From Department of Oncology-Pathology
Karolinska Institutet, Stockholm, Sweden

STUDIES ON HUMAN PAPILLOMAVIRUS (HPV) AND OTHER MARKERS IN THE DEVELOPMENT AND PROGNOSIS OF HPV ASSOCIATED CANCER

Andreas Ährlund-Richter

Karolinska Institutet

Stockholm 2022

Published by Karolinska Institutet.
Printed by Universitetsservice US-AB, 2022

ISBN **978-91-8016-644-7**

Studies on human papillomavirus (HPV) and other markers in the development and prognosis of HPV associated cancer

THESIS FOR DOCTORAL DEGREE (Ph.D.)

By

Andreas Ährlund-Richter

The thesis will be defended in public at J3:20 Sven Ivar Seldinger, 22022-06-14

Principal Supervisor:
Anders Näsman, Associate professor
Karolinska Institutet
Department of Oncology-Pathology

Co-supervisor(s):
Du Juan, PhD
Karolinska Institutet
Department of Microbiology, Tumour Biology, and Cell Biology

Michael Mints MD., PhD
Karolinska Institutet
Department of Medicine Solna
Immunology and Allergy

Stefan Holzhauser, PhD
Karolinska Institutet
Department of Oncology-Pathology

Opponent:
Associate professor Hans Brunnström
Lunds University
Lunds University Cancer Center

Examination Board:
Professor Sören Andersson
Örebro University
School of Medical Sciences

Professor em. Anders Hjerpe
Karolinska Institutet
Department of Laboratory Medicine

Associate professor Christofer Juhlin
Karolinska Institutet
Department of Oncology-Pathology

To my family, inside and outside the lab.

POPULAR SCIENCE SUMMARY OF THE THESIS

Background and previous findings. Human papilloma virus (HPV) is a virus family with more than 200 different members, called HPV types of which infect skin and others infect the mucous membranes in the genital and oral area in humans. Most HPV types are harmless and the infection is either cleared, or in some cases the virus may trigger the development of warts or papillomas, similar to what related papilloma viruses can cause in animals, and this is how they got their name.

However, in humans some HPV types can cause cancers, and these HPV types are referred to as high-risk (HR) HPV types. HR-HPVs can cause anogenital cancers, such as cancer of the cervix as well as specific cancers also in the head and neck region. It has the past decades been shown that some cancers in the upper throat region have increased in prevalence, and these are in fact mainly attributed to HPV-positive (HPV^+) tonsillar and base of tongue cancer (TSCC and BOTSCC).

In previous studies, before HPV vaccination was introduced into the school-based vaccination program in Sweden, we therefore looked at the prevalence rates of oral and cervical HPV infections in patients 15-23 years of age at a youth clinic in Stockholm. The sampling was initially done during the time period between 2008-2010, and followed up during the period of 2013-2015 and later with the aim to follow HPV oral and cervical prevalence before and after HPV vaccination. As HPV vaccine rates increased, HPV vaccine type prevalence went down in youth.

In parallel, during the past decades, several studies have shown a possible link between the bacteria in the gut as well as in the vaginal region and several diseases. Bacteria and microorganisms are also referred to as *microbiota.* In fact, during the mid-2010s several studies showed possible linkage between vaginal microbiota and obstetric outcomes, inflammatory disease, as well as sexually transmitted disease. Later, several meta-studies and one DNA sequencing study abroad, showed an association between HPV prevalence and microbiota. So this thesis also deals with microbiota and HPV.

As mentioned above HPV associated TSCC and BOTSCC have increased in incidence during the past decades and therefore more information about these tumours is important and also whether it is possible to screen has been of importance. Further, it has previously been shown that most HPV associated anogenital cancers go through three stages of development; pre-malignant stages, dysplasia and high-grade dysplasia/cancer in situ, and later transformations into malignant stages; invasive cancer, and metastatic cancer, so for cervical cancer screening is available. However, these stages are not well studied in HPV^+ TSCC or BOTSCC so more information on this subject would be useful.

In addition, it was shown that patients with HPV^+ TSCC and BOTSCC have a better overall and disease-free survival as compared to patients with corresponding HPV-negative (HPV^-) cancer, when treated the same way. So to better individualize treatments many studies have focused on identifying prognostic biomarkers and markers useful for targeted therapy to better individualize patient treatment.

The aim of this thesis was therefore to follow up some of the above earlier findings and the results are presented below.

In *Paper I*, we followed up HPV vaccination coverage and changes in cervical HPV prevalence at a youth clinic in Stockholm, Sweden, during a decade after the introduction

of the HPV vaccine. The proportion of HPV vaccinated women increased from 10.7% 2008-2010 to 82.1% 2017-2018. An overall reduction of the HPV types included in the vaccine was observed, more pronounced in vaccinated than in unvaccinated women. However, other high-risk HPV (HR-HPV), not included in the vaccine strategy, still remained high.

In *Paper II*, we investigated possible influence of HPV status, age and vaccination status on the vaginal microbiota in a cohort from Uppsala and Stockholm. Microbial alpha-diversity was found to be much higher in the HPV infected group compared to the HPV negative group, in particular if the women were infected with HR-HPV types and had multiple HPV types. Moreover, as roughly double the number of women with non-*lactobacilli* dominant vaginal microbiota were infected with HR-HPV types as compared to those that had *L. crispatus* dominated vaginal microbiota after adjustment for age and vaccination status. Consequently, infection with oncogenic HPVs was associated with non-*lactobacilli* dominant vaginal microbiota.

In *Paper III,* we compared gene expression in HPV^+ versus HPV^- high-grade dysplasia and invasive TSCC and BOTSCC and found them to be similarly differentiated to invasive cancer stages. Using immunohistochemistry (IHC) and RNA-panels, forty genes showed differential expression (e.g. SPARC, psoriasin I, collagen-1 and galectin-1).

In *Paper IV*, we performed whole-exome-sequencing on primary and relapsed HPV^+ TSCC/BOTSCC, in an attempt to identify genetic markers possibly useful for prognosis or treatment in the two groups. Specifically, for the *CDC27* gene a deletion of high impact was found only in tumours of patients that had relapsed but in no primary tumours of patients without recurrence. Three variants, two with deletions in *BCLAF1* and *OVCH2* and one with a substitution in *OR2T35*, and 26 mutated genes were disclosed as mutated in >30% of all cases, thereby possibly consisting part of a global mutational signature for HPV^+ TSCC/BOTSCC.

Conclusions. The presented studies in this thesis reaffirm the efficacy of the HPV vaccine programs in Stockholm, but with a remaining continuous prevalence of non-vaccine HR-HPV types. The results also suggest an influence from the HPV status on the vaginal microbiota make up. Furthermore, the data suggest that that HPV^+ and HPV^- TOSCC/BOTSCC although not identical also likely have similar dysplastic cancer stages. Finally, there are differences between the mutational profiles of HPV^+ TSCC/BOTSCC that re-occur compared to those that do not re-occur, but also here there are genes that are similarly altered in primary tumours of patients that are cured or that relapse.

ABSTRACT

Background and aims: Human papilloma virus (HPV) is a risk factor for anogenital and oropharyngeal cancer (OPSCC) and commonly transmitted sexually although most infections are cleared without adverse effects. Notably, the past decades the incidences of HPV positive (HPV^+), but not HPV negative (HPV^-) tonsillar and base of tongue cancer (TSCC and BOTSCC), the two major OPSCC subtypes have both increased. For this reason, we wanted to follow HPV-prevalence. Before HPV vaccination a high prevalence of HPV was shown in the cervix and oral cavity of youth aged 15-23 years at a youth clinic in Stockholm 2008-2010, but later, with rising vaccine coverage, a decrease of HPV vaccine types was observed between 2013-2015. In parallel, in the mid-2010s many studies showed a link between vaginal microbiota and obstetric outcomes, inflammatory disease, as well as sexually transmitted disease. A few years later, several meta-studies and one DNA sequencing study abroad, showed an association between HPV prevalence and microbiota. In this context, of note, most HPV associated anogenital cancers go through three stages of development, pre-malignant stages, dysplasia, high-grade dysplasia/cancer in situ, to malignant stages, invasive cancer, and metastatic cancer. However, these stages are not well studied in HPV^+ TSCC and BOTSCC and although there have been many biomarker studies in these tumours additional ones would be of use to better individualize treatment of these cancers. *The aim of this thesis* was therefore to follow up some of these findings.

Approaches. In paper I, we followed up the HPV vaccination coverage and HPV prevalence at a youth clinic in Stockholm, to investigate vaccine effects. In paper II, we investigated possible effects of HPV status, age and vaccination status on the vaginal microbiota of women in a cohort from Uppsala and Stockholm. In paper III, we analysed and compared gene expression in high-grade dysplasia and invasive cancer in HPV^+ and HPV^- TSCCC/BOTSCC with particular emphasis on HPV status. In paper IV, we did whole-exome-sequencing on primary tumours of HPV^+ TSCC/BOTSCC patients with and without recurrences, to identify similarities and differences between the groups as well as to identify markers of prognostic significance or candidates for targeted therapy.

Results: In paper I, the proportion of HPV vaccinated women increased from 10.7% 2008-2010 to 82.1% 2017-2018. HPV-vaccine types were reduced overall and more in vaccinated than in unvaccinated women, but other high-risk HPV types still remained high. In paper II, microbial alpha-diversity was significantly higher for HPV^+ compared to HPV^- patients. Twice as many HPV^+ than HPV^- women had non-*lactobacillus* dominant vaginal microbiota compared to *L. crispatus* dominated vaginal microbiota and oncogenic HPVs were associated with non-*lactobacillus* dominant vaginal microbiota. In paper III, invasive and non-invasive tumours gene-expression were compared using immunohistochemistry (IHC) and RNA-panels. Forty genes showed differential expression, e.g. SPARC, psoriasin I, collagen-1 and galectin-1 and HPV^+ and HPV^- dysplasia was similarly differentiated from invasive cancer. In paper IV, a high-impact deletion on CDC27 was observed only in primaries of patients with relapse and 3 variants and 26 mutated genes, were present > 30% of all primaries regardless of prognosis.

Conclusions. The presented studies in this thesis reaffirm the efficacy of the HPV vaccine programs in Stockholm, but with a remaining continuous prevalence of non-vaccine HR-HPV types. The results also suggest an influence from the HPV status on the vaginal microbiota make up. Furthermore, the data suggest that that HPV^+ and HPV^- TSCC/BOTSCC although not identical also likely have similar dysplastic cancer stages. Finally, there are differences between the mutational profiles of HPV^+ TSCC/BOTSCC that re-occur compared to those that do not re-occur, but also here there are genes that are similarly altered in primary tumours of patients that are cured or that relapse.

LIST OF SCIENTIFIC PAPERS INCLUDED IN THESIS

I. **Ährlund-Richter A,** Cheng L, Hu YOO, Svensson M, Pennhag AAL, Ursu RG, Haeggblom L, Grün N, Ramqvist T, Engstrand L, Dalianis T, Du J. Changes in Cervical Human Papillomavirus (HPV) Prevalence at a Youth Clinic in Stockholm, Sweden a Decade After the Introduction of the HPV Vaccine. *Front Cell Infect Microbiol. 2019 Mar 20;9:59.*

II. Liqin Cheng, Johanna Norenhag, Yue O. O. H, Nele Brusselaers, Emma Fransson, **Andreas Ährlund-Richter**, Unnur Guðnadóttir, Pia Angelidou, Yinghua Zha, Marica Hamsten, Ina Schuppe Koistinen, Matts Olovsson, Lars Engstrand, Juan Du. Vaginal microbiota and human papillomavirus infection among young Swedish women. NPJ Biofilms Microbiomes 2020 Oct 12;6(1):39

III. Haeggblom L, **Ährlund-Richter A,** Mirzaie L, Farrajota Neves da Silva P, Ursu RG, Ramqvist T, Näsman A. Differences in gene expression between high-grade dysplasia and invasive HPV^+ and HPV^- TSCC/BOTSCC / Differences in gene expression between high-grade dysplasia and invasive HPV+ and HPV- tonsillar and base of tongue cancer *Cancer Med. 2019 Oct;8(14):6221-6232.*

IV. **Ährlund-Richter A**, Holzhauser S, Dalianis T, Näsman A, Mints M. Whole-exome sequencing of HPV positive tonsillar and base of tongue squamous cell carcinomas reveals a global mutational pattern along with relapse-specific somatic variants. Cancers (Basel). 2021 Dec 24;14(1):77.

List of scientific papers not included in thesis

Papers related to HPV prevalence

Du J, Nordfors C, **Ährlund-Richter A,** Sobkowiak M, Romanitan M, Näsman A, Andersson S, Ramqvist T, Dalianis T. Prevalence of oral human papillomavirus infection among youth, Sweden. Emerg Infect Dis. 2012 Sep;18(9):1468-71. doi: 10.3201/eid1809.111731. PMID: 22932445;

Nordfors C, Vlastos A, Du J, **Ährlund-Richter A,** Tertipis N, Grün N, Romanitan M, Haeggblom L, Roosaar A, Dahllöf G, Donà MG, Benevolo M, Ramqvist T, Munck-Wikland E, Dalianis T. Human papillomavirus prevalence is high in oral samples of patients with tonsillar and base of tongue cancer. Oral Oncol. 2014 May;50(5):491-7. doi: 10.1016/j.oraloncology.2014.02.012. Epub 2014 Mar 7. PMID: 24613649.

Grün N, **Ährlund-Richter A,** Franzén J, Mirzaie L, Marions L, Ramqvist T, Dalianis T. Oral human papillomavirus (HPV) prevalence in youth and cervical HPV prevalence in women attending a youth clinic in Sweden, a follow up-study 2013-2014 after gradual introduction of public HPV vaccination. Infect Dis (Lond). 2015 Jan;47(1):57-61. doi: 10.3109/00365548.2014.964764. Epub 2014 Nov 7. PMID: 25378085.

Grün N, **Ährlund-Richter A,** Franzén J, Mirzaie L, Marions L, Ramqvist T, Dalianis T. Follow-up on oral and cervical human papillomavirus prevalence 2013-2015 in youth at a youth clinic in Stockholm, Sweden. Infect Dis (Lond). 2016 Feb;48(2):169-70. doi: 10.3109/23744235.2015.1094573. Epub 2015 Oct 7. PMID: 26536907.

Papers related to HPV and biomarkers

Lindquist D, **Ährlund-Richter A,** Tarján M, Tot T, Dalianis T. Intense CD44 expression is a negative prognostic factor in tonsillar and base of tongue cancer. Anticancer Res. 2012 Jan;32(1):153-61. PMID: 22213301.

Nordfors C, Grün N, Tertipis N, **Ährlund-Richter A,** Haeggblom L, Sivars L, Du J, Nyberg T, Marklund L, Munck-Wikland E, Näsman A, Ramqvist T, Dalianis T. CD8+ and CD4+ tumour infiltrating lymphocytes in relation to human papillomavirus status and clinical outcome in tonsillar and base of tongue squamous cell carcinoma. Eur J Cancer. 2013 Jul;49(11):2522-30. doi: 10.1016/j.ejca.2013.03.019. Epub 2013 Apr 6. PMID: 23571147.

Tertipis N, Hammar U, Näsman A, Vlastos A, Nordfors C, Grün N, **Ährlund-Richter A,** Sivars L, Haeggblom L, Marklund L, Hammarstedt-Nordenvall L, Chaturvedi AK, Munck-Wikland E, Ramqvist T, Bottai M, Dalianis T. A model for predicting clinical outcome in patients with human papillomavirus-positive tonsillar and base of tongue cancer. Eur J Cancer. 2015 Aug;51(12):1580-7. doi: 10.1016/j.ejca.2015.04.024. Epub 2015 May 26. PMID: 26025766.

Haeggblom L, Attoff T, Yu J, Holzhauser S, Vlastos A, Mirzae L, **Ährlund-Richter A,** Munck-Wikland E, Marklund L, Hammarstedt-Nordenvall L, Ye W, Ramqvist T, Näsman A, Dalianis T. Changes in incidence and prevalence of human papillomavirus in tonsillar and base of tongue cancer during 2000-2016 in the Stockholm region and Sweden. Head Neck. 2019 Jun;41(6):1583-1590. doi: 10.1002/hed.25585. Epub 2018 Dec 24. PMID: 30584688.

Bersani C, Mints M, Tertipis N, Haeggblom L, Sivars L, **Ährlund-Richter Ä,** Vlastos A, Smedberg C, Grün N, Munck-Wikland E, Näsman A, Ramqvist T, Dalianis T. A model using concomitant markers for predicting outcome in human papillomavirus positive oropharyngeal cancer. Oral Oncol. 2017 May;68:53-59. doi: 10.1016/j.oraloncology.2017.03.007. Epub 2017 Mar 23. PMID: 28438294.

Bersani C, Sivars L, Haeggblom L, DiLorenzo S, Mints M, **Ährlund-Richter A,** Tertipis N, Munck-Wikland E, Näsman A, Ramqvist T, Dalianis T. Targeted sequencing of tonsillar and base of tongue cancer and human papillomavirus positive unknown primary of the head and neck reveals prognostic effects of mutated FGFR3. Oncotarget. 2017 May 23;8(21):35339-35350. doi: 10.18632/oncotarget.15240. PMID: 28525363; PMCID: PMC5471059.

Landin D, **Ährlund-Richter A,** Mirzaie L, Mints M, Näsman A, Kolev A, Marklund L, Dalianis T, Munck-Wikland E, Ramqvist T. Immune related proteins and tumor infiltrating CD8+ lymphocytes in hypopharyngeal cancer in relation to human papillomavirus (HPV) and clinical outcome. Head Neck. 2020 Nov;42(11):3206-3217. doi: 10.1002/hed.26364. Epub 2020 Jul 1. PMID: 32613643.

LIST OF ABBREVIATIONS

AC	Adenocarcinoma
AIN	Anal intraepithelial neoplasia
AS	Amplicon sequencing
BOTSCC	Base of tongue squamous cell carcinoma
BV	Bacterial vaginosis
BVAB	Bacterial vaginosis associated bacteria
CC	Cervical cancer
CD	Cluster of differentiation
CDI	*Clostridium difficile* infection
CDKN2/AB	Cyclin dependent kinase 2/A/B
CIN	Cervical intraepithelial neoplasia
CRT	Chemoradiotherapy
CT	Chemotherapy
DFS	Disease free survival
EBV	Epstein-Barr virus
EGFR	Epidermal growth factor receptor
EMA	European Medicines Agency
ES	Exome sequencing
E1-7	Early regions proteins 1-7
FFPE	Formalin fixed and paraffin embedded
FDA	Food and Drug Administration
FGFR	Fibroblast growth factor receptor gene/s
FMT	Faecal Microbiota Transplantation
FoxP3	Forkhead box P3
HLA	Human leukocyte antigen
HNSCC	Head and neck squamous cell carcinoma
HPV	Human papillomavirus
HPV^{+}	Human papillomavirus positive
HPV^{-}	Human papillomavirus negative
HR	High-risk
HSIL	High-grade intraepithelial lesion
HTS	High throughput sequencing
IARC	International Agency for Cancer Research
IBD	Inflammatory Bowel Disease
ICD	International classification of diseases

IHC	Immunohistochemistry
LCR	Long control region
L1-2	Late region protein 1-2
LSIL	Low-grade squamous intraepithelial lesion
miRNA	MicroRNA
mRNA	Messenger RNA
NGS	Next generation sequencing
OR	Open reading frame
p16	$P16^{INK4a}$ protein
PAL	Position of the late polyadenylation site
PAP-test	Papanicoloau-test
PD-1	Programmed cell death protein 1
PD-L1	Programmed cell death-ligand 1
PE	Early promoter
PI3K	Phosphoinositide 3-kinases
PIK3CA	Phosphatidylinositol-4,5-Bisphosphate 3-Kinase Catalytic Subunit Alpha gene
PL	Late promoter
Rb	Retinoblastoma protein
RRP	Recurrent respiratory papillomatosis
RT	Radiotherapy
SCC	Squamous cell carcinoma
TILs	Tumour infiltrating lymphocytes
TNM	Tumour-node-metastasis-TNM classification of malignant tumours
TS	Targeted sequencing
VaIN	Vaginal intraepithelial neoplasia
VEGFA	Vascular endothelial growth factor A
VIN	Vulval intraepithelial neoplasia
VLPs	Virus like particles
VMT	Vaginal microbiome transplantation
WGS	Whole genome sequencing
WES	Whole exome sequencing
WHO	World Health Organization

CONTENTS

POPULAR SCIENCE SUMMARY OF THE THESIS 4
ABSTRACT 6
LIST OF SCIENTIFIC PAPERS INCLUDED IN THESIS 7
List of scientific papers not included in thesis 8
LIST OF ABBREVIATIONS 9
1 INTRODUCTION 14
2 LITERATURE REVIEW 16
2.1 Human papillomavirus (HPV) 16
2.1.1 Classification 16
2.1.2 The viral genome and its proteins 17
2.1.3 The long control region (LCR) 18
2.1.4 The early region (E) proteins 18
2.1.5 The late region (L) proteins 19
2.1.6 Transmission 20
2.1.7 HPV prevalence in the general population 22
2.1.8 HPV and disease 22
2.1.9 HPV and benign lesions 22
2.1.10 HPV and malignancies 23
2.1.11 HPV screening 31
2.1.12 HPV prevalence in the general population 32
2.1.13 HPV vaccines 33
2.2 Microbiota 34
2.2.1 Gut microbiota 34
2.2.2 Vaginal microbiota 35
2.2.3 Oral Microbiota 37
3 RESEARCH AIMS 38
4 STUDY INDIVIDUALS 39
4.1 Young adults attending a youth health care center 39
4.2 Uppsala Maternity health clinic patients 40
4.3 Patients with tonsillar and base of tongue squamous cell carcinoma 40
4.4 Considerations regarding the study individuals 41
5 MATERIALS 43
5.1 Cervical swabs 43
5.2 Cancer tissues 43
6 METHODS 45
6.1 Preparation of cervical swabs and DNA extraction 45
6.2 TumoUr selection, microdissection and RNA and DNA extraction 45
6.3 RNA extraction and multiplex gene expression analysis 47
6.4 HPV DNA extraction of TSCC, BOTSCC and metastatic lesions 48
6.5 HPV Genotyping 48

6.6 Bead based analysis on a Magpix instrument 49
6.7 Statistical analysis 50
6.8 IMMUNOHISTOCHEMISTRY (IHC) 50
6.9 DNA Sequencing 52
6.9.1 Introduction 52
6.9.2 General NGS Workflow 53
6.9.3 NGS Analysis 54
6.10 Types of DNA Sequencing 55
7 RESULTS AND DISCUSSION 59
7.1 Paper I. Changes in Cervical Human Papillomavirus (HPV) Prevalence at a Youth Clinic in Stockholm, Sweden a Decade After the Introduction of the HPV Vaccine 59
7.1.1 Aim 59
7.1.2 Background 59
7.1.3 Material and Methods 59
7.1.4 Results 60
7.1.5 Discussion 62
7.1.6 Conclusion 63
7.2 Paper II. Vaginal microbiota and human papillomavirus (HPV) infection among young Swedish women 64
7.2.1 Aim 64
7.2.2 Background 64
7.2.3 Material and Methods 64
7.2.4 Results 65
7.2.5 Discussion 67
7.2.6 Conclusion 68
7.3 Paper III. Differences in gene expression between high-grade dysplasia and invasive HPV+ and HPV- tonsillar and base of tongue cancer. 69
7.3.1 Aim 69
7.3.2 Background 69
7.3.3 Material and Methods 69
7.3.4 Results 70
7.3.5 Discussion 73
7.3.6 Conclusion 74
7.4 Paper IV. Whole-Exome Sequencing of HPV positive tonsillar and base of tongue squamous cell carcinomas reveals a global mutational pattern along with relapse-specific somatic variants 75
7.4.1 Aim 75
7.4.2 Background 75
7.4.3 Material and Methods 75
7.4.4 Results 76
7.4.5 Discussion 79

7.4.6 Conclusions 81
8 CONCLUSIONS 82
VI. FUTURE PERSPECTIVES 83
9 ACKNOWLEDGEMENTS 84
10 REFERENCES 87

1 INTRODUCTION

The physician Rigoni-Stern investigated the death certificates of women from Verona during the period of 1760-1839, and noted a lower incidence of cervical cancers (CCs) in virgins and nuns, compared to women that had been married, widowed or were prostitutes (DiMaio D 2015). Much later, in the early 20th century, Payton Rous performed an experiment to test the transferability of cancer through viruses (Rous P 1911). To test this theory, he extracted a tumour from a hen with spindle-cell sarcoma, mixed it, filtered it from whole cells, and injected the mixture into a healthy hen. The hen subsequently developed a similar tumour, showing the possibility of a filterable agent causing cancer. It could not be proven in mammals at the time, but he received the Nobel-prize in 1966 as the viral-transmission theory gained ground. However, 30 years after the first discoveries of Payton Rous, in 1941, the inventor of the pap smear, Georgios Papanicolaou found that cellular changes preceded invasive cancers in patients (Papanicolaou GN and Traut H 1941). Nevertheless, the first association of a particular virus with cancer in humans would not be a genital virus, but the Epstein-Barr virus (EBV), named after and discovered by Michael Anthony Epstein and Yvonne Barr 1964 (Epstein A et al., 1964). This virus was isolated from Burkitts Lymphoma cells cultured *in vitro* (Epstein A et al., 1964).

The association between human papillomavirus (HPV) a main topic of this thesis, and cancer was not established until later and one reason was that HPV was shown to be a family consisting of many HPV types, for review see e.g. Tomassino M 2014. It was in 1983 when the viral transmission theory was established well enough, that Harald Zur Hausen disclosed a connection between HPV and CC which was in conflict with an earlier suggested linkage between CC and Herpes simplex virus II (Dürst M et al., 1983). Harald Zur Hausen was later awarded the Nobel-prize in 2008.

Notably, it was in 1995 that HPV16 first was acknowledged as being a carcinogenic by the International Agency on Research against Cancer (IARC 1995, Vol 64). Not too much later, in 2006, the very first vaccine directed against HPV16, 18, 6 and 11 (Gardasil) and CC was approved by the Food and Drug Administration (FDA) and this was followed in 2007 by a vaccine against HPV16 and 18 (Cervarix) (Harper DM and DeMars LR 2017). That HPV also was associated with oropharyngeal squamous cell carcinoma (OPSCC) and especially tonsillar and base of tongue squamous cell carcinoma (TSCC and BOTSCC) its dominant subsites, a disease primarily affecting men was observed in 2000, (Gillison ML et

al., 2000, Mellin H et al., 2000) and acknowledged by IARC in 2007 (IARC 2007, Vol 104). At the time increases in the incidences of TSCC/BOTSCC/OPSCC were being reported and suggested to be due to rises in HPV positive (HPV^+) TSCC and BOTSCC (Conway DI et al., 2006, Sturgis EM et al., 2007, Näsman A et al., 2009, Attner P et al., 2010). In 2014, an additional HPV vaccine, (Gardasil 9) (against HPV16, 18, 31, 33, 45, 52, 58, and 6 and 11) was FDA approved (Harper DM and DeMars LR 2017). Currently, HPV vaccination of girls is performed in very many countries, while for boys HPV vaccination is still not the rule and unfortunately rarely distributed (Harper DM and DeMars LR 2017).

In Sweden, in 2010, 10-12-year old girls had the possibility to be HPV vaccinated free of charge, but it was first in 2012, that Gardasil was offered to them through the school based-vaccination program, and this decision making is described by Tegnell A et al., 2009. In parallel, young women 18-26 years were also offered catch-up vaccination for free e.g. in Stockholm and other counties in Sweden. In order to obtain information on base-line HPV prevalence as well as study the effects of HPV vaccination, researchers in our group have followed cervical and oral HPV prevalence for more than 10 years at a large youth clinic in Stockholm, for review see Du J et al., 2021.

My own curiosity and engagement in the field started in 2011 but was soon intensified with interest in studies on HPV associated TSCC and BOTSCC and HPV in the oral cavity when I started to study dentistry, but my life turned into a more data science profile, and this thesis may reflect part of this transition and development.

2 LITERATURE REVIEW

In this literature review, a brief background of human papillomavirus (HPV), microbiota, dysplasia and cancer will be presented.

2.1 HUMAN PAPILLOMAVIRUS (HPV)

2.1.1 Classification

Historically, the papillomaviruses (PV) were grouped together with the polyomaviruses, into the family, *papoviridae* due to both having a circular double-stranded DNA genome and nonenveloped capsids (Melnick JL et al., 1974). Later studies showed different genome organizations, and completely different genome sizes (van Doorslaer K et al., 2017). The nucleotide and amino acid sequences also did not overlap substantially. PV can be grouped into 16 families, with HPV being comprised of the five families *Alpha*, *Beta, Gamma, Mu* and *Nu-papillomavirus* (Figure 1) (Doorbar J et al., 2012, Tommasino M 2014).

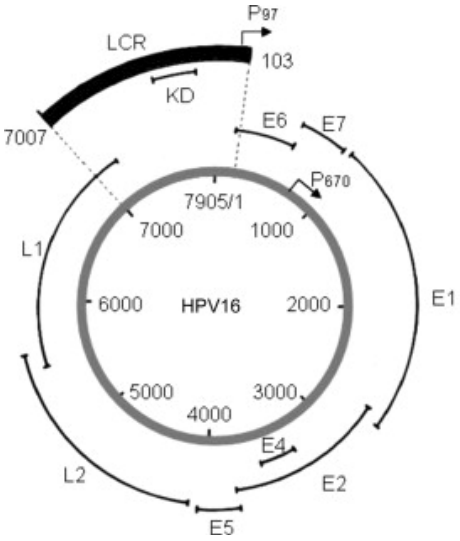
		HPV type	Disease (% attributed cases)
Alpha	mucosal high-risk	HPV16	Cervical squamous cell carcinoma (~50) Cervical adenocarcinoma (~35) Oropharyngeal cancer (~25)
		HPV18	Cervical squamous cell carcinoma (~20) Cervical adenocarcinoma (~35)
		HPV31, 33, 35, 39, 45, 51, 52, 56, 58, 59	Cervical squamous cell carcinoma (~30)
	mucosal low-risk	HPV6, 11	Benign genital lesions Respiratory papillomatosis
		HPV13, 32	Oral focal epithelial hyperplasia
	cutaneous benign	HPV2,3, 27, 57	Skin warts
Mu		HPV1	Skin warts
Beta	cutaneous	HPV5 and 8	First beta HPV types isolated from SSC of EV indiviuals
		HPV9, 12, 14, 15, 17, 19-25, 36-38, 47, 49, 75, 76, 80, 92, 93, 96, 98-100, 104, 105, 107, 110, 111, 113, 115, 118, 120, 122, 124, 143, 145, 150- 152, 159	Likely associated with SCC in EV patients as well as immuno-compromised and immuno-competent individuals
Gamma		HPV4, 48, 50, 60, 65, 88, 95, 101, 103, 108, 109, 112, 115, 116, 119, 121, 123, 126-142, 144, 146-149, 153-158, 161-170	Unknown

Figure. 1. Many of the identified HPV types that belong to different genera (i.e. alpha, beta, gamma and mu) of the HPV are shown. In addition, the main diseases that have been associated with different HPV types are described in the left panels of the figure. (From Tommasino M 2014, with permission from the publisher).

