# VOYAGER: an international consortium investigating the role of human papilloma virus and genetics in oral and oropharyngeal cancer risk and survival

**Author list**

Gormley M [1,2,3*], Adhikari A [3], Dudding T [1,2], Pring M [2,3], Hurley K [3], Macfarlane GJ [4], Lagiou P [5], Lagiou A [6], Polesel J [7], Agudo A [8], Alemany L [9,10], Ahrens W [11], Healy CM [12], Conway DI [13], Canova C [14], Holcatova I [15], Richiardi L [16], Znaor A [17], Olshan AF [18], Hung RJ [19,20], Liu G [20, 21], Bratman S [22], Zhao X [23], Holt J [23], Cortez R [24], Gaborieau V [24], McKay JD [24], Waterboer T [25], Brennan P [24], Hayes N [23], Diergaarde B [26], Virani S [24*]

* Corresponding authors

**Affiliations**

[1] MRC Integrative Epidemiology Unit, Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, UK.

[2] Bristol Dental School, University of Bristol, Bristol, UK.

[3] University Hospitals Bristol NHS Foundation Trust Bristol Dental Hospital, Bristol, UK

[4] School of Medicine, Medical Sciences and Nutrition, University of Aberdeen, UK

[5] School of Medicine, National and Kapodistrian University of Athens, Greece

[6] School of Public Health, University of West Attica, Greece

[7] Unit of Cancer Epidemiology, Centro di Riferimento Oncologico di Aviano (CRO) National Cancer Institute, IRCCS, Italy

[8] Nutrition and Cancer Unit, Cancer Epidemiology Research Program, Catalan Institute of Oncology/IDIBELL, Barcelona, Spain

[9] Infections and Cancer Unit, Cancer Epidemiology Research Program, Catalan Institute of Oncology/IDIBELL, Barcelona, Spain

[10] Centro de Investigación Biomédica en Red: Epidemiología y Salud Pública (CIBERESP CB06/02/0073), Madrid, Spain.

[11] Epidemiological Methods and Etiological Research, Leibniz Institute for Prevention Research and Epidemiology – BIPS, Germany

[12] School of Dental Science, Dublin Dental University Hospital, Trinity College Dublin, Ireland

[13] School of Medicine, Dentistry, and Nursing, University of Glasgow, UK

[14] Department of Cardiac, Thoracic and Vascular Sciences University of Padova, Italy

[15] Institute of Hygiene and Epidemiology, Charles University Prague, Czech Republic

[16] Reference Centre for Epidemiology and Cancer Prevention in Piemonte, Italy

[17] Cancer Surveillance, International Agency for Research on Cancer, France

[18] Department of Epidemiology, Gillings School of Global Public Health, University of North Carolina, US.

[19] Prosserman Centre for Population Health Research, Lunenfeld-Tanenbaum Research Institute, Sinai Health System, Toronto, Canada.

[20] Dalla Lana School of Public Health, University of Toronto, Toronto, Canada.

[21] Computational Biology and Medicine Program, Princess Margaret Cancer Centre, Toronto Canada

[22] Department of Radiation Oncology, Princess Margaret Cancer Centre, University Health Network, University of Toronto, Toronto, Canada.

[23] Department of Medicine, University of Tennessee, USA

[24] Genomic Epidemiology Group, World Health Organization, International Agency for Research on Cancer, Lyon, France.

[25] Infections and Cancer Epidemiology, German Cancer Research Center (DKFZ), Heidelberg, Germany.

[26] Department of Human Genetics, School of Public Health, University of Pittsburgh, and UPMC Hillman Cancer Center, Pittsburgh, US.

**Word count: 3,224**

**Abstract**

Head and neck cancer (HNC) is the sixth most common cancer globally. Incidence and survival rates vary significantly across geographic regions and tumor subsites. This is partly due to differences in risk factor exposure, which includes tobacco smoking, alcohol consumption and human papillomavirus (HPV) infection, alongside detection and treatment strategies. The VOYAGER (human papillomaVirus, Oral and oropharYngeal cAncer GEnomic Research) consortium is a collaboration between five large North American and European studies which generated data on 10,530 participants (7,233 cases and 3,297 controls). The primary goal of the collaboration was to improve understanding of the role of HPV and genetic factors in oral cavity and oropharyngeal cancer risk and outcome. Demographic and clinical data collected by the five studies were harmonized, and HPV status was determined for the majority of cases. In addition, 999 tumors were sequenced to define somatic mutations. These activities generated a comprehensive biomedical resource that can be utilized to answer critical outstanding research questions to help improve HNC prevention, early detection, treatment, and surveillance.

**Key words**

Head and neck cancer, oral cancer, oropharyngeal cancer, human papilloma virus, risk factors, survival.

## Background

### Head and Neck Cancer Genomic Epidemiology

Head and neck cancer (HNC) which is primarily squamous cell carcinoma, includes cancers of the oral cavity, pharynx and larynx.(1, 2) Globally, the incidence of oral and oropharyngeal cancer is estimated at 8.0 and 2.0 per 100,000, respectively, and is predicted to increase by 30% by 2030.(3, 4) Five-year survival remains poor, averaging between 40 – 50% with hypopharynx cases experiencing the worst outcomes.(5) Incidence and survival rates vary significantly across geographic regions and tumor subsites, partly due to differences in risk factor exposure. Established HNC risk factors include tobacco smoking and alcohol intake, which together account for a similar population attributable risk for both oral (64%) and oropharyngeal (72%) cancer.(6) However, human papilloma virus (HPV) infection, particularly high-risk subtype 16, has emerged as another major risk factor for oropharyngeal cancer.(7-9) Worldwide, it is estimated that around 52,000 incident HNC cases are caused by a persistent HPV infection each year, with attributable fractions highest in high-income countries in North America and Europe.(10-13)

Given the decline of tobacco use in developed countries, the incidence rate of HPV driven [HPV(+)] oropharyngeal cancer is now surpassing that of oral cancer.(9-12, 14) HPV(+) oropharyngeal tumors are considered distinct entities, demonstrating more favorable treatment response and prognosis compared to non-HPV related oropharyngeal cancer [HPV(-)].(8, 12, 15-17) This is likely due to differences in etiology, patient and tumor characteristics, with HPV(+) oropharyngeal tumors presenting more frequently in younger individuals (<65 years), and in those reporting higher numbers of sexual partners with reduced cumulative tobacco exposure compared to HPV(-) cases.(12, 18) However, only a small proportion of those with an oral HPV infection will develop HNC and despite better long-term survival, up to 25% of patients still develop disease recurrence within 5 years after initial diagnosis.(19) To improve prevention, early detection and prognosis, a better understanding of the role of host genetics and interactions with modifiable risk factors, such

as tobacco and alcohol use in oral and oropharyngeal cancer risk and survival is required.(20)

HPV driven carcinogenesis is characterized by increased expression of the viral oncogenes E6 and E7, leading to increased degradation of tumor suppressor proteins p53 and Rb, respectively and loss of cell cycle activation. This can result in genomic instability and resistance to apoptosis.(21, 22) HPV(+) and HPV(-) head and neck tumors harbor a similar burden of somatic variants. However, HPV(+) oropharyngeal tumors carry fewer copy-number alterations, suggesting a higher degree of genomic stability.(23-26) Genome profiling studies have provided a list of genes that are recurrently mutated in HNC, including *TP53*, *CDKN2A* (which encodes for $p16^{INK4}$), *NOTCH1* and *PIK3CA*.(23, 24, 27-29) Genes recurrently mutated in HPV(+) oropharyngeal cancer are related to epithelial structure and differentiation, in addition to *RB1* (encoding the Rb protein).(12, 23-26) The presence or absence of particular somatic alterations in tumors may be good markers of cancer prognosis and response to treatment, but there is still a need to identify novel somatic driver alterations, particularly as relatively few HPV(+) oropharyngeal cancer cases have been sequenced to date.(24, 30, 31) Identification of molecular markers associated with prognosis could facilitate better monitoring and clinical decision making, including the use of de-escalation treatment strategies among those at lower risk of recurrence or progression as a means to improve quality of life and, conversely, more aggressive treatment in those deemed at higher risk.

Compared to other major cancer sites such as breast, lung and colorectal, HNC is relatively rare, hampering research efforts. Collaboration between studies to form extensive biomedical databases and resources plays a crucial role in driving progress across various domains of cancer research.

