Oral Oncology 49 (2013) 525–533

Contents lists available at SciVerse ScienceDirect

Oral Oncology

journal homepage: www.elsevier.com/locate/oraloncology

Oral cavity tumors in younger patients show a poor prognosis and do not contain viral RNA

J. Brägelmann [a,b], I. Dagogo-Jack [a], M. El Dinali [a], T. Stricker [j], C.D. Brown [k], Z. Zuo [a], A. Khattri [a], M. Keck [a], M.E. McNerney [c,d], R. Longnecker [g], K. Bieging [g], M. Kocherginsky [h], K. Alexander [f,i], R. Salgia [a,i], M.W. Lingen [d,i], E.E. Vokes [a,i], K.P. White [c,e,i], E.E.W. Cohen [a,i,*], T.Y. Seiwert [a,i,*]

[a] Department of Medicine, Section of Hematology/Oncology, The University of Chicago, Chicago, IL, USA
[b] Rheinische Friedrich-Wilhelms University Bonn, Bonn, Germany
[c] Institute for Genomics and Systems Biology, The University of Chicago, Chicago, IL, USA
[d] Department of Pathology, The University of Chicago, USA
[e] Argonne National Laboratories, Argonne, IL, USA
[f] Department of Pediatrics, Section of Pediatric Infectious Disease, The University of Chicago, Chicago, IL, USA
[g] Department of Microbiology and Immunology, Feinberg School of Medicine Northwestern University, Chicago, IL, USA
[h] Department of Medicine, Health Studies, The University of Chicago, Chicago, IL, USA
[i] The University of Chicago Cancer Research Center, Chicago, IL, USA
[j] Department of Pathology, Microbiology and Immunology, Vanderbilt University, Nashville, TN, USA
[k] Department of Genetics, University of Pennsylvania Philadelphia, PA, USA

ARTICLE INFO

Article history:
Received 26 December 2012
Accepted 6 February 2013
Available online 13 March 2013

Keywords:
Oral cavity squamous cell carcinoma
RNA-Seq
Oncovirus

SUMMARY

Background: Oral cavity and in particular oral tongue cancers occur with a rising incidence in younger patients often lacking the typical risk factors of tobacco use, alcohol use, and human papilloma virus (HPV) infection. Their prognosis when treated with chemoradiation has not been well studied and responsible risk factors remain elusive. A viral etiology (other than HPV) has been hypothesized.
Methods: First we analyzed outcomes from 748 head and neck cancer patients with locoregionally advanced stage tumors treated with curative-intent chemoradiation by anatomic site. Second, we analyzed seven oral tongue (OT) tumors from young, non-smokers/non-drinkers for the presence of viral mRNA using short-read massively-parallel sequencing (RNA-Seq) in combination with a newly-developed digital subtraction method followed by viral screening and discovery algorithms. For positive controls we used an HPV16-positive HNC cell line, a cervical cancer, and an EBV-LMP2A transgene lymphoma.
Results: Younger patients with oral cavity tumors had worse outcomes compared to non-oral cavity patients. Surprisingly none of the seven oral tongue cancers showed significant presence of viral transcripts. In positive controls the expected viral material was identified.
Conclusion: Oral cavity tumors in younger patients have a poor prognosis and do not appear to be caused by a transcriptionally active oncovirus.

Introduction

Head and Neck Cancer (HNC) is the 6th most common cancer worldwide.[1,2] Tobacco and alcohol use are the primary etiologic factors of HNC and account for the majority of head and neck tumors.[2,3] Recently oropharyngeal cancer has been linked to infection with human papilloma virus (HPV). The incidence of these virus-associated cancers has been rising rapidly, suggesting increased HPV exposure and infection rates in the past two decades.[3–5] HPV-associated HNC typically occur in non-smokers and non-drinkers, and are limited to the oropharynx.[2,4,6]

Unlike oropharyngeal cancers, tumors of the oral cavity, and in particular oral tongue tumors (which comprise a major subgroup of oral cavity tumors), rarely harbor HPV,[7–10] even though p16 expression is frequently present.[7,10] In addition, a significant proportion of oral tongue tumors occur in younger patients who lack classic HNC risk factors such as tobacco and/or alcohol use.[11] These tumors tend to be clinically and histologically distinct from other head and neck cancers,[12,13] and are reported to have an aggressive clinical phenotype often requiring intensive multimodality treatments.[9,14] Nevertheless the clinical outcome of patients with oral

* Corresponding authors. Address: The University of Chicago, 5841 S. Maryland Ave., MC2115, Chicago, IL 60637, USA. Tel.: +1 773 702 2452; fax: +1 773 702 3163.
E-mail addresses: ecohen@medicine.bsd.uchicago.edu (E.E.W. Cohen), tseiwert@medicine.bsd.uchicago.edu (T.Y. Seiwert).