PVs are generally classified by investigating the L1 gene nucleotide sequence, as this sequence is relatively well conserved. Any HPV 10% dissimilar in this region compared to other HPV is counted as a novel type (Bernard HU et al., 2010). Over 200 types of HPV have been identified and roughly 150 have been whole genome sequenced (Doorbar J et al., 2012, Tommasino M 2014). The Alpha branch contains low risk HPV variants that cause genital warts and are deemed low risk together as are members of the Beta and Gamma branch (Doorbar J et al., 2012). Notably however, the Alpha branch also contains high-risk (HR) mucosal HPV types which can cause neoplasms and cancer (Doorbar J et al., 2012).

2.1.2 The viral genome and its proteins

All HPV types have a double stranded circular genome consisting of roughly 8 kbp double-stranded DNA (Figure 2) (Tommasino M 2014) The genome is for practical reasons arbitrarily divided into an early a late and a regulatory region (Tommasino M 2014, van Doorslaer K et al., 2017). The regulatory region contains regulatory elements, the early region encodes the regulatory proteins E1-E7 and the late region the structural viral capsid proteins L1 and L2 (Tommasino M 2014).



Viral Protein	Functions and Features
E1	It forms a heterodimer complex with E2 and controls viral replication
E2	It regulates early gene promoter and together with E1 viral DNA replication
E4	It may mediate the viral particle realease by destabilizing cytokeratin network
E5	It stimulates mitogenic signals of growth factors
E6	It inactivates many cellular proteins and is one of the major viral oncoproteins
E7	It inactivates many cellular proteins and is one of the major viral oncoproteins
L1	It is the major capsid protein and is the component of the HPV prophylactic vaccine
L2	It is the minor capsid protein

Figure. 2. The double-stranded DNA HPV16 genome is represented by a grey circle annotated with the nucleotide numbers. The positions of the long control region (LCR) and the early genes (E1–7) and late genes (L1 and L2) are also shown. The early and late promoters, P97 and P670, respectively, are indicated by arrows. The main functions and features of the early and late gene products are listed in the table. (From Tommasino M 2014, with permission from the publisher)

2.1.3 The long control region (LCR)

The long control region (LCR) is located in the area between the open reading frames (ORFs) of E6 and L1 (Figure 2). It includes the majority of the regulatory elements necessary for viral DNA replication and transcription and the LCRs of the various HPV types can vary considerably in size (Tommasino M 2014).

2.1.4 The early region (E) proteins

The E1 and E2 viral proteins are important during the period of early infection and are active during initial replication for establishing an infection (Tommasino 2014). E2 contains a DNA-binding and a protein-binding domain and can this way form a homodimer that can bind to the regulatory region. E2 has the capacity to bind to E1 and as a dimer they bind to the viral origin of replication, and the DNA replication machinery of the cells of the host (Tommasino M 2014, Graham SV 2017) . E2 is also an important key regulator of E6/E7 abundance (see below)

E4 is a protein expressed from the early region of the genome and has been suggested to be produced later in the viral cycle (Tommasino M 2014). It can associate and disrupt the cytoplasmic keratin network and very many studies have shown that E4 is actively involved in viral release (Doorbar J 2013).

The E5 viral protein is assumed to indirectly contribute to genome amplification by modifying the cellular environment and it has also been suggested to be an oncoprotein, and down-regulate expression of the major histocompatibility complex, thereby hindering immune recognition (Doorbar J et al., 2012, Venuti A et al., 2011).

The E6 and E7 viral proteins are together in HR-HPV types regarded as oncogenes. E7 binds to the retinoblastoma protein (Rb) and abrogates cell cycle control and pushes the cell to proliferate, and this indirectly activates the $p16^{INK4a}$ (p16) protein leading to p16 overexpression (Doorbar J et al., 2012, Tommasino M 2014). E6 binds to p53 and inhibits cell repair before proliferation and indirectly promotes proliferation of cells with mutations (Doorbar J et al., 2012, Tommasino M 2014). In many tumours, HPV is found integrated into the host chromosome, and in those cases the viral integration site frequently lies within the E1 and E2 genes, and loss of E2, leads to loss of E6/E7 regulation, and upon a persistent

high expression of E6 and E7 an accumulation of genetic errors may occur that in turn eventually may lead to the development of cancer (Doorbar J et al., 2012, Tommasino M 2014).

Several additional ORFs have been identified for E3, E5 and E8, but their expression is not frequently observed through all PVs.

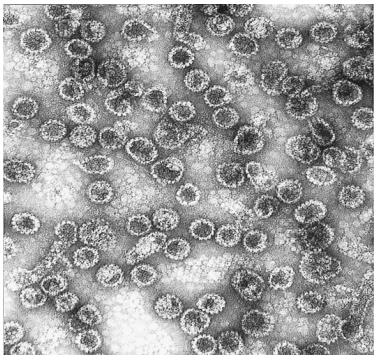


Figure 3. Transmission electron micrograph of HPV16 L1/L2 virus-like particles (VLPs) isolated from recombinant baculovirus infected Sf-9 insect cells. (From Schiller JT and Lowy RD 1996 with permission from the publisher)

2.1.5 The late region (L) proteins

The major capsid protein L1 and the minor capsid protein L2 form together the viral capsid (Tommasino M 2014). Around 360 L1 molecules are situated on the outside of the viral capsid and are mainly formed as pentamers and around 20 L2 molecules are contained in the capsid (Tomassino M 2014). For long it has been known that under specific conditions the major capsid proteins of the *papoviridae* can self-assemble and form virus like particles (VLPs) (Figure 3) and these are the basis of today's HPV vaccines (Lowry RD and Schiller JT 2012, Harper DM and DMars LR 2017).

These VLPs are usually produced in large scale either in insect cells or in yeast (Cervarix and Gardasil respectively, for further details see below).

2.1.6 Transmission

HPV is generally assumed to be transmitted via microtears in the skin or mucosa and in addition HPV infection is suggested to be one of the most common sexually transmitted disease in the USA (Tommasino M 2014, Burd EM 2003). Experiments on cottontail rabbits have shown that light wounding of the site of infection does increase papilloma infections compared to a control (Cladel NM et al., 2008). Furthermore, suggested activities of skin contact are sexual contact, but also possibly kissing (Syrjänen S 2003, Syrjänen S 2004, D'Souza G et al., 2009).

Notably, risk of anogenital and oral infection is correlated to high sexual activity and early time of sexual debut (Anaya Saavedra G et al., 2008, D' Souza et al., 2009). However, in clinical experiments blood has been shown to be an effective vector, but how important this is as a native mode of transmission is unclear (Cladel NM et al., 2008).

There are other ways of transmission as well. One other way of transmission is vertical transmission as reviewed by Syrjänen S 2010. HPV can also be detected in the blood of cancer patients, and today assaying for HPV in the blood has been discussed as a possible tool for monitoring successful treatment or relapse after HPV related cancers (Routman DM et al., 2022). However, the latter and similar studies need to be followed up further.

How the virion enters the cell is not completely clear, but several studies propose that after the capsid makes contact with the basal lamina (Figure 4) (for review see Doorbar J et al., 2012). The L1 and L2 proteins on the viral capsid react with heparin sulphate proteoglycans and likely also laminin followed by structural changes of the capsid and furin cleavage of L2 thereby facilitating the transfer of the capsid to a second receptor on the keratinocyte and internalisation and subsequent transfer to the nucleus (Doorbar J et al., 2012).

The virus then delivers its genome into the cell, which is subsequently transported into the nucleus of the host cell. Studies suggest that it is during the healing process of wounds, as the infected keratinocytes divide, that the viral DNA actually is incorporated into the host nucleus (Doorbar J et al., 2012, Brianti P et al., 2017).

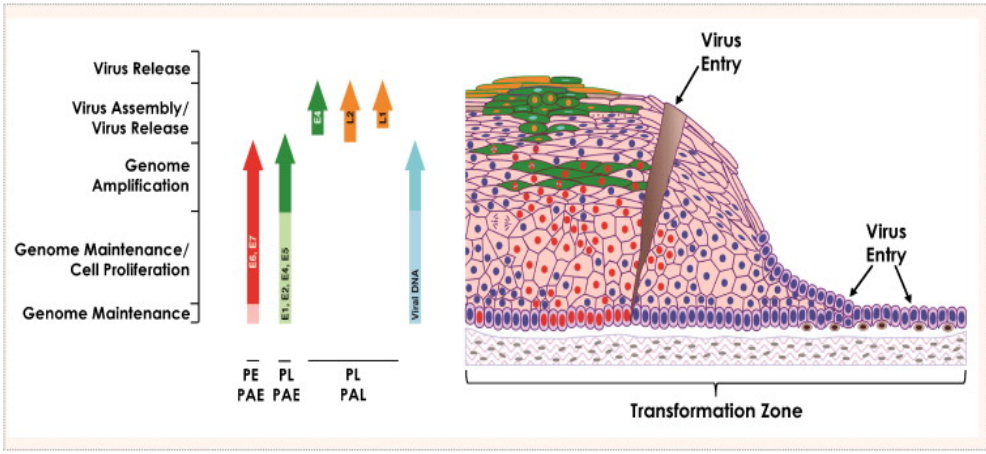


Figure 4. Life Cycle of High-Risk HPVs in Cervical Epithelium. **(**From Doorbar J et al., 2012, with permission from the publisher).

In multi-layered stratified epithelium, such as the ectocervix, infection is thought to require the presence of a microwound that allows the infectious virions to access the basal lamina. The infected basal cells form the reservoir of infection, and in these cells, the viral genome is maintained as a low copy number episome. As these cells divide, they produce daughter cells that are pushed outwards towards the epithelial surface. Different events in the virus life cycle are triggered at different stages during this migration. In lesions (such as CIN1) caused by high-risk HPV types (such as HPV16), cells in the lower layers express E6 and E7 and are driven through the cell cycle and are stimulated to divide (cycling cells marked with red nuclei). In the mid layers, proteins necessary for genome amplification become elevated in these cells, allowing genome amplification to occur. These cells express the viral E4 protein and are typically in the S or G2 phases of the cell cycle (E4 presence marked in green, with red nuclei indicating replication competence). In the upper epithelial layers, the cells leave the cell cycle, and in a subset of the E4-positive cells, the virus L2 and L1 proteins are made, allowing packaging of the amplified viral genomes. The site of expression of the different viral gene products is shown to the left of the image, with the key stages during productive infection listed alongside. At the cervical transformation zone and the endocervix, it is thought that HPV may also be able to infect columnar epithelial cells, the epithelial reserve cells, and cells at the squamo-columnar junction. Infection of these cell types may be associated with different patterns of disease progression and with the development of adenocarcinoma. IARC: International Agency for Research on Cancer; PAE: Position of the early polyadenylation site; PAL: Position of the late polyadenylation site; PE: Early promoter, also referred to as p97; PL: late promoter, also referred to as p670.

The viral DNA can be found both in a non-integrated form inside the nucleus as well as an integrated form. The integrated form is more common in HPV associated cancers (Arias-Pulido H et al., 2006). After infecting the cells in the basal lamina, the HPV DNA copies will remain relatively stable (Doorbar J 2005). As the infected cells differentiate to close the wound, the E6 and E7 proteins in high-risk HPV degrade TP53 and Rb respectively, creating rapid division of the cell and amplification of the virus, and possibly inducing neoplasia (Doorbar J 2005). Virus particles from the initially infected region then infect other host cells.

Genital warts are also manifestations of sexually transmitted HPV infection, but only few will develop visible warts and only about 10% of those infected will transmit the virus (Leslie SW et al., 2021).

2.1.7 HPV prevalence in the general population

The global prevalence of HPV infections vary largely. Comparing cervical cancer (CC) caused by HPV, a study from 2012 estimated the population attributable fraction of CC to be 4.8% of all cancers globally, with 2.5% in the more developed regions of the world and 15% in less developed regions (Forman D et al., 2012). Newer research from 2021 looking at CC overall between 1990-2019, saw a decrease in incidence in the developed world, with post-soviet states with weaker health care seeing a strong increase in incidence (Sung H et al., 2021). HPV screening and vaccine is seen as one of the stronger contributing factors to decrease for some of the regions in the study (Ma X et al., 2021).

2.1.8 HPV and disease

As mentioned before most HPV types are situated in the skin and are asymptomatic, while some can cause skin common skin warts, while others are mucosal. The latter can cause genital warts, recurrent respiratory papillomatosis (RRP) and different types of cancer, e.g. cancer of the cervix uteri (cervical cancer-CC), other anogenital cancer and oropharyngeal squamous cell carcinoma (OPSCC), more specifically tonsillar and base of tongue squamous cell carcinoma (TSCC and BOTSCC) (Tommasino M 2014). The focus of my thesis is on mucosal HPV prevalence in the oral and cervical locations and dysplasia and biomarkers in HPV^{+} TSCC and BOTSCC. Below however, some words will be mentioned about both benign and malignant diseases associated with mucosal HPVs.

2.1.9 HPV and benign lesions

Genital warts

Types HPV 6 and 11 cause genital warts that present both separately and in clusters, and are found in the genital and anal area for more details see Yanofsky VR et al., 2012. About 10% of those infected with HPV will develop genital warts. Smoking is suggested to be correlated with the risk of getting genital warts and the recurrence rate of the warts depend on health, immune status as well as previous HPV vaccinations. However, 30% of most genital warts disappear within 4 months. Disfiguration is the most common complication, with possible malign transformation as an additional risk (Leslie SW et al., 2021). Genital warts are extremely common, and affect up to one million individuals in the United States

each year (Yanofsky VR et al., 2012). About 90% of these are estimated to be caused by HPV.

Recurrent respiratory papillomatosis (RRP)

In addition to genital warts, HPV6 and 11 also cause Recurrent Respiratory Papillomatosis (RRP) (Katsenos S and Becker HD 2011). RRP are papilloma's in the upper aero-digestive tract, and are like their genital counterpart very unreliable. RRP can cause a range of complications, to mild symptoms such as voice change, to serious airway of blockage, and can both spontaneously recede and reoccur. Much like with HPV-caused genital warts, there is a potential for malign transformation (Fortes HR et al., 2017). The lesions in RRP are typically cauliflower like, and usually occur in the transitional zone between the squamous epithelium and the ciliated columnar epithelium. This is similar to the pattern in the transformation zone in CC (Fortes HR et al., 2017, Elson DA et al., 2000). About 1.4-1.8 per 100 000 individuals are affected each year in the UK according to two studies (Armstrong LR et al., 1999, Donne AJ et al., 2017).

2.1.10 HPV and malignancies

Below a figure denotes the association of HPV and the prevalence of HPV to different types of cancers (Figure 5). This is followed by a short presentation of some types of mucosal malignancies caused by HPVs. Information on treatment will only be presented for cervical and oropharyngeal cancer.

Cancer of the uterine cervix and its treatment

Cancer of the cervix uteri, or cervical cancer (here abbreviated CC) is the best-known HPV induced cancer, with in 2006, >500 000 and ~250 000 deaths per year (Parkin DM and Bray F 2006). CC is found to be caused by HPV in >99% of all cases, but additional factors may also contribute to cancer development (Burd EM 2003, Herrington CS 1999). Moreover, CC is the 4th most frequently occurring cancer in women, with in 2018, around 570,000 new cancer cases and which at the time was 6.6% of all female cancers. There are two major types of CC, squamous cell carcinoma (SCC) and adenocarcinoma (AC), where the former accounts for the majority of the cases. The most

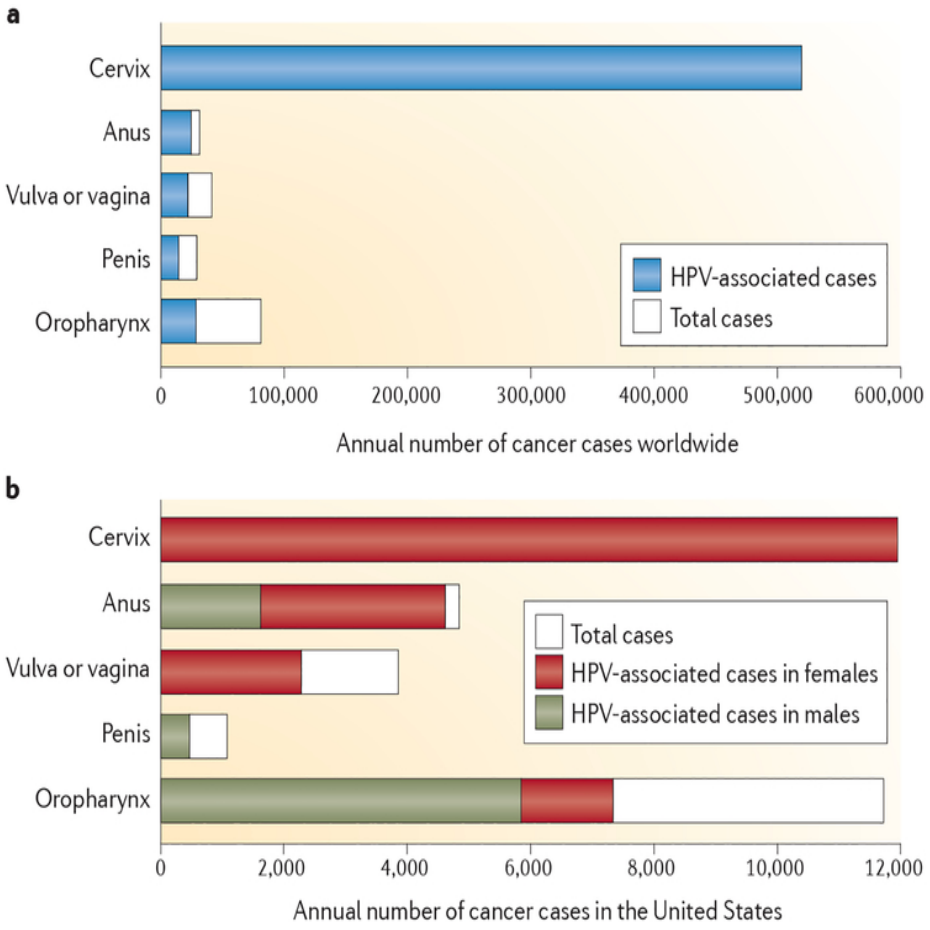


Figure 5. Annual numbers of HPV associated cancer cases. A, worldwide incidence and distribution of HPV-associated cancers. Red, HPV-positive cancer; white, HPV-negative cancer. B, incidence and distribution of HPV-associated cancers in the United States (data from refs. 3 and 4). The approximate percentage of HPV-associated cancer attributable to HPV16 and -18 is also shown. Red, HPV-positive cancer; white, HPV-negative cancer. (From Lowy DR and Schiller JT 2012, with permission from the publisher).

frequent HPV type associated to CC is HPV16 and this HPV type was responsible for half of all cases in Europe and the US before the introduction of HPV vaccination (Burd EM 2003). In parallel, HPV18, 31 and 45 together accounted for roughly 25-30% of CC cases (Burd EM 2003).

The process from infection to dysplasia to transformation takes place during a period of several years (Figure 6), thereby allowing for the possibility of screening and considerable but not complete prevention (von Knoebel M and Vinokurova S 2009, Dillner J 2019).

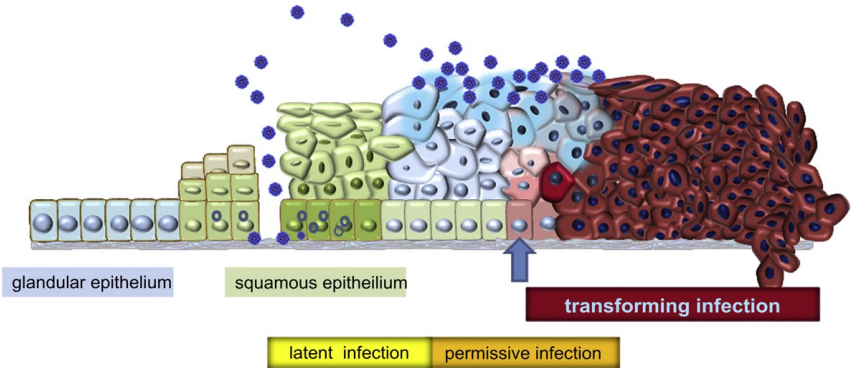


Figure 6. Schematic representation of the various modes of an HPV infection.
Minor lacerations of the squamous epithelium permit the virus to meet its natural host cell at the bottom of the squamous epithelium. Upon viral uptake, transport to the nucleus and release of the circular episomal viral genome genetic activity of the virus appears to be blocked (latent infection, dark green cells). Viral gene expression starts in individual cell, and, for unknown reasons, permits the local expansion of the infected cells into a permissive infection that results in viral replication and release of replicated viral particles at the surface of the squamous epithelium (light green basal cells, blue intermediate and superficial cells). In some instances and, particularly in basal cells at the transformation zone between squamous and glandular epithelial cells, the permissive or replicating mode of viral gene expression may shift into the trans- forming mode of viral gene expression. The latter is characterized by high-level expression of the E6 and E7 genes (red cells) (for further details please refer to the text of this review). (From von Knebel Doeberitz M and Vinokurova S, 2009, with permission from the publisher).

Precancerous lesions can be graded in different ways. For example, pre-neoplastic lesions in cervix can be divided into cervical intraepithelial neoplasia (CIN) with their grading (1-3), depending on the grade of epithelial dysplasia. CIN 1, represents mild abnormal cell growth encompassing maximum 1/3 of the basal epithelium, CIN 2, abnormal cell growth encompassing 2/3 of the basal epithelium and CIN 3 spans more than 2/3 of the epithelium and can effectively be classified as carcinoma in situ (Dillner J 2019). Similar grades can be applied in anus i.e. anal intraepithelial neoplasia (AIN1-3), vulva i.e. vulvar intraepithelial neoplasia (VIN1-3) and vagina, i.e. vaginal intraepithelial neoplasia (VaIN1-3). Since 2014 the world health organization (WHO), however, recommends a two-tier system in cervix, in order to increase histological reproducibility. Now the terms low-grade squamous intraepithelial lesion (LSIL), roughly corresponding to CIN1, and high-grade squamous intraepithelial lesion (HSIL) roughly corresponding to CIN2-3) are recommended in the cervix. The important fact is that they exist and are of use in order to prevent many cases of CC and this approach was first introduced by Papanicoloau G and Traut H, 1941. More about screening will be presented below.

Approximately, 90% of all deaths caused by CC occur in low- and middle-income countries, suggesting that education of the population, in order to obtain early diagnosis,

prevention, effective screening methods and therapy should be further improved in such countries to decrease the numbers of patient deaths (World Health Organization. Cervical Cancer. 2018).

Hopefully, HPV vaccination will eventually prevent the most CC cases, but great efforts will be needed before vaccination coverage will be sufficient enough to cover women in the poorest areas (Lowy DR and Schiller JT 2012). However, this is also dependent to a high degree on different attitudes with regard to HPV vaccination in various areas.

Treatment. The typical treatment for CC is chemotherapy (CT) combined with concomitant radiotherapy (RT) and around half of the patients survive (Eifel PJ 2006). However, this is not entirely true since 80% of deaths are in less developed parts of the world and suggested to be due to lack of screening and later detection (Arbyn M et al., 2020). There is some support for this as patients at diagnosis often present with local disease at an advanced stage, or with metastasis, possibly contributing to the diseases high mortality rates. In 2005, the survival rate was at 16% over a five-year period for patients with metastasis (Tewari KS and Monk BJ 2005). As the survival rates are poor, there seems to be less focus on quality-of-life improvement after treatment.

Vulvar, Vaginal and Penile Cancer

Vulvar, vaginal and penile cancers are uncommon, and account for >1% of malignant tumours in Europe and North America (Parkin DM and Bray F, 2006).

Anal cancer

HPV infection is an important risk-factor for anal cancer, with roughly 90% of anal cancers caused by HPV (Parkin DM and Bray F 2006).

Tonsillar and base of tongue cancer (TSCC and BOTSCC), treatment and biomarkers

Tonsillar and base of tongue squamous cell carcinoma (TSCC and BOTSCC). In the 1980-1990s HPV was proposed to be causative also of head neck cancer (Syrjanen KJ and Surjanen SM 1981, McKaig RG et al., 1998). Subsequently, HPV was reported to primarily be correlated to the development of oropharyngeal squamous cell carcinoma

(OPSCC), and then more specifically TSCC and BOTSCC the major subsites of OPSCC and accounting for more than 80% of the OPSCC cases (Gillison ML et al., 2000, Mellin H et al., 2000, Dahlgren L et al., 2004). However, it was in 2007, that IARC first acknowledged HPV as a risk factor for OPSCC and TSCC (IARC 2007).

In parallel, several reports have shown a rise in the incidence of TSCC and BOTSCC the past four to five decades (Conway DI and Stockton DL 2006, Sturgis EM et al., 2007) and closely after that it was revealed that this increase depended on HPV infection and a specific increase of mainly HPV^+ TSCC and BOTSCC (Näsman A et al., 2009, Attner P et al., 2010). Furthermore, patients with HPV^+ cancer had a notably better prognosis than those with HPV^- tumours (80% compared to 40% 5-year disease free as well as overall survival) (Mellin H et al., 2000, Attner P et al., 2011). Therefore, having information regarding the HPV status of a tumour prior to therapy would likely be very useful for individualizing treatment, since patients with HPV^+ cancer in general have a more favourable outcome. HPV16 is currently reported to account for roughly 90% of all HPV^+ TSCC and BOTSCC, which notably also are mostly found to affect men (Ramqvist T, Grün N et al., 2015).

Most carcinomas develop in a multistep fashion with different stages of dysplasia that is followed by high-grade dysplasia/carcinoma in situ and subsequently invasive disease. Grading of dysplasia in the head and neck region can however be difficult and the different stages of dysplasia are not that well-defined, as e.g. in CC (described above). As an example of the difficulties with grading of dysplasia, the WHO classification 2017 of staging only acknowledges different staging systems in the larynx and the oral cavity but do not recommend any particular system, including not recommending 2-tier or 3-tier scales. However, while the occurrence of dysplasia as a pre-cancerous lesion in head and neck carcinomas in general is adopted, the occurrence of dysplasia in HPV mediated OPSCC is debated. Interestingly, some authors have even postulated that HPV mediated premalignant phases of OPSCC do not exist. Notably, the WHO 2017 edition states that “dysplasia of the surface epithelium is rarely identified” in HPV mediated oropharyngeal cancers, as compared to HPV negative tumours.

Treatment. TSCC and BOTSCC therapy is like that of other head neck squamous cell carcinoma (HNSCC) often aggressive, as these tumours especially the HPV-positive ones, have a nodal spread (Näsman A et al 2021). Patients with stage III-IV disease are treated with

chemotherapy (CT) and radiotherapy (RT) i.e. chemoradiotherapy (CRT) and sometimes supplemented with epidermal growth factor receptor (EGFR) blockers (Licitra L et al., 2002). Treatment also depends on the patient's condition (Licitra L et al., 2002). Severe side effects such as swallowing, nausea, mucositis, and systemic infections and fatigue are often a result of these treatments and patients with persistent lymph node metastases are often treated with surgical dissection in the neck region with even more side effects. Consequently, CRT with surgery causes even more side-effects, with e.g. extensive stiffness of the neck, and more severe issues with swallowing. Other long-term adverse effects are taste alterations, partial deafness, and possibly osteonecrosis requiring reconstructive surgery all affecting long term quality of life.

This intensified therapy has despite changes in TNM classification of malignant tumours (TNM7 to TNM8), not increased survival for patients with HPV^+ TSCC/BOTSCC with worse prognosis, so individualizing therapy if possible is paramount for these increasing numbers of patients with HPV^+ TSCC and BOTSCC, and most do not need aggressive therapy (Näsman A et al., 2017, Näsman A et al., 2021).

Therefore much of our work has been to find biomarkers that could be of assistance for predicting prognosis for HPV^+ TSCC and BOTSCC and in addition, to find biomarkers which could predict sensitivity to specific targeted therapy. Below a chapter on biomarkers is presented.

Biomarkers in tonsillar and base of tongue cancer according to HPV status. Initial efforts to disclose biomarkers in these tumours, were done to characterize or to find potential differences between HPV^+ and HPV^- TSCC and BOTSCC or to find similarities between different HPV^+ tumours. Early studies showed e.g. similarities between HPV^+ TSCC and BOTSCC and HPV^+ cervical and vulvar cancer, such as e.g. p16 overexpression, the rare presence of a p53 mutation and/or the amplification of specific parts of chromosome 3q (for review see Näsman A et al., 2021, and also Dahlgren L et al., 2003, Crook T et al., 1992 and Wilting SM et al., 2009).

Investigating for prognostic and targetable markers came later, with focus on using immunohistochemistry (IHC) and studying immunological and stem cell markers. Today, many consistently report that the presence $CD8^+$ lymphocytes infiltrating, or in the vicinity of the tumour, are numerically higher in HPV^+ TSCC, BOTSCC than in corresponding HPV^-

tumours and that having high numbers of $CD8^+$ cells was associated to better prognosis and survival (Näsman A et al., 2012, Nordfors C, Grün N, Tertipis N et al., 2013, Oguejiofor K et al., 2015, Oguejiofor K et al., 2017, Tertipis N et al., 2015, Welters MJP et al., 2020). In addition, finding a low $CD4^+/CD8^+$ ratio, or a high $CD8^+/FoxP3^+$ ratio in TSCC/BOTSCC was correlated to improved clinical outcome (Näsman A et al., 2012, Nordfors C, Grün N, Tertipis N et al., 2013). In other studies, $CD68^+$ $CD163^+$ M2-macrophages were also examined and the degree of infiltration of these macrophages was correlated with worse outcome in HNSCC, with the majority of the cases being OPSCC (Balermpas P et al., 2014, Cioni B et al., 2019, Santegoets SJ et al., 2020). In another report, the expression of the programmed death ligand 1 (PD-L1) as a biomarker was investigated in $CD68^+$ macrophages, and in HPV^- TSCC/BOTSCC, and it was shown that high numbers of $CD68^+$ macrophages expressing PD-L1 tended to present a positive immune environment (Oguejiofor K et al., 2017). In a similar way, in HPV^+ cancers, infiltrated by $CD8^+$ and $CD68^+$ immune cells with a high PD-L1 expression, clinical outcome was positive (Young RJ et al., 2020).

During this period, we found that some markers had high sensitivity and covered a limited number of patients with good prognosis, while others had lower sensitivity and covered more patients, so we tried to apply different models for optimizing prognostication (Tertipis N et al., 2015, Bersani C, Mints M et al., 2017). These were useful to some extent, but the conclusion was that additional markers could still be use.

Additional studies involved molecular techniques, using next generation sequencing (NGS) in HNSCC and many different features were observed when comparing HPV^+ and HPV^- cancers, including the possible use of targeted therapies (Lui VW et al., 2013, Sewell A et al., 2014, Gaykalova DA et al 2014, Chung CH et al., 2015, Rusan M et al., 2015, Tinhofer I et al., 2016, Bersani C, Sivars L et al., 2017, Cancer Genome Atlas, N. 2015 and for review see Näsman A et al., 2021).