## Construction and content

### The VOYAGER Consortium: Design

In 2016, the VOYAGER (human papillomaVirus, Oral and oropharYngeal cAncer GEnomic Research) consortium was established, bringing together five large North American and European studies (**Figure 1**) with a focus on oral and oropharyngeal cancers. The primary goal of the collaboration was to improve understanding of the role of HPV and genomic factors in oral cavity and oropharyngeal cancer risk and outcome. The project was funded by the US National Institute of Dental and Craniofacial Research (NIDCR; R01DE025712).

The VOYAGER consortium includes 10,530 participants in total (6,489 oral and oropharyngeal cancer cases, 744 other head and neck cancer cases, and 3,297 controls), with detailed demographic, risk factor and clinical data. The five studies comprising VOYAGER have been previously described and are: (a) the Alcohol-related cancers and genetic susceptibility in Europe (ARCAGE) study (32), (b) the Toronto Mount Sinai Hospital-Princess Margaret (MSH-PMH) study (Toronto) (33), (c) the University of Pittsburgh case-control study on head and neck cancer (Pittsburgh) (34), (d) the Carolina Head and Neck Cancer Epidemiology (CHANCE) study (35), and (e) the Head and Neck 5000 study (HN5000) (36). Ethical approval was obtained as described in the **Declarations** section below.

Each study contributed oral cavity and oropharyngeal cancer cases as defined by the following ICD-10 codes: oral cavity (C00.3-C00.6, C00.8-C00.9, C02.0-C02.3, C02.8, C02.9, C03.0-C03.9, C04.0-C04.9, C05.0, C05.8, C05.9, C06.0-C06.9) oropharynx (C01-C01.9, C02.4, C05.1-C05.2, C09.0-C10.9), head and neck not otherwise specified (NOS) (C76.0), pharynx NOS (C14.0), or any cases with overlap of these sites. Demographic information, including age, sex, ethnicity, geographic region and education level, and information on established risk factors, e.g., smoking and alcohol history, were shared. Clinical variables

included ICD code, tumor, nodal and metastasis status, HPV status (defined by p16 or HPV DNA), vital status, follow-up time, and treatment information. Given all cases were diagnosed between 2002 – 2018, the American Joint Committee on Cancer (AJCC) 7$^{th}$ edition was used, mirroring the staging used in clinical practice.(37)

**Data Generation and Harmonization**

The VOYAGER consortium generated several different types of data including clinical, demographic and behavior variables. Data harmonization was conducted for variables that were centralized across all studies, as illustrated in **Tables 1** and **2**. Upon receipt of data, cleaning and validation checks were conducted to identify inconsistencies, outliers, and missing values. Follow-up data was carefully collated and harmonized to capture follow-up times, progression events, vital status and cause of death to facilitate high-quality prognostic research. A comprehensive data dictionary was developed to specify final definitions, formats, permissible values, coding schemes and classifications (i.e., for categorical variables). Version control was implemented to ensure consistency of data use across analyses. Further detail on the statistical analyses performed is available in the **Supplementary Methods**.

HPV16 E6 serology was prioritized as a marker of HPV(+) oropharyngeal cancer as this has been shown to be a highly sensitive and specific marker of HPV oncogenic infection in oropharyngeal cancer and can be easily assayed from blood.(38, 39) Multiplex serology was performed on 75% (n= 5,294) of all HNC cases and 92% (n= 3,250) of oropharyngeal cancer cases using a previously developed Luminex assay.(40, 41) Multiplex serology generates quantitative data expressed in median fluorescence intensity (MFI) units for each pathogen-specific antigen and serum. Seropositivity for every antigen was based on previously determined standardized cut-offs in order to optimize sensitivity and specificity.(42, 43) When serology was missing, two concordant tumor markers, p16 immunohistochemistry (IHC) and high-risk HPV DNA in-situ hybridization (ISH) were required to determine HPV

194 status. If p16 IHC and HPV DNA ISH were discordant or only one marker was available,
195 then HPV status was unknown. This algorithm was based on evaluation of biomarkers
196 performance compared to molecular reference method (serology) from known data, led by
197 consortium members.(39, 44)
198
199 Genotyping data was generated at the Center for Inherited Disease Research (CIDR) in
200 several rounds. The first round (X01HG007780) was performed using the Illumina
201 OncoArray, which was custom designed for cancer studies by the OncoArray Consortium
202 part of the Genetic Associations and Mechanisms in Oncology (GAME-ON) Network. All
203 samples (6,034 cases and 6,585 controls) were genotyped as part of the oral and pharynx
204 cancer OncoArray study, except for 1,023 controls from the Toronto study which were
205 genotyped as part of the Lung OncoArray. This genotyping data was used to conduct the
206 first genome-wide association study (GWAS) on head and neck cancer in 2017.(45) With the
207 confirmation of HPV status, genotyping data were also used to run a GWAS of oral and
208 oropharyngeal cancer, stratified by HPV status.(46) A second round of genotyping was
209 undertaken (X01HG010743) for an additional 1,491 samples in VOYAGER. This was
210 conducted on the All of Us Array, an Illumina array customized for the All of Us Consortium
211 and designed to include multiethnic context.(47) The genotyping data from OncoArray and
212 the All of Us Array, has contributed to the largest HNC GWAS to date including 19,073
213 cases and 38,357 controls identifying 29 independent genetic loci.(48)
214
215 DNA tumor sequencing data were also generated for 999 samples in the VOYAGER
216 consortium using a custom cancer gene panel that has been previously reported.(49-53)
217 Next generation sequencing was performed using the Agilent SureSelect protocol and
218 reagents according to manufacturer's specifications. The assay targets all genes of the
219 Cancer Gene Census, in addition to clinically relevant targets such as drug metabolizing
220 enzymes. The assay also performs whole genome sequencing of HPV16 and 18 using
221 methodology previously reported to offer clinical diagnostic accuracy comparable or better to

conventional approaches while at the same time capturing base level resolution across the HPV genome.(50) The resulting libraries were sequenced on Illumina sequencers, primarily NovaSeq according to manufacturer's specifications. Further details of the assay can be found elsewhere.(54) Target depth was mean exon coverage of 500x coverage. All analytic tools and the pipelines for integrating steps have been publicly reported and are available as open source software including: BWA for sequence alignment (55), NGS Copy (54) for copy number assessment, Strelka (56), ABRA for realignment and structural variant detection (57), and UNMASC for variant prioritization and filtering.(49)

**VOYAGER Data Resource: Clinical and Demographic Profile**

The VOYAGER consortium comprises 10,530 participants in total (6,489 oral and oropharyngeal cancer cases, 744 other head and neck cancer cases, and 3,297 controls). The primary focus was on oral cavity and oropharyngeal cancers, with systematically planned and standardized inclusion criteria using ICD-10 coding. However, additional cases from other head and neck subsites were also included opportunistically. In total, 3,514 oropharynx cancers and 2,975 oral cavity cancers, alongside 744 cases from several other head and neck cancer subsites were included. The controls included were used for the genotyping studies. To facilitate the use of this resource, we present here a brief description of the oral cavity and oropharyngeal cancers and controls, while noting that this resource contains additional head and neck subsites, described in Error! Reference source not found.**,** that may be valuable to the scientific community.

Cancer cases were contributed to by HN5000 (41%), followed by Toronto (23%), Pittsburgh (13%), CHANCE (12%) and ARCAGE (11%). There were significant differences ($p < 0.0001$) in anatomical site and staging across studies, with HN5000 contributing the highest number of oropharyngeal cancer cases (45%) and therefore higher numbers of late (stage III and IV) disease (**Supplementary Table 1**). Differences were also detected between cases versus controls across all clinical and demographic variables, except age ($p = 0.308$) (**Table 1**).

HPV16 E6 serology status was available for 3,250 (92%) oropharyngeal cancer cases, of which 61% were HPV(+) (**Figure 1; Table 1**). When serology was missing, p16 and high-risk HPV DNA ISH concordance determined HPV oropharyngeal cancer status in the remaining ~8% of cases. There was good concordance between p16 and HPV DNA ISH. The proportion of HPV(+) and HPV(-) oropharyngeal tumors was similar across North American and European regions (p= 0.156), but these cases varied in terms of all other demographic, clinical and risk factor behavior variables (**Table 1; Table 2**).

