCAACG	20	50,7			100%		
HPV16_47_F	CAAACCGTTTTCGGTTACAC	20	52,4	E6	302	100%	0	0
HPV16_47_R	ATGCATAAATCCCGAAAAGC	20	51,4			100%		

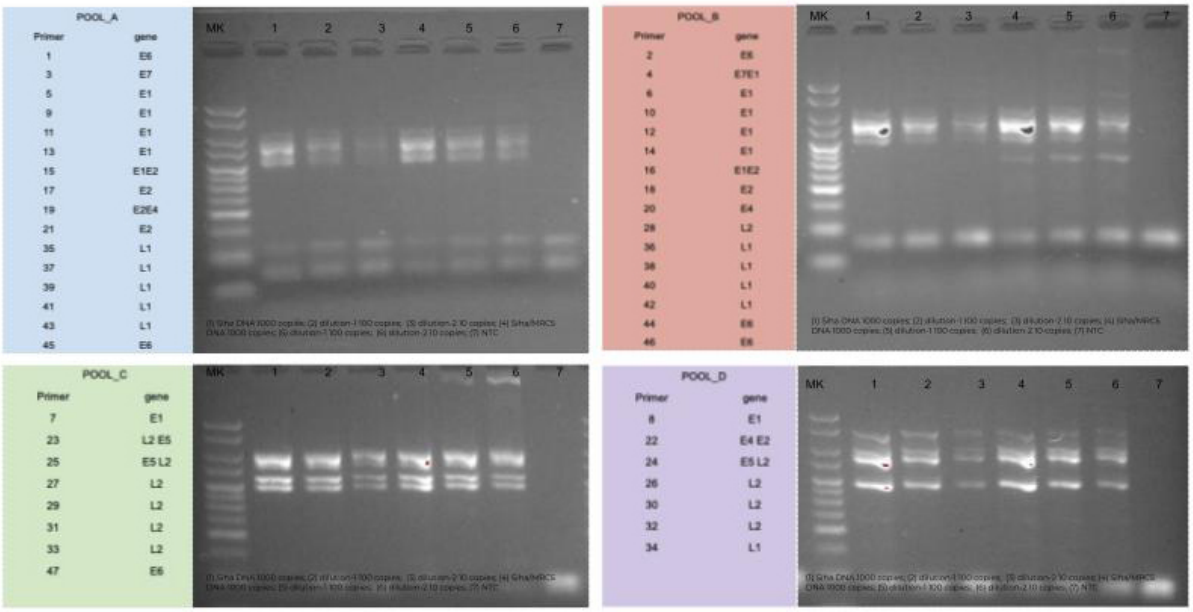
SUPPLEMENT 3 POOL PRIMER DESIGN

Figure S3. Pool Primer Design



Legend: Pool Primer aligned with the HPV genome. The figure shows the HPV genes coloured by each E6 (red), E7 (pink), E1 (blue), E2 (orange), E4 (green), E5 (purple), L2 (light blue) and L1 (grey) which are the target of binding by the set of primers (1 to 47) classified according to the respective pool A (blue), B (brown), C (green) and D (purple) where they are combined. Our design takes a Whole-genome approach by primer set fragment amplification.

Figure S4. Primer's sensibility study of SiHa cell line in agarose gel



Legend: Pool Primer sensibility analysis by each Pool A amplifies fragments in the E6, E7, E1, E2, E4, and L1 region; pool B fragments from the E6, E1, E2, E4, L2, and L1 region; pool C fragments from the E1, L2, E5, and E6 region; and pool D the fragments from the E1, E4, E2, E5, L2, and L1 region. Agarose gel data show the fragments size from respective pool with 3 dilutions for 100 ng Siha cell's DNA (well 1, 2 and 3) and 3 dilutions for 100ng Siha cell's DNA diluted MRC5 cells DNA (well 4, 5, and 6). The well 7 show data from NTC. All pools obtained a signal for the series of dilutions, varying in intensity. In all of them, fragments of different sizes are observed.

APROVAÇÃO DO PROJECTO

Exma. Senhora
Dra. Daniela Cochicho Tavares de Sousa
Rua de São Bento, 576 – 1º Esq.
1250-223 Lisboa

SUA REFERÊNCIA	SUA COMUNICAÇÃO DE	NOSSA REFERÊNCIA	DATA
		0002	04 JAN 2019

ASSUNTO: **INTENÇÃO DE DOUTORAMENTO EM BIOMEDICINA.**

Comunico a V. Exa. que o Presidente do Conselho Científico da Faculdade de Ciências Médicas | NOVA Medical School da Universidade NOVA de Lisboa (FCM|NMS/UNL), por delegação de poderes do plenário, deliberou em 27 de dezembro 2018:

"O Conselho Científico aprova a Intenção de Doutoramento em Biomedicina, tendo em conta o parecer do Coordenador do Ciclo de Estudos".