HPV^+ TSCC, BOTSCC or OPSCC mostly presented mutations in the Phosphatidylinositol-4,5-Bisphosphate 3-Kinase Catalytic Subunit Alpha (*PIK3CA),* notch homolog 1 translocation-associated *(NOTCH1)*, and Fibroblast growth receptor (*FGFR) 3* genes, while HPV^- tumours frequently presented mutated *TP53* and cyclin dependent kinase inhibitor 2A/B (*CDKN2A/B)* (Lui VW et al., 2013, Sewell A et al., 2014, Gaykalova DA et al., 2014, Chung CH et al., 2015, Rusan M et al., 2015, Seiwert T et al., 2015, Tinhofer I et al., 2016, Bersani C et al., 2017, Cancer Genome Atlas, N. 2015 and for review see Näsman A et al.,

2021). The value for predicting prognosis with regard to the mutations in the above genes varied with discrepancies between the different studies (see review Näsman A et al., 2021). Nonetheless, many inhibitors target *PIK3CA* and *FGFR3*, and the Food and Drug Agency (FDA) has recently approved the phosphoinositide 3-kinase inhibitor (PI3K inhibitor, alpelisib (BYL719) for use clinically for advanced breast cancer and the FGFR inhibitor, erdafitinib (JNJ-42756493) for advanced bladder cancer (Isaacson VPH et al., 2015, Leenhardt F et al., 2021, Tabernero J et al., 2015). These two inhibitors categories, may very well be of use in HPV^+ TSCC, BOTSCC and OPSCC, due to that around 20 % and 10 % respectively of them have *PIK3CA* and/or *FGFR3* mutations respectively (Tinhofer I et al., 2016, Bersani C, Sivars L et al., 2017).

MicroRNA (miR) expression in HPV^+ and HPV^- TSCC, BOTSCC, has also been studied but the data are often variable. Nevertheless, miR-9, 155 and 163b were in a number of studies shown to be overexpressed in HPV^+ when compared to their presence in HPV^- OPSCC, while in contrast miR-31 and 193b were downregulated (Hui AB et al., 2013, Gao G et al., 2013, Lajer CB et al., 2011, Lajer CB et al., 2012, Miller DL et al., 2015, for review see Näsman A et al 2021). In another study, overexpression of miR-142-3p, 146a, 26b was associated to better clinical outcome, while the opposite was shown for expression of miR- 31, 24, 193b (Gao G et al., 2013). Our group has also reported that high miR 155 expression was correlated to a favourable outcome in patients with HPV^+ TSCC and BOTSCC, while the converse was found for patients with tumours with a high miR 185 expression, but in that study miR 193b expression was not associated to survival (Bersani C et al., 2018). To conclude, there are several studies on miR expression and clinical outcome and in TSCC, BOTSCC and OPSCC but the data vary considerably, and we therefore suggest that more extensive knowledge is needed prior to the clinical use of miRs as predictive biomarkers.

There are also reports on the transcriptome of HPV^+ and HPV^- TSCC, BOTSCC and OPSCC. Some focus on expression of specific HPV mRNAs in HNSCC, while others examine all mRNA types, with a few already mentioned above (Ramqvist T, Mints M et al., 2015, Campo MS et al., 2010, Li H et al., 2006). Others find that there are differences in immune responses, proliferation, apoptosis and the cell cycle when comparing HPV^+ and HPV^- TSCC, BOTSCC and OPSCC, which could be expected to some extent (Mirghani H et al., 2014, Martinez I et al., 2007, Wichmann G et al., 2015).

There are also a few but a limited number of studies on protein profiling in TSCC, BOTSCC and OPSCC (Sewell A et al., 2014, Slebos RJ et al., 2013, Ramqvist T et al., 2018). In one study, utilizing Olink multiplex immunoassays in fresh frozen samples from 42 HPV^+ and 17 HPV^- TSCC and BOTSCC in comparison to normal tissue, researchers in our group found some proteins related to angiogenesis and hypoxia tended to be associated to survival (Ramqvist T et al., 2018). For example, a high expression of vascular endothelial growth factor A (VEGFA) was correlated to worse prognosis in HPV^+ cancer and it is possible that angiogenesis related proteins could be potential targets for future treatment in HPV^+ TSCC and BOTSCC (Ramqvist T et al 2018).

Lately, the oral microbiome has been found to act as an interesting marker in HNSCC (Chen Z et al., 2020). Moreover, a lower diversity of microbiota was found in HNSCC than that observed in healthy controls and in addition a different bacterial taxonomy was suggested to possibly be used for distinguishing oral cancer samples from OPSCC and normal samples (Guerrero-Preston, R et al., 2016).

In addition, bacteria species such as *Fusobacterium nucleatum* and *Actinobacteria,* often shown to contribute to carcinogenesis upon *in vitro* experiments, are often disclosed to be more frequently observed in the HNSCC oral microbiome (Chen Z et al., 2020, Guerrero-Preston, R et al., 2016, Hayes RB et al., 2018). Therefore, despite that there are huge variations in the significant changed bacteria, similar signatures for example the enrichment of pro-inflammatory features have been found to be more prominent in oral squamous cell carcinoma patients (Chen Z et al., 2020).

2.1.11 HPV screening

HPV and cervical cancer (CC)

Screening programs for detecting cancer pre-stages in the cervix are well established and have reduced the cancer burden significantly (McGraw SL and Ferrante JM 2014). Furthermore, over the years these programs have changed considerably. From previously, performing (Papanicoloau-test) PAP-test by doctors and midwives, today more and more screening programs have focused on HPV testing and self-tests and many efficient possibilities have been presented (Kyrgiou M et al., 2020, Dillner J 2019, Dillner J et al.,

2021, Hortlund M et al., 2021, Partanen VM et al., 2021). The different screening methods have a lot of benefits, and may improve prevention, however, none of these methods are perfect. The introduction of HPV vaccination will also improve prevention. However, HPV vaccination coverage varies a lot by region and trends in trust towards vaccines, screenings for cervical pre-cancerous lesions will likely still stay relevant in the future.

HPV and tonsillar and base of tongue cancer (TSCC and BOTSCC)

Screening programs for pre-stages in the tonsillar and base of tongue regions are not established (Paper III). Although one can suspect pre-cancerous lesions should be present to some extent similar to that observed in CC they are likely not so readily observed. Furthermore, to screen for presence of HPV in the oral cavity as a risk factor for TSCC and BOTSCC is likely much more complicated as compared to screening for CC. This is due to that detection of HPV in the oral cavity is much more complex since the production of 0.5-1.5 litres of saliva per day dilutes the signal (Nordfors C et al., 2014).

Other publications have shown the presence of antibodies against HPV16 early antigen in serum before the development of an oropharyngeal cancer, but whether this is practical today remains to be further explored, for a systematic review see Hibbert J et al., 2021.

2.1.12 HPV prevalence in the general population

The global prevalence of HPV induced cancers varies largely. Comparing CCs caused by HPV, a study from 2012 estimated the population attributable fraction to be 4.8% globally, around 2.5% in the more developed regions of the world and about 15% in less developed regions (Forman D et al., 2012). Newer research from 2021 looking at CC overall between 1990-2019, saw a decrease in incidence in the developed world, with post-soviet states with weaker health care seeing a strong increase in incidence (Sung H et al., 2021). HPV screening and vaccine is seen as one of the stronger contributing factors to decrease and increase respectively for the regions in the study (Ma X et al., 2021).

2.1.13 HPV vaccines

Knowledge of L1s capacity to self-assemble into virus-like-particles (VLPs) similar to that of VP1 of polyomaviruses has been around since some decades (Salunke DM et al., 1989, Kirnbauer R et al., 1992). However, at the time industrial production was still not developed. The first commercially available vaccine for HPV was Gardasil (Merck&CO. I. Package Insert - Gardasil. 2006). In 2006, it was licensed by the US Food and Drug Administration (FDA). A bivalent vaccine, Cervarix (GlaxoSmithKline. Package Insert - Cervarix. 2009) was approved shortly after by the European Medicines Agency (EMA) in 2007. Gardasil and Cervarix cover HPV16 and 18 the most common HPV types in CCs (causing 70% of CCs) and moreover Gardasil covers HPV6 and 11 that cause considerable number of condylomas and RRPs (de Sanjose S et al., 2010).

The third and most recent HPV vaccine, Gardasil 9 (Merck&CO. I. Package Insert - Gardasil 9. 2014) is nonavalent vaccine and covers the same types as Gardasil, and an additional five types i.e. in total HPV6, 11, 16, 18, 31, 33, 45, 52, and 58. Gardasil 9 would cover an additional 20% of CCs, totalling to about 90% of CC cases (Yang DY et al., 2016, Cheng L, Wang Y et al., 2020). The three vaccines all consist of synthetically manufactured VLPs made up of the L1 epitope of different HPV types, and with different adjuvants and antigenic load (Merck&CO.I Package Insert – Gardasil. 2006, GlaxoSmithKline. Package Insert – Cervarix 2009, Merck&CO.I. Package Insert – Gardasil-9. 2014). The current method of production of L1 is to use yeast cells (in Gardasil) or baculovirus (in Cervarix). All vaccines are suggested to be administered from 9 years of age (https://www.who.int/teams/immunization-vaccines-and-biologicals/diseases/human-papillomavirus-vaccines-(HPV)). The current recommendation is two doses for patients up to 16 years of age, and three doses for older patients (Harper DM and DeMars LR 2017).

In Sweden 2012, a school-base vaccination program was introduced for girls aged between 10-12 years. Until 2020, when the HPV vaccine was introduced also for boys, the vaccination programs in Sweden were girl-exclusive, but parents could get vaccines for their boys privately. A Finnish population based trial found support for gender-neutral vaccination programs, based on a significant increase in herd-immunity for the gender-neutrally vaccinated group (Lehtinen M et al., 2018).

Gender-neutral vaccination programs are rare, but found in e.g. Denmark, Argentina, as well as regions of the USA. As of June 2020, 55% (107/194) of the WHO members states have established HPV vaccination (Bruni L et al., 2021). The global target for CC elimination set by WHO is at 2030, and has three goals. For 90% of girls fully vaccinated with HPV by 15 years of age, for 70% of women to be screened twice before the age 45 years, and for 90% of those with identified cervical disease to receive an appropriate treatment. Unfortunately, reduced confidence in the vaccines have reduced the vaccine coverage in some countries, with more people opting out of vaccination programs. However, a review in 2018 found any side effects to be outweighed by the benefits of the vaccine (Philips A et al., 2018). A set of 109 safety studies were analysed, with patients from 6 different countries, making up 2.5 million vaccinated individuals in total. Gardasil 9 was found to be slightly more prone to give side-effects (Philips A et al., 2018). Common side effects from Gardasil, Gardasil 9 and Cervarix are injection-site swelling, pain, while serious effects are uncommon (Cheng L, Wang Y et al., 2020).

2.2 MICROBIOTA

For long, there has been a discussion, why some individuals develop cancer after HPV infection, while others do not. There have been discussions about co-factors such as immunity and persistence, but more recently a new player has been introduced and below some words about microbiota.

2.2.1 Gut microbiota

Microbiota is the composition of the micro-organisms inhabiting an environment. Many microbiotas have an intimate relationship with multicellular organisms, causing disease but more often developing a symbiotic relationship with the host (De Sordi L et al., 2017). In the gut, high species diversity of the microbiota is often correlated with health and significant lower microbiota diversity and richness are linked to diseases (Sommer F et al., 2017). Many diseases including cancer had been correlated with altered microbiota, although causation versus correlation is not always clear (Cullin N et al., 2021). On a species level, how a micro-organism impacts the host is often multi-factorial and there are multitude of ways of interaction between the micro-organism with the host, and with other members of the microbiota.

One important effect of healthy microbiota is colonization resistance towards other species (especially pathogenic ones). Colonization resistance can be both direct and indirect. Direct can be niche-competition such as competing for the same nutrition in the gut, or direct antagonism, such as bacteriophages attacking bacteria, while indirect can be via metabolites or immune molecules (De Sordi L et al., 2017). Some bacteria could cause pathogenesis throughout several microbiotic regions, like *porphyromonas gingivalis,* that is associated with periodontitis in the oral region, inflammatory bowel disease in the gut (Sohn J et al., 2021).

Gut microbiota is one of the more explored regions in microbiota research (De Sordi L et al., 2017). Patients suffering from inflammatory bowel disease (IBD) having a different microbiota from individuals with a healthy gut. An example of a well-studied mutualistic bacteria protecting against disease is *Akkermansia muciniphilia.* In studies on inflammatory bowel disease (IBD) patients, AM was found at a higher amount in the gut microbiota of healthy patients than IBD patients (van Passel MW et al 2011). Another common pathology associated with gut microbiota is antibiotic-associated-diarrhea. About 20% of these are caused by *Clostridiodes Difficile* Infection (CDI) (Lessa FC et al., 2015)). The burden of CDI is quite high, with 400 000 cases and 29 000 deaths in 2011 (Lessa FC et al., 2015). Recently, faecal microbiota transplantation (FMT) has showed very promising result in treating the CDI, suggesting an attractive approach of microbiota-based therapy and essential role of gut microbiota (Sunkara T et al., 2018).

2.2.2 Vaginal microbiota

Composition of healthy vaginal microbiota

The vaginal microbiome is usually inhabited by one dominant species of *Lactobacillus* (e.g., *Lactobacillus crispatus, Lactobacillus gasseri, Lactobacillus iners,* or *Lactobacillus jensenii*) (Coudray MS and Madhivanan P, 2020, Cheng L, Norehag et al., 2020, Serrano MG et al., 2019). However, there are women whose vaginal microbiome is dominated by other non-*Lactobacillus* species, and this occurs more often in Black or Hispanic women (Ravel J et al., 2011, Serrano MG et al., 2019). Vaginal microbiota had been shown play an important role in women's health, and different to the gut microbiota where diversity is

considered healthy, the healthy vaginal microbiota has a low diversity (Coudray MS and Madhivanan, P 2020, Cheng L, Norenhag J et al., 2020, Serrano MG et al., 2019).

Bacterial vaginosis (BV)

Bacterial vaginosis (BV) is a disease in which the vaginal microbiome shows an overgrowth of obligate and facultative anaerobes, which thereby disturbs the vaginal microbiome balance (Cheng L, Norenhag J et al., 2020). BV is a frequent and often recurring vaginal infection in women of reproductive age and is not only correlated with adverse health outcomes but also with a decreased quality of life.

The bacterial families commonly associated with BV are *Gardnerella, Atopobium, Prevotella, Bacteroides, Peptostreptococcus, Mobiluncus, Sneathia, Leptotrichia,* and *Mycoplasma* (Oakley et al., 2008). Several studies have confirmed a group of bacteria of the *Clostridiae* family as associated with BV so called "Bacterial Vaginosis Associated Bacteria" (BVAB) (Fredricks et al., 2005, Ugur et al., 2019). Women with BV may present symptoms such as vaginal discharge, increased vaginal pH, fish-like odour, and burning when urinating, but they may also lack symptoms (Donders D et al., 2010).

The mechanisms for acquiring BV are not entirely known, however, risk factors have been shown to have many sex partners and douching, i.e. cleansing of the vagina or a natural *lack of lactobacilli (Mayo clinic data).*

High vaginal microbial composition has been shown to be more susceptible to sexually transmitted diseases, including human immunodeficiency virus (HIV), herpes simplex virus, chlamydia, gonorrhoea and HPV (Torcia MG 2019, Petrova MI et al., 2013, Cheng L, Norenhag J et al., 2020, Brusselaers N et al., 2018). BV is also related to poor fertility treatment outcome (Koedooder R et al., 2019), premature deliveries and low weight babies and preterm birth (Fettwiss JM et al., 2019).

Furthermore, there are studies showing that HR-HPV clearance was less efficient in women harbouring *Lactobacillus iners* or a subgroup of vaginal microbiota dominated by *Gardnerella, Prevotella, Atopobium, Megasphoera* as compared to *Lactobacillus crispatus* dominated group (Torcia MG 2019).

The currently recommended treatment for BV is antibiotics, in spite a high recurrence rate over one year (Coudray MS and Madhivanan, P 2020). However, live biopharmaceutical products/probiotics as well as vaginal microbiome transplantation (VMT) have also been explored in clinical trials for BV (Coudray MS and Madhivanan, P 2020).

2.2.3 Oral Microbiota

Like is the case with vaginal and gut microbiota, there are commensal opportunistically pathogenic microbiota in this region as well. *Leptotrichia Buccalis* is a normal constituent of oral flora, but in patients with neutropenia (lack of neutrophile immune cells in blood) this bacterium can cause severe infections (Morgenstein AA et al., 1980). Oral microbiotas are often associated with periodontitis or maladies of the gum-tissue and bone surrounding teeth. *Prevotella* is a gram-negative bacterium that is involved in periodontal disease (Tanaka S et al 2008). The genera *Phorphorymonas* is found both in gut and oral microbiota. As previously mentioned (see Gut Microbiota) *Porphyromonas* is associated with periodontal disease as well as possibly also oropharyngeal cancers (Sohn J et al., 2021, Sun J et al., 2021).

Interestingly *Prevotella* is not only associated with dental ligaments, but also rheumatoid arthritis (Scher JU et al., 2013). The most common HPV type in the oral cavity is HPV16 according to a systematic review by Shigeshi H et al., 2016. It is not entirely clear however if the correlation with HPV16 in the oral cavity and bacteria is a causal one. HPV16 is more frequently found in the oropharynx, which is argued to be because of capture into the tonsillar crypts (Faraji F et al., 2017). Inflammatory periodontal pockets may serve as a similar capture point for HPV16, rather than any interaction with the microbiota and HPV itself. A study found mixed results when looking at elderly patients in Japan (Shighesi H et al., 2021). *Prevotella and Veilonella* were prevalent in patients with deep pockets and no HPV16, while *Phorphorymonas* was more common in patients with HPV16. However, the patient group was quite small, and how and if HPV16 interacts with certain genus of bacteria in the oral cavity remains to be seen.

The oral microbiome is not discussed further in this thesis but presented since we investigate HPV in the oral cavity.

3 RESEARCH AIMS

- To follow up the degree coverage of HPV vaccination at a youth clinic in Stockholm one decade after the introduction of HPV vaccination, and to investigate possible vaccine effects on different HPV types.

- To analyse possible associations between HPV status, age and vaccination status and the vaginal microbiota, of women in cohorts from Uppsala and Stockholm.

- To analyse gene expression patterns in high-grade dysplasia and invasive cancer in HPV^+ and HPV^- TSCCC/BOTSCC, and compare with particular emphasis to the HPV status.

- To perform whole-exome-sequencing (WES) on primary tumours of patients with HPV^+ TSCC/BOTSCC with and without recurrences, with the aim to identify similarities and differences as well as to identify markers of prognostic significance or candidates for targeted treatments.

4 STUDY INDIVIDUALS

Study individuals were enrolled from a youth clinic in the central of Stockholm, and maternity clinics in Uppsala and from the Karolinska University Hospital.

4.1 YOUNG ADULTS ATTENDING A YOUTH HEALTH CARE CENTER

Young women and men aged 15-23 years attending a youth clinic in the centre of Stockholm were included in papers I and II (Ährlund-Richter A et al., 2019 and Cheng L Norenhag J et al., 2020 resp.). At the time this investigation was initiated, roughly 4000 women and 800 men attended this specific youth clinic in Stockholm, for birth control advice and/or for therapy because of sexually transmitted diseases. Participation was anonymous and voluntary and the written consent included limited data, more specifically only data on the year and month of birth and HPV vaccination and for the microbiota studies, antibiotic status were included. The numbers of individuals included can seem relatively low, but enrolment was not possible in the periods with a high workload for the midwives, however when approached most individuals did in fact participate.

In paper I, data from almost1000 individuals examined in previous studies 2008-2010 and 2013-2015 were included and compared to more recent data from 2017-2018, where cervical were collected from 178 young women. In paper II, from the same youth clinic in Stockholm above, >200 were enrolled and 169 women and girls were eventually included in the study, after excluding a few due to antibiotic usage or for incomplete clinical information (for further details see paper II).The age of those enrolled was between 14 and 23 years.

In the ethical application, we applied for the exception of acquiring approval from the parents of the youth below 18 years of age, since we anticipated it may complicate possible sensitive situations among the youth and this was approved according to the below permissions. The permissions for the studies of HPV prevalence in the cervical tract and for the studies of microbiota for the youth clinic were for paper I 2008/813-31/2, 2008/870 and 2017/725-31 and for paper II 2012/1756-31/2 with addition approved 2016-07-04, 2017/725-3, 2019-04201 all approved by the Stockholm Regional Ethics Committee.

4.2 UPPSALA MATERNITY HEALTH CLINIC PATIENTS

In paper II, from a gynaecological clinic in Uppsala, a total of 139 patients were enrolled from standard cervical screenings. Standard cervical screenings at the maternal health clinic provided 133 patients, and follow-up screenings at the maternal clinic provided 6 patients, however eventually 88 samples were finally included in the complete analysis due to incomplete clinical information on antibiotic usage or to presence of too little DNA after extraction. Collection of samples was performed by the regular clinical staff. The Uppsala patients were between 23 and 29 years of age.

Permission 2016/517 was obtained for these studies from the Uppsala Regional Ethics Committee

4.3 PATIENTS WITH TONSILLAR AND BASE OF TONGUE SQUAMOUS CELL CARCINOMA

Patients with tonsillar and base of tongue squamous cell carcinoma diagnosed between 2000 and 2015 in Stockholm, were found by using the Swedish Cancer Registry and utilizing the International Classification of Diseases system (ICD-10: C09.0, C09.1, C09.8, C09.9, C01.9) In paper III (Haeggblom L et al., 2019) a total, 24 patients diagnosed 2007-2015 were a included and in Paper IV (Ährlund-Richter A et al., 2021) a total of 40 patients diagnosed between 2000-2014 were initially included. Tumour classification and staging had been done previously (Table 1) and was taken from patients case reports.

Table 1. Tumour classification and staging[1]

[1]Classification of tumour stage in these papers was done according to the International Union against Cancer. The mandatory parameters are listed below. (Adapted also partly from Dr Juan Du's thesis with her permission)
T: size or direct extent of the primary tumour
Tx: tumour cannot be evaluated Tis: carcinoma in situ T0: no signs of tumour T1, T2, T3, T4: size and/or extension of the primary tumour
N: degree of spread to regional lymph nodes
Nx: lymph nodes cannot be evaluated N0: tumour cells absent from regional lymph nodes N1: regional lymph node metastasis present; (at some sites: tumour spread to closest or small number of regional lymph nodes) N2: tumour spread to an extent between N1 and N3 (N2 is not used at all sites) N3: tumour spread to more distant or numerous regional lymph nodes (N3 is not used at all sites)
M: presence of metastasis
Mx: distant metastasis cannot be evaluated M0: no distant metastasis M1: metastasis to distant organs (beyond regional lymph nodes)

More specifically, in paper III pre-treatment tumour samples from a large consecutive cohort including patients with TSCC diagnosed 2007-2015 previously tested for presence of HPV DNA and p16 overexpression (here defined as having HPV^+ status), were immunohisto-morphologically reviewed. In total, 13 HPV^+ positive cases were identified harbouring both high-grade dysplasia and invasive tumour in the same tumour section. In addition, 13 cases with HPV^- TSCC and presence of high-grade dysplasia and invasive tumour were also randomly selected. However, due to technical issues (not enough material left in tissue block for research purposes) two HPV^- cases were lost, resulting in 13 + 11 cases of HPV^+ and HPV^- cases respectively.

In paper IV, all patients were identified from a large consecutive cohort of patients diagnosed with TSCC and BOTSCC, previously clinically characterized and tested for presence of HPV DNA and p16 status. Patients with HPV DNA and p16 positive TSCC/BOTSCC having local/distant recurrence within 5 years after treatment and with available material both from primary tumour, as well as the recurrent tumour, were included as "cases" (n=20). From the same large consecutive cohort, 20 additional patients with no evidence of disease 5 years after treatment and with available tumour material were included as "controls".

The studies were conducted according to ethical permissions 2005/431-31/4, 2005/1330-32, 2009/1278-31/4 2017/1035-31/2 and 2018/870-32 for the head and neck cancer region from the Regional Ethical Committee at Karolinska Institutet, Stockholm, Sweden.

4.4 CONSIDERATIONS REGARDING THE STUDY INDIVIDUALS

The Stockholm youth clinic. Individuals here were aged 15-23 years of age. It can be argued that the participation was low, but this was due to the heavy workload of the midwives most of the time, with few opportunities for informing the youth and sample collection. It can also be argued that this youth population had more sexually transmitted diseases than average youth. However, we followed this youth clinic over time so the relative decrease should be reasonable. In addition, studies on HPV prevalence were also conducted elsewhere during the same period of in a school room setting and similar results were obtained (Nordfors C, Grün N Haeggblom L et al., 2013)

The Uppsala clinic. Individuals here were aged 23-29 years of age, i.e. slightly older than the group above. However, this is due to the feature of the age cut-off from different clinics in Sweden. Furthermore, in paper II we found correlation between vaginal microbiota and HPV infection even after adjusting the age.

The Karolinska University Hospital. Individuals here were middle aged and cancer patients. In paper III, two cases were lost due to technical reasons and patients with HPV^+ and HPV^- tumours were not matched on any clinical parameter.

In paper IV, we only included patients with available FFPE material both from the primary as well as the recurrent tumour. The reason was to also assess differences between primary and local/distant relapse. However, most patients with distant metastasis were diagnosed with cytology only. Besides, patients with metastasis to bone tissue were also excluded due to the heavy decalcification treatment of tumour samples at the pathology unit. However, as described below, the DNA quality of the recurrent tumours were often poor and were therefore in some cases excluded from further analysis. In retrospect, it would perhaps have been better to include all patients with local/distant relapse independent of if they had material available in order to increase the sample size of the study.

5 MATERIALS

5.1 CERVICAL SWABS

Study individuals were enrolled from a youth clinic in the central of Stockholm, and maternity clinics in Uppsala (Papers I and II). The majority of the vaginal swab samples collected in the youth clinic were self-collected, while collected by clinic staff in the Uppsala clinics. However, a previous study 2018, found no difference between the microbiota in self-collected and staff-collected samples (Hugerth LW et al., 2018). Likewise, in a study by our group, no major differences were found regarding the presence of HPV depending on whether the samples were collected by the staff or if they were self-sampled (Ramqvist T et al., 2011).

In both paper I and II, vaginal swabs (FLOQSwabs™, Copan Flock Technologies, Brescia, Italy) were used. The cervical swabs were inserted about 2-3 cm into the vagina and swirled for about 30 seconds and then distributed into FluidX tubes (Brooks Life Sciences, Chelmsford, MA, USA) containing 0.8 ml of DNA/RNA Shield (Zymo Research Corp, Irvine, CA). The tubes were stored at 4°C and collected within three days and stored at -20 until further DNA extraction (See methods).

5.2 CANCER TISSUES

All primary tumour samples and samples from recurrences or distant metastasis from patients with TSCC or BOTSCC were pre-treatment samples and collected as formalin fixed embedded (FFPE) tissues. Moreover, for verification of the initial diagnosis a second pathologist reviewed all cancer specimens.

Considerations regarding the study materials

Cervical swabs. The same procedures for collection were applied 2017-208, to that of the initially collected swabs and the in early consecutive studies, with the exception of that in the earlier studies the cervical swabs were instead inserted into 5 ml of SurePath preservation solution (Ramqvist T et al., 2011).

In paper I and II, all cervical swabs were instead preserved in DNA/RNA Shield (Zymo Research Corp, Irvine, CA, which is mainly to lyse bacteria and human cells and at the same time preserving both DNA and RNA. This should however not have affected the outcome of the extracted human, bacterial or HPV DNA preparation (Paper II).

Cancer tissues. All samples were less than 20 years old, and most less than 10 years old, which should therefore not affect the extraction of DNA/RNA immensely (described below).

6 METHODS

6.1 PREPARATION OF CERVICAL SWABS AND DNA EXTRACTION

In paper I and II, upon obtaining the cervical tubes, subsequently the beads were beaten with ZR Bashing Bead Lysis Tubes (0.1 and 0.5 mm from Zymo Research Corp, Irvine, CA) and centrifuged at 1,600 rpm with the 96 FastPrep machine (MP Biomedicals, Santa Ana, CA, USA) for 1min. Thereafter, the sample buffer was spun at 4,400 rpm for 4 min in order to separate out the beads. After that, the supernatant was incubated at 37° C for 3 h while shaking at 1,000 rpm with lysozyme buffer (20 mM Tris-Cl, 2 mM sodium-EDTA, 100 g/ml lysozyme; Sigma, St. Louis, MO, USA).

Thereafter the above, DNA was extracted using the ZR-96 Genomic DNA MagPrep kit (Zymo Research Corp, Irvine, CA) according to the manufacturer's instruction. Finally, the extracted DNA was eluted from the magnetic beads with 70 µl Elution Buffer (10 mM Tris-Cl, pH 8.5; Qiagen, Venlo, Netherlands) and the purified DNA was stored at -20 C before HPV genotyping. For every batch of DNA extraction, collection fluid and Zymo Microbial Community DNA Standard (Cat. no. D6300, Zymo Research, USA) were included as the negative and positive extraction controls, respectively.

Furthermore, owing to the sample size which was large and collection period which was long, two versions of the ZR-96 Genomic DNA MagPrep kits were used. Both versions are from product upgrades validated to produce comparable microbiota results (Paper II).

Methodological consideration

Despite some changes in the different protocols over the years, we anticipate that the DNA that was extracted from the different samples had sufficient quality and did not differ to that extent that it would have affected the obtained data.

6.2 TUMOUR SELECTION, MICRODISSECTION AND RNA AND DNA EXTRACTION

In paper III, haematoxylin/eosin stained FFPE tumour sections were histologically scrutinized by three experienced researchers (L Haeggblom, P Farrajota Neves da Silva and

A Näsman). The samples were then selected for harbouring sufficient amounts of high-grade dysplasia/cancer in situ as well as invasive cancer. Adequate and available biopsies were thereafter sectioned (six 5 µm-sections per sample) by microtome with RNase-free water. These sections were then put on Membrane Slides PEN-Membrane 2.0 µm (No.11505158) (Leica Microsystems AB). Subsequentially, the tissue sections were deparaffinized for 2 × 15 minutes in fresh xylene. This procedure was then followed by rehydration using subsequently lower percentages of fresh ethanol (5 minutes each in 100%, 95%, 70%, 50%, 0% ethanol) and thereafter again stained with haematoxylin for 30 seconds.

Finally, on a Leica LMD 7000 microscope (Leica Microsystems AB), utilizing the Laser Microdissection System (version 7.6.5684) the separate dysplastic and invasive carcinoma areas of the six replicate tumour sections were laser micro dissected. They were then collected separately in microcentrifuge tubes containing PKD buffer from the RNeasy FFPE kit (Qiagen).

In paper IV, similar procedures were used to distinguish tumour from normal tissue and metastasis or recurrence from normal tissue. However, the histological examination was then performed by one researcher (A Näsman).

Methodological consideration

In both paper III-IV, laser micro dissection was used to separate different compartments (dysplasia *vs.* invasive disease and invasive cancer *vs* normal tissue) of interest for the specific study.

In paper III, we estimate that the "purity" of the dysplastic and the invasive epithelial "compartment" is almost 100%. However, due to technical reasons with laser micro dissection, it is likely that stromal cells have been included in both "compartments". Similarly, it is also likely that lymphoid cells – especially intratumorally tumour infiltrating lymphocytes (TILs) – have also been included in both "compartments".