There are in total 2,975 oral cavity, and 3,514 oropharyngeal cancer cases (2,138 HPV(+), 1,146 HPV(-) and 230 HPV status unknown oropharyngeal cancer cases included in VOYAGER (**Table 1**). Most cases presented in males (73%), outnumbering females across all subsites. The overall mean age at diagnosis in all cases was 60 years (SD= 10.7), with the lowest mean age observed in the HPV(+) oropharyngeal cancer group (58 years (SD= 8.9)). Overall age at diagnosis for oropharyngeal cancer (59 (SD= 9.6)) was significantly younger than that of oral cavity (61 (SD= 12.2)) (p< 0.0001) (**Table 1**). Only 3% of oral and oropharyngeal squamous cell carcinoma cases (n= 224) presented under 40 years old, and these were predominantly patients with cancer of the oral cavity (71%). The highest frequency of postsecondary education was observed in the HPV(+) oropharyngeal cancer group (42%), with the lowest in the HPV(-) oropharyngeal group (28%) (**Table 1**).

Most cases were current (33%) or former smokers (42%), except for the HPV(+) oropharyngeal group which had the largest never smoker population (34%) with significant differences in smoking found between subsites (p< 0.0001) (**Table 2**). Similarly, more than half (63%) of cases reported current alcohol drinking, which was consistent across subsites. The proportion of never drinkers was different in oral cavity (16%) compared to oropharyngeal cancer (12%), and significantly lower in total cases (14%) versus controls (22%) (p< 0.0001) (**Table 2**). Overweight BMI was most common in the HPV(+) oropharyngeal cancer subsite (40%), with underweight BMI being the least frequently

recorded category across oral and oropharyngeal cancer cases. There were significant differences between total cases and controls across all BMI categories (p<0.0001) (**Table 2**).

Overall, significantly more oropharyngeal cancer cases presented at stage IV (72% compared to oral cavity cancers which presented earlier, at stage I (28%) and II (22%) (p<0.0001) (**Table 1**). Non-surgical treatment using radiotherapy, with or without chemotherapy, was the most common treatment modality for oropharyngeal cancer (66%). Surgery was the most common treatment modality for cases of oral cavity cancer, with almost half (47%) of these patients receiving surgery alone and another 40% undergoing surgery plus adjuvant radiotherapy with or without chemotherapy. Collecting information on disease outcome was a primary focus of the consortium given the high risk of recurrence and poor survival associated with HNC. The median length of follow-up time for oral and oropharyngeal cancer cases was 5.3 years. Overall median survival time was 9.8 years, and the five-year survival rate was 66% across all oral cavity and oropharyngeal cancer sites (**Table 3**).

Patients with HPV(+) oropharyngeal cancer had the best overall survival, with a median survival time of 14.3 years and 5-year survival at 81% (**Table 3** and **Figure 2**). For patients where HPV status was not available, probability of survival at 5 years fell between that of HPV(+) and HPV(-) oropharyngeal cancer patients, at 63%, indicating these were likely a mix of patients with and without HPV(+) oropharyngeal cancer. The probability of being progression-free at 5 years after diagnosis was highest for HPV(+) oropharyngeal cancer patients at 75%, and similar for all other subsites, ranging from 44% – 50%. Overall, the probability of being progression-free at 5 years was 57%. Disease-specific survival reflected similar trends across subsites (**Table 3** and **Figure 2**). HPV(+) oropharyngeal cancer patients had the highest probability of disease-specific survival at 5 years (88%) and HPV(-) oropharyngeal cancer had the lowest (67%). Across all cancers, the probability of disease-specific survival was 78% (**Table 3** and **Figure 2**).

Median length of follow-up time across studies varied, but not significantly so. Median length of follow-up time was 8.4 years for ARCAGE, 10.0 years for CHANCE, 4.9 years for HN5000, 5.7 years for Pittsburgh and 5.2 years for Toronto (p= 0.4). Patients from the ARCAGE study have lower overall, progression-free and disease-specific survival compared to other studies for HPV(+) oropharyngeal cancers, however, there were only 63 HPV(+) oropharyngeal cancer cases in this cohort (**Supplementary Figures 1 - 3**). In addition, the smoking rates of patients in ARCAGE is much higher compared to other studies and this has been shown to affect interpretation of its HPV(+) oropharyngeal cancer profiles(38).

The clinical, demographic and survival profile of the VOYAGER cohort are characteristic of what we know of HNC patients, epidemiologically and clinically.(4, 58, 59) We show the predominance of male patients, the established differences in age at diagnosis, education and survival of HPV driven and non-HPV cancers, and the expected distributions of risk factor behaviors.(4, 58) Therefore, VOYAGER provides a reliable representative resource for further research into this patient population. Importantly, the overlap of data with follow-up data, somatic tumor sequencing and germline genotyping is a strength of this resource (**Figure 3**).

**Utility and discussion**

Several important findings have come from the VOYAGER consortium. Previous GWAS studies of oral and oropharyngeal cancer conducted by our group highlighted the important role of the human leukocyte antigen (HLA) region (6p21.3) in susceptibility to oropharyngeal cancer.(45) In particular, a two-fold protective effect was observed for oropharyngeal cancer and the HLA haplotype DRB1*1301-DQA1*0103-DQB1*0603. This haplotype was previously reported to also be protective for cervical cancer, a cancer type that is primarily driven by HPV infection. In VOYAGER, HPV status was determined in 92% of oropharyngeal tumors via HPV16 E6 serology enabling the first GWAS focused on HPV driven HNCs using fine

mapping techniques.(46) Within the 6p21.3 locus, there were two specific loci (rs4713462 and rs9269942) independently associated with reduced risk of oropharyngeal cancer (**Figure 4**). These loci were separately associated with antibodies against specific HPV16 proteins which implicates specific germline variants in the natural immune response against HPV(+) oropharyngeal cancer, supporting the use of therapeutic vaccines to protect against this disease.

A recent publication utilized VOYAGER data to develop a risk prediction model for HNC including genetic markers, HPV serostatus, demographic and lifestyle risk factors in populations of European ancestry. The addition of HPV serology provided substantial predictive accuracy for oropharyngeal cancer (AUC= 0.94, 95%CI: 0.92 – 0.95 in men and AUC= 0.92, 95%CI: 0.88 – 0.95 in women) above that of previously published models, highlighting the need to consider primary prevention and intensive surveillance for oropharyngeal cancer subgroups.(60) Importantly, however, while HPV serology is a marker for HPV driven oropharyngeal cancer, the use of HPV serology needs to be carefully evaluated among smokers as demonstrated in another VOYAGER publication.(38) Diagnostic accuracy of HPV serology was evaluated utilizing VOYAGER data and found to be highly sensitive and specific independent of age, sex, year of diagnosis, BMI at diagnosis, current alcohol use and primary tumor size, but exhibited some variation in diagnostic accuracy for heavy smokers and by lymph node involvement. This work provides further evidence that this additional HPV biomarker can be used for early diagnosis.

Randomized clinical trials remain the ‘gold standard’ to ascertain the causal effect of interventions or modifiable exposures. However, they are not always feasible in terms of cost, time or ethics (61). Conversely, observational studies are subject to confounding, bias and reverse causality. This has significantly limited study design in HNC research, given that it is a relatively rare cancer with correlated risk factors, requiring long-term follow-up from the point of exposure to disease onset. To overcome such limitations, Mendelian randomization

(MR) uses measured genetic variation to examine the causal effect of potentially modifiable exposures on health outcomes in observational data.(62-64) The genotyping data available in VOYAGER has contributed to multiple MR studies (65-70) investigating a wide range of genetically proxied exposures. These studies have among other things demonstrated an independent causal effect for both smoking and alcohol on both HPV(+) and HPV(-) HNC, suggesting that the effect of alcohol may have been previously underestimated.(65, 69) This work further strengthens the evidence to support public health messaging around prevention in HNC.

**Conclusion and future prospects**

The value of HNC data generated as part of VOYAGER highlights its use for studies focused on prognosis, particularly overall survival where data is most complete. The variation across centers is an important consideration, as studies contributing to VOYAGER were conducted across different geographical settings and health systems. The VOYAGER data resource is particularly suited for genomic studies on risk factors and outcome.

## References

1. Zhou T, Huang W, Wang X, Zhang J, Zhou E, Tu Y, et al. Global burden of head and neck cancers from 1990 to 2019. iScience. 2024;27(3):109282.