Com os melhores cumprimentos,

Dra. Sónia Tavares
Coordenadora da Secção de Pós-Graduação da Divisão Académica

ST/ars

www.nms.unl.pt

Decisão final sobre o projecto "Human Papilomavirus in Head and Neck Squamous cell carcinomas: the relevance of HPV16 variants"

A Comissão de Ética da NMS|FCM-UNL (CEFCM) decidiu, por unanimidade, aprovar, do ponto de vista ético, o projecto de investigação intitulado "***Human Papilomavirus in Head and Neck Squamous cell carcinomas: the relevance of HPV16 variants***" (nº70/2019/CEFCM), submetido pela Mestre Daniela Cochicho Tavares de Sousa.

Lisboa, 23 de janeiro de 2020

O Presidente da Comissão de Ética,

(Prof. Doutor Diogo Pais)

TO WHOM IT MAY CONCERN

The Ethics Research Committee NMS|FCM-UNL (CEFCM) has unanimously, approved the Project entitled "***Human Papilomavirus in Head and Neck Squamous cell carcinomas: the relevance of HPV16 variants***" (nr.70/2019/CEFCM), submitted by the Master Daniela Cochicho Tavares de Sousa.

Lisbon, January, 23st, 2020

The Chairman of the Ethics Research Committee,

(Diogo Pais, MD, PhD)

Campo dos Mártires da Pátria, 130 | 1169-056 Lisboa | Portugal | T. +351 218 803 000 – Ext. 20447 E-mail cefcm@fcm.unl.pt

Consentimento para participação no projecto de investigação

"DETECÇÃO E GENOTIPAGEM DO HPV EM ESFREGAÇOS e BIÓPSIAS BUCO-FARINGEOS DE INDIVIDUOS REFERENCIADOS AO IPOLFG COM LESÃO DIAGNOSTICADA NAS VIAS AEREAS DIGESTIVAS SUPERIORES".

O IPOLFG é um centro multidisciplinar de referência para a prestação de serviços de saúde no domínio da oncologia com actividade nas áreas de investigação, ensino, prevenção, diagnóstico, tratamento, reabilitação e continuidade de cuidados. O desenvolvimento de actividades de investigação é fundamental para que se possa progredir no conhecimento dos factores que condicionam o aparecimento da doença e para que se encortrem respostas mais adequadas ao seu tratamento. O material biológico que é colhido para realização de diagnósticos pode constituir-se como fonte de informação da maior importância para o desenvolvimento de projectos de investigação.

O procedimento médico a que vai ser submetido envolve a recolha de amostra que irá ser utilizada para diagnóstico. Contudo esse mesmo material poderá ser também utilizado no âmbito de um projecto de investigação actualmente em curso no IPOLFG através do qual se pretende determinar a prevalência do Vírus do Papiloma Humano (HPV) em lesões na orofaringe e cavidade oral.

No caso de concordar em participar neste projecto, da amostra biológica fresca ou parafina [esfregaço ou biópsia buco-faríngeo] colhida para diagnóstico ser-lhe-á retirada uma porção que será enviada ao Laboratório de Virologia para a detecção de agentes virais.

O risco e o desconforto associado à colheita em questão serão mínimos ou inexistentes, o dador será sempre antecipadamente informado dos riscos e grau de desconforto associados aos procedimentos pelo médico.

Serão cumpridas todas as normas éticas de garantia do anonimato, aceites internacionalmente: todas as informações clínicas, demográficas e outras relacionadas compiladas na base de dados como todos os resultados deste estudo estarão sujeitas a sigilo médico e os quais permanecerão anónimos durante todo o estudo.

O consentimento de armazenamento pode ser retirado em qualquer altura e essa amostra será destruída.

O projecto só se concretizará após obtenção de parecer favorável da Comissão de Ética do IPOLFG e se estiverem cumpridas todas as normas em vigor sobre realização de projectos de investigação e publicação/apresentação de resultados.

A participação é voluntária e a sua recusa em participar não envolverá qualquer penalização ou perda de benefícios. Esta é uma doação altruísta, não havendo por isso qualquer compensação para o dador. Contudo, a sua participação proporcionará a aquisição de conhecimentos que poderão vir a beneficiá-lo a si ou a terceiros no futuro.

O consentimento é revogável pelo que no caso de consentir participar no projecto e mais tarde pretender retirar esse consentimento poderá, a todo o tempo e sem sujeição a qualquer formalidade, fazê-lo.

_____

Nome _____, declaro que:

- li e compreendi a informação que consta do presente formulário;

- fui esclarecido relativamente a todas as dúvidas que coloquei sobre a participação em projectos de investigação;

- foi-me concedido tempo que considero suficiente para reflectir sobe a decisão de participar neste projecto de investigação,

pelo que **autorizo** a utilização, no âmbito do presente projecto, do material biológico que será colhido no acto cirúrgico a que irei ser submetido.

Assinatura……

Data……

Elementos identificativos do médico que obteve o consentimento

Nome……

Nº de cédula profissional/nº mecanográfico……

Assinatura ……

Data ……