Likewise, in paper IV we estimate having obtained >90% tumour tissue (due to stromal cells and TILs) and around 100% normal tissue in the individual samples prior to that DNA was extracted.

We also anticipate that the amount and quality of the DNA that was extracted was of sufficient quality to pursue the analysis, as also indicated in most of the cases. However, we did in Paper IV, have some cases that did not have sufficient material to proceed with whole exome sequencing (WES), especially in material from local/distant relapse.

6.3 RNA EXTRACTION AND MULTIPLEX GENE EXPRESSION ANALYSIS

In paper III, samples were when possible kept on ice and after laser microdissection RNA was extracted immediately to warrant the best RNA quality, utilizing the RNeasy FFPE kit (Qiagen) according to the manufacturer's instructions. The RNA concentration was measured using a Qubit 4 Fluorometer (Thermo Fisher Scientific) and the Qubit™ RNA HS Assay Kit (Thermo Fisher Scientific).

In total, 12/24 samples (6 $HPV^{+}p16^{+}$ and 6 $HPV^{-}p16^{-}$ cancers) with adequate RNA concentrations (>2 ng/µL) were utilized for the multiplex gene expression assay. However, because of low RNA concentrations in the FFPE material, the nCounter Low RNA Input Amplification Kit (NanoString Technologies) was used. After that the hybridization reactions were initiated by utilizing the provided master kit and progression primers necessary for the nCounter PanCancer Progression Panel (NanoString Technologies). The gene expression assay was then done in accordance with the manufacturer's instructions using the nCounter® Sample Prep Station with FLEX configuration (NanoString Technologies) and the nCounter® Digital Analyzer 5s (NanoString Technologies) for reading the samples. For the analysis of the data the nSolver Analysis Software (version 4.0) was used, and this then included the nCounter Advanced Analysis add-on software (version 2.0.115). In more detail, fold changes and *P*-values were calculated utilizing nCounter default settings. Pathway scores were calculated by measurements of genes included in 34 different pathways, utilizing the nCounter Advanced Analysis Software.

Methodological consideration RNA extractions

Preparation of RNA from FFPE material can be challenging due to for example crosslinking and fragmentation, and the RNA concentrations that were obtained were low. For these reasons a low RNA Input Amplification Kit (NanoString Technologies) was used.

The nanoString technology allows a very low RNA input (less than 100 ng) and is suitable in FFPE material. Here, our material was limited – especially the material from the dysplastic “compartment” – making the method suitable for our aims.

6.4 HPV DNA EXTRACTION OF TSCC, BOTSCC AND METASTATIC LESIONS

In paper IV, after microdissection, for DNA extraction the FFPE AllPrep DNA/RNA kit (Qiagen) was utilized according to the manufacturer's instructions. The reason for this was that we anticipated we may later also assay for RNA expression. The amount of DNA was measured using a Qubit 4 Fluorometer (Thermo Fisher Scientific) and the Qubit™ dsDNA HS Assay Kit (Thermo Fisher Scientific).

Methodological consideration DNA extractions

FFPE AllPrep DNA/RNA kit (Qiagen) is in fact optimised for RNA and DNA retrieval from FFPE material. The obtained DNA was tested by a quality control at SciLife Stockholm and adequate for further analysis and sequencing.

6.5 HPV GENOTYPING

PCR and PCR amplification

In papers I-IV, for amplification by PCR, the Qiagen multiplex master mix (QIAGEN AB, Sollentuna, Sweden) was utilized. A number of broad spectrum GP5+/6+ primers, as previously described, for amplifying the HPV L1 gene from 27 mucosal HPV types as well as cellular beta-globin primers as positive control were used, for details see (Ramqvist T et al., 2011).

As controls in all individual tests, DNA from the SiHa cell line, in quantities that corresponded to 1, 10 and 100 HPV16 genomes per 5 µl, were used as positive controls and 10 µl of RNAse free water was used as a negative control instead of DNA.

Amplification was performed in a T100 PCR BioRad (Sweden) PCR thermal cycler, with denaturation for 15 min, subsequently followed by 40 amplification cycles. Every

amplification included 20 sec denaturation at 94°C, 90 sec annealing at 38°C and 80 sec of elongation at 71°C. Upon completed amplification, the 5 µl PCR product was mixed with 2 µl of beadmix, including beads correlating to the L1 gene of the different HPV types, HPV16 E6 and beta-globin (Ramqvist T et al., 2011, Du et al., 2011, Paper I).

The above PCR product and bead mix were set into 96-well plates, and put in a thermal cycler at 95°C for 10 min for DNA denaturation. Subsequently, after 1 min on ice, the 96-well plate was put in a thermomixer at 41°C for 30 min, 500 rpm. The samples were then washed by vacuum-filtration. Thereafter, 70 µl conjugate, streptavidin R phycoerythrin 1:300, was added, followed by an incubation period of 30 min at room temp. The samples were then washed with 100 µl wash-buffer 3 times, and then suspended in 100 µl wash-buffer and set into new 96-well plates, and analysed in the Magpix.

The assay used in Papers I and II covered 27 types, which covered all high-risk HPV types 16, 18, 30, 31, 33, 34, 35, 39, 45, 51, 52, 56, 58, 59, 66, 68, and 70, as well as some low-risk types, 6, 11, 26, 42, 43, 44, 53, 67, 69, 73. This assay was initially developed by the Michael Pawlita group (Schmitt M et al., 2006). However, since we have introduced the method in our laboratory the number of HPV types included have in 2014 been increased from 24 to 27, and e.g. HPV 30, 67 and 69 were included which they were not initially (Ramqvist T et al., 2011, Nordfors C et al., 2014).

6.6 BEAD BASED ANALYSIS ON A MAGPIX INSTRUMENT

All studies in this thesis used Magpix for detection of HPV. After DNA extraction and PCR amplification, detailed HPV type analysis was done using the Magpix instrument (Luminex Inc, TX, USA). After the samples had been washed, they were mixed with magnetic beads that have a unique fluorescent colour for each assay HPV type. This was followed by a denaturation step (heating) to release the DNA strands from each other and allow different types of beads, corresponding to each assay-HPV type to bind with any matching DNA sequence. This was followed by ligation. More specifically, the wells are one at a time loaded into a chamber for analysis. A magnet arranges the beads in a monolayer for easier analysis, and non-magnetic-bead bound DNA products are washed away. Any bead bound with material is retained. The samples with potentially bound beads are then exposed to red and blue light, that identify the type of bead, and the amount of bound DNA respectively, ensuring that each bead is present, and how much material was available. The result is

reported back as a median fluorescent value or MFI. MFI is calculated as the median fluorescent measurement from a time series recorded when exposing the beads to light. The cut-off for positivity was 1.5 times the background fluorescence, a bit more margin than the cut-off used in the original paper of the method used here (Schmitt M et al., 2006).

Methodological consideration

Of note, typing of HPV30, 67, and 69 was not initially included in the corresponding assays of the very first cohort sample collection 2008–2010 from the youth clinic or the 2003–2008 cervical cancer (CC) sample collection, as the HPV DNA detection was done using the protocols described earlier (Ramqvist T et al., 2011, Du J et al., 2011). In paper I and II the analysis also included HPV types 30, 67 and 69 and the analysis of HPV16 E6, making the analysis more thorough. Nevertheless, the data are presented in such a way that one can still investigate the results of the HPV types included in the vaccines.

6.7 STATISTICAL ANALYSIS

After obtaining the data, these were compared in different ways. Any HPV type infection/overall HPV infection (i.e. including all 27 HPV types); or the 15 HR-HPV types indicated above; infection with the quadrivalent-Gardasil® HPV types (HPV16, 18, 6, and 11) together or alone, were compared between the HPV vaccinated and unvaccinated cohorts by Chi-square analysis. Likewise, differences between HPV types collected in this study were also compared with our previous reports using Chi-square analysis.

6.8 IMMUNOHISTOCHEMISTRY (IHC)

Immunohistochemistry (IHC) was used to evaluate protein expression. In total, 4µm thick sections from all 24 TSCC/BOTSCC biopsies were stained in study III. The sections were de-paraffinized in xylene, and then in several steps submerged in decreasing concentrations of ethanol to rehydrate them.

Antigen retrieval was achieved by boiling the sections in citrate or EDTA buffer (see considerations). To inactivate endogenous peroxidase, hydrogen peroxidase was applied. Unspecific binding was blocked by applying horse serum.

Six different monoclonal antibodies were used targeting psoriasin, periplakin, Type 1 collagen, Galectin-1, Galectin-1 TILS, and SPARC. For details see also paper III.

Antibodies were incubated on slides overnight in a humid chamber and subsequentially incubated with a secondary antibody (see paper III) with a biotinylation, which was then incubated with avidin-biotin enzyme complex (Vectastain Elite ABC kit (HRP), Vector Laboratories, Burlingame, USA). Substrate chromogen-39-diaminobenzydine (DAB) (Vector Laboratories, Burlingame, USA) was then applied, which gives the complex (if bound) a brown colour. Subsequently, all slides were counterstained with haematoxylin. Before mounting, the slides were de-hydrated with increasing concentrations of first ethanol, and then xylene.

Evaluation. Sections were evaluated by researchers blinded for HPV status. Psoriasin expression was assessed as percentage positive cells and a difference in staining pattern more than 30% were considered as differential expression. For Periplakin, Galectin-1, IL1-RA staining, the intensity of invasive and dysplastic tumour areas were graded as 0=absent, 1=weak, 2=medium and 3 = strong. In addition, Galectin-1 was also considered by TILs present, with 0 = absent, 1 = few, 2 = intermediate, and 3 = many. For Type I collagen and SPARC the staining intensity was graded in the same way as above, but the extracellular matrix was assessed here instead.

Methodological considerations immunohistochemistry (IHC)

A general consideration concerning IHC, is that only a thin section of the tumour is evaluated. This can lead to a misconception if a tumour is very heterogenous. Moreover, evaluating staining intensity can in some cases be problematic since the staining may vary between samples/batches etc. In this case, however, we evaluated staining intensity in two different “compartments” (dysplasia and invasive tumour) in the same tumour tissue stained at the same time under exactly the same conditions. We also only binary assessed if there were differences or not in staining between high grade dysplasia and invasive cancer per tumour/tumour slide. Therefore, staining intensity should be reliable tool in this setting.

6.9 DNA SEQUENCING

6.9.1 Introduction

DNA sequencing is the process used to identify the sequence of the nucleic acids in a DNA strand and this can be done for a variety of applications. Comparisons of sequenced data between species can give some information on the phylogeny and evolution of an organism. The sequence data can also be used to identify markers of inherited traits or common mutations related to disease, such as cancer. It can also give us some understanding of regulatory regions. For a general review see reference Zhou X et al., 2010, as well as Ng SB et al., 2009, Ng SB et al., 2010 and Biancalana V and Laporte J, 2015.

Thus, DNA sequencing can be used to identify new biomarkers and treatment targets in cancer and inherited diseases, or to identify symbiotic and pathogenic human microbiota. Since, DNA sequencing has been made more readily available the past years, both with regard to speed and economical affordability, here some technologies are described in more detail, also since they were not covered in the literature review above.

Sequencing technologies in current use are referred to as **Next-generation** sequencing (NGS) or **High-throughput** sequencing (HTS) technologies. While there are technical differences between different platforms, all approaches translate DNA fragments from a sample of interest into nucleotide sequences called *reads.* The number of reads covering an area of the DNA is referred to as *depth*, while the length of DNA covered by the reads is referred to as *coverage* (Figure 7). When sequencing is performed on an organism with a well-studied genome, reads are mapped to a *reference genome* from that organism to find their correct position in the genome (*alignment*). In novel, or less well-studied organisms, *de-novo genome assembly* is used to reconstruct the genome from the read data.

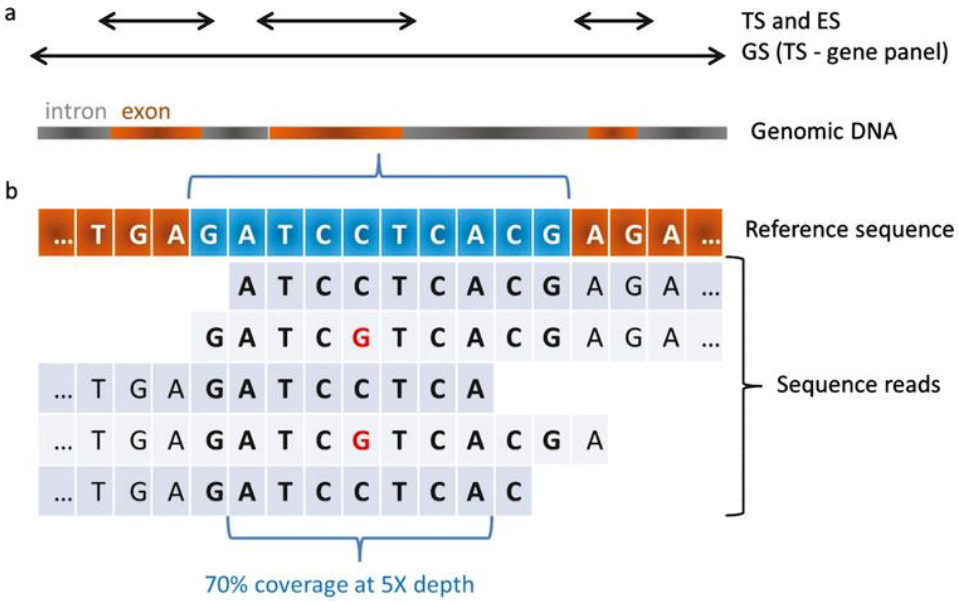


Figure 7. Coverage and depth in MPS. a) Targeted Sequencing (TS) and Exome Sequencing (ES) target exons of a panel of genes or of most genes respectively. Genome Sequencing (GS) targets the genome and in particular exons and introns of all genes. TS may be used for the study of genomic regions. b) Each region of interest is theoretically sequenced several times by overlapping sequence fragments called reads. When considering a sequence of 10 bases(in blue), in case 7 out of the 10 bases are covered by 5 different reads, the coverage of this region is 70% at 5X depth. A potential single nucleotide variant (SNV; C to G transition) is shown at the heterozygous state. (From Biancalana V and Laporte J, 2015, with permission from the publisher according to the Creative Commons Attribution Non-Commercial License)

6.9.2 General NGS Workflow

NGS includes three basic steps: library preparation, sequencing and analysis. Library preparation is performed to prepare fragments that can be read in the sequencing step. Depending on sequencing technology or amount of DNA in the sample, the sample DNA can be PCR-amplified or be left as is. The DNA is then fragmented, either chemically or mechanically. The fragments can then have adapters added, to amplify certain regions for amplicon or exome sequencing, to track the sample of origin in multiplexing setups or for compatibility with sequencing machines (such as Illumina). Sequencing generally involves some sort of clustering and PCR amplification of DNA fragments, followed by creation of single-strand DNA templates and addition of complementary nucleotides to the template fragments, where the sequence of the added complementary bases for each fragment is recorded through methods such as pyrosequencing (Roche) or fluorescent nucleotide probes (Illumina).

Due to the nature of the sequencing methods, the output at this point will describe light intensity in various forms. This data is converted to nucleotide sequences through *base calling*, where the output is a string of nucleotide sequences (*FASTA* file). Depending on the sequencing machine's estimation of base calling accuracy, the nucleotide gets a quality score, showing how confident the base call is. This is called a *FASTQ* file. For downstream analysis, *FASTQ* files are aligned to a reference genome, showing the nucleotide sequence of the studied sample across the genome. The alignment files will also have a quality score, dependent on genome position and read depth, showing how confident the alignment is.

6.9.3 NGS Analysis

Depending on the research question, various types of information can be extracted from the aligned data. In general, *germline* and *somatic* variants – both single-nucleotide and structural variants – and *copy number variations* can be studied. Germline variants can be thought of as mutations present in all cells of the specific organism studied, a result of natural variation. They can be identified when analysing non-cancer samples, such as peripheral blood, as diverging from the canonical genome sequence. When analysing a cancer sample, variants are compared against databases of human genetic variation. Here, germline variants will tend to have a relatively high *allele frequency* – i.e., they are present in a subset of the normal population.

Somatic variants are specific to a tumour, and not present in non-cancer cells of the patient. Because of this, when analysing DNA sequences in cancer samples, both a cancer sample and a normal sample from the same patient are typically sequenced, and only variants that are neither found in the normal sample nor in normal variant databases are considered cancer-specific somatic variants.

Copy number variations can be inferred by comparing the read depth at the same loci between the cancer and the normal sample. In this way, amplifications and deletions of both single genes and entire chromosomes can be identified.

6.10 TYPES OF DNA SEQUENCING

Whole Genome Sequencing (WGS) is a comprehensive method of analysing the entire genomic DNA of a sample. WGS is also known as “full genome sequencing” or “complete genome sequencing”.

WGS can provide more types of data than other, more targeted approaches. For example, mitochondrial DNA and regulatory intronic regions will be covered by WGS. The coverage of exons is also higher than with exome-targeted approaches.

Additionally, to create new reference genomes or sequence organisms without a reference genome (de-novo sequencing), WGS is required. The wide coverage also makes it more likely to call de novo mutations and provides higher accuracy in calling single-nucleotide polymorphisms (SNPs) and insertions-deletions (INDELs).

Whole Exome Sequencing (WES) is a type of targeted sequencing, focusing on the coding regions of the genome. The exome is estimated to represent roughly 1-2% of the human genome, but accounts for most mutations identified in Mendelian disorders (Ng SB et al., 2009, Ng SB et al., 2010). In WES, exome regions of the DNA are captured with probes targeted to these regions. The captured DNA then undergoes multiple cycles of PCR amplification before sequencing. Illumina-based WES is illustrated in Figure 8.

Output data for WES is greatly reduced to about 5% of WGS. Sequencing of only the exons saves reagents and computational analysis time. WES costs about a half to a third as much as WGS and the output is substantially less computationally heavy to analyse. This makes WES an attractive alternative to WGS in a clinical setting, where mainly protein-coding regions are of interest.

In Paper IV we used WES. One reason for this choice of method was cost and time limitations. An added benefit however is that the results in the study suggest a potential future possibility to use WES cost-effectively in a clinical setting to study outcome-related mutations and plan treatment accordingly to each patient’s mutational profile.

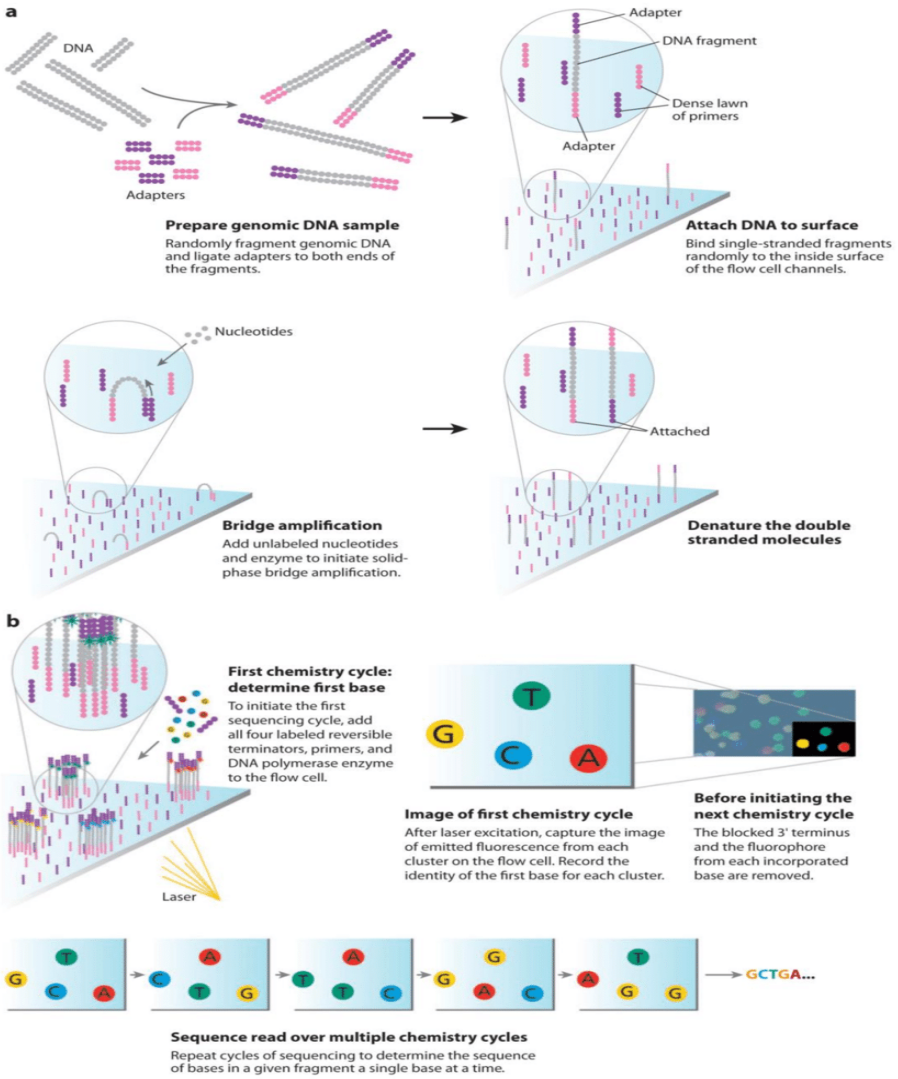


Figure 8. Schematic illustration of whole exome sequencing (Illumina) according to Masoudi-Nejad et al 2013, with permission from the publisher.

Amplicon Sequencing (AS) is a highly specific targeted approach that typically only sequences one or a limited set of genomic regions or genes. This can be panels of specific genes or well-known high-risk areas for mutations. In amplicon sequencing, loci of interest are amplified using specific PCR primers before sequencing. AS allows a very high read depth and can save even more time than WES. The high read depth also sequences GC-rich regions more accurately. The speed and read depth allows for efficient sequencing of microbiota, and several genera of microbes can be identified in a mixed sample. The high read-depth can also allow sequencing of samples contaminated by host DNA and can

separate normal tissue from tumour tissue in human samples. The amount of reads for each gene can also be used to interpret the relative abundance of organisms in the sample.

Typical genes sequenced to differentiate between species are house-keeping genes such as 16S ribosomal genes common to prokaryotes. Apart from microbial sequencing, custom panels can also be useful to cheaply and quickly identify known high-risk cancer mutations in a clinical setting. 16S amplicon sequencing was used in paper II to detect species of bacteria in the human samples. In clinical settings, AS could potentially replace culturing for the detection of microbes, as bacterial culture has a bias to easily lab-grown bacteria, while sequencing lacks this bias. A study comparing AS of the 16S gene from faecal samples compared to traditional bacterial culture methods showed that about two thirds of uniquely detected bacteria were detected by AS (Gupta S et al., 2019).

Next generation Illumina dye sequencing (used in paper II, Microbiota)

In paper II, for sequencing of microbiota the V3-V4 regions of the 16S rRNA genes were amplified from the extracted DNA using Illumina index-binding primer pairs 341F/805R, as described in detail (Hugerth LW et al., 2018, Paper II).

Bioinformatics analysis. After base calling and de-multiplexing, the *FASTQ* files generated from each run were as routinely done, subjected to quality trimming, de-noising, merging, and chimera removal, and thereafter processed into sequencing tables. Further details are described in Paper II, and by Martin M 2011, Callahan BJ et al., 2016. Following decontamination, sequence annotation, **phylogenetic tree construction and alpha- and beta- diversity calculation steps** are described in more detail in Paper II and by et al Davis NM et al., 2018.

Methodological considerations DNA sequencing

Current microbiota sequencing analysis is based on relative abundance comparison, which is not considered quantitative. Novel methods like including spikes in species have been tested and may better quantify bacterial quantity. However, we compared targeted regions of the 16S rRNA genes, using the V3-V4 region which provides the best resolution on the species level for vaginal microbiota. This is a well-established method, adequate for the purposes of the study. Due to the methodological limitations, we however did not consider other viruses than HPV, fungi nor parasites in the microbiota.

Whole Exome Sequencing (WES) of cancer samples (used in paper IV)

For sequencing, the Illumina NovaSeq 6000 (Illumina, San Diego, CA, USA) was used, with sequencing by synthesis and initially 40-250 ng DNA was used for library preparation (For further details see paper IV). The primers xGen Library Amp Primer (0.5 mM, Integrated DNA Technologies, Coralville, IA, USA) were used to facilitate DNA synthesis, and PCR performed for 10 cycles during base calling. Around 60 millions of read pairs were generated per sample. Downstream analysis focused solely on somatic variants predictive of relapse

Notably, before sequencing, tumour tissue samples were separated into normal and cancer tissue through laser microdissection and calling of somatic variants was used with the normal tissue samples as reference. In a few cases, due to quality issues, normal samples were not available. In these cases, allele frequency was used to separate germline from somatic mutations. In these cases, allele frequency was used to separate germline from somatic mutations. For specific details related to library preparation (using 40-250 ng DNA) and target enrichment with a multiplex fashion (using 375 ng DNA) as well as variant calling and analysis, please see details paper IV.

Methodological considerations

Since the study was focused on identifying biomarkers predictive of relapse, it only analysed somatic variants. Finally, an important consideration is the use of normal tissue for somatic variant calling. Many studies use peripheral blood, where there is no risk of contamination from surrounding cancer cells. Since our study used FFPE material, we reasoned that a better control would be normal tissue from the same FFPE slides as the cancer cells, to avoid FFPE artifacts biasing the analysis. With laser microdissection, very high purity can be achieved, avoiding problems with cross-contamination between cancer and non-cancer cells. The fact that not all samples had corresponding normal tissue is a definite problem, addressed in the paper. To not lose these samples, they were kept for analysis and variant calling used allele frequency to select somatic variants. However, in downstream analysis, variants of interest in these samples were only kept if they were also present in samples with normal control tissue.

7 RESULTS AND DISCUSSION

7.1 PAPER I. CHANGES IN CERVICAL HUMAN PAPILLOMAVIRUS (HPV) PREVALENCE AT A YOUTH CLINIC IN STOCKHOLM, SWEDEN A DECADE AFTER THE INTRODUCTION OF THE HPV VACCINE

Ährlund-Richter A, Cheng L, Hu YOO, Svensson M, Pennhag AAL, Ursu RG, Haeggblom L, Grün N, Ramqvist T, Engstrand L, Dalianis T, Du J.
Front Cell Infect Microbiol. 2019 Mar 20;9:59.

7.1.1 Aim

The goal of this paper was to examine the effect of human papillomavirus (HPV) catch-up vaccination as well as school-based HPV vaccination on the base line high cervical HPV-prevalence, at a youth clinic in Stockholm upon the introduction of the HPV vaccine during the years 2008-2018.

7.1.2 Background

During the period of 2008-2010 there was a high base-line cervical HPV-prevalence (in total 69.5%) and in particular a high HPV16 variant prevalence (34.7%) in unvaccinated women at a large youth clinic in Stockholm. However, during 2013-2015, a few years after the initiation of the quadrivalent-Gardasil® HPV-vaccine, the prevalence of HPV16 and HPV6 decreased. In this paper, we aimed to investigate the prevalence of HPV in cervical samples 10 years after primary sampling. Below some of the details of Paper I are included, but for a detailed description, see the paper itself.

7.1.3 Material and Methods

More specifically during 2017-2018, in total 178 cervical swabs, from 15-23 year old women, were analysed for 27 HPV types using a bead-based multiplex method. Cervical HPV prevalence data were then correlated to age, vaccination status and compared to cervical HPV prevalence in a cohort of 615 samples from 2008 - 2010 and a cohort of 338 samples from 2013 to 2015 from the same clinic, in addition to HPV types in 143 cervical cancer (CC)

cases during 2003-2008 in Stockholm. For details of the analysis see, the Material and Methods sections above and Paper I.

7.1.4 Results

Compared to the first study the numbers of HPV vaccinated women had risen considerably, more specifically from 10.7% (2008-2010) to 82.1% (2017-2018). Furthermore, in the 2017-2018 cohort, the combined prevalence of all 27 HPVs; all high-risk HPVs (HR-HPVs) and the combined presence of HPV16, 18, 6 and 11 (the quadrivalent-Gardasil® types), was lower in HPV vaccinated in comparison to non-vaccinated women (67.4 vs. 93.3%, $p = 0.0031$, 60.1 vs. 86.7%, $p = 0.0057$ and 5.8 vs. 26.7%, $p = 0.002$, respectively). This is illustrated in Figure. 9 below.

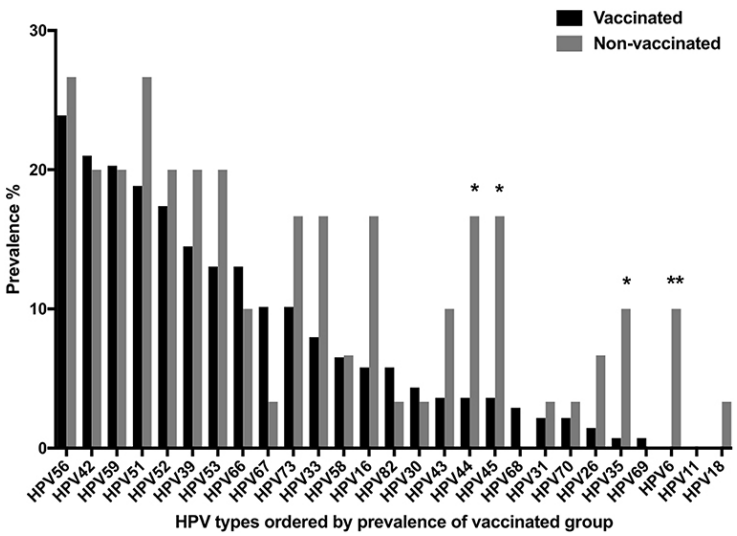


Figure 9. Cervical human papillomavirus (HPV) prevalence of all tested HPV types according to vaccination status (Vaccinated or non-vaccinated). *p < 0.05, **p>0.01. (From paper I, with permission from the publisher).

In addition, notably e.g. HPV16 prevalence in HPV vaccinated women was 5.8%, and HPV16 prevalence in non-vaccinated women 2017-2018 was much lower than that in 2008-2010 (16.7 and 34.7%, respectively, $p = 0.0471$). In addition, similar trends were observed for HPV18 and 11. Furthermore, we found that the prevalence of HR-HPV35 and 45 decreased in HPV vaccinated as compared non-vaccinated women. Lastly, similar to previous studies the highest HPV prevalence was observed at the age 21 years and that the majority of women, with an HPV infection, were infected with multiple HPV types.

Of note, however, in all the women, irrespective of HPV vaccination status, the most frequent non-quadrivalent-Gardasil® vaccine HR-HPV types were HPV39, 51, 52, 56, and 59 (Figure 10). These latter HPV types contributed together to roughly 9.8% of CC cases diagnosed 2003-2008 in Stockholm, and their relative frequency tended to have risen in the period of 2017-2018 as compared to the period 2008-2010.