2. Fitzmaurice C, Allen C, Barber RM, Barregard L, Bhutta ZA, Brenner H, et al. Global, Regional, and National Cancer Incidence, Mortality, Years of Life Lost, Years Lived With Disability, and Disability-Adjusted Life-years for 32 Cancer Groups, 1990 to 2015: A Systematic Analysis for the Global Burden of Disease Study. JAMA Oncol. 2017;3(4):524-48.

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

4. Johnson DE, Burtness B, Leemans CR, Lui VWY, Bauman JE, Grandis JR. Head and neck squamous cell carcinoma. Nat Rev Dis Primers. 2020;6(1):92.

5. Bosetti C, Carioli G, Santucci C, Bertuccio P, Gallus S, Garavello W, et al. Global trends in oral and pharyngeal cancer incidence and mortality. International Journal of Cancer. 2020;147(4):1040-9.

6. Hashibe M, Brennan P, Chuang SC, Boccia S, Castellsague X, Chen C, et al. Interaction between Tobacco and Alcohol Use and the Risk of Head and Neck Cancer: Pooled Analysis in the International Head and Neck Cancer Epidemiology Consortium. Cancer Epidem Biomar. 2009;18(2):541-50.

7. Heck JE, Berthiller J, Vaccarella S, Winn DM, Smith EM, Shan'gina O, et al. Sexual behaviours and the risk of head and neck cancers: a pooled analysis in the International Head and Neck Cancer Epidemiology (INHANCE) consortium. Int J Epidemiol. 2010;39(1):166-81.

8. Ang KK, Harris J, Wheeler R, Weber R, Rosenthal DI, Nguyen-Tân PF, et al. Human papillomavirus and survival of patients with oropharyngeal cancer. N Engl J Med. 2010;363(1):24-35.

9. Chaturvedi AK, Engels EA, Pfeiffer RM, Hernandez BY, Xiao W, Kim E, et al. Human papillomavirus and rising oropharyngeal cancer incidence in the United States. J Clin Oncol. 2011;29(32):4294-301.

10. Chaturvedi AK, Anderson WF, Lortet-Tieulent J, Curado MP, Ferlay J, Franceschi S, et al. Worldwide trends in incidence rates for oral cavity and oropharyngeal cancers. J Clin Oncol. 2013;31(36):4550-9.

11. Mehanna H, Beech T, Nicholson T, El-Hariry I, McConkey C, Paleri V, et al. Prevalence of human papillomavirus in oropharyngeal and nonoropharyngeal head and neck cancer--systematic review and meta-analysis of trends by time and region. Head Neck. 2013;35(5):747-55.

12. Lechner M, Liu J, Masterson L, Fenton TR. HPV-associated oropharyngeal cancer: epidemiology, molecular biology and clinical management. Nature Reviews Clinical Oncology. 2022;19(5):306-27.

13. de Martel C, Georges D, Bray F, Ferlay J, Clifford GM. Global burden of cancer attributable to infections in 2018: a worldwide incidence analysis. Lancet Glob Health. 2020;8(2):e180-e90.

14. Lauritzen BB, Grønlund MW, Jakobsen KK, Justesen MM, Garset-Zamani M, Carlander AF, et al. Epidemiological trends and survival of oropharyngeal cancer in a high HPV-prevalent area: A Danish population-based study from 2000 to 2020. Int J Cancer. 2024;155(12):2169-79.

15. Fakhry C, Westra WH, Li S, Cmelak A, Ridge JA, Pinto H, et al. Improved survival of patients with human papillomavirus-positive head and neck squamous cell carcinoma in a prospective clinical trial. J Natl Cancer Inst. 2008;100(4):261-9.

16. Fakhry C, Zhang Q, Nguyen-Tan PF, Rosenthal D, El-Naggar A, Garden AS, et al. Human papillomavirus and overall survival after progression of oropharyngeal squamous cell carcinoma. J Clin Oncol. 2014;32(30):3365-73.

17. Zanoni DK, Patel SG, Shah JP. Changes in the 8th Edition of the American Joint Committee on Cancer (AJCC) Staging of Head and Neck Cancer: Rationale and Implications. Curr Oncol Rep. 2019;21(6):52.

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

19. INTEGRATE. Post-Treatment Head and Neck Cancer Care: National Audit and Analysis of Current Practice in the United Kingdom. Clin Otolaryngol. 2021;46(1):284-94.

20. Brenner N, Mentzer AJ, Hill M, Almond R, Allen N, Pawlita M, et al. Characterization of human papillomavirus (HPV) 16 E6 seropositive individuals without HPV-associated malignancies after 10 years of follow-up in the UK Biobank. EBioMedicine. 2020;62:103123.

21. Huibregtse JM, Scheffner M, Howley PM. A cellular protein mediates association of p53 with the E6 oncoprotein of human papillomavirus types 16 or 18. Embo j. 1991;10(13):4129-35.

22. Scheffner M, Werness BA, Huibregtse JM, Levine AJ, Howley PM. The E6 oncoprotein encoded by human papillomavirus types 16 and 18 promotes the degradation of p53. Cell. 1990;63(6):1129-36.

448 23. Seiwert TY, Zuo Z, Keck MK, Khattri A, Pedamallu CS, Stricker T, et al. Integrative
449 and comparative genomic analysis of HPV-positive and HPV-negative head and neck
450 squamous cell carcinomas. Clin Cancer Res. 2015;21(3):632-41.

451 24. The Cancer Genome Atlas Network. Comprehensive genomic characterization of
452 head and neck squamous cell carcinomas. Nature. 2015;517(7536):576-82.

453 25. Gillison ML, Akagi K, Xiao W, Jiang B, Pickard RKL, Li J, et al. Human
454 papillomavirus and the landscape of secondary genetic alterations in oral cancers. Genome
455 Res. 2019;29(1):1-17.

456 26. Dogan S, Xu B, Middha S, Vanderbilt CM, Bowman AS, Migliacci J, et al.
457 Identification of prognostic molecular biomarkers in 157 HPV-positive and HPV-negative
458 squamous cell carcinomas of the oropharynx. Int J Cancer. 2019;145(11):3152-62.

459 27. Agrawal N, Frederick MJ, Pickering CR, Bettegowda C, Chang K, Li RJ, et al. Exome
460 sequencing of head and neck squamous cell carcinoma reveals inactivating mutations in
461 NOTCH1. Science. 2011;333(6046):1154-7.

462 28. Pickering CR, Zhang J, Neskey DM, Zhao M, Jasser SA, Wang J, et al. Squamous
463 cell carcinoma of the oral tongue in young non-smokers is genomically similar to tumors in
464 older smokers. Clin Cancer Res. 2014;20(14):3842-8.

465 29. Stransky N, Egloff AM, Tward AD, Kostic AD, Cibulskis K, Sivachenko A, et al. The
466 mutational landscape of head and neck squamous cell carcinoma. Science.
467 2011;333(6046):1157-60.

468 30. Gauthier MA, Kadam A, Leveque G, Golabi N, Zeitouni A, Richardson K, et al. Long-
469 read sequencing of oropharyngeal squamous cell carcinoma tumors reveal diverse patterns
470 of high-risk Human Papillomavirus integration. Front Oncol. 2023;13:1264646.

31. Lechner M, Frampton GM, Fenton T, Feber A, Palmer G, Jay A, et al. Targeted next-generation sequencing of head and neck squamous cell carcinoma identifies novel genetic alterations in HPV+ and HPV- tumors. Genome Med. 2013;5(5):49.

32. Lagiou P, Georgila C, Minaki P, Ahrens W, Pohlabeln H, Benhamou S, et al. Alcohol-related cancers and genetic susceptibility in Europe: the ARCAGE project: study samples and data collection. Eur J Cancer Prev. 2009;18(1):76-84.

33. Thomas S, Carroll JC, Brown MC, Chen Z, Mirshams M, Patel D, et al. Nicotine dependence as a risk factor for upper aerodigestive tract (UADT) cancers: A mediation analysis. PLoS One. 2020;15(8):e0237723.

34. Troy JD, Grandis JR, Youk AO, Diergaarde B, Romkes M, Weissfeld JL. Childhood passive smoke exposure is associated with adult head and neck cancer. Cancer Epidemiol. 2013;37(4):417-23.

35. Divaris K, Olshan AF, Smith J, Bell ME, Weissler MC, Funkhouser WK, et al. Oral health and risk for head and neck squamous cell carcinoma: the Carolina Head and Neck Cancer Study. Cancer Causes Control. 2010;21(4):567-75.

36. Ness AR, Waylen A, Hurley K, Jeffreys M, Penfold C, Pring M, et al. Establishing a large prospective clinical cohort in people with head and neck cancer as a biomedical resource: head and neck 5000. BMC Cancer. 2014;14(1):973.

37. Edge SB, Compton CC. The American Joint Committee on Cancer: the 7th edition of the AJCC cancer staging manual and the future of TNM. Ann Surg Oncol. 2010;17(6):1471-4.

38. Kusters JMA, Diergaarde B, Ness A, Schim van der Loeff MF, Heijne JCM, Schroeder L, et al. Diagnostic accuracy of HPV16 early antigen serology for HPV-driven oropharyngeal cancer is independent of age and sex. Int J Cancer. 2024;154(2):389-402.

39. 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;13(12).

40. Waterboer T, Sehr P, Pawlita M. Suppression of non-specific binding in serological Luminex assays. Journal of Immunological Methods. 2006;309(1):200-4.

41. Waterboer T, Sehr P, Michael KM, Franceschi S, Nieland JD, Joos TO, et al. Multiplex human papillomavirus serology based on in situ-purified glutathione s-transferase fusion proteins. Clin Chem. 2005;51(10):1845-53.

42. Clifford GM, Shin H-R, Oh J-K, Waterboer T, Ju Y-H, Vaccarella S, et al. Serologic Response to Oncogenic Human Papillomavirus Types in Male and Female University Students in Busan, South Korea. Cancer Epidemiology, Biomarkers & Prevention. 2007;16(9):1874-9.

43. Kreimer AR, Johansson M, Waterboer T, Kaaks R, Chang-Claude J, Drogen D, et al. Evaluation of Human Papillomavirus Antibodies and Risk of Subsequent Head and Neck Cancer. Journal of Clinical Oncology. 2013;31(21):2708-15.

44. Morais E, Kothari S, Chen YT, Roberts CS, Gómez-Ulloa D, Fenoll R, et al. The BROADEN study: The design of an observational study to assess the absolute burden of HPV-related head and neck cancers. Contemp Clin Trials. 2022;115:106631.

45. Lesseur C, Diergaarde B, Olshan AF, Wünsch-Filho V, Ness AR, Liu G, et al. Genome-wide association analyses identify new susceptibility loci for oral cavity and pharyngeal cancer. Nat Genet. 2016;48(12):1544-50.

46. Ferreiro-Iglesias A, McKay JD, Brenner N, Virani S, Lesseur C, Gaborieau V, et al. Germline determinants of humoral immune response to HPV-16 protect against oropharyngeal cancer. Nat Commun. 2021;12(1):5945.

519 47. Denny JC, Rutter JL, Goldstein DB, Philippakis A, Smoller JW, Jenkins G, et al. The
520 "All of Us" Research Program. N Engl J Med. 2019;381(7):668-76.

521 48. Ebrahimi E, Sangphukieo A, Park H, Gaborieau V, Ferreiro-Iglesias A, Diergaarde B,
522 et al. Cross-ancestral GWAS identifies 29 novel variants across Head and Neck Cancer
523 subsites. medRxiv. 2024:2024.11.18.24317473.

524 49. Little P, Jo H, Hoyle A, Mazul A, Zhao X, Salazar AH, et al. UNMASC: tumor-only
525 variant calling with unmatched normal controls. NAR Cancer. 2021;3(4):zcab040.

526 50. Montgomery ND, Parker JS, Eberhard DA, Patel NM, Weck KE, Sharpless NE, et al.
527 Identification of Human Papillomavirus Infection in Cancer Tissue by Targeted Next-
528 generation Sequencing. Appl Immunohistochem Mol Morphol. 2016;24(7):490-5.

529 51. Seifert BA, O'Daniel JM, Amin K, Marchuk DS, Patel NM, Parker JS, et al. Germline
530 Analysis from Tumor-Germline Sequencing Dyads to Identify Clinically Actionable
531 Secondary Findings. Clin Cancer Res. 2016;22(16):4087-94.

532 52. Patel NM, Michelini VV, Snell JM, Balu S, Hoyle AP, Parker JS, et al. Enhancing
533 Next-Generation Sequencing-Guided Cancer Care Through Cognitive Computing.
534 Oncologist. 2018;23(2):179-85.

535 53. Patel NM, Jo H, Eberhard DA, Yin X, Hayward MC, Stein MK, et al. Improved Tumor
536 Purity Metrics in Next-generation Sequencing for Clinical Practice: The Integrated
537 Interpretation of Neoplastic Cellularity and Sequencing Results (IINCaSe) Approach. Appl
538 Immunohistochem Mol Morphol. 2019;27(10):764-72.

539 54. Zhao X, Wang A, Walter V, Patel NM, Eberhard DA, Hayward MC, et al. Combined
540 Targeted DNA Sequencing in Non-Small Cell Lung Cancer (NSCLC) Using UNCseq and
541 NGScopy, and RNA Sequencing Using UNCqeR for the Detection of Genetic Aberrations in
542 NSCLC. PLoS One. 2015;10(6):e0129280.

55. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 2009;10(3):R25.

56. Saunders CT, Wong WS, Swamy S, Becq J, Murray LJ, Cheetham RK. Strelka: accurate somatic small-variant calling from sequenced tumor-normal sample pairs. Bioinformatics. 2012;28(14):1811-7.

57. Mose LE, Wilkerson MD, Hayes DN, Perou CM, Parker JS. ABRA: improved coding indel detection via assembly-based realignment. Bioinformatics. 2014;30(19):2813-5.

58. Rettig EM, D'Souza G. Epidemiology of head and neck cancer. Surg Oncol Clin N Am. 2015;24(3):379-96.

59. Maxwell JH, Grandis JR, Ferris RL. HPV-Associated Head and Neck Cancer: Unique Features of Epidemiology and Clinical Management. Annu Rev Med. 2016;67:91-101.

60. Budhathoki S, Diergaarde B, Liu G, Olshan A, Ness A, Waterboer T, et al. A risk prediction model for head and neck cancers incorporating lifestyle factors, HPV serology and genetic markers. Int J Cancer. 2023;152(10):2069-80.

61. Lawlor DA, Tilling K, Davey Smith G. Triangulation in aetiological epidemiology. International Journal of Epidemiology. 2017;45(6):1866-86.

62. Smith GD, Ebrahim S. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? International Journal of Epidemiology. 2003;32(1):1-22.

63. Davey Smith G, Hemani G. Mendelian randomization: genetic anchors for causal inference in epidemiological studies. Hum Mol Genet. 2014;23(R1):R89-98.

64. Sanderson E, Glymour MM, Holmes MV, Kang H, Morrison J, Munafò MR, et al. Mendelian randomization. Nature Reviews Methods Primers. 2022;2(1):6.

65. Gormley M, Dudding T, Sanderson E, Martin RM, Thomas S, Tyrrell J, et al. A multivariable Mendelian randomization analysis investigating smoking and alcohol consumption in oral and oropharyngeal cancer. Nature Communications. 2020;11(1).

66. Gormley M, Dudding T, Kachuri L, Burrows K, Chong AHW, Martin RM, et al. Investigating the effect of sexual behaviour on oropharyngeal cancer risk: a methodological assessment of Mendelian randomization. BMC Medicine. 2022;20(1).

67. Gormley M, Yarmolinsky J, Dudding T, Burrows K, Martin RM, Thomas S, et al. Using genetic variants to evaluate the causal effect of cholesterol lowering on head and neck cancer risk: A Mendelian randomization study. PLoS Genetics. 2021;17(4 April 2021).

68. Gormley M, Dudding T, Thomas SJ, Tyrrell J, Ness AR, Pring M, et al. Evaluating the effect of metabolic traits on oral and oropharyngeal cancer risk using Mendelian randomization. eLife. 2023;12:e82674.

69. Thakral A, Lee JJW, Hou T, Hueniken K, Dudding T, Gormley M, et al. Smoking and alcohol by HPV status in head and neck cancer: a Mendelian randomization study. Nature Communications. 2024;15(1):7835.

70. Kachuri L, Saarela O, Bojesen SE, Davey Smith G, Liu G, Landi MT, et al. Mendelian Randomization and mediation analysis of leukocyte telomere length and risk of lung and head and neck cancers. Int J Epidemiol. 2019;48(3):751-66.