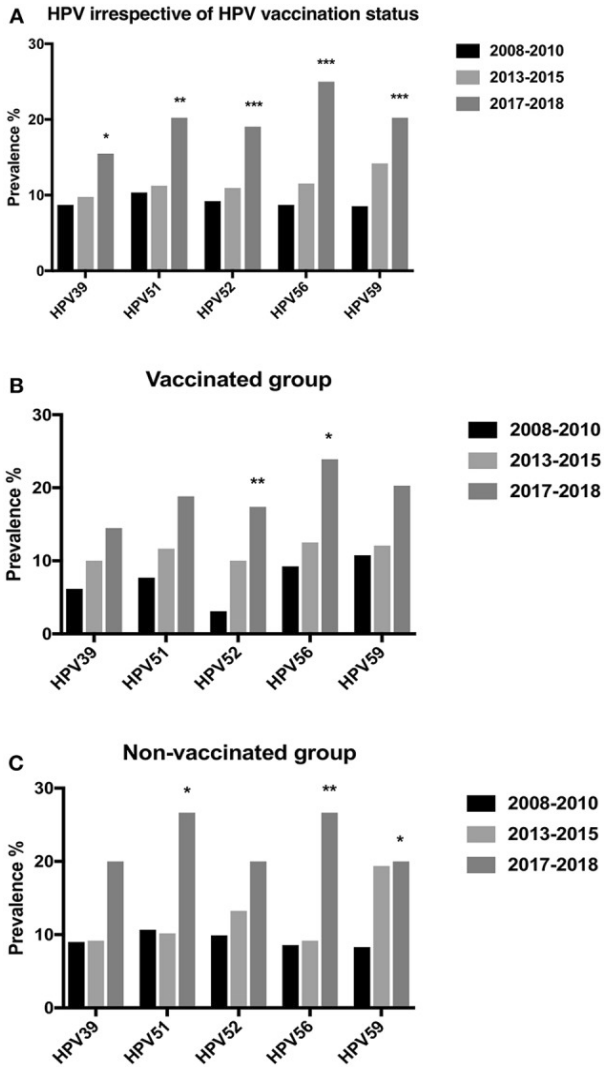


Figure 10. Prevalence of specific (HPV) types over the years 2008-2018. (A) Cervical prevalence of HPV39, 51, 52, 56, and 59 irrespective of vaccination over the years 2008-2010, 2013-2015, and 2017-2018. (B) Cervical prevalence of HPV39, 51, 52, 56 and 59 in vaccinated women over the years 2008-2010, 2013-2015, and 2017-2018. (C) Cervical prevalence of HPV39, 51, 52, 56 and 59 in non-vaccinated vaccinated women over the years 2008-2010, 2013-2015, and 2017-2018. *p < 0.05, **p>0.01, ***p<0.001 (Paper I, with permission from the publisher).

7.1.5 Discussion

In this paper, we have shown that during the time period 2008 and 2018, that the proportion of HPV vaccinated women rose from 10.7%, 2008–2010, to 82.1%, 2017–2018, while the prevalence of HPV16, 18, 6 and 11 (in the quadrivalent-Gardasil® vaccine) decreased substantially. In addition, we observed that in the 2017–2018 cohort, their combined presence had decreased significantly in HPV vaccinated when compared to that in unvaccinated women. These findings are in concordance with other studies indicating that a high coverage of HPV vaccinated individuals has a high impact on HPV prevalence and HPV associated diseases, lowering the burden of HPV infection, genital warts and cervical disease (Grün N et al., 2016, Dehelendorff C et al., 2018, Drolet M et al., 2019) . This study, was however, very illustrative and had the benefit of studying HPV prevalence in a cohort with a generally high prevalence, and thereby allowing for the analysis of the fate of some individual HPV types.

Of note, HPV16 prevalence had decreased in non-vaccinated cohorts in 2013–2015 as well as in 2017–2018 when compared to the 2008–2010 cohort (Grün N et al., 2016). We also found that in the 2017–2018 cohort, HR-HPVs in total and in particularly HR-HPV35 and 45 declined in HPV vaccinated vs. in non-vaccinated women. Notably however, this did not apply to the in 2017-2018 most commonly occurring HR-HPVs HPV 39, 51, 52, 56, and 59, which instead tended to increase as compared to 2008–2010 (Paper I).

That a handful of HR-HPV types tended to have become more common could be due to a variety of reasons, including chance. One possibility, not yet confirmed, could be increased sexual activity in recent years. Another possibility could be that cross-immunization potentially induced by the previously naturally occurring HPV-vaccine types e.g., HPV16, 18, was superior than the vaccine against these non-HR HPV types. However, the latter was not supported by a study showing cross-immunization against types genetically related to HPV16 in vaccinated, but not in non-vaccinated women (Saccucci M et al., 2018). It has also been proposed that eliminating HPV16, could have enhanced detection of HR-HPV types that theoretically could be competing for the same primers (Saccucci M et al., 2018). However, irrespective of why, that specific HR-HPV types, not included in the vaccines tend to rise in prevalence is in accordance to that shown by others and should be followed (Saccucci M et al., 2018, Machalek DA et al., 2018).

The latest HPV vaccine, Gardasil®9 covers HPV6, 11, 16, 18, 31, 33, 45, 52, and 58 (Schiller J and Lowy D, 2018). It thereby includes 1/5 (HPV52) of the most common HR-HPVs 39, 51, 52, 56, and 59 found in the 2017–2018 cohort as well as 5/6 of most common HR-HPV types (HPV16, 18, 31, 33, 45, and 56) that were present in the CC diagnosed 2003–2008 in Stockholm (Du J et al., 2011).

Therefore, Gardasil®9, should be very useful and inhibit many new CC cases, however one must still bear in mind that 4/5 HR-HPV types HPV39, 51, 56, and 59 (not included in Gardasil®9), during 2003–2008 contributed to 11/143 (7.7%) of the CC cases in Stockholm (Du J et al., 2011). *Still, of these 11 cases, only six were single infections*, whereas four were co-infections, so several HPV types could be transcribed at the same time (Du J et al., 2011, Halec G et al., 2013).

Paper I has limitations. The women here attend the youth centre for birth control advice and possible sexually transmitted diseases and thereby introduce a selection bias (Ramqvist T et al., 2011, Grün N et al., 2016). In addition, HR-HPV types HPV 30, 67, 69 currently examined, were not included initially so the total HPV prevalence may have been affected, but this should not have influenced the analysis of the HPV vaccine or non-vaccine HR-HPV types since these were included before (Du J et al., 2011, Ramqvist T et al., 2011).

To summarize, HPV vaccination has introduced a decline in the prevalence of vaccine HPV types HPV16, 18, 6, and 11, but five non-vaccine specific HR-HPV types HPV39, 51, 52, 56, and 59 still remain high at a Stockholm youth clinic. Of these, only one HPV52 is included in Gardasil®9. Screening and follow up of HPV infections remain important also in the future and one could consider additional HPV types for the next generation of HPV-vaccines.

7.1.6 Conclusion

HPV vaccination has increased in coverage and cervical combined HPV16, 18, 6 and 11 type prevalence has declined significantly the past decade in women at a youth clinic in Stockholm. However, non-vaccine HR-HPV types remain so screening remains important and additional HPV types could be considered for the next generation of vaccines.

7.2 PAPER II. VAGINAL MICROBIOTA AND HUMAN PAPILLOMAVIRUS (HPV) INFECTION AMONG YOUNG SWEDISH WOMEN

Liqin Cheng, Johanna Norenhag, Yue O. O. H, Nele Brusselaers, Emma Fransson, **Andreas Ährlund-Richter,** Unnur Guðnadóttir, Pia Angelidou, Yinghua Zha, Marica Hamsten, Ina Schuppe Koistinen, Matts Olovsson, Lars Engstrand, Juan Du. V.
NPJ Biofilms Microbiomes 2020 Oct 12;6(1):39

7.2.1 Aim

To define the HPV-associated microbial community in a young population with a high HPV vaccine coverage.

7.2.2 Background

Human papillomavirus (HPV) infection is one of the most frequently sexually transmitted diseases globally and despite that the HPV vaccine has aided in reducing the prevalence of HPV infections and HPV associated tumours, not all high risk (HR) HPV types are covered (Bruni L et al., 2010, Bruni L et al., 2021). Thereby further interventions are needed. Furthermore, notably variations in vaginal microbiota, with increased microbial diversity and a notable the absence of *Lactobacilli*, have been demonstrate to enhance sexually transmitted diseases in some countries (Mitra A et al., 2016, Ondordonk AB et al., 2016). For this reason we attempted to identify whether there were distinctions in the microbial community in individuals with an HPV infection or not and investigated a young cohort with a known high HPV prevalence and a high HPV vaccine coverage.

7.2.3 Material and Methods

To if possible specify the HPV associated microbial community in a Scandinavian country with a high vaccination coverage, we performed the largest cross-sectional study with 345 young healthy Swedish females (youth clinic: n=206, age 14-22 years; cervical screening: n=139, 23-29 years). The microbial community and its potential correlation with HPV infection status (by covering 27 HPV subtypes) was examined. For details of the analysis see, the Material and Methods sections above and Paper II.

7.2.4 Results

Microbial alpha-diversity was found to be statistically significantly higher in HPV infected women when compared to that in HPV negative women, in particular if the women were infected with HR-HPV types and had multiple HPV types. This was e.g. reflected by differences between the women in the youth clinic and the cervical screening group (Figure 11). In addition, we observed a small but statistically significantly increased microbiota diversity in participants infected with oncogenic HPV39 and 56 (See paper II).

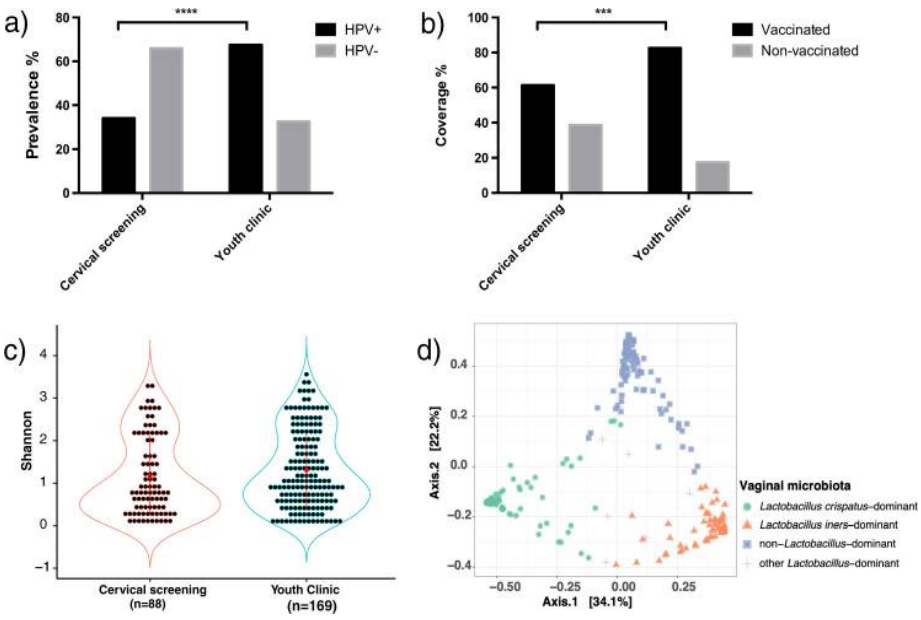


Figure 11. Comparison of HPV prevalence, HPV vaccination status, and microbial diversity in the youth clinic and the cervical screening samples.

a Significantly higher HPV prevalence was observed from the youth clinic samples than the cervical screening samples. **b** Significantly higher HPV vaccination coverage was shown in samples from the youth clinic than samples from the cervical screening. **c** Microbial alpha diversity based on Shannon analysis did not show the difference between samples from the youth clinic and the cervical screening. Every dot in the violin plot represents one individual. Data were presented as mean values with standard deviations. **d** Principal coordinates analysis (PCoA) of microbial species data based on Bray-Curtis distance matrix demonstrated three main vaginal microbiota clusters. Statistical significance between the groups was tested by Fisher's exact test in **a** and **b**, and by Wilcoxon rank-sum one-sided test in **c** ($p = 0.108$). ***$p < 0.001$ and ****$p < 0.0001$. HPV+: HPV-infected, HPV−: HPV-uninfected. (Paper II, with permission from the publisher).

The vaginal microbiota of the HPV infected participants included a larger number of bacterial vaginosis-associated bacteria (BVAB), such as e.g. *Sneathia*, *Prevotella* and *Megasphaera*, indicating that these could potentially be used as biomarkers and/or treatment targets (Figure 12).

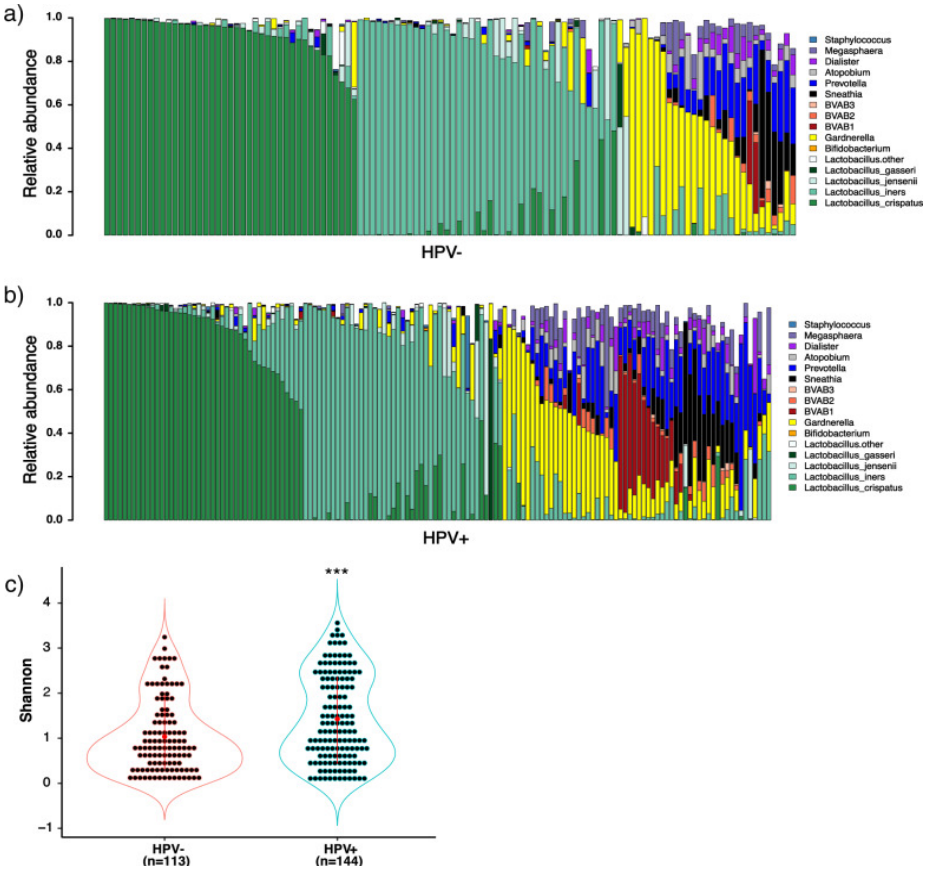


Figure 12. Difference in vaginal microbiota of HPV-uninfected and HPV-infected young women.

a Vaginal microbiota at the genus/species level from HPV-uninfected young women. Except BVABs, the following criteria were used in order to show the important and abundant taxa clearly: (1) Bacteria with over 1% mean relative abundance in all the samples. (2) *Lactobacillus* species that have more than 10% of reads in any sample. (3) Non-*Lactobacillus* genera that have over 30% of reads in any sample. **b** Vaginal microbiota at genus/species level from HPV-infected young women. Same criteria were used as in **a**. **c** Microbial alpha diversity (Shannon) comparison between groups of HPV-uninfected and HPV-infected young women demonstrated a significantly higher vaginal microbiota diversity among HPV-infected women by Wilcoxon rank-sum one-sided test. Data were presented as mean values with standard deviations. ***$p < 0.001$. HPV+: HPV-infected, HPV−: HPV-uninfected. (Paper II, with permission from the publisher).

Moreover, the analysis showed that double the number of women with non-*lactobacilli* dominant vaginal microbiota had infection with HR-HPV types compared to those with *L. crispatus* dominated vaginal microbiota (odds ratio 2.0, 95% confidence intervals 1.0-3.9), when adjusting for vaccination status, age, and population.

7.2.5 Discussion

In this study, we disclosed that in women infected with any HPV, oncogenic HPV, and multiple HPV types that there was a higher microbiota diversity when compared to that of women not infected with HPV. In addition, a small but still significantly higher microbiota diversity was observed in women presenting HR-HPV39 and 56 infection. Furthermore, from our data we propose that BVABs, (that have not been studied HPV-related studies before), together with *Sneathia*, *Prevotella*, and *Megasphaera*, are associated with HPV infection. Finally, although, HPV vaccination had a strong protective effect against the vaccine HPV types, it did not influence vaginal microbiota.

As also revealed previously in Paper I, this study showed the strong protection of HPV vaccination against HPV infection, in particular against having multiple HPV types, thereby showing the efficacy of the Swedish national HPV vaccine program (Paper I, Grün N et al., 2016, Ramqvist T et al., 2011).

However, there was still a high HPV infection rate due to other HPV types not included the vaccine and this was in particular observed in the age group 19-24 years. The latter finding could be due to that this age group was frequent in the youth clinic, where youth visit for advice on sexually transmitted disease or birth control. Our analysis may for this reason be regarded as biased by including sexually active young women that may have presented a sexually transmitted disease.

Nevertheless, despite that the sample source, vaccination status and age may have influenced HPV prevalence, these parameters did not seem to have affected the variation in the vaginal microbiota which is of note for future studies.

It has been proposed that vaginal microbiota and associated changes at an early life stage may have fundamental effects later in life, such as developing cancer (Mitra A et al., 2016 . Here, we have performed large cross-sectional study of vaginal microbiota focusing on young women (14–29 years) utilizing the sequencing method and although the study lacks, data on vaginal pH or vaginal infections other than HPV, when compared to other vaginal microbiota studies it has other benefits. The current study namely limits the influence of confounders e.g. such as variability of hormone and immunity levels due to that we do not

include a large age span. Furthermore, the young cohort allows us to study a population with a high prevalence of HPV infection (Paper I).

One can discuss, what future options our study may imply. There are recent studies that have indicated that some of *Lactobacillus* species, such as *L. gasseri*, may possibly be favourable for clearance HPV (Broteman RM et al., 2014, Brusselaers N et al., 2019). Also, other studies along the same line propose a possible correlation of some compositions of vaginal microbiota and HPV clearance or progression to cervical dysplasia and cancer (Mitra A et al., 2016, Norenhag J et al., 2020).

Nonetheless, despite that few longitudinal reports suggest microbiota may influence HPV persistent infection (Shannon B et al., 2017), HPV may also contribute to the alteration of vaginal microbiota stability and composition, but variations in HPV prevalence and vaginal microbiota may occur at the same time due to sexual activities. These possible bi-directional effects need to be further investigated, with special emphasis on longitudinal ones. In addition, possible underlying mechanisms with regard to potential interactions between microbiota and HPV would be of interest to explore.

7.2.6 Conclusion

To summarise, in this cross-sectional study of a large cohort of young Swedish women with a high HPV vaccination ratio, we show a significant correlation between non-*Lactobacillus*-dominated vaginal microbiota and HPV infection. The latter including any HPV, HR-HPV and multiple HPV types.

Furthermore, we demonstrate that the HPV vaccine exhibits a minor effect on vaginal microbiota, and on bacterial species e.g. BVABs, where the latter may possibly be utilized e.g. as a marker and/or target for therapy of infection with HPV.

7.3 PAPER III. DIFFERENCES IN GENE EXPRESSION BETWEEN HIGH-GRADE DYSPLASIA AND INVASIVE HPV+ AND HPV- TONSILLAR AND BASE OF TONGUE CANCER.

Haeggblom L, **Ährlund-Richter A,** Mirzaie L, Farrajota Neves da Silva P, Ursu RG, Ramqvist T, Näsman A.
Cancer Med. 2019 Oct;8(14):6221-6232.

7.3.1 Aim

To in HPV-positive (HPV^+) and HPV-negative (HPV^-) TSCC/BOTSCC analyse potential similarities and dissimilarities in gene and protein expression between high-grade dysplasia and invasive cancer in both.

7.3.2 Background

HPV can cause TSCC and BOTSCC, and not only cervical cancer (CC) and other anogenital cancers. However, while premalignant stages in e.g. CC have been investigated thoroughly, extremely little is investigated and clarified regarding premalignant stages in TSCC and BOTSCC or the influence of HPV. In fact, some authorities have postulated that there are no pre-stages in HPV mediated TSCC/BOTSCC. Here we therefore explored and analysed for possible similarities and differences in gene and protein expression in high-grade dysplasia and invasive cancer in HPV^+ and HPV^- TSCC and BOTSCC.

7.3.3 Material and Methods

From a larger material twenty-four tumours with high-grade dysplasia and invasive carcinoma were identified and selected for further analysis. From these 24 tumours, formalin fixed paraffin embedded (FFPE) samples with lesions, were then laser micro-dissected from HPV^+ and HPV^- TSCC and BOTSCC tumour sections.

Differential gene expression was examined using nanoString RNA-panels and genes of interest were validated on the protein level by applying immunohistochemistry. For further details see the Material and Methods sections above and Paper III.

7.3.4 Results

After microdissection and differential gene expression analysis the data were compared between the HPV$^+$ and HPV$^-$ samples. Differences in gene expression between in situ and invasive cancer were found to be similar in HPV$^+$ and HPV$^-$ TSCC and BOTSCC (Figure 13). In total, 40 genes in the HPV$^+$ tumours displayed a statistically significant different expression in high-grade dysplasia as compared to invasive cancer and this was true for 33 genes in the HPV$^-$ cancers (Figure 13). Furthermore, 5/9 of the most noteworthy pathways presented a corresponding increased activity in invasive cancer when compared to high-grade dysplasia in not only the HPV$^+$ but also the HPV$^-$ cancers.

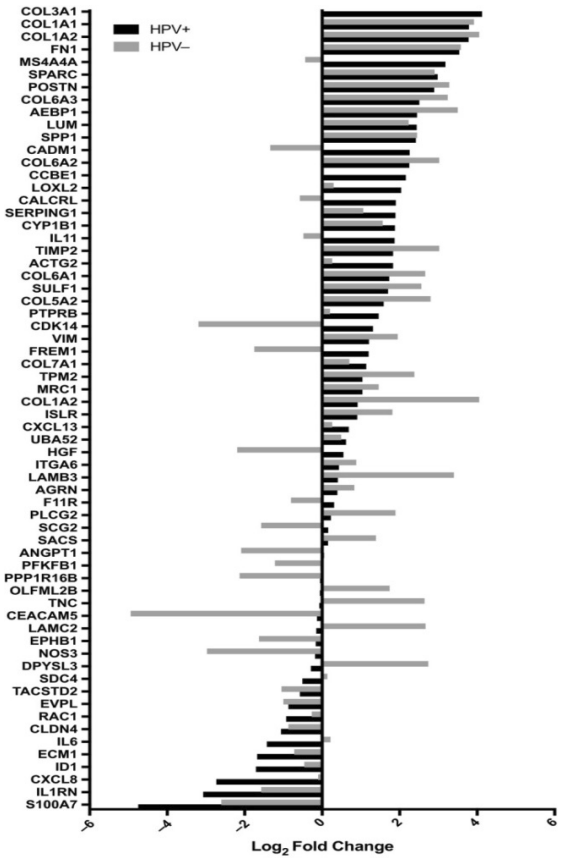
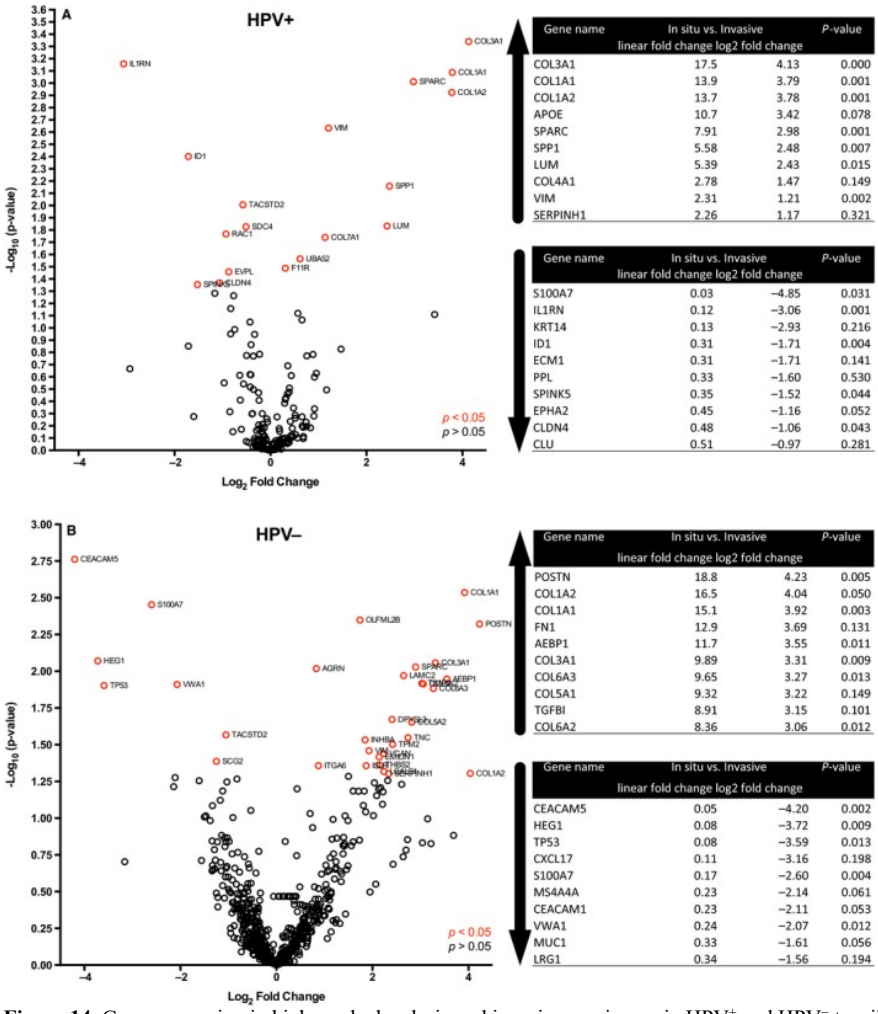


Figure 13. Comparison between HPV$^+$ and HPV$^-$ TSCC and BOTSCC with regard to in situ and invasive cancer. Log_2 fold change difference between high-grade dysplasia and invasive carcinoma comparing statistically significant differences ($P < .05$) in mRNA expression in the HPV$^+$ and HPV$^-$ samples separately. All genes showing a positive fold change value have a higher expression in invasive carcinoma compared to high-grade dysplasia, and genes showing a negative fold change value have a lower expression in invasive carcinoma compared to high-grade dysplasia. (From Paper III, with permission from the publisher).

Notably, significant differential mRNA expressions were disclosed in some genes such as for *COL1A1, SPARC,* and *LGALS1* that were more active in invasive carcinoma in comparison to high-grade dysplasia, while the contrary was observed for *S100A7*, that was more highly active in high-grade dysplasia (Figure 14).



Gene name	In situ vs. Invasive linear fold change	log2 fold change	*P*-value
COL3A1	17.5	4.13	0.000
COL1A1	13.9	3.79	0.001
COL1A2	13.7	3.78	0.001
APOE	10.7	3.42	0.078
SPARC	7.91	2.98	0.001
SPP1	5.58	2.48	0.007
LUM	5.39	2.43	0.015
COL4A1	2.78	1.47	0.149
VIM	2.31	1.21	0.002
SERPINH1	2.26	1.17	0.321

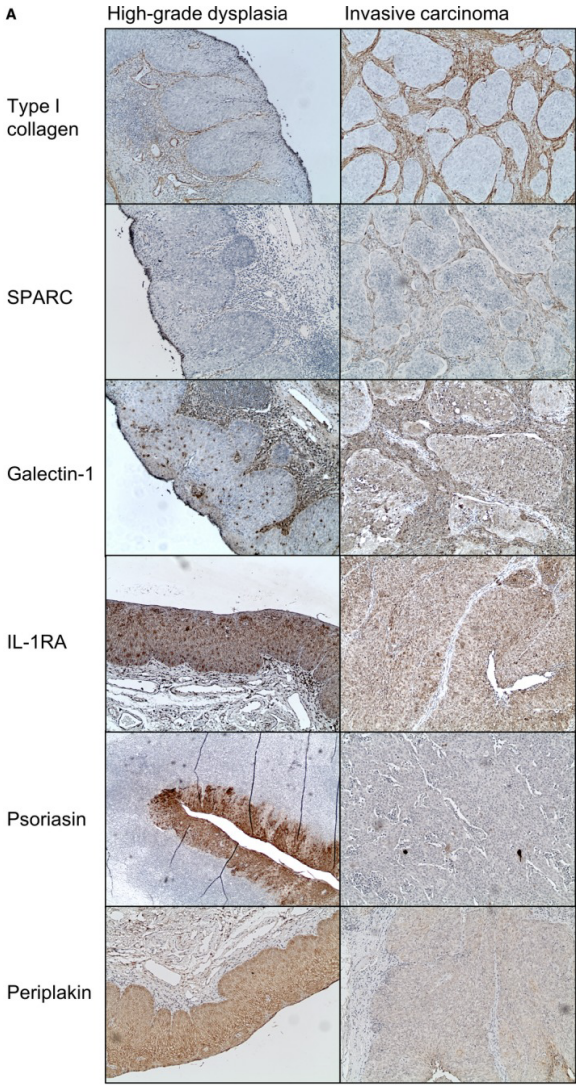
Gene name	In situ vs. Invasive linear fold change	log2 fold change	*P*-value
S100A7	0.03	−4.85	0.031
IL1RN	0.12	−3.06	0.001
KRT14	0.13	−2.93	0.216
ID1	0.31	−1.71	0.004
ECM1	0.31	−1.71	0.141
PPL	0.33	−1.60	0.530
SPINK5	0.35	−1.52	0.044
EPHA2	0.45	−1.16	0.052
CLDN4	0.48	−1.06	0.043
CLU	0.51	−0.97	0.281

Gene name	In situ vs. Invasive linear fold change	log2 fold change	*P*-value
POSTN	18.8	4.23	0.005
COL1A2	16.5	4.04	0.050
COL1A1	15.1	3.92	0.003
FN1	12.9	3.69	0.131
AEBP1	11.7	3.55	0.011
COL3A1	9.89	3.31	0.009
COL6A3	9.65	3.27	0.013
COL5A1	9.32	3.22	0.149
TGFBI	8.91	3.15	0.101
COL6A2	8.36	3.06	0.012

Gene name	In situ vs. Invasive linear fold change	log2 fold change	*P*-value
CEACAM5	0.05	−4.20	0.002
HEG1	0.08	−3.72	0.009
TP53	0.08	−3.59	0.013
CXCL17	0.11	−3.16	0.198
S100A7	0.17	−2.60	0.004
MS4A4A	0.23	−2.14	0.061
CEACAM1	0.23	−2.11	0.053
VWA1	0.24	−2.07	0.012
MUC1	0.33	−1.61	0.056
LRG1	0.34	−1.56	0.194

Figure 14. Gene expression in high-grade dysplasia and invasive carcinoma in HPV$^+$ and HPV$^-$ tonsillar and base of tongue squamous cell carcinoma separately. (A) Volcano plot showing differentially expressed genes between high-grade dysplasia and invasive carcinoma by $\log_2$ fold change (*x*-axis) and $-\log_{10}$of *P*-value (*y*-axis) in HPV$^+$ tonsillar and base of tongue squamous cell carcinoma cases. The top 10 genes with increased and decreased linear fold change as well as $\log_2$ fold change expression between high-grade dysplasia and invasive carcinoma are together with *P*-values presented in the tables to the right. (B) Same as in (A), here however with HPV$^-$ tonsillar and base of tongue squamous cell carcinoma cases. (With permission from the publisher Paper III).