## Declarations

### Ethics approval and consent to participate

The VOYAGER project was approved by the International Association for Research on Cancer (IARC) Ethics Committee (IEC) (Project No.: 16 – 34). The five parent studies were all approved by the respective institutional review boards (IRBs) at each of the participating center: University of North Carolina at Chapel Hill (Study No.: 01-0390), University of Pittsburgh (STUDY19120160), The University Health Network (Project No.: 07-0521), Sinai Health System (REB No.: 08-0191) and the National Health Service, Health Research Authority (Project ID: 24028).

### Consent for publication

Not applicable.

### Availability of data and materials

Information on the VOYAGER Consortium can be found at https://voyager.iarc.who.int/. The contact page can be used for data and collaboration requests. Non-commercial research projects are generally approved if the proposed research complies with the signed agreements between studies and their research participants.

Genotype data for the oral and pharynx cancer OncoArray study have been deposited at the database of Genotypes and Phenotypes (dbGaP) and are available under controlled access under accession phs001202.v1.p1. Genotype data for the All of Us study are also available via dbGaP under controlled access under accession phs003225.v1.p1.

The oral and pharyngeal GWAS summary statistics by cancer site and world region have been deposited in the IEU Open GWAS platform (https://gwas.mrcieu.ac.uk/) under the

GWAS IDs: ieu-b-89, ieu-b-90, ieu-b-94, ieu-b-96, ieu-b-93, ieu-b-97, ieu-b-91, ieu-b-95 and ieu-b-98.

**Competing interests**

Scott Bratman reports grants from AstraZeneca, personal fees and equity from Adela, patents licensed to Adela and Roche, and service on advisory board for EMD Serono. Where members are identified as personnel of the International Agency for Research on Cancer/ World Health Organization, the authors alone are responsible for the views expressed in this article and they do not necessarily represent the decisions, policy or views of the International Agency for Research on Cancer / World Health Organization.

**Funding**

VOYAGER was funded by US National Institute of Dental and Craniofacial Research (NIDCR) grant R01DE025712 (PIs: Brennan, Diergaarde, and Hayes: The role of germline and somatic DNA mutations in oral and oropharyngeal cancers). Genotyping using the OncoArray and the All of Us array was performed at the Center for Inherited Disease (CIDR) and funded by NIDCR 1X01HG007780-0 and NIDCR/NCI X01HG010743, respectively. The Alcohol-Related Cancers and Genetic Susceptibility Study in Europe (ARCAGE) was funded by the European Commission's fifth frame-work program (QLK1-2001-00182), the Italian Association for Cancer Research, Compagnia di San Paolo/FIRMS, Region Piemonte and Padova University (CPDA057222). The Carolina Head and Neck Cancer Epidemiology (CHANCE) study was supported in part by the National Cancer Institute (R01CA90731). The Head and Neck 5000 study was a component of independent research funded by the National Institute for Health Research (NIHR) under its Programme Grants for Applied Research scheme (RP-PG-0707-10034). The views expressed in this publication are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health. Core funding was also provided through awards from Above and Beyond, University Hospitals Bristol and Weston Research Capability Funding and the NIHR Senior Investigator

award to Professor Andy Ness. Human papilloma virus (HPV) serology was supported by a Cancer Research UK Programme Grant, the Integrative Cancer Epidemiology Programme (grant number: C18281/A19169). The University of Pittsburgh head and neck cancer case-control study was supported by US National Institutes of Health grants P50CA097190 and P30CA047904. The MSH-PMH study was supported by the Canadian Cancer Society Research Institute and the Princess Margaret Head & Neck Translational Research Program, with philanthropic funds from Joe's Team and the Wharton, Elia, Riley, and Tozer families. G.L. is funded by the Alan B. Brown Chair in Molecular Genomics and the Lusi Wong Foundation Fund.

## Author contributions

Analyses for this manuscript were conducted by M.G. and S.V. The manuscript was drafted by M.G., S.V. and A.A. All authors contributed to the interpretation of the results and critical revision of the manuscript. All authors read and approved the final manuscript.

## Acknowledgements

The authors would like to thank all study participants.

## MAIN TABLES

**Table 1.** Clinical and demographic characteristics of the oral and oropharyngeal cancer cases and controls included in VOYAGER, stratified by subsite and HPV status

| | | Total OPC N= 3,514 | HPV(+) OPC N= 2,138 | HPV(-) OPC N= 1,146 | HPV(+) *vs.* HPV(-) OPC P-value | OC N= 2,975 | Total OPC *vs.* OC P-value | Total OPC + OC cases N= 6,489 | Controls N= 3,297 | Total OPC + OC cases *vs.* controls P-value |
|---|---|---|---|---|---|---|---|---|---|---|
| **Sex (%)** | | | | | | | | | | |
| | Female | 705 (20) | 370 (17) | 285 (25) | | 1,071 (36) | | 1,776 (27) | 1,213 (37) | |
| | Male | 2,809 (80) | 1,768 (83) | 861 (75) | <0.0001 | 1,903 (64) | <0.0001 | 4,712 (73) | 2,084 (63) | <0.0001 |
| | Unknown | 0 | 0 | 0 | | 1 | | 1 | 0 | |
| **Age at diagnosis (in yrs)** | | | | | | | | | | |
| | Mean (SD) | 59 (9.6) | 58 (8.9) | 60 (9.8) | | 61 (12.2) | | 60 (10.8) | 59 (11.9) | |
| | [min, max] | [33, 84] | [26, 85] | [21, 94] | <0.0001 | [18, 94] | <0.0001 | [18, 94] | [17, 91] | 0.308 |
| **Ethnicity (%)** | | | | | | | | | | |
| | White | 2,947 (94) | 1,963 (97) | 828 (88) | | 2,271 (91) | | 5,218 (93) | 2,374 (91) | |
| | Black | 93 (3) | 14 (<1) | 68 (7) | | 103 (4) | | 196 (3) | 183 (7) | |
| | Asian | 53 (1) | 20 (1) | 31 (3) | | 89 (4) | | 142 (3) | 4 (<1) | |
| | Latino/ Hispanic | 7 (<1) | 5 (<1) | 2 (<1) | | 2 (<1) | | 9 (<1) | 0 (0) | |
| | American Indian | 2 (<1) | 1 (<1) | 1 (<1) | | 1 (<1) | | 3 (<1) | 2 (<1) | |
| | Other | 28 (1) | 14 (<1) | 11 (<1) | <0.001 | 16 (<1) | <0.0001 | 44 (<1) | 6 (<1) | <0.0001 |
| | Unknown | 384 | 121 | 205 | | 493 | | 877 | 728 | |
| **Geographic region (%)** | | | | | | | | | | |
| | Canada | 903 (26) | 542 (25) | 263 (23) | | 567 (19) | | 1,470 (23) | 975 (30) | |
| | USA | 735 (21) | 436 (21) | 222 (19) | | 841 (28) | | 1,576 (24) | 1,596 (48) | |
| | Europe | 1,876 (53) | 1,160 (54) | 661 (58) | 0.156 | 1,567 (53) | <0.0001 | 3,443 (53) | 726 (22) | <0.0001 |
| **Education (%)** | | | | | | | | | | |
| | None/ some school | 880 (30) | 472 (26) | 379 (42) | | 847 (34) | | 1,727 (32) | 368 (12) | |
| | High school | 925 (32) | 566 (32) | 275 (30) | | 827 (33) | | 1,752 (32) | 753 (23) | |
| | Postsecondary | 1,081 (38) | 755 (42) | 260 (28) | <0.0001 | 834 (33) | 0.003 | 1,915 (36) | 2,114 (65) | <0.0001 |
| | Unknown | 628 | 345 | 232 | | 467 | | 1,095 | 62 | |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| **Stage (AJCC 7th ed.) (%)** | I | 165 (5) | 45 (2) | 112 (10) | | 800 (28) | | 965 (15) | NA | NA |
| | II | 292 (8) | 115 (6) | 159 (14) | | 636 (22) | | 928 (15) | | |
| | III | 504 (15) | 280 (13) | 191 (17) | | 326 (11) | | 830 (13) | | |
| | IV | 2,481 (72) | 1,675 (79) | 652 (59) | <0.0001 | 1,110 (39) | <0.0001 | 3,591 (57) | | |
| | Unknown | 72 | 23 | 32 | | 103 | | 175 | | |
| **HPV status (%)** | | | | | | | | | | |
| | Negative | 1,146 (35) | 0 (0) | 1,146 (100) | | 1,702 (96) | | 2,848 (56) | NA | NA |
| | Positive | 2,138 (65) | 2,138 (100) | 0 (0) | <0.0001 | 66 (4) | <0.0001 | 2,204 (44) | | |
| | Unknown | 230 | 0 (0) | 0 (0) | | 1,207 | | 1,437 | | |
| **Treatment (%)** | | | | | | | | | | |
| | Radio + chemo | 1,716 (49) | 1,155 (55) | 462 (40) | | 168 (6) | | 1,884 (30) | NA | NA |
| | Surgery | 209 (6) | 64 (3) | 124 (11) | | 1,148 (47) | | 1,357 (22) | | |
| | Surgery + radio + chemo | 562 (16) | 401 (18) | 130 (11) | | 449 (15) | | 1,011 (16) | | |
| | Surgery + radio | 356 (10) | 195 (9) | 142 (12) | | 742 (25) | | 1,098 (18) | | |
| | Radiotherapy | 594 (17) | 298 (14) | 250 (22) | | 116 (4) | | 710 (11) | | |
| | Palliative care | 28 (<1) | 10 (<1) | 12 (1) | | 12 (<1) | | 40 (<1) | | |
| | Chemotherapy | 12 (<1) | 3 (<1) | 8 (<1) | | 6 (<1) | | 18 (<1) | | |
| | No treatment | 37 (<1) | 12 (<1) | 18 (2) | <0.0001 | 64 (2) | <0.0001 | 101 (2) | | |
| **Vital status (%)** | | | | | | | | | | |
| | Alive | 2,407 (68) | 1,711 (80) | 569 (50) | | 1,677 (56) | | 4,084 (63) | NA | NA |
| | Dead | 1,107 (32) | 427 (20) | 577 (50) | <0.0001 | 1,298 (44) | <0.0001 | 2,405 (37) | | |
| **Cause of death (%)** | HNC | 620 (60) | 239 (60) | 325 (61) | | 711 (59) | | 1,331 (60) | NA | NA |
| | Other cancer | 111 (11) | 33 (8) | 65 (12) | | 130 (11) | | 241 (11) | | |
| | Other disease | 296 (29) | 129 (32) | 143 (27) | 0.057 | 359 (30) | 0.843 | 655 (29) | | |
| | Unknown | 2,487 | 1,737 | 613 | | 1,775 | | 4,262 | | |

Key: HPV, human papilloma virus; OPC, oropharyngeal cancer; OC, oral cancer; radio, radiotherapy; chemo, chemotherapy.

HPV16 E6 serology was prioritized as a marker of HPV(+) oropharyngeal cancer. When serology was missing, two concordant tumor markers, p16 immunohistochemistry (IHC) and high-risk HPV DNA in-situ hybridization (ISH) were required to determine HPV status.

**Table 2.** Information on established risk factor behavior in VOYAGER, stratified by subsite and HPV status

| | | Total OPC | HPV(+) OPC | HPV(-) OPC | HPV(+) *vs.* HPV(-) OPC | OC | Total OPC *vs.* OC | Total OPC + OC cases | Controls | Total OPC + OC cases vs. controls |
|---|---|---|---|---|---|---|---|---|---|---|
| | | N= 3,514 | N= 2,138 | N= 1,146 | P-value | N= 2,975 | P-value | N= 6,489 | N= 3,297 | P-value |
| **Smoking status (%)** | Current | 862 (28) | 277 (15) | 498 (51) | | 988 (37) | | 1,850 (32) | 545 (16) | |
| | Former | 1,425 (46) | 949 (51) | 381 (39) | | 1,001 (38) | | 2,426 (42) | 1,298 (40) | |
| | Never | 787 (26) | 638 (34) | 103 (10) | <0.0001 | 671 (25) | <0.0001 | 1,458 (26) | 1,437 (44) | <0.0001 |
| | Unknown | 440 | 274 | 164 | | 315 | | 755 | 17 | |
| **Alcohol drinking status (%)** | Current | 1,984 (64) | 1,212 (65) | 643 (65) | | 1,618 (61) | | 3,602 (62) | 1,888 (58) | |
| | Former | 463 (15) | 233 (12) | 180 (18) | | 387 (14) | | 850 (15) | 666 (20) | |
| | Never | 376 (12) | 230 (12) | 95 (10) | | 437 (16) | | 813 (14) | 722 (22) | |
| | Never/ Former * | 278 (9) | 211 (11) | 67 (7) | <0.0001 | 228 (9) | <0.0001 | 506 (9) | 0 (0) | <0.0001 |
| | Unknown | 413 | 252 | 161 | | 305 | | 718 | 21 | |
| **Body mass index, in kg/m$^2$ (%)** | Healthy (18.5 to <25) | 1,034 (35) | 508 (28) | 452 (47) | | 1,057 (42) | | 2,091 (38) | 433 (27) | |
| | Underweight (<18.5) | 91 (3) | 18 (<1) | 68 (7) | | 96 (4) | | 187 (3) | 12 (<1) | |
| | Overweight (25.0 to <30) | 1,080 (36) | 744 (40) | 280 (29) | | 820 (33) | | 1,900 (35) | 623 (39) | |
| | Obese (>30) | 782 (26) | 579 (31) | 159 (17) | <0.0001 | 523 (21) | <0.0001 | 1,305 (24) | 528 (33) | <0.0001 |
| | Unknown | 527 | 289 | 187 | | 479 | | 1,006 | 1,701 | |

Key: HPV, human papilloma virus; OPC, oropharyngeal cancer; OC, oral cancer.

* Never/Former category used when participants were only asked if they were a current drinker (HN5000 study). For these participants there was no information available to determine whether they were a former or a never drinker if they were not a current drinker.

HPV16 E6 serology was prioritized as a marker of HPV(+) oropharyngeal cancer**.** When serology was missing, two concordant tumor markers, p16 immunohistochemistry (IHC) and high-risk HPV DNA in-situ hybridization (ISH) were required to determine HPV status.

**Table 3.** Survival outcomes for the oral and oropharyngeal cancer cases in VOYAGER, stratified by subsite and HPV status

| | OPC HPV Unknown | HPV(+) OPC | HPV(-) OPC | OC | Total OPC + OC |
|---|---|---|---|---|---|
| **Overall survival** | | | | | |
| **N (number of cases)** | 230 | 2,138 | 1,146 | 2,975 | 6,489 |
| **Median Survival time, yrs** | 10.3 years | 14.3 years | 5.6 years | 7.8 years | 9.8 years |
| **Probability of survival at 5y (95%CI), %** | 62.8 (56.7, 69.5) | 81.1 (79.4, 82.9) | 53.1 (50.2, 56.2) | 60.7 (58.9, 62.5) | 66.1 (65.0, 67.4) |
| **Log-rank p-value** | | | <0.001 | | |
| **Progression-free survival** | | | | | |
| **N (number of cases)** | 164 | 1,731 | 894 | 2,272 | 5,061 |
| **Median progression-free survival time, yr** | 3.8 years | 10.5 years | 3.5 years | 5.1 years | 6.3 years |
| **Probability of progression-free survival at 5y (95%CI), %** | 47.2 (40.1, 55.7) | 74.9 (72.7, 77.1) | 43.7 (40.2, 47.3) | 50.5 (48.3, 52.7) | 57.4 (56.0, 58.9) |
| **Log-rank p-value** | | | <0.001 | | |
| **Disease-specific survival*** | | | | | |
| **N (number of cases)** | 220 | 2,112 | 1,102 | 2,877 | 6,311 |
| **Disease-specific survival at 5y (95%CI), %** | 72.7 (66.8, 79.2) | 88.4 (87.0, 89.8) | 67.5 (64.6, 70.6) | 74.6 (72.9, 76.3) | 78.1 (77.0, 79.2) |
| **Gray's p-value** | | | <0.001 | | |

Key: HPV, human papilloma virus; OPC, oropharyngeal cancer; OC, oral cancer; Overall, including oral and oropharyngeal cancers.

Progression-free survival data was not available for the CHANCE study.
*Median not reached for disease-specific survival.

HPV16 E6 serology was prioritized as a marker of HPV(+) oropharyngeal cancer. When serology was missing, two concordant tumor markers, p16 immunohistochemistry (IHC) and high-risk HPV DNA in-situ hybridization (ISH) were required to determine HPV status.