Moreover, obvious differences in protein expression were observed for SPARC, psoriasin, type I collagen and galectin-1 in not only HPV$^+$, but also in HPV$^-$ tumours (Figure 15).



B

	Number of samples with higher expression in high-grade dysplasia	Number of samples with higher expression invasively	Number of samples with no difference between high-grade dysplasia and invasive carcinoma	P-value
SPARC	0	17	6	<0.0001
Psoriasin	14	0	10	0.0001
Type I Collagen	0	13	10	0.0002
Galectin-1	1	9	13	0.0215
IL1-RA	5	0	19	0.0625
Galectin-1 TILs	1	6	16	0.1250
Periplakin	8	3	12	0.2266

Figure 15. (A) Examples of immunohistochemistry stainings of type I collagen, SPARC, galectin-1, IL-1RA, psoriasin, and periplakin, in high-grade dysplasia/carcinoma in situ an in invasive carcinoma. Expression of type I collagen and SPARC is higher in the tissue surrounding invasive carcinoma, and galectin-1 has a higher expression in invasive carcinoma, whereas IL-1RA, psoriasin, and periplakin show a higher expression in high-grade dysplasia compared to invasive carcinoma. (B) Summary of immunohistochemical protein evaluation scores and *P*-values for selected proteins, HPV^+, and HPV^- tumors combined. (Paper III, with permission from the publisher)

.

7.3.5 Discussion

In paper III, differences in gene expression in high-grade dysplasia/cancer in situ as compared to invasive HPV^{+} and HPV^{-} TSCC/BOTSCC were reported. Noteworthy differential mRNA expressions were observed for *COL1A1, SPARC*, and *LGALS1* which were expressed to a higher degree in invasive carcinoma in comparison to high-grade dysplasia, while the contrary was observed for *S100A7*, that was more expressed in high-grade dysplasia. By IHC on the protein level these findings were confirmed. Also, many resemblances in signal pathway activities and gene expression in both HPV^{+} and HPV^{-} TSCC/BOTSCC were observed.

Nonetheless, in HPV^{+} TSCC/BOTSCC dysplasia is rarely disclosed and entirely premalignant stages have been debated, although found frequently in HPV^{-} TSCC/BOTSCC. For this reason, some researchers have proposed that there are no HPV^{+} premalignant phases in TSCC/BOTSCC (Holmes BJ et al 2019), but on the other hand they have been described by others, who also pursued some molecular studies of HPV^{+} TSCC/BOTSCC dysplasia (Mooren JJ et al., 2014, Masterson L et al., 2015).

Here, however we aimed to investigate several genes and noted several similarities in both HPV^{+} and HPV^{-} tumours both in dysplasia and invasive cancer, indicating that indeed that dysplastic HPV^{+} TSCC/BOTSCC is distinguishable from invasive HPV^{+} TSCC/BOTSCC, and we propose that premalignant stages can exist in HPV^{+} TSCC/BOTSCC.

With regard to signalling pathways, an upregulation of ECM receptor interaction, ECM structure, collagen family, and cellular growth factor pathways would likely be expected in HPV^{+} as well as HPV^{-} invasive carcinoma. In addition, a higher stem cell associated score was also disclosed in invasive cancer when compared to dysplasia, and this was detected to a higher extent in HPV^{+} than in HPV^{-} samples. This would be in concordance with a showing of that the cancer stem cell pool more extensive in HPV^{+} than in HPV^{-} OPSCC (Zhang M et al., 2014).

With regard to the transcriptional level, here collagen genes *COL1A1, COL1A2, COL3A1*, were the ones most significantly active in invasive cancer when correlated to their activity in dysplasia in both HPV^{+} and HPV^{-} cancers. This is likely not surprising since they are abundant in the ECM and they are also important with regard to tumour progression and

cell proliferation and the association to outcome in various tumour types (Kirkland SC 2009).

SPARC (secreted protein acidic and rich in cysteine) another matricellular protein, which regulates cell-matrix interactions and signalling pathways in cells, was also found increased in invasive cancer in comparison to its expression in dysplasia. This protein is assumed to be supplied by cancer associated fibroblasts in the tumour stroma and proposed to be involved in the regulation of tumour cell growth and metastasis (Framson PE et al., 2009, Paulsson J and Micke P 2014).

Also of note was *S100A7* encoding the protein psoriasin part of the S100 family harbouring calcium-binding motifs and a significant mediator of cell maturation and survival, with its expression correlated to tumour progression and survival and progression with opposing functions (Hattinger E et al 2013). Notably, psoriasin was also found significantly more overexpressed in oral leucoplakia lesions with squamous cell hyperplasia or dysplasia and in HNSCC, when compared to normal tissue, and the nuclear accumulation in the latter was associated with worse clinical outcome (Tripathi SC et al., 2010) . Yet another gene of note is *IL1RN* which codes for interleukin-1 receptor antagonist (IL-1RA) which in turn inhibits the IL-1 receptor. IL-1RA protein expression has been proposed to inhibit angiogenesis, tumour growth and metastasis in vivo in mice (Lewis AM et al., 2006). Here, we found that *IL1RN* gene expression is decreased in invasive cancer in comparison to dysplastic epithelium, particularly in HPV^+ TSCC/BOTSCC. Notably, corresponding data have been demonstrated in CIN3 vs. invasive cervical carcinoma (den Boon JA et al., 2015).

There were limitations in this study, few samples were included and the amount of RNA that could be extracted was low.

7.3.6 Conclusion

As far as we know, this is the first study showing similarities and differences in gene and protein expression in invasive and dysplastic HPV^+ and HPV^- TSCC/BOTSCC.

7.4 PAPER IV. WHOLE-EXOME SEQUENCING OF HPV POSITIVE TONSILLAR AND BASE OF TONGUE SQUAMOUS CELL CARCINOMAS REVEALS A GLOBAL MUTATIONAL PATTERN ALONG WITH RELAPSE-SPECIFIC SOMATIC VARIANTS

Ährlund-Richter A, Holzhauser S, Dalianis T, Näsman A, Mints M.
Cancers (Basel). 2021 Dec 24;14(1):77.

7.4.1 Aim

To disclose more biomarkers in HPV^+ and HPV^- TSCC/BOTSCC to improve prediction of prognosis and for possible use in prospective trials for de-escalation or targeted therapy.

7.4.2 Background

Patients with HPV^+ TSCC/BOTSCC with IHC biomarkers e.g. loss of/low HLA class I, high numbers of $CD8^+$ tumour infiltrating lymphocytes (TILs) present a high 3-year survival (92-100%) compared to 80% for HPV^+TSCC/BOTSCC patients in general (for review see Näsman A et al., 2021). Also, molecular markers such as HPV16 E2 mRNA expression were also shown to be prognostic (Ramqvist T, Mints M et al., 2015). By combining some of them we identified 56% of all patients with an excellent prognosis (Bersani C, Mints M et al., 2017). To disclose more markers, hotspot gene sequencing of 50 cancer related genes was done, and we disclosed mutated FGFR3 as a predictive marker (Bersani C, Sivars L et al., 2017). However, to better tailor patient treatment more prognostic markers or markers useful for targeted therapy would be of benefit. For this reason whole exome sequencing (WES) was done on formalin fixed paraffin embedded (FFPE) material of patients with HPV^+ TSCC/BOTSCC that had a recurrence/relapse or not.

7.4.3 Material and Methods

Forty patients, 20 with and 20 without recurrence with HPV^+ TSCC/BOTSCC were selected. From FFPE samples from their primary tumours, their recurrences or metastasis, were micro-dissected in order to separate tumour from normal tissue and WES was performed at SciLife Stockholm. Bioinformatics was be performed using by a bioinformatic pipe line developed by Valtteri Wirta, further analysed as described in more detail above and in Paper IV.

7.4.4 Results

Dataset Summary

After omitting all samples with insufficient quality, 17 primary cancer samples from 17 individuals with relapse (14 TSCC and 3 BOTSCC), 10 relapses (i.e., 5 local recurrences and 5 distant metastases) and 18 primary cancers from individuals (16 TSCC and 2 BOTSCC) without relapses, could be used for continued analysis. To aid somatic variant calling, normal material was available with adequate quality from 13 individuals with recurrence and 12 individuals without recurrence. After WES and variant filtering performed as described in more detail (see Paper I), totally 6147 unique variants (SNVs and structural variants) harboured in 4184 genes in the dataset were disclosed. Per sample, an average of 236 variants harboured in 201 genes were disclosed (For details see Paper IV).

Per-Variant Analyses

Specific variants differentially existing in primary tumours from patients with and without recurrences, were analysed first and the analysis was restricted to variants observed in the tumours of at least four patients (>20%) of either group while being absent in the other group. Five variants fulfilled these criteria and are depicted in Supplementary Table S1 of Paper IV. A high-impact deletion in the *CDC27* gene, was the only variant enriched ($p < 0.05$) in recurrent samples and was present in 5/17 primary tumours of individuals with recurrence, together with in one local relapse. This variant was then notably absent in all nonrecurrent patient samples.

In the primary tumours of patients without recurrence, a substitution in *KCNJ12* was enriched ($p < 0.05$), and disclosed in 5/18 tumour samples. In four primary tumour samples of patients without recurrence, variants in *KRTAP4-11*, *NBPF20* and *LILRB3* were also found and these were absent in all primaries cancers of individuals that had relapses and they were also absent in the local/distant relapses. However, these were not significant ($p = 0.1$ for all variants).

Concentrating on common occurring variants, 98 variants harboured within 80 unique genes (Supplementary Table S2 of Paper IV were disclosed in >25% of all patient samples irrespective of relapse or not (nine or more samples). *BCLAF1* and *OVCH2*, both harbouring deletions, were the most common variants and they were both demonstrated in

12 samples, and they were accompanied in frequency by a variant of *OR2T35* with a specific substitution that was found in 11 primary tumour samples. In Figure 16 the most relevant variants are depicted.

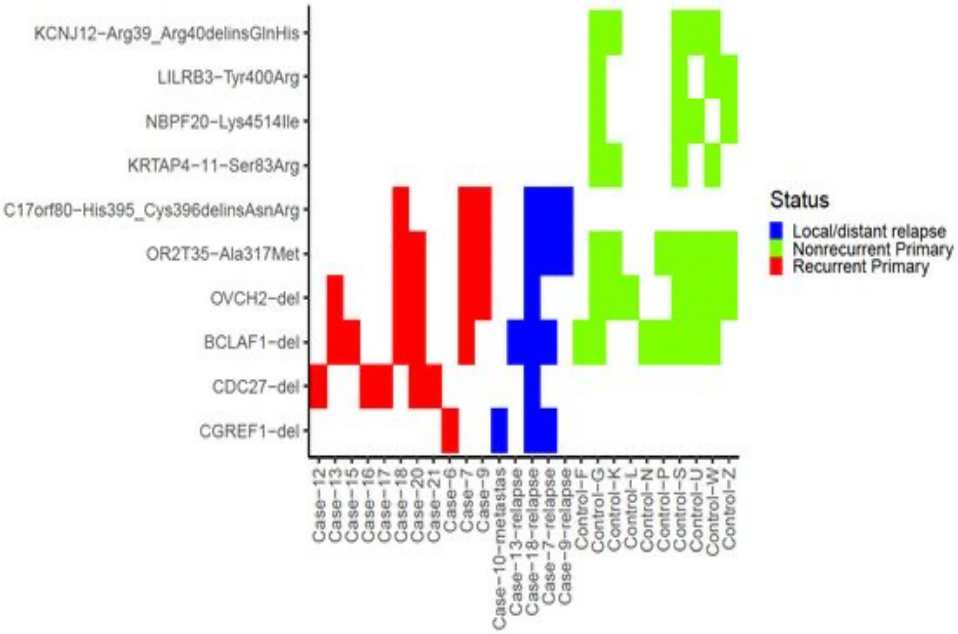


Figure 16. Heatmap of the most commonly occurring variants, variants specific to either primary of patients with local/distant recurrence (denoted as cases) or primaries of nonrecurrent patients (denoted as controls) and variants unique to relapses (denoted metastasis, if distant metastasis, or relapse, if local recurrence). Variants are on the y-axis, samples are on the x-axis. Color indicates sample category. Samples where none of the variants were present were excluded from the plot. (From paper IV, with permission from the publisher).

Per-Gene Analysis

The corresponding analyses done for the variants were also done on the gene level. All above noted unique variants per gene were collated and when any of these variants was found in a sample, the gene was calculated as mutated in that sample. There were no genes that were mutated specifically in any one of the primary cancers of individuals with a relapse, but in 4/18 primaries of patients with no recurrences, *HERC2*, was uniquely mutated in 4/18 of their primaries, while this was not the case in any of the primaries or recurrences of individuals with relapses ($p = 0.1$).

In total, 26 genes were found to be mutated in >30% of all primary cancers irrespective of clinical outcome and these are shown in Figure 17.

AQP7 was the most frequently mutated gene. In addition, *BCLAF1*, *OVCH2* and *OR2T35*, having the most frequent unique variants, were also disclosed as genes mutated among largest number of samples.

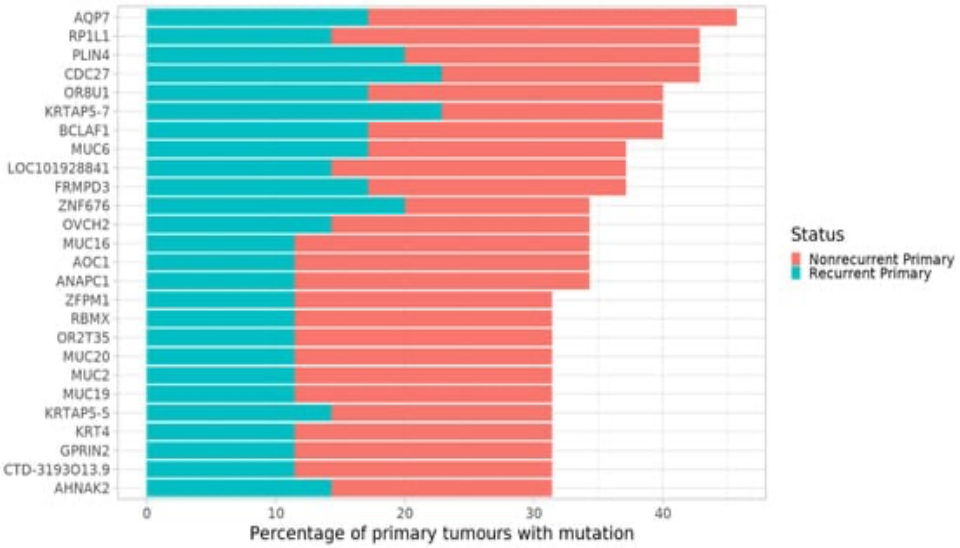


Figure 17. Bar plot of all genes mutated in >30% of primary tumours. Genes are on the y-axis. The x-axis shows the percentage of all primary tumors where a gene is mutated. Colour indicates recurrent or nonrecurrent status. (From Paper IV, with permission from the publisher).

The unique variant of *CDC27*, found only in primary tumours of individuals with recurrence, was also included in the group of the commonly mutated genes. More specifically *CDC27* was found to be mutated in primaries of 8/17 individuals with and in 7/18 individuals without recurrence, as well as in three relapsed samples.

Among the 26 genes above, three *KRT4*, *KRTAP5-5* and *KRTAP5-7* were keratin-related genes, and another five were mucins. When performing enrichment analysis, pathways correlated to carbohydrates and the extracellular matrix were enriched significantly. In addition, notably mucins were found in all these enriched pathways.

Mutations in Hotspot Genes

In HNSCC some genes are known to be frequently mutated, but these were not disclosed among those above so we especially analysed *FGFR3* and *PIK3CA* known to be commonly affected in HPV$^+$ TSCC/BOTSCC/OPSCC and we also focused on *TP53* mutations.

FGFR3 was found to be mutated in 2/17 primary cancers of individuals with relapse, and in 3/18 individuals that had not relapsed, while *PIK3CA* was found mutated in the primary cancers of one individual with a recurrence as well as in 4/18 individuals without recurrence. *TP53* was, rarely mutated in this HPV$^+$ cohort, as could be anticipated and mutations were identified only in two primary cancers of individuals with a recurrence.

7.4.5 Discussion

In paper IV, WES was completed in primary tumours of 17 HPV$^+$ TSCC/BOTSCC patients with and 18 patients without relapses, and in 10 local/distant recurrences of the former patients, all previously micro-dissected to distinguish tumour from normal material. In the *CDC27* gene a deletion of high impact was found only in tumours of patients that had relapsed but in no primary tumours of patients without recurrence.

In >30% of all cases, three variants—more specifically two in *BCLAF1* and *OVCH2* with deletions and one with a substitution in *OR2T35* as well as 26 mutated genes were disclosed, thereby possibly consisting part of a global mutational signature for HPV$^+$ TSCC/BOTSCC. *AQP7* was the most frequently mutated gene since it was found in >45% of all primary tumours. In addition, several keratin-associated genes and mucins were also often mutated. Upon performing enrichment analysis pathways associated to extracellular matrix and carbohydrates were disclosed. Previously described mutations in *PIK3CA* and *FGFR3* were noted, but these were not among the most frequent ones, and mutations of *TP53* were rare.

The specific deletion in *CDC27* frequently found (5/17, 29%) and specific only to tumours of patients with a recurrence was a novel finding. Upon validation, this variant may possibly be utilized as prognostic or as a therapeutic target. *CDC27* is namely a component of the anaphase-promoting complex/cyclosome and it has been shown that variations and overexpression in *CDC27* can influence the cell cycle, mitosis, cancer pathogenesis and clinical outcome (Kazemi-Sefat et al., 2021, Melloy PG 2020). Targeted therapy against *CDC27* is currently not available, but antibodies directed towards *CDC27* may be of potential diagnostic interest.

Among the most frequently mutated genes in the above cohort, we found keratin-associated proteins and mucins thereby indicating them to be part of a global HPV$^+$ OPSCC mutational signature and both groups have associations to e.g. assisting motility and

invasion (Berens EB et al., 2017, Hollingsworth M.A et al 2004). For tailored therapy of special interest is that non-small cell lung cancer (NSCLC) presenting *MUC19* mutations are regarded to respond well to inhibitors against PD1 (Zhou L et al., 2021), a gene thereby interesting to study in this context for HPV$^+$ OPSCC. *BCLAF1* was a commonly affected gene, due to many *BCLAF* variants, and in the cancer setting described as an associated transcription factor for Bcl2 (Ma Z et al., 2021). Furthermore, variants of this gene may also be of interest for both diagnostic and targeting purposes.

In our cohort, *AQP7* was the most mutated gene and of interest since it codes for a membrane channel with known metabolic roles and it has also been suggested as a target for some tumours e.g. breast cancer (Dai C et al., 2020).

In a report similar ours, WES was done in primary cancers of 51 HPV$^+$ OPSCC, where 16 relapsed and 35 did not recur, and *KMT2D* was demonstrated to be the most often mutated gene in not only primary (14%) but also recurrent (42%) HPV$^+$ OPSCC (Harbison RA et al., 2018). *KMT2D* mutations were detected in 3/35 (9%) primary cancers in our study, but not in the relapses, but upon removing our filtering for protein-altering variants, the numbers of detected *KMT2D* mutations increased to 14% of primaries and 20% of recurrences, indicating that the more stringent variant filtering in our study, could explain the differences in the two studies.

This study has its limitations, where one is the limited number of patients and this is due to that relapses are uncommon. Other limitations were that we used FFPE samples and that we did not obtain normal tissue for all the cases. Nevertheless, some of our major findings (such as the specific *CDC27* deletion, and that *BCLAF1* was commonly affected) were detected in samples irrespective if these samples had or did not have paired normal material. Nevertheless, one benefit in this study was that normal tissue was laser micro-dissected from the same tissue block as the tumour. Possible paraffin-related artefacts were thereby existent in the both the tumour and control tissue, thereby circumventing false calling of somatic variants.

To summarize, a specific *CDC27* variant was disclosed as specific for cancers of patients with HPV$^+$ OPSCC with recurrence. In addition, a mutational signature common for HPV$^+$ OPSCC patients irrespective of prognostic status, including mucins and keratin-associated proteins, as well as some variants, e.g. a *BCLAF1* variant were observed.

7.4.6 Conclusions

In this paper, a specific *CDC27* variant that was unique in HPV$^+$ TSCC/BOTSCC primaries of patients with relapse was disclosed. In addition, a common mutational signature for HPV$^+$ TSCC/BOTSCC patients including mucin and keratin-associated proteins as well as specific variants, e.g. a *BCLAF1* variant were disclosed. The data could be of importance both for assessing outcome and initiating possible future targeted therapy. However, before this is the case the data require validation.

8 CONCLUSIONS

I. Human papillomavirus (HPV) vaccination has reached a high coverage at a Stockholm youth clinic during the years 2008-2018 and has resulted in a significant decrease of HPV16, 18, 6 and 11 cervical prevalence especially in vaccinated individuals. However, prevalence of some high risk (HR) - HPV types e.g. HPV39, 51, 52, 56, and 59 still remained.

II. Infection with HR-HPVs were correlated with some vaginal microbiota bacterial vaginosis-associated bacteria (BVAB), such as *Sneathia, Prevotella*, and *Megasphaera* were associated with HPV infection regardless of vaccination status and age. Moreover, women with non-*Lactobacilli* dominant vaginal microbiota much more likely to present HR-HPV types, as compared to those with *L. crispatus* dominated vaginal microbiota, after adjustment for vaccination status and age. These findings provide additional support to the close correlation between HPV status and vaginal microbiota. Importantly, they may also indicate an intriguing possibility to decrease HPV vaginal prevalence by influencing vaginal microbiota.

III. HPV^+ and HPV^- dysplasia were both similarly differentiated from invasive tonsillar and base of tongue squamous cell carcinomas (TSCC/BOTSCC). Furthermore, both differences and similarities were detected in gene expression between dysplastic and invasive HPV^+ and HPV^- TSCC/BOTSCC.

IV. A variant with a specific *CDC27* deletion was observed in roughly 30% of samples of patients with recurrent disease and could upon validation be a prospective marker to predict clinical outcome. In addition, a number of commonly mutated genes, including mucins and keratin-associated proteins as well as variants of e.g. *BCLAF1*, were identified as potential markers and should be studied more extensively for use as targets for targeted therapy.

V. FUTURE PERSPECTIVES

The present work can be continued in several ways. Below some suggestions indicated in chronological order of Papers I-IV.

I. It is of great importance to work for a high HPV vaccine coverage in Sweden and globally. Furthermore, it is of great importance to keep screening for HPV, as well as to improve today's HPV vaccines in order to eliminate as many HR-HPV types as possible.

II. Further studies on vaginal microbiota and HPV are needed to investigate bi-directional effects between HPV status and vaginal microbiota. More specifically, longitudinal studies with large cohorts would provide valuable information regarding how vaginal microbiota play roles in HPV infection and cancer development. In addition, studies on potential mechanisms for why certain vaginal microbiota are related to HPV infection could be essential in understanding these interactions and of use for future clinical trials.

III. Some of the detected genes in paper III, e.g. *S100A7* and its protein psoriasin and IL1RN encoding the inerleukin-1 receptor antagonist IL-1RA may be of interest to follow up further in TSCC/BOTSCC lesions or other lesions within the head and neck region as diagnostic markers/diagnostic patterns of invasive disease.

IV. Further studies are needed to validate the detected CDC27 deletion variant in recurrent TSCC/BOTSCC in additional TSCC/BOTSCC cohorts to confirm whether it is a prognostic marker or not. Further studies are also needed to validate whether the commonly detected variants or mutated genes indeed are frequently present in HPV^+ TSCC/BOTSCC irrespective of their prognosis

9 ACKNOWLEDGEMENTS

Big thanks to many people for encouraging me with the work on this thesis. Below some specifications.

My supervisors

First of all a big thanks to my supervisor *Anders Näsman* for taking me on, and for especially making histology exciting. Big thanks also to my co-supervisor *Du Juan* that always took things with an easy stride and has been there all along. Big thanks, also especially to my co-supervisor *Michael Mints* for encouraging and motivating me to pursue and write this thesis in the first place. Big thanks, to *Stefan Holzhauser* for his genuine curiousity and for always being happy to try out different approaches when everything seemed stuck.

To present and previous group members

Big thanks to my team-mate Mark Zupancic for always keeping spirits up in the team and always helping out. Big thanks Nathalie Grün that always has the best career advice. Big thanks to Rania Kostopoulou for her wisdom on working in academic-life, for Ramona Ursu for getting the lab group together and to Nikos Tertipis for his amazing after-work cooking.

Big thanks also to Linnea Haeggblom that got me pipetting and dancing properly, and to Leila Mirzaie's amazing focus yet being able to have fun, to Joar Franzén that made us laugh ,as well as to Lars Sivars for teaching me anatomy.

Big thanks to Torbjörn Ramqvist for teaching me to solve things on the fly, to Cinzia Bersani for teaching me to word myself properly and to Cecilia Nordfors (Smedman) for her viral passion for the subject.

Big thanks, also to our clinical colleagues Eva Munck-Wikland, Linda Marklund and Lalle Hammarstedt Nordenwall and more recently David Landin for always keeping the association to our patients.

To members outside the group

Big thanks to Liqin Cheng for working in parallel with me on Papers I and II.

Big thanks to Weimin Ye and his group for teaching me about medical databases and epidemiology and to all my friends for their curious questions and support.

Big thanks to all the midwives at the youth clinic in Stockholm, without your help this thesis would not have been possible.

To my immediate family

A big thanks to the encouraging words of my sisters Irene and Sofie and my nieces Nina and Paola. Likewise a big thanks to uncle Hercules my great brother-in-laws Mattias and Iakovos. A big thanks to my parents Tina and Lars, who always have my back, and always ask the best questions. Finally, a big thanks to my girlfriend Layla for inspiring me every day.

10 REFERENCES

Ährlund-Richter A, Cheng L, Hu YOO, Svensson M, Pennhag AAL, Ursu RG, Haeggblom L, Grün N, Ramqvist T, Engstrand L, Dalianis T, Du J. Changes in Cervical Human Papillomavirus (HPV) Prevalence at a Youth Clinic in Stockholm, Sweden, a Decade After the Introduction of the HPV Vaccine. Front Cell Infect Microbiol. 2019 Mar 20;9:59. doi: 10.3389/fcimb.2019.00059.

Ährlund-Richter A, Holzhauser S, Dalianis T, Näsman A, Mints M. Whole-Exome Sequencing of HPV Positive Tonsillar and Base of Tongue Squamous Cell Carcinomas Reveals a Global Mutational Pattern along with Relapse-Specific Somatic Variants. Cancers (Basel). 2021 Dec 24;14(1):77. doi: 10.3390/cancers14010077.

Anaya-Saavedra G, Ramírez-Amador V, Irigoyen-Camacho ME, García-Cuellar CM, Guido-Jiménez M, Méndez-Martínez R, García-Carrancá A. High association of human papillomavirus infection with oral cancer: a case-control study. Arch Med Res. 2008 Feb;39(2):189-97. doi: 10.1016/j.arcmed.2007.08.003.

Arbyn M, Weiderpass E, Bruni L, de Sanjosé S, Saraiya M, Ferlay J, Bray F. Estimates of incidence and mortality of cervical cancer in 2018: a worldwide analysis. Lancet Glob Health. 2020 Feb;8(2):e191-e203. doi: 10.1016/S2214-109X(19)30482-6. Erratum in: Lancet Glob Health. 2022 Jan;10(1):e41.

Arias-Pulido H, Peyton CL, Joste NE, Vargas H, Wheeler CM. Human papillomavirus type 16 integration in cervical carcinoma in situ and in invasive cervical cancer. J Clin Microbiol. 2006 May;44(5):1755-62. doi: 10.1128/JCM.44.5.1755-1762.2006.

Armstrong LR, Derkay CS, Reeves WC. Initial results from the national registry for juvenile-onset recurrent respiratory papillomatosis. RRP Task Force. Arch Otolaryngol Head Neck Surg. 1999 Jul;125(7):743-8. doi: 10.1001/archotol.125.7.743.

Attner P, Du J, Näsman A, Hammarstedt L, Ramqvist T, Lindholm J, Marklund L, Dalianis T, Munck-Wikland E. The role of human papillomavirus in the increased incidence of base of tongue cancer. Int J Cancer. 2010 Jun 15;126(12):2879-84. doi: 10.1002/ijc.24994.

Attner P, Du J, Näsman A, Hammarstedt L, Ramqvist T, Lindholm J, Marklund L, Dalianis T, Munck-Wikland E. Human papillomavirus and survival in patients with base of tongue cancer. Int J Cancer. 2011 Jun 15;128(12):2892-7. doi: 10.1002/ijc.25625. Erratum in: Int J Cancer. 2012 Nov 1;131(9):E1182.

Balermpas P, Rödel F, Liberz R, Oppermann J, Wagenblast J, Ghanaati S, Harter PN, Mittelbronn M, Weiss C, Rödel C, Fokas E. Head and neck cancer relapse after chemoradiotherapy correlates with CD163+ macrophages in primary tumour and CD11b+ myeloid cells in recurrences. Br J Cancer. 2014 Oct 14;111(8):1509-18. doi: 10.1038/bjc.2014.446.

Berens EB, Sharif GM, Schmidt MO, Yan G, Shuptrine CW, Weiner LM, Glasgow E, Riegel AT, Wellstein A. Keratin-associated protein 5-5 controls cytoskeletal function and cancer cell vascular invasion. Oncogene. 2017 Feb 2;36(5):593-605. doi: 10.1038/onc.2016.234.

Bersani C, Mints M, Tertipis N, Haeggblom L, Sivars L, Ährlund-Richter A, Vlastos A, Smedberg C, Grün N, Munck-Wikland E, Näsman A, Ramqvist T, Dalianis T. A model using concomitant markers for predicting outcome in human papillomavirus positive oropharyngeal cancer. Oral Oncol. 2017 May;68:53-59. doi: 10.1016/j.oraloncology.2017.03.007.

Bersani C, Sivars L, Haeggblom L, DiLorenzo S, Mints M, Ährlund-Richter A, Tertipis N, Munck-Wikland E, Näsman A, Ramqvist T, Dalianis T. Targeted sequencing of tonsillar and base of tongue cancer and human papillomavirus positive unknown primary of the head and neck reveals prognostic effects of mutated FGFR3. Oncotarget. 2017 May 23;8(21):35339-35350. doi: 10.18632/oncotarget.15240.

Bersani C, Mints M, Tertipis N, Haeggblom L, Näsman A, Romanitan M, Dalianis T, Ramqvist T. MicroRNA-155, -185 and -193b as biomarkers in human papillomavirus positive and negative tonsillar and base of tongue squamous cell carcinoma. Oral Oncol. 2018 Jul;82:8-16. doi: 10.1016/j.oraloncology.2018.04.021.