## MAIN FIGURES

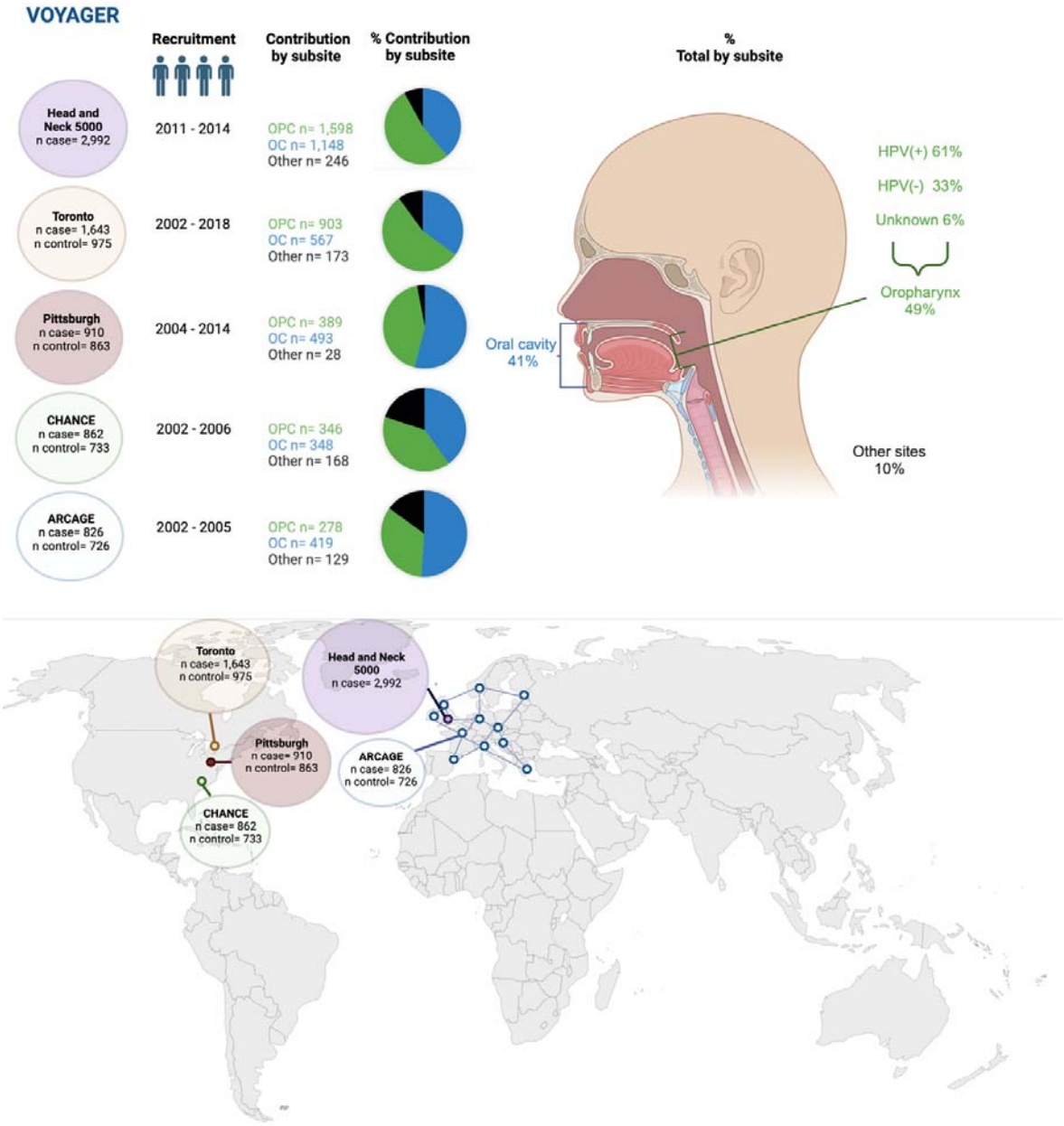


**Figure 1.** Overview of the five studies included in VOYAGER.

Key: HPV, human papilloma virus; OPC, oropharyngeal cancer (green); OC, oral cancer (blue); Other sites (black).
Created in BioRender. Gormley, M. (2025) https://BioRender.com/j07p735

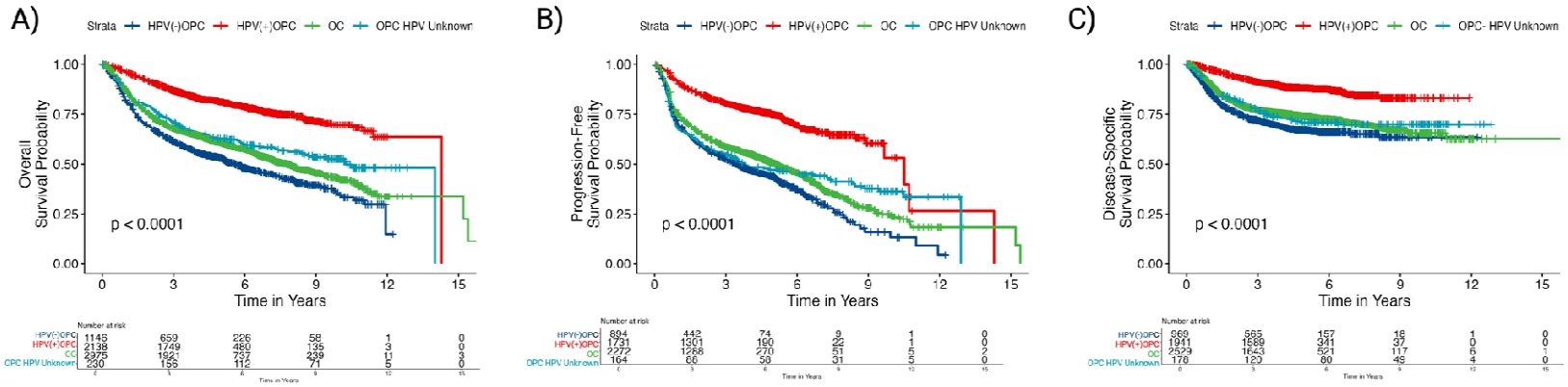


**Figure 2.** a) Overall survival, b) progression-free survival, and c) disease-specific survival for oral and oropharyngeal cancer cases in VOYAGER.

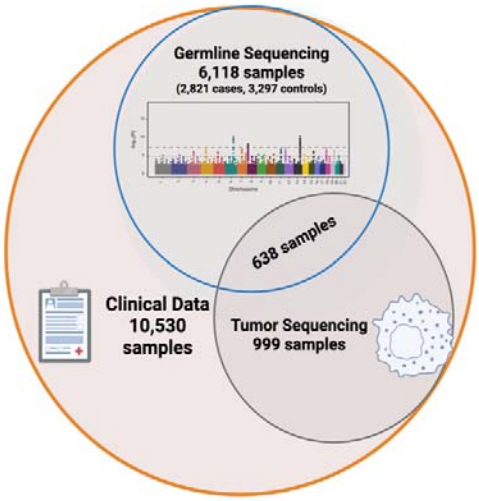


**Figure 3.** Overlapping data available in VOYAGER. Genotyping data includes cases and controls, tumor sequencing samples are for oropharyngeal and oral cavity cases only and clinical data includes all available cases and controls

Created in BioRender. (2025) https://BioRender.com/h48b673

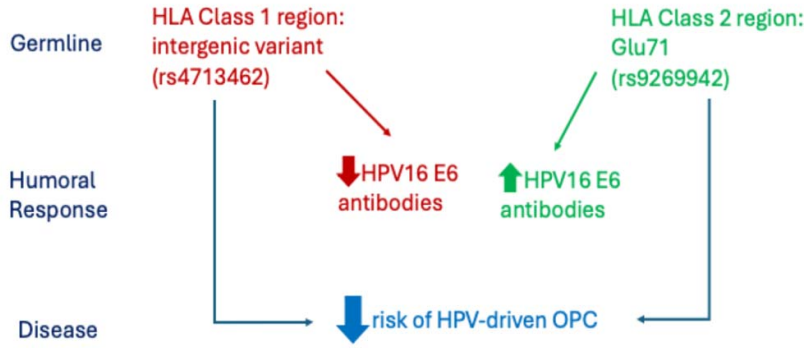


**Figure 4.** Viral host interactions suggest HLA loci that are specific for HPV16 viral proteins