Bernard HU, Burk RD, Chen Z, van Doorslaer K, zur Hausen H, de Villiers EM. Classification of papillomaviruses (PVs) based on 189 PV types and proposal of taxonomic amendments. Virology. 2010 May 25;401(1):70-9. doi: 10.1016/j.virol.2010.02.002.

Biancalana V, Laporte J. Diagnostic use of Massively Parallel Sequencing in Neuromuscular Diseases: Towards an Integrated Diagnosis. J Neuromuscul Dis. 2015 Sep 2;2(3):193-203. doi: 10.3233/JND-150092.

Brianti P, De Flammineis E, Mercuri SR. Review of HPV-related diseases and cancers. New Microbiol. 2017 Apr;40(2):80-85. (no doi). PMID: 28368072.

Brotman RM, Shardell MD, Gajer P, Tracy JK, Zenilman JM, Ravel J, Gravitt PE. Interplay between the temporal dynamics of the vaginal microbiota and human papillomavirus detection. J Infect Dis. 2014 Dec 1;210(11):1723-33. doi: 10.1093/infdis/jiu330.

Bruni L, Diaz M, Castellsagué X, Ferrer E, Bosch FX, de Sanjosé S. Cervical human papillomavirus prevalence in 5 continents: meta-analysis of 1 million women with normal cytological findings. J Infect Dis. 2010 Dec 15;202(12):1789-99. doi: 10.1086/657321.

Bruni L, Saura-Lázaro A, Montoliu A, Brotons M, Alemany L, Diallo MS, Afsar OZ, LaMontagne DS, Mosina L, Contreras M, Velandia-González M, Pastore R, Gacic-Dobo M, Bloem P. HPV vaccination introduction worldwide and WHO and UNICEF estimates of national HPV immunization coverage 2010-2019. Prev Med. 2021 Mar;144:106399. doi: 10.1016/j.ypmed.2020.106399. Erratum in: Prev Med. 2022 Feb;155:106925.

Brusselaers N, Shrestha S, van de Wijgert J, Verstraelen H. Vaginal dysbiosis and the risk of human papillomavirus and cervical cancer: systematic review and meta-analysis. Am J Obstet Gynecol. 2019 Jul;221(1):9-18.e8. doi: 10.1016/j.ajog.2018.12.011.

Burd EM. Human papillomavirus and cervical cancer. Clin Microbiol Rev. 2003 Jan;16(1):1-17. doi: 10.1128/CMR.16.1.1-17.2003.
Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJ, Holmes SP. DADA2: High-resolution sample inference from Illumina amplicon data. Nat Methods. 2016 Jul;13(7):581-3. doi: 10.1038/nmeth.3869.

Campo MS, Graham SV, Cortese MS, Ashrafi GH, Araibi EH, Dornan ES, Miners K, Nunes C, Man S. HPV-16 E5 down-regulates expression of surface HLA class I and reduces recognition by CD8 T cells. Virology. 2010 Nov 10;407(1):137-42. doi: 10.1016/j.virol.2010.07.044.

Cancer Genome Atlas Network. Comprehensive genomic characterization of head and neck squamous cell carcinomas. Nature. 2015 Jan 29;517(7536):576-82. doi: 10.1038/nature14129.

Chen Z, Wong PY, Ng CWK, Lan L, Fung S, Li JW, Cai L, Lei P, Mou Q, Wong SH, Wu WKK, Li RJ, Meehan K, Lui VWY, Chow C, Lo KW, Chan ABW, Boon SS, Lau EHL, Yeung Z, Chan KCA, Wong EWY, Cheng ASL, Yu J, Chan PKS, Chan JYK. The Intersection between Oral Microbiota, Host Gene Methylation and Patient Outcomes in Head and Neck Squamous Cell Carcinoma. Cancers (Basel). 2020 Nov 18;12(11):3425. doi: 10.3390/cancers12113425.

Cheng L, Norenhag J, Hu YOO, Brusselaers N, Fransson E, Ährlund-Richter A, Guðnadóttir U, Angelidou P, Zha Y, Hamsten M, Schuppe-Koistinen I, Olovsson M, Engstrand L, Du J. Vaginal microbiota and human papillomavirus infection among young Swedish women. NPJ Biofilms Microbiomes. 2020 Oct 12;6(1):39. doi: 10.1038/s41522-020-00146-8.

Cheng L, Wang Y, Du J. Human Papillomavirus Vaccines: An Updated Review. Vaccines (Basel). 2020 Jul 16;8(3):391. doi: 10.3390/vaccines8030391.

Chung CH, Guthrie VB, Masica DL, Tokheim C, Kang H, Richmon J, Agrawal N, Fakhry C, Quon H, Subramaniam RM, Zuo Z, Seiwert T, Chalmers ZR, Frampton GM, Ali SM, Yelensky R, Stephens PJ, Miller VA, Karchin R, Bishop JA. Genomic alterations in head and neck squamous cell carcinoma determined by cancer gene-targeted sequencing. Ann Oncol. 2015 Jun;26(6):1216-1223. doi: 10.1093/annonc/mdv109.

Cioni B, Jordanova ES, Hooijberg E, van der Linden R, de Menezes RX, Tan K, Willems S, Elbers JBW, Broeks A, Bergman AM, Zuur CL, de Boer JP. HLA class II expression on tumor cells and low numbers of tumor-associated macrophages predict clinical outcome in oropharyngeal cancer. Head Neck. 2019 Feb;41(2):463-478. doi: 10.1002/hed.25442.

Cladel NM, Hu J, Balogh K, Mejia A, Christensen ND. Wounding prior to challenge substantially improves infectivity of cottontail rabbit papillomavirus and allows for standardization of infection. J Virol Methods. 2008 Mar;148(1-2):34-9. doi: 10.1016/j.jviromet.2007.10.005.

Conway DI, Stockton DL, Warnakulasuriya KA, Ogden G, Macpherson LM. Incidence of oral and oropharyngeal cancer in United Kingdom (1990-1999) -- recent trends and regional variation. Oral Oncol. 2006 Jul;42(6):586-92. doi: 10.1016/j.oraloncology.2005.10.018.

Coudray MS, Madhivanan P. Bacterial vaginosis-A brief synopsis of the literature. Eur J Obstet Gynecol Reprod Biol. 2020 Feb;245:143-148. doi: 10.1016/j.ejogrb.2019.12.035.

Crook T, Wrede D, Tidy JA, Mason WP, Evans DJ, Vousden KH. Clonal p53 mutation in primary cervical cancer: association with human-papillomavirus-negative tumours. Lancet. 1992 May 2;339(8801):1070-3. doi: 10.1016/0140-6736(92)90662-m.

Cullin N, Azevedo Antunes C, Straussman R, Stein-Thoeringer CK, Elinav E. Microbiome and cancer. Cancer Cell. 2021 Oct 11;39(10):1317-1341. doi: 10.1016/j.ccell.2021.08.006.

Dahlgren L, Mellin H, Wangsa D, Heselmeyer-Haddad K, Björnestål L, Lindholm J, Munck-Wikland E, Auer G, Ried T, Dalianis T. Comparative genomic hybridization analysis of tonsillar cancer reveals a different pattern of genomic imbalances in human papillomavirus-positive and -negative tumors. Int J Cancer. 2003 Nov 1;107(2):244-9. doi: 10.1002/ijc.11371.

Dahlgren L, Dahlstrand HM, Lindquist D, Högmo A, Björnestål L, Lindholm J, Lundberg B, Dalianis T, Munck-Wikland E. Human papillomavirus is more common in base of tongue than in mobile tongue cancer and is a favorable prognostic factor in base of tongue cancer patients. Int J Cancer. 2004 Dec 20;112(6):1015-9. doi: 10.1002/ijc.20490.

Dai C, Charlestin V, Wang M, Walker ZT, Miranda-Vergara MC, Facchine BA, Wu J, Kaliney WJ, Dovichi NJ, Li J, Littlepage LE. Aquaporin-7 Regulates the Response to Cellular Stress in Breast Cancer. Cancer Res. 2020 Oct 1;80(19):4071-4086. doi: 10.1158/0008-5472.CAN-19-2269.

Davis NM, Proctor DM, Holmes SP, Relman DA, Callahan BJ. Simple statistical identification and removal of contaminant sequences in marker-gene and metagenomics data. Microbiome. 2018 Dec 17;6(1):226. doi: 10.1186/s40168-018-0605-2.

Dehlendorff C., Sparén P., Baldur-Felskov B., Herweijer E., Arnheim-Dahlström L., Ploner A., et al. . (2018). Effectiveness of varying number of doses and timing between doses of quadrivalent HPV vaccine against severe cervical lesions. *Vaccine* 36, 6373–6378. 10.1016/j.vaccine.2018.09.011

den Boon JA, Pyeon D, Wang SS, et al. Molecular transitions from papillomavirus infection to cervical precancer and cancer: role of stromal estrogen receptor signaling. Proc Natl Acad Sci. 2015;112:E3255-64. doi: 10.1073/pnas.1509322112.

de Sanjose S, Quint WG, Alemany L, Geraets DT, Klaustermeier JE, Lloveras B, Tous S, Felix A, Bravo LE, Shin HR, Vallejos CS, de Ruiz PA, Lima MA, Guimera N, Clavero O, Alejo M, Llombart-Bosch A, Cheng-Yang C, Tatti SA, Kasamatsu E, Iljazovic E, Odida M, Prado R, Seoud M, Grce M, Usubutun A, Jain A, Suarez GA, Lombardi LE, Banjo A, Menéndez C, Domingo EJ, Velasco J, Nessa A, Chichareon SC, Qiao YL, Lerma E, Garland SM, Sasagawa T, Ferrera A, Hammouda D, Mariani L, Pelayo A, Steiner I, Oliva E, Meijer CJ, Al-Jassar WF, Cruz E, Wright TC, Puras A, Llave CL, Tzardi M, Agorastos T, Garcia-Barriola V, Clavel C, Ordi J, Andújar M, Castellsagué X, Sánchez GI, Nowakowski AM, Bornstein J, Muñoz N, Bosch FX; Retrospective International Survey and HPV Time Trends Study Group. Human papillomavirus genotype attribution in invasive cervical cancer: a retrospective cross-sectional worldwide study. Lancet Oncol. 2010 Nov;11(11):1048-56. doi: 10.1016/S1470-2045(10)70230-8.

De Sordi L, Khanna V, Debarbieux L. The Gut Microbiota Facilitates Drifts in the Genetic Diversity and Infectivity of Bacterial Viruses. Cell Host Microbe. 2017 Dec 13;22(6):801-808.e3. doi: 10.1016/j.chom.2017.10.010.

Dillner J. Early detection and prevention. Mol Oncol. 2019 Mar;13(3):591-598. doi: 10.1002/1878-0261.12459.

Dillner J, Elfström KM, Baussano I. Prospects for accelerated elimination of cervical cancer. Prev Med. 2021 Dec;153:106827. doi: 10.1016/j.ypmed.2021.106827.

DiMaio D. Nuns, warts, viruses, and cancer. Yale J Biol Med. 2015 Jun 1;88(2):127-9. (no doi)

Drolet M, Bénard É, Pérez N, Brisson M; HPV Vaccination Impact Study Group. Population-level impact and herd effects following the introduction of human papillomavirus vaccination programmes: updated systematic review and meta-analysis. Lancet. 2019 Aug 10;394(10197):497-509. doi: 10.1016/S0140-6736(19)30298-3.

D'Souza G, Agrawal Y, Halpern J, Bodison S, Gillison ML. Oral sexual behaviors associated with prevalent oral human papillomavirus infection. J Infect Dis. 2009;199:1263–1269. Doi: 10.1086/597755

Donders G. Diagnosis and management of bacterial vaginosis and other types of abnormal vaginal bacterial flora: a review. Obstet Gynecol Surv. 2010 Jul;65(7):462-73. doi: 10.1097/OGX.0b013e3181e09621.

Donne AJ, Keltie K, Cole H, Sims AJ, Patrick H, Powell S. Prevalence and management of recurrent respiratory papillomatosis (RRP) in the UK: cross-sectional study. Clin Otolaryngol. 2017 Feb;42(1):86-91. doi: 10.1111/coa.12683.

Doorbar J. The papillomavirus life cycle. J Clin Virol. 2005 Mar;32 Suppl 1:S7-15. doi: 10.1016/j.jcv.2004.12.006.

Doorbar J, Quint W, Banks L, Bravo IG, Stoler M, Broker TR, Stanley MA. The biology and life-cycle of human papillomaviruses. Vaccine. 2012 Nov 20;30 Suppl 5:F55-70. doi: 10.1016/j.vaccine.2012.06.083.

Doorbar J. The E4 protein; structure, function and patterns of expression. Virology. 2013 Oct;445(1-2):80-98. doi: 10.1016/j.virol.2013.07.008.

Du J, Näsman A, Carlson JW, Ramqvist T, Dalianis T. Prevalence of human papillomavirus (HPV) types in cervical cancer 2003-2008 in Stockholm, Sweden, before public HPV vaccination. Acta Oncol. 2011 Nov;50(8):1215-9. doi: 10.3109/0284186X.2011.584556.

Du J, Ährlund-Richter A, Näsman A, Dalianis T. Human papilloma virus (HPV) prevalence upon HPV vaccination in Swedish youth: a review based on our findings 2008-2018, and perspectives on cancer prevention. Arch Gynecol Obstet. 2021 Feb;303(2):329-335. doi: 10.1007/s00404-020-05879-7.

Dürst M, Gissmann L, Ikenberg H, zur Hausen H. A papillomavirus DNA from a cervical carcinoma and its prevalence in cancer biopsy samples from different geographic regions. Proc Natl Acad Sci U S A. 1983 Jun;80(12):3812-5. doi: 10.1073/pnas.80.12.3812.

Eifel PJ. Chemoradiotherapy in the treatment of cervical cancer. Semin Radiat Oncol. 2006 Jul;16(3):177-85. doi: 10.1016/j.semradonc.2006.02.007.

Elson DA, Riley RR, Lacey A, Thordarson G, Talamantes FJ, Arbeit JM. Sensitivity of the cervical transformation zone to estrogen-induced squamous carcinogenesis. Cancer Res. 2000 Mar 1;60(5):1267-75. (no doi) PMID: 10728686.

Epstein MA Achong BG, Barr YM. Virus particles in cultured lymphoblasts from Burkitts lymphoma. Lancet. 1964 Mar 28;1(7335):702-3. doi: 10.1016/s0140-6736(64)91524-7.

Faraji F, Zaidi M, Fakhry C, Gaykalova DA. Molecular mechanisms of human papillomavirus-related carcinogenesis in head and neck cancer. Microbes Infect. 2017 Sep-Oct;19(9-10):464-475. doi: 10.1016/j.micinf.2017.06.001.

Fettweis JM, Serrano MG, Brooks JP, Edwards DJ, Girerd PH, Parikh HI, Huang B, Arodz TJ, Edupuganti L, Glascock AL, Xu J, Jimenez NR, Vivadelli SC, Fong SS, Sheth NU, Jean S, Lee V, Bokhari YA, Lara AM, Mistry SD, Duckworth RA 3rd, Bradley SP, Koparde VN, Orenda XV, Milton SH, Rozycki SK, Matveyev AV, Wright ML, Huzurbazar SV, Jackson EM, Smirnova E, Korlach J, Tsai YC, Dickinson MR, Brooks JL, Drake JI, Chaffin DO, Sexton AL, Gravett MG, Rubens CE, Wijesooriya NR, Hendricks-Muñoz KD, Jefferson KK, Strauss JF 3rd, Buck GA. The vaginal microbiome and preterm birth. Nat Med. 2019 Jun;25(6):1012-1021. doi: 10.1038/s41591-019-0450-2.

Forman D, de Martel C, Lacey CJ, Soerjomataram I, Lortet-Tieulent J, Bruni L, Vignat J, Ferlay J, Bray F, Plummer M, Franceschi S. Global burden of human papillomavirus and related diseases. Vaccine. 2012 Nov 20;30 Suppl 5:F12-23. doi: 10.1016/j.vaccine.2012.07.055.

Fortes HR, von Ranke FM, Escuissato DL, Araujo Neto CA, Zanetti G, Hochhegger B, Souza CA, Marchiori E. Recurrent respiratory papillomatosis: A state-of-the-art review. Respir Med. 2017 May;126:116-121. doi: 10.1016/j.rmed.2017.03.030.

Framson PE, Sage EH. SPARC and tumor growth: where the seed meets the soil? J Cell Biochem. 2004 Jul 1;92(4):679-90. doi: 10.1002/jcb.20091.

Fredricks DN, Fiedler TL, Marrazzo JM. Molecular identification of bacteria associated with bacterial vaginosis. New Engl. J. Med. 2005;353:1899–1911. doi: 10.1056/NEJMoa043802

Gao G, Gay HA, Chernock RD, Zhang TR, Luo J, Thorstad WL, Lewis JS Jr, Wang X. A microRNA expression signature for the prognosis of oropharyngeal squamous cell carcinoma. Cancer. 2013 Jan 1;119(1):72-80. doi: 10.1002/cncr.27696.

Gaykalova, D.A.; Mambo, E.; Choudhary, A.; Houghton, J.; Buddavarapu, K.; Sanford, T.; Darden, W.; Adai, A.; Hadd, A.; Latham, G.; et al. Novel insight into mutational landscape of head and neck squamous cell carcinoma. *PLoS One* 2014, *9*, e93102, doi:10.1371/journal.pone.0093102.

GlaxoSmithKline. Package Insert - Cervarix. 2009. https://www.fda.gov/downloads/biologicsbloodvaccines/vaccines/approvedproducts/ucm186981.pdf. Accessed 25 Sep 2018.

Graham SV. The human papillomavirus replication cycle, and its links to cancer progression: a comprehensive review. Clin Sci (Lond). 2017 Aug 10;131(17):2201-2221. doi: 10.1042/CS20160786.

Gillison ML, Koch WM, Capone RB, Spafford M, Westra WH, Wu L, Zahurak ML, Daniel RW, Viglione M, Symer DE, Shah KV, Sidransky D. Evidence for a causal association between human papillomavirus and a subset of head and neck cancers. J Natl Cancer Inst. 2000 May 3;92(9):709-20. doi: 10.1093/jnci/92.9.709.

Grün N, Ährlund-Richter A, Franzén J, Mirzaie L, Marions L, Ramqvist T, Dalianis T. Follow-up on oral and cervical human papillomavirus prevalence 2013-2015 in youth at a youth clinic in Stockholm, Sweden. Infect Dis (Lond). 2016 Feb;48(2):169-70. doi: 10.3109/23744235.2015.1094573.

Guerrero-Preston R, Godoy-Vitorino F, Jedlicka A, Rodríguez-Hilario A, González H, Bondy J, Lawson F, Folawiyo O, Michailidi C, Dziedzic A, Thangavel R, Hadar T, Noordhuis MG, Westra W, Koch W, Sidransky D. 16S rRNA amplicon sequencing identifies microbiota associated with oral cancer, human papilloma virus infection and surgical treatment. Oncotarget. 2016 Aug 9;7(32):51320-51334. doi: 10.18632/oncotarget.9710.

Gupta S, Mortensen MS, Schjørring S, Trivedi U, Vestergaard G, Stokholm J, Bisgaard H, Krogfelt KA, Sørensen SJ. Amplicon sequencing provides more accurate microbiome information in healthy children compared to culturing. Commun Biol. 2019 Aug 5;2:291. doi: 10.1038/s42003-019-0540-1.

Haeggblom L, Ährlund-Richter A, Mirzaie L, Farrajota Neves da Silva P, Ursu RG, Ramqvist T, Näsman A. Differences in gene expression between high-grade dysplasia and invasive HPV^{+} and HPV^{-} tonsillar and base of tongue cancer. Cancer Med. 2019 Oct;8(14):6221-6232. doi: 10.1002/cam4.2450.

Halec G, Schmitt M, Dondog B, Sharkhuu E, Wentzensen N, Gheit T, Tommasino M, Kommoss F, Bosch FX, Franceschi S, Clifford G, Gissmann L, Pawlita M. Biological activity of probable/possible high-risk human papillomavirus types in cervical cancer. Int J Cancer. 2013 Jan 1;132(1):63-71. doi: 10.1002/ijc.27605.

Harbison RA, Kubik M, Konnick EQ, Zhang Q, Lee SG, Park H, Zhang J, Carlson CS, Chen C, Schwartz SM, Rodriguez CP, Duvvuri U, Méndez E. The mutational landscape of recurrent versus nonrecurrent human papillomavirus-related oropharyngeal cancer. JCI Insight. 2018 Jul 26;3(14):e99327. doi: 10.1172/jci.insight.99327.

Harper DM, DeMars LR. HPV vaccines - A review of the first decade. Gynecol Oncol. 2017 Jul;146(1):196-204. doi: 10.1016/j.ygyno.2017.04.004. Epub 2017 Apr 22. Erratum in: Gynecol Oncol. 2017 Nov;147(2):489.

Hattinger E, Zwicker S, Ruzicka T, Yuspa SH, Wolf R. Opposing functions of psoriasin (S100A7) and koebnerisin (S100A15) in epithelial carcinogenesis. Curr Opin Pharmacol. 2013 Aug;13(4):588-94. doi: 10.1016/j.coph.2013.04.007.

Hayes RB, Ahn J, Fan X, Peters BA, Ma Y, Yang L, Agalliu I, Burk RD, Ganly I, Purdue MP, Freedman ND, Gapstur SM, Pei Z. Association of Oral Microbiome With Risk for Incident Head and Neck Squamous Cell Cancer. JAMA Oncol. 2018 Mar 1;4(3):358-365. doi: 10.1001/jamaoncol.2017.4777.

Herrington CS. Do HPV-negative cervical carcinomas exist?--revisited. J Pathol. 1999 Sep;189(1):1-3. doi: 10.1002/(SICI)1096-9896(199909)189:1<1::AID-PATH432>3.0.CO;2-P.

Hibbert J, Halec G, Baaken D, Waterboer T, Brenner N. Sensitivity and Specificity of Human Papillomavirus (HPV) 16 Early Antigen Serology for HPV-Driven Oropharyngeal Cancer: A Systematic Literature Review and Meta-Analysis. Cancers (Basel). 2021 Jun 16;13(12):3010. doi: 10.3390/cancers13123010.

Hollingsworth M.A., Swanson B.J. Mucins in cancer: Protection and control of the cell surface. Nat. Rev. Cancer. 2004;4:45–60. doi: 10.1038/nrc1251.

Holmes BJ, Wenig BM. Virus-associated carcinomas of the head & neck: Update from the 2017 WHO classification. *Ann Diagn Pathol*. 2019;38:29-42. 10.1016/j.anndiagpath.2018.10.008.

Hortlund M, Mühr LSA, Lagheden C, Hjerpe A, Dillner J. Audit of laboratory sensitivity of human papillomavirus and cytology testing in a cervical screening program. Int J Cancer. 2021 Dec 15;149(12):2083-2090. doi: 10.1002/ijc.33769.

Hugerth, LW, Seifert M, Pennhag AAK, Du J, Hamsten MC, Schuppe-Koistinen I, Engstrand LA comprehensive automated pipeline for human microbiome sampling, 16S rRNA gene sequencing and bioinformatics processing. BioRxiv 2018. https://doi.org/10.1101/286526

Hui AB, Lin A, Xu W, Waldron L, Perez-Ordonez B, Weinreb I, Shi W, Bruce J, Huang SH, O'Sullivan B, Waldron J, Gullane P, Irish JC, Chan K, Liu FF. Potentially prognostic miRNAs in HPV-associated oropharyngeal carcinoma. Clin Cancer Res. 2013 Apr 15;19(8):2154-62. doi: 10.1158/1078-0432.CCR-12-3572.

(https://www.who.int/teams/immunization-vaccines-and-biologicals/diseases/human-papillomavirus-vaccines-(HPV).

International Agency for Research on Cancer. IARC Monographs on the Evaluation of Carcinogenic Risks to Humans. Volume 64, Human papillomaviruses. Lyon, France; 1995.

International Agency for Research on Cancer. IARC Monographs on the Evaluation of Carcinogenic Risks to Humans. Volume 90, Human papillomaviruses. Lyon, France; 2007.

Isaacsson Velho PH, Castro G Jr, Chung CH. Targeting the PI3K Pathway in Head and Neck Squamous Cell Carcinoma. Am Soc Clin Oncol Educ Book. 2015:123-8. doi: 10.14694/EdBook_AM.2015.35.123.

Katsenos S, Becker HD. Recurrent respiratory papillomatosis: a rare chronic disease, difficult to treat, with potential to lung cancer transformation: apropos of two cases and a brief literature review. Case Rep Oncol. 2011 Mar 23;4(1):162-71. doi: 10.1159/000327094.

Kazemi-Sefat GE, Keramatipour M, Talebi S, Kavousi K, Sajed R, Kazemi-Sefat NA, Mousavizadeh K. The importance of CDC27 in cancer: Molecular pathology and clinical aspects. Cancer Cell Int. 2021;21:160. doi: 10.1186/s12935-021-01860-9.

Kirkland SC. Type I collagen inhibits differentiation and promotes a stem cell-like phenotype in human colorectal carcinoma cells. Br J Cancer. 2009 Jul 21;101(2):320-6. doi: 10.1038/sj.bjc.6605143.

Kirnbauer R, Booy F, Cheng N, Lowy DR, Schiller JT. Papillomavirus L1 major capsid protein self-assembles into virus-like particles that are highly immunogenic. Proc Natl Acad Sci U S A. 1992 Dec 15;89(24):12180-4. doi: 10.1073/pnas.89.24.12180.

Koedooder R, Singer M, Schoenmakers S, Savelkoul PHM, Morré SA, de Jonge JD, Poort L, Cuypers WJSS, Beckers NGM, Broekmans FJM, Cohlen BJ, den Hartog JE, Fleischer K, Lambalk CB, Smeenk JMJS, Budding AE, Laven JSE. The vaginal microbiome as a predictor for outcome of in vitro fertilization with or without intracytoplasmic sperm injection: a prospective study. Hum Reprod. 2019 Jun 4;34(6):1042-1054. doi: 10.1093/humrep/dez065. Erratum in: Hum Reprod. 2019 Oct 2;34(10):2091-2092.

Kyrgiou M, Arbyn M, Bergeron C, Bosch FX, Dillner J, Jit M, Kim J, Poljak M, Nieminen P, Sasieni P, Kesic V, Cuzick J, Gultekin M. Cervical screening: ESGO-EFC position paper of the European Society of Gynaecologic Oncology (ESGO) and the European Federation of Colposcopy (EFC). Br J Cancer. 2020 Aug;123(4):510-517. doi: 10.1038/s41416-020-0920-9.

Lajer CB, Nielsen FC, Friis-Hansen L, Norrild B, Borup R, Garnæs E, Rossing M, Specht L, Therkildsen MH, Nauntofte B, Dabelsteen S, von Buchwald C. Different miRNA signatures of oral and pharyngeal squamous cell carcinomas: a prospective translational study. Br J Cancer. 2011 Mar 1;104(5):830-40. doi: 10.1038/bjc.2011.29.

Lajer CB, Garnæs E, Friis-Hansen L, Norrild B, Therkildsen MH, Glud M, Rossing M, Lajer H, Svane D, Skotte L, Specht L, Buchwald C, Nielsen FC. The role of miRNAs in human papilloma virus (HPV)-associated cancers: bridging between HPV-related head and neck cancer and cervical cancer. Br J Cancer. 2012 Apr 24;106(9):1526-34. doi: 10.1038/bjc.2012.109. Erratum in: Br J Cancer. 2017 Jul 13;:

Leenhardt F, Alexandre M, Jacot W. Alpelisib for the treatment of PIK3CA-mutated, hormone receptor-positive, HER2-negative metastatic breast cancer. Expert Opin Pharmacother 2021, 1-9, doi:10.1080/14656566.2021.1873952.

Lehtinen M, Söderlund-Strand A, Vänskä S, Luostarinen T, Eriksson T, Natunen K, Apter D, Baussano I, Harjula K, Hokkanen M, Kuortti M, Palmroth J, Petäjä T, Pukkala E, Rekonen S, Siitari-Mattila M, Surcel HM, Tuomivaara L, Paavonen J, Dillner J, Dubin G, Garnett G. Impact of gender-neutral or girls-only vaccination against human papillomavirus-Results of a community-randomized clinical trial (I). Int J Cancer. 2018 Mar 1;142(5):949-958. doi: 10.1002/ijc.31119.

Leslie SW, Sajjad H, Kumar S. Genital Warts. 2021 Sep 17. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2022 Jan–. PMID: 28722914.

Lessa FC, Mu Y, Bamberg WM, Beldavs ZG, Dumyati GK, Dunn JR, Farley MM, Holzbauer SM, Meek JI, Phipps EC, Wilson LE, Winston LG, Cohen JA, Limbago BM, Fridkin SK, Gerding DN, McDonald LC. Burden of Clostridium difficile infection in the United States. N Engl J Med. 2015 Feb 26;372(9):825-34. doi: 10.1056/NEJMoa1408913.

Lewis AM, Varghese S, Xu H, Alexander HR. Interleukin-1 and cancer progression: the emerging role of interleukin-1 receptor antagonist as a novel therapeutic agent in cancer treatment. J Transl Med. 2006 Nov 10;4:48. doi: 10.1186/1479-5876-4-48.

Li, H.; Ou, X.; Xiong, J.; Wang, T. HPV16E7 mediates HADC chromatin repression and downregulation of MHC class I genes in HPV16 tumorigenic cells through interaction with an MHC class I promoter. Biochem Biophys Res Commun 2006, 349, 1315-1321, doi:10.1016/j.bbrc.2006.08.182.

Licitra L, Bernier J, Grandi C, Merlano M, Bruzzi P, Lefebvre JL. Cancer of the oropharynx. Crit Rev Oncol Hematol. 2002 Jan;41(1):107-22. doi: 10.1016/s1040-8428(01)00129-9.

Lui VW, Hedberg ML, Li H, Vangara BS, Pendleton K, Zeng Y, Lu Y, Zhang Q, Du Y, Gilbert BR, Freilino M, Sauerwein S, Peyser ND, Xiao D, Diergaarde B, Wang L, Chiosea S, Seethala R, Johnson JT, Kim S, Duvvuri U, Ferris RL, Romkes M, Nukui T, Kwok-Shing Ng P, Garraway LA, Hammerman PS, Mills GB, Grandis JR. Frequent mutation of the PI3K pathway in head and neck cancer defines predictive biomarkers. Cancer Discov. 2013 Jul;3(7):761-9. doi: 10.1158/2159-8290.CD-13-0103.

Lowy DR, Schiller JT. Reducing HPV-associated cancer globally. Cancer Prev Res (Phila). 2012 Jan;5(1):18-23. doi: 10.1158/1940-6207.CAPR-11-0542.

Ma X, Harripersaud K, Smith K, Fairley CK, Zou H, Zou Z, Wang Y, Zhuang G, Zhang L. Modeling the epidemiological impact and cost-effectiveness of a combined schoolgirl HPV vaccination and cervical cancer screening program among Chinese women. Hum Vaccin Immunother. 2021 Apr 3;17(4):1073-1082. doi: 10.1080/21645515.2020.1832835.

Ma Z, Wang H, Meng F, Han Y, Chen Y, Xiao M, Jiang H, Yu Z, Xu B. Role of BCLAF-1 in PD-L1 stabilization in response to ionizing irradiation. Cancer Sci. 2021 Oct;112(10):4064-4074. doi: 10.1111/cas.15056.

Machalek DA, Garland SM, Brotherton JML, Bateson D, McNamee K, Stewart M, Rachel Skinner S, Liu B, Cornall AM, Kaldor JM, Tabrizi SN. Very Low Prevalence of Vaccine Human Papillomavirus Types Among 18- to 35-Year Old Australian Women 9 Years Following Implementation of Vaccination. J Infect Dis. 2018 Apr 23;217(10):1590-1600. doi: 10.1093/infdis/jiy075.

Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal.* 2011;17:10. doi: 10.14806/ej.17.1.200.

Martinez I, Wang J, Hobson KF, Ferris RL, Khan SA. Identification of differentially expressed genes in HPV-positive and HPV-negative oropharyngeal squamous cell carcinomas. Eur J Cancer. 2007 Jan;43(2):415-32. doi: 10.1016/j.ejca.2006.09.001.

Masterson L, Sorgeloos F, Winder D, Lechner M, Marker A, Malhotra S, Sudhoff H, Jani P, Goon P, Sterling J. Deregulation of SYCP2 predicts early stage human papillomavirus-positive oropharyngeal carcinoma: A prospective whole transcriptome analysis. Cancer Sci. 2015 Nov;106(11):1568-75. doi: 10.1111/cas.12809.

Masoudi-Nejad, A., Narimani, Z., Hosseinkhan, N. (2013). De Novo Assembly Algorithms. In: Next Generation Sequencing and Sequence Assembly. SpringerBriefs in Systems Biology, vol 4. Springer, New York, NY. https://doi.org/10.1007/978-1-4614-7726-6_4

McGraw SL, Ferrante JM. Update on prevention and screening of cervical cancer. World J Clin Oncol. 2014 Oct 10;5(4):744-52. doi: 10.5306/wjco.v5.i4.744.

McKaig RG, Baric RS, Olshan AF. Human papillomavirus and head and neck cancer: epidemiology and molecular biology. Head Neck. 1998 May;20(3):250-65. doi: 10.1002/(sici)1097-0347(199805)20:3<250::aid-hed11>3.0.co;2-o.

Melloy PG. The anaphase-promoting complex: A key mitotic regulator associated with somatic mutations occurring in cancer. *Genes Chromosomes Cancer.* 2020;59:189–202. doi: 10.1002/gcc.22820.

Mellin H, Friesland S, Lewensohn R, Dalianis T, Munck-Wikland E. Human papillomavirus (HPV) DNA in tonsillar cancer: clinical correlates, risk of relapse, and survival. Int J Cancer. 2000 May 20;89(3):300-4. (no doi.) PMID: 10861508.

Melnick JL, Allison AC, Butel JS, Eckhart W, Eddy BE, Kit S, Levine AJ, Miles JA, Pagano JS, Sachs L, Vonka V. Papovaviridae. Intervirology. 1974;3(1-2):106-20. doi: 10.1159/000149746.

Merck&CO. I. Package Insert - Gardasil. 2006. https://www.fda.gov/downloads/biologicsbloodvaccines/vaccines/approvedproducts/ucm111263.pdf. Accessed 25 Sep 2018.

Merck&CO. I. Package Insert - Gardasil 9. 2014. https://www.fda.gov/downloads/biologicsbloodvaccines/vaccines/approvedproducts/ucm426457.pdf. Accessed 25 Sep 2018.

Miller DL, Davis JW, Taylor KH, Johnson J, Shi Z, Williams R, Atasoy U, Lewis JS Jr, Stack MS. Identification of a human papillomavirus-associated oncogenic miRNA panel in human oropharyngeal squamous cell carcinoma validated by bioinformatics analysis of the Cancer Genome Atlas. Am J Pathol. 2015 Mar;185(3):679-92. doi: 10.1016/j.ajpath.2014.11.018.

Mirghani H, Ugolin N, Ory C, Lefèvre M, Baulande S, Hofman P, St Guily JL, Chevillard S, Lacave R. A predictive transcriptomic signature of oropharyngeal cancer according to HPV16 status exclusively. Oral Oncol. 2014 Nov;50(11):1025-34. doi: 10.1016/j.oraloncology.2014.07.019.

Mitra A, MacIntyre DA, Marchesi JR, Lee YS, Bennett PR, Kyrgiou M. The vaginal microbiota, human papillomavirus infection and cervical intraepithelial neoplasia: what do we know and where are we going next? Microbiome. 2016 Nov 1;4(1):58. doi: 10.1186/s40168-016-0203-0.

Mooren JJ, Gültekin SE, Straetmans JM, Haesevoets A, Peutz-Kootstra CJ, Huebbers CU, Dienes HP, Wieland U, Ramaekers FC, Kremer B, Speel EJ, Klussmann JP. P16(INK4A) immunostaining is a strong indicator for high-risk-HPV-associated oropharyngeal carcinomas and dysplasias, but is unreliable to predict low-risk-HPV-infection in head and neck papillomas and laryngeal dysplasias. Int J Cancer. 2014 May 1;134(9):2108-17. doi: 10.1002/ijc.28534.

Morgenstein AA, Citron DM, Orisek B, Finegold SM. Serious infection with Leptotrichia buccalis. Report of a case and review of the literature. Am J Med. 1980 Nov;69(5):782-5. doi: 10.1016/0002-9343(80)90452-0.

Näsman A, Attner P, Hammarstedt L, Du J, Eriksson M, Giraud G, Ahrlund-Richter S, Marklund L, Romanitan M, Lindquist D, Ramqvist T, Lindholm J, Sparén P, Ye W, Dahlstrand H, Munck-Wikland E, Dalianis T. Incidence of human papillomavirus (HPV) positive tonsillar carcinoma in Stockholm, Sweden: an epidemic of viral-induced carcinoma? Int J Cancer. 2009 Jul 15;125(2):362-6. doi: 10.1002/ijc.24339.

Nasman, A.; Romanitan, M.; Nordfors, C.; Grun, N.; Johansson, H.; Hammarstedt, L.; Marklund, L.; Munck-Wikland, E.; Dalianis, T.; Ramqvist, T. Tumor infiltrating CD8+ and Foxp3+ lymphocytes correlate to clinical outcome and human papillomavirus (HPV) status in tonsillar cancer. *PLoS One* 2012, *7*, e38711, doi:10.1371/journal.pone.0038711.

Näsman A, Bersani C, Lindquist D, Du J, Ramqvist T, Dalianis T. Human Papillomavirus and Potentially Relevant Biomarkers in Tonsillar and Base of Tongue Squamous Cell Carcinoma. Anticancer Res. 2017 Oct;37(10):5319-5328. doi: 10.21873/anticanres.11958.

Näsman A, Holzhauser S, Kostopoulou ON, Zupancic M, Ährlund-Richter A, Du J, Dalianis T. Prognostic Markers and Driver Genes and Options for Targeted Therapy in Human-Papillomavirus-Positive Tonsillar and Base-of-Tongue Squamous Cell Carcinoma. Viruses. 2021 May 14;13(5):910. doi: 10.3390/v13050910. PMID: 34069114;

Ng SB, Turner EH, Robertson PD, Flygare SD, Bigham AW, Lee C, Shaffer T, Wong M, Bhattacharjee A, Eichler EE, Bamshad M, Nickerson DA, Shendure J. Targeted capture and massively parallel sequencing of 12 human exomes. Nature. 2009 Sep 10;461(7261):272-6. doi: 10.1038/nature08250.

Ng SB, Buckingham KJ, Lee C, Bigham AW, Tabor HK, Dent KM, Huff CD, Shannon PT, Jabs EW, Nickerson DA, Shendure J, Bamshad MJ. Exome sequencing identifies the cause of a mendelian disorder. Nat Genet. 2010 Jan;42(1):30-5. doi: 10.1038/ng.499.

Nordfors C, Grün N, Tertipis N, Ährlund-Richter A, Haeggblom L, Sivars L, Du J, Nyberg T, Marklund L, Munck-Wikland E, Näsman A, Ramqvist T, Dalianis T. CD8+ and CD4+ tumour infiltrating lymphocytes in relation to human papillomavirus status and clinical outcome in tonsillar and base of tongue squamous cell carcinoma. Eur J Cancer. 2013 Jul;49(11):2522-30. doi: 10.1016/j.ejca.2013.03.019.

Nordfors C, Grün N, Haeggblom L, Tertipis N, Sivars L, Mattebo M, Larsson M, Häggström-Nordin E, Tydén T, Ramqvist T, Dalianis T. Oral human papillomavirus prevalence in high school students of one municipality in Sweden. Scand J Infect Dis. 2013 Nov;45(11):878-81. doi: 10.3109/00365548.2013.821626.

Nordfors C, Vlastos A, Du J, Ahrlund-Richter A, Tertipis N, Grün N, Romanitan M, Haeggblom L, Roosaar A, Dahllöf G, Donà MG, Benevolo M, Ramqvist T, Munck-Wikland E, Dalianis T. Human papillomavirus prevalence is high in oral samples of patients with tonsillar and base of tongue cancer. Oral Oncol. 2014 May;50(5):491-7. doi: 10.1016/j.oraloncology.2014.02.012.

Norenhag J, Du J, Olovsson M, Verstraelen H, Engstrand L, Brusselaers N. The vaginal microbiota, human papillomavirus and cervical dysplasia: a systematic review and network meta-analysis. BJOG. 2020 Jan;127(2):171-180. doi: 10.1111/1471-0528.15854.

Oakley BB, Fiedler TL, Marrazzo JM, Fredricks DN. Diversity of human vaginal bacterial communities and associations with clinically defined bacterial vaginosis. *Appl. Environ. Microbiol.* 2008;74:4898–4909. doi: 10.1128/AEM.02884-07.

Oguejiofor K, Galletta-Williams H, Dovedi SJ, Roberts DL, Stern PL, West CM. Distinct patterns of infiltrating CD8+ T cells in HPV+ and CD68 macrophages in HPV-oropharyngeal squamous cell carcinomas are associated with better clinical outcome but PD-L1 expression is not prognostic. Oncotarget. 2017 Feb 28;8(9):14416-14427. doi: 10.18632/oncotarget.14796.

Oguejiofor K, Hall J, Slater C, Betts G, Hall G, Slevin N, Dovedi S, Stern PL, West CM. Stromal infiltration of CD8 T cells is associated with improved clinical outcome in HPV-positive oropharyngeal squamous carcinoma. Br J Cancer. 2015 Sep 15;113(6):886-93. doi: 10.1038/bjc.2015.277.

Onderdonk AB, Delaney ML, Fichorova RN. The human microbiome during bacterial vaginosis. *Clin. Microbiol. Rev.* 2016;29:223–238. doi: 10.1128/CMR.00075-15.

Papanicolaou GN, Traut HF. The diagnostic value of vaginal smears in carcinoma of the uterus. 1941. Arch Pathol Lab Med. 1997 Mar;121(3):211-24.

Parkin DM, Bray F. Chapter 2: The burden of HPV-related cancers. Vaccine. 2006 Aug 31;24 Suppl 3:S3/11-25. doi: 10.1016/j.vaccine.2006.05.111.

Partanen VM, Dillner J, Tropé A, Ágústsson ÁI, Pankakoski M, Heinävaara S, Sarkeala T, Wang J, Skare GB, Anttila A, Lönnberg S. Comparison of cytology and human papillomavirus-based primary testing in cervical screening programs in the Nordic countries. J Med Screen. 2021 Dec;28(4):464-471. doi: 10.1177/0969141321992404.

Paulsson J, Micke P. Prognostic relevance of cancer-associated fibroblasts in human cancer. Semin Cancer Biol. 2014 Apr;25:61-8. doi: 10.1016/j.semcancer.2014.02.006.

Petrova MI, van den Broek M, Balzarini J, Vanderleyden J, Lebeer S. Vaginal microbiota and its role in HIV transmission and infection. FEMS Microbiol Rev. 2013 Sep;37(5):762-92. doi: 10.1111/1574-6976.12029.

Phillips A, Patel C, Pillsbury A, Brotherton J, Macartney K. Safety of Human Papillomavirus Vaccines: An Updated Review. Drug Saf. 2018 Apr;41(4):329-346. doi: 10.1007/s40264-017-0625-z.

Ramqvist T, Du J, Lundén M, Ahrlund-Richter S, Ferreira J, Marions L, Sparén P, Andersson S, Dalianis T. Pre-vaccination prevalence of human papillomavirus types in the genital tract of 15-23-year-old women attending a youth health clinic in Stockholm, Sweden. Scand J Infect Dis. 2011 Feb;43(2):115-21. doi: 10.3109/00365548.2010.526957.

Ramqvist T, Grün N, Dalianis T. Human papillomavirus and tonsillar and base of tongue cancer. Viruses. 2015 Mar 20;7(3):1332-43. doi: 10.3390/v7031332.

Ramqvist T, Mints M, Tertipis N, Näsman A, Romanitan M, Dalianis T. Studies on human papillomavirus (HPV) 16 E2, E5 and E7 mRNA in HPV-positive tonsillar and base of tongue cancer in relation to clinical outcome and immunological parameters. Oral Oncol. 2015 Dec;51(12):1126-31. doi: 10.1016/j.oraloncology.2015.09.007.

Ramqvist T, Näsman A, Franzén B, Bersani C, Alexeyenko A, Becker S, Haeggblom L, Kolev A, Dalianis T, Munck-Wikland E. Protein Expression in Tonsillar and Base of Tongue Cancer and in Relation to Human Papillomavirus (HPV) and Clinical Outcome. Int J Mol Sci. 2018 Mar 25;19(4):978. doi: 10.3390/ijms19040978.

Ravel J, Gajer P, Abdo Z, Schneider GM, Koenig SS, McCulle SL, Karlebach S, Gorle R, Russell J, Tacket CO, Brotman RM, Davis CC, Ault K, Peralta L, Forney LJ. Vaginal microbiome of reproductive-age women. Proc Natl Acad Sci U S A. 2011 Mar 15;108 Suppl 1(Suppl 1):4680-7. doi: 10.1073/pnas.1002611107.

Rous P. A sarcoma of the fowl transmissible by an agent separable from the tumor cells. J Exp Med. 1911 Apr 1;13(4):397-411. doi: 10.1084/jem.13.4.397.

Routman DM, Kumar S, Chera BS, Jethwa KR, Van Abel KM, Frechette K, DeWees T, Golafshar M, Garcia JJ, Price DL, Kasperbauer JL, Patel SH, Neben-Wittich MA, Laack NL, Chintakuntlawar AV, Price KA, Liu MC, Foote RL, Moore EJ, Gupta GP, Ma DJ. Detectable Post-operative Circulating Tumor Human Papillomavirus (HPV) DNA And Association with Recurrence in Patients with HPV-Associated Oropharyngeal Squamous Cell Carcinoma. Int J Radiat Oncol Biol Phys. 2022 Feb 11:S0360-3016(22)00158-4. doi: 10.1016/j.ijrobp.2022.02.012.

Rusan M, Li YY, Hammerman PS. Genomic landscape of human papillomavirus-associated cancers. *Clin Cancer Res* 2015, *21*, 2009-2019, doi:10.1158/1078-0432.CCR-14-1101.

Saccucci M, Franco EL, Ding L, Bernstein DI, Brown D, Kahn JA. Non-Vaccine-Type Human Papillomavirus Prevalence After Vaccine Introduction: No Evidence for Type Replacement but Evidence for Cross-Protection. Sex Transm Dis. 2018 Apr;45(4):260-265. doi: 10.1097/OLQ.0000000000000731.

Salunke DM, Caspar DL, Garcea RL. Polymorphism in the assembly of polyomavirus capsid protein VP1. Biophys J. 1989 Nov;56(5):887-900. doi: 10.1016/S0006-3495(89)82735-3.

Santegoets SJ, Duurland CL, Jordanova EJ, van Ham VJ, Ehsan I, Loof NM, Narang V, Dutertre CA, Ginhoux F, van Egmond SL, J P Welters M, van der Burg SH. $CD163^{+}$cytokine-producing cDC2 stimulate intratumoral type 1 T cell responses in HPV16-induced oropharyngeal cancer. J Immunother Cancer. 2020 Aug;8(2):e001053. doi: 10.1136/jitc-2020-001053.

Scher JU, Sczesnak A, Longman RS, Segata N, Ubeda C, Bielski C, Rostron T, Cerundolo V, Pamer EG, Abramson SB, Huttenhower C, Littman DR. Expansion of intestinal Prevotella copri correlates with enhanced susceptibility to arthritis. Elife. 2013 Nov 5;2:e01202. doi: 10.7554/eLife.01202.

Schiller JT, Lowy DR. Papillomavirus-like particles and HPV vaccine development. Semin Cancer Biol. 1996 Dec;7(6):373-82. doi: 10.1006/scbi.1996.0046.

Schiller J, Lowy D. Explanations for the high potency of HPV prophylactic vaccines. Vaccine. 2018 Aug 6;36(32 Pt A):4768-4773. doi: 10.1016/j.vaccine.2017.12.079.

Schmitt M, Bravo IG, Snijders PJ, Gissmann L, Pawlita M, Waterboer T. Bead-based multiplex genotyping of human papillomaviruses. J Clin Microbiol. 2006 Feb;44(2):504-12. doi: 10.1128/JCM.44.2.504-512.2006.

Seiwert TY, Zuo Z, Keck MK, Khattri A, Pedamallu CS, Stricker T, Brown C, Pugh TJ, Stojanov P, Cho J, Lawrence MS, Getz G, Brägelmann J, DeBoer R, Weichselbaum RR, Langerman A, Portugal L, Blair E, Stenson K, Lingen MW, Cohen EE, Vokes EE, White KP, Hammerman PS. Integrative and comparative genomic analysis of HPV-positive and HPV-negative head and neck squamous cell carcinomas. Clin Cancer Res. 2015 Feb 1;21(3):632-41. doi: 10.1158/1078-0432.CCR-13-3310.

Serrano MG, Parikh HI, Brooks JP, Edwards DJ, Arodz TJ, Edupuganti L, Huang B, Girerd PH, Bokhari YA, Bradley SP, Brooks JL, Dickinson MR, Drake JI, Duckworth RA 3rd, Fong SS, Glascock AL, Jean S, Jimenez NR, Khoury J, Koparde VN, Lara AM, Lee V, Matveyev AV, Milton SH, Mistry SD, Rozycki SK, Sheth NU, Smirnova E, Vivadelli SC, Wijesooriya NR, Xu J, Xu P, Chaffin DO, Sexton AL, Gravett MG, Rubens CE, Hendricks-Muñoz KD, Jefferson KK, Strauss JF 3rd, Fettweis JM, Buck GA. Racioethnic diversity in the dynamics of the vaginal microbiome during pregnancy. Nat Med. 2019 Jun;25(6):1001-1011. doi: 10.1038/s41591-019-0465-8.

Sewell A, Brown B, Biktasova A, Mills GB, Lu Y, Tyson DR, Issaeva N, Yarbrough WG. Reverse-phase protein array profiling of oropharyngeal cancer and significance of PIK3CA mutations in HPV-associated head and neck cancer. Clin Cancer Res. 2014 May 1;20(9):2300-11. doi: 10.1158/1078-0432.CCR-13-2585.

Shannon B, Yi TJ, Perusini S, Gajer P, Ma B, Humphrys MS, Thomas-Pavanel J, Chieza L, Janakiram P, Saunders M, Tharao W, Huibner S, Shahabi K, Ravel J, Rebbapragada A, Kaul R. Association of HPV infection and clearance with cervicovaginal immunology and the vaginal microbiota. Mucosal Immunol. 2017 Sep;10(5):1310-1319. doi: 10.1038/mi.2016.129.

Shigeishi H, Sugiyama M. Risk Factors for Oral Human Papillomavirus Infection in Healthy Individuals: A Systematic Review and Meta-Analysis. J Clin Med Res. 2016 Oct;8(10):721-9. doi: 10.14740/jocmr2545w.

Shigeishi H, Su CY, Kaneyasu Y, Matsumura M, Nakamura M, Ishikawa M, Saito A, Ohta K, Sugiyama M. Association of oral HPV16 infection with periodontal inflammation and the oral microbiome in older women. Exp Ther Med. 2021 Feb;21(2):167. doi: 10.3892/etm.2020.9598.

Slebos RJ, Jehmlich N, Brown B, Yin Z, Chung CH, Yarbrough WG, Liebler DC. Proteomic analysis of oropharyngeal carcinomas reveals novel HPV-associated biological pathways. Int J Cancer. 2013 Feb 1;132(3):568-79. doi: 10.1002/ijc.27699.

Sohn J, Li L, Zhang L, Settem RP, Honma K, Sharma A, Falkner KL, Novak JM, Sun Y, Kirkwood KL. Porphyromonas gingivalis indirectly elicits intestinal inflammation by altering the gut microbiota and disrupting epithelial barrier function through IL9-producing $CD4^+$ T cells. Mol Oral Microbiol. 2022 Apr;37(2):42-52. doi: 10.1111/omi.12359.

Sommer F, Bäckhed F. Know your neighbor: Microbiota and host epithelial cells interact locally to control intestinal function and physiology. Bioessays. 2016 May;38(5):455-64. doi: 10.1002/bies.201500151. Epub 2016 Mar 18. PMID: 26990415.

Sommer F, Anderson JM, Bharti R, Raes J, Rosenstiel P. The resilience of the intestinal microbiota influences health and disease. Nat Rev Microbiol. 2017 Oct;15(10):630-638. doi: 10.1038/nrmicro.2017.58.

Sturgis EM, Cinciripini PM. Trends in head and neck cancer incidence in relation to smoking prevalence: an emerging epidemic of human papillomavirus-associated cancers? Cancer. 2007 Oct 1;110(7):1429-35. doi: 10.1002/cncr.22963.

Sun J, Tang Q, Yu S, Xie M, Xie Y, Chen G, Chen L. Role of the oral microbiota in cancer evolution and progression. Cancer Med. 2020 Sep;9(17):6306-6321. doi: 10.1002/cam4.3206.

Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, Bray F. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA Cancer J Clin. 2021 May;71(3):209-249. doi: 10.3322/caac.21660.

Sunkara T, Rawla P, Ofosu A, Gaduputi V. Fecal microbiota transplant - a new frontier in inflammatory bowel disease. J Inflamm Res. 2018 Aug 31;11:321-328. doi: 10.2147/JIR.S176190.

Syrjänen KJ, Surjänen SM. Histological evidence for the presence of condylomatous epithelial lesions in association with laryngeal squamous cell carcinoma. ORL J Otorhinolaryngol Relat Spec. 1981;43(4):181-94. doi: 10.1159/000275538.

Syrjänen S. Human papillomavirus infections and oral tumors. Med Microbiol Immunol. 2003 Aug;192(3):123-8. doi: 10.1007/s00430-002-0173-7.

Syrjänen S. HPV infections and tonsillar carcinoma. J Clin Pathol. 2004 May;57(5):449-55. doi: 10.1136/jcp.2003.008656.

Syrjänen S. Current concepts on human papillomavirus infections in children. APMIS. 2010 Jun;118(6-7):494-509. doi: 10.1111/j.1600-0463.2010.02620.x.

Tabernero J, Bahleda R, Dienstmann R, Infante JR, Mita A, Italiano A, Calvo E, Moreno V, Adamo B, Gazzah A, Zhong B, Platero SJ, Smit JW, Stuyckens K, Chatterjee-Kishore M, Rodon J, Peddareddigari V, Luo FR, Soria JC. Phase I Dose-Escalation Study of JNJ-42756493, an Oral Pan-Fibroblast Growth Factor Receptor Inhibitor, in Patients With Advanced Solid Tumors. J Clin Oncol. 2015 Oct 20;33(30):3401-8. doi: 10.1200/JCO.2014.60.7341.

Tanaka S, Yoshida M, Murakami Y, Ogiwara T, Shoji M, Kobayashi S, Watanabe S, Machino M, Fujisawa S. The relationship of Prevotella intermedia, Prevotella nigrescens and Prevotella melaninogenica in the supragingival plaque of children, caries and oral malodor. J Clin Pediatr Dent. 2008 Spring;32(3):195-200. doi: 10.17796/jcpd.32.3.vp65717781561811. PMID: 18524268.

Tegnell A, Dillner J, Andrae B. Introduction of human papillomavirus (HPV) vaccination in Sweden. Euro Surveill. 2009 Feb 12;14(6):19119.(no doi) PMID: 19215721.

Tertipis N, Hammar U, Näsman A, Vlastos A, Nordfors C, Grün N, Ährlund-Richter A, Sivars L, Haeggblom L, Marklund L, Hammarstedt-Nordenvall L, Chaturvedi AK, Munck-Wikland E, Ramqvist T, Bottai M, Dalianis T. A model for predicting clinical outcome in patients with human papillomavirus-positive tonsillar and base of tongue cancer. Eur J Cancer. 2015 Aug;51(12):1580-7. doi: 10.1016/j.ejca.2015.04.024.

Tewari KS, Monk BJ. Gynecologic oncology group trials of chemotherapy for metastatic and recurrent cervical cancer. Curr Oncol Rep. 2005 Nov;7(6):419-34. doi: 10.1007/s11912-005-0007-z. .

Tinhofer I, Budach V, Saki M, Konschak R, Niehr F, Jöhrens K, Weichert W, Linge A, Lohaus F, Krause M, Neumann K, Endris V, Sak A, Stuschke M, Balermpas P, Rödel C, Avlar M, Grosu AL, Abdollahi A, Debus J, Belka C, Pigorsch S, Combs SE, Mönnich D, Zips D, Baumann M; DKTK-ROG. Targeted next-generation sequencing of locally advanced squamous cell carcinomas of the head and neck reveals druggable targets for improving adjuvant chemoradiation. Eur J Cancer. 2016 Apr;57:78-86. doi: 10.1016/j.ejca.2016.01.003.

Tommasino M. The human papillomavirus family and its role in carcinogenesis. Semin Cancer Biol. 2014 Jun;26:13-21. doi: 10.1016/j.semcancer.2013.11.002.

Torcia MG. Interplay among Vaginal Microbiome, Immune Response and Sexually Transmitted Viral Infections. Int J Mol Sci. 2019 Jan 11;20(2):266. doi: 10.3390/ijms20020266.

Tripathi SC, Matta A, Kaur J, Grigull J, Chauhan SS, Thakar A, Shukla NK, Duggal R, DattaGupta S, Ralhan R, Siu KW. Nuclear S100A7 is associated with poor prognosis in head and neck cancer. PLoS One. 2010 Aug 3;5(8):e11939. doi: 10.1371/journal.pone.0011939.

Ugur AR, Tuncer E I, Findik D. Identification of bacterial vaginosis-associated bacteria in male urethra: Co-occurrence of Atopobium vaginae and Gardnerella vaginalis. *Marshall J. Med.*10.21161/mjm.180025 (2019). (no doi).

Van Doorslaer K, Chen D, Chapman S, Khan J, McBride AA. Persistence of an Oncogenic Papillomavirus Genome Requires *cis*Elements from the Viral Transcriptional Enhancer. mBio. 2017 Nov 21;8(6):e01758-17. doi: 10.1128/mBio.01758-17.

van Passel MW, Kant R, Zoetendal EG, Plugge CM, Derrien M, Malfatti SA, Chain PS, Woyke T, Palva A, de Vos WM, Smidt H. The genome of Akkermansia muciniphila, a dedicated intestinal mucin degrader, and its use in exploring intestinal metagenomes. PLoS One. 2011 Mar 3;6(3):e16876. doi: 10.1371/journal.pone.0016876.

Venuti A, Paolini F, Nasir L, Corteggio A, Roperto S, Campo MS, Borzacchiello G. Papillomavirus E5: the smallest oncoprotein with many functions. Mol Cancer. 2011 Nov 11;10:140. doi: 10.1186/1476-4598-10-140.

von Knebel Doeberitz M, Vinokurova S. Host factors in HPV-related carcinogenesis: cellular mechanisms controlling HPV infections. Arch Med Res. 2009 Aug;40(6):435-42. doi: 10.1016/j.arcmed.2009.06.00

Welters MJP, Santegoets SJ, van der Burg SH. The Tumor Microenvironment and Immunotherapy of Oropharyngeal Squamous Cell Carcinoma. Front Oncol. 2020 Dec 21;10:545385. doi: 10.3389/fonc.2020.545385.

Wichmann G, Rosolowski M, Krohn K, Kreuz M, Boehm A, Reiche A, Scharrer U, Halama D, Bertolini J, Bauer U, Holzinger D, Pawlita M, Hess J, Engel C, Hasenclever D, Scholz M, Ahnert P, Kirsten H, Hemprich A, Wittekind C, Herbarth O, Horn F, Dietz A, Loeffler M; Leipzig Head and Neck Group (LHNG). The role of HPV RNA transcription, immune response-related gene expression and disruptive TP53 mutations in diagnostic and prognostic profiling of head and neck cancer. Int J Cancer. 2015 Dec 15;137(12):2846-57. doi: 10.1002/ijc.29649.

Wilting SM, Smeets SJ, Snijders PJ, van Wieringen WN, van de Wiel MA, Meijer GA, Ylstra B, Leemans CR, Meijer CJ, Brakenhoff RH, Braakhuis BJ, Steenbergen RD. Genomic profiling identifies common HPV-associated chromosomal alterations in squamous cell carcinomas of cervix and head and neck. BMC Med Genomics. 2009 Jun 1;2:32. doi: 10.1186/1755-8794-2-32.

World Health Organization. Head and Neck Cancer 2017 (Does this exist, page 23)

World Health Organization. Cervical Cancer. 2018

Yanofsky VR, Patel RV, Goldenberg G. Genital warts: a comprehensive review. J Clin Aesthet Dermatol. 2012 Jun;5(6):25-36. (no doi) PMID: 22768354; PMCID: PMC3390234.

Yang DY, Bracken K. Update on the new 9-valent vaccine for human papillomavirus prevention. Can Fam Physician. 2016 May;62(5):399-402. Epub 2016 May 12. (no doi) PMID: 27255620; PMCID: PMC4865336.

Young RJ, Bressel M, Porceddu S, Cernelc J, Savas P, Liu H, Urban D, Thai AA, Cooper C, Fua T, Neeson P, Rischin D, Solomon B. Validation and characterisation of prognostically significant PD-L1^{+} immune cells in HPV+ oropharyngeal squamous cell carcinoma. Oral Oncol. 2020 Feb;101:104516. doi: 10.1016/j.oraloncology.2019.104516.

Zhang M, Kumar B, Piao L, Xie X, Schmitt A, Arradaza N, Cippola M, Old M, Agrawal A, Ozer E, Schuller DE, Teknos TN, Pan Q. Elevated intrinsic cancer stem cell population in human papillomavirus-associated head and neck squamous cell carcinoma. Cancer. 2014 Apr 1;120(7):992-1001. doi: 10.1002/cncr.28538.

Zhou L, Huang L, Xu Q, Lv Y, Wang Z, Zhan P, Han H, Shao Y, Lin D, Lv T, Song Y. Association of *MUC19* Mutation With Clinical Benefits of Anti-PD-1 Inhibitors in Non-small Cell Lung Cancer. Front Oncol. 2021 Mar 22;11:596542. doi: 10.3389/fonc.2021.596542.

Zhou X, Ren L, Meng Q, Li Y, Yu Y, Yu J. The next-generation sequencing technology and application. Protein Cell. 2010 Jun;1(6):520-36. doi: 10.1007/s13238-010-0065-3. Epub 2010 Jul 7.