1368-8375/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.oraloncology.2013.02.003

cavity tumors, in particular those occurring in younger patients, compared to other head and neck tumors is not well studied.

An increasing overall incidence of oral tongue/oral cavity cancers, paralleling the trend for oropharyngeal tumors, has been noted.[14] Furthermore oral cavity squamous cell carcinomas (OCSCC) frequently occur in transplant recipients.[15,16] Due to these reasons it has been hypothesized that such tumors, located at a site of primary host-environment interaction with contact to foreign material (including bodily fluids) may be associated with an infectious agent such as a virus.[5] Human papilloma virus-associated cancers are uncommon in the oral cavity and occur at a rate of 5.9%.[7] Instead infectious agents such as Herpesviruses have been postulated as contributors to oral tongue tumors, but this association remains unproven.[9]

All known oncoviruses are associated with expression of viral proteins.[17] The detection of viral transcripts or proteins is a primary mode of detection of oncoviruses in tumors and may also be used to distinguish an active viral influence on tumor maintenance from incidental viral infection.[18–20] It is therefore essential to evaluate mRNA transcripts rather than just viral DNA.[21,22]

Until recently, screening tissues for potential viruses required candidate approaches.[23–26] With the advent of massively parallel sequencing technology, unbiased discovery of new, unknown viruses has become possible.[27] To handle the large amounts of data, cloud computing-based methods have become available, but require large and expensive resources.[28]

The primary goal of our study was to investigate differences in clinical outcome between tumors of the oral cavity compared to tumors in other sites of the head and neck region, and to discover a potential viral etiology of oral tongue tumors. To achieve this goal we screened seven oral tongue tumors for known and unknown viruses using short-read massively parallel sequencing. In addition, we also wanted to establish the bioinformatic tools necessary to detect unknown viral transcripts. By using tools readily available in a standard laboratory we obviate the need for cloud computing facilities or expensive multiprocessor workstations.[28]

In this report we demonstrated significantly decreased survival rates for young patients (<45 years) with oral cavity tumors compared to patients with tumors at other locations in a cohort of 748 HNC patients. To our knowledge, this is the largest cohort of HNC patients on which a study like this has been done. In addition we screened seven oral tongue tumors from non-smokers/non-drinkers and three virus-positive controls using a novel two-algorithm pipeline on a desktop computer. Using next-generation RNA-Seq data, we demonstrate that oral tongue cancers from non-smokers/non-drinkers do not show evidence of production of viral transcripts, making a viral etiology for the tumors studied unlikely.

Methods

Survival analysis

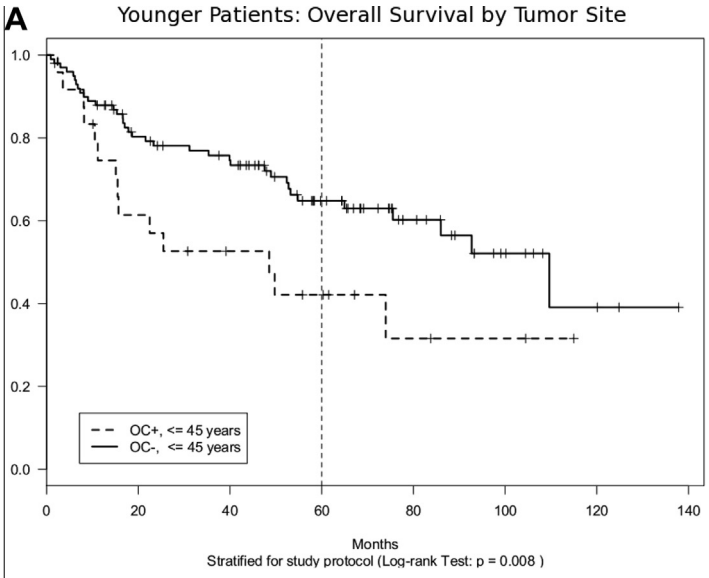
Seven hundred and forty eight patients with advanced head and neck squamous cell carcinoma (HNSCC) were identified from the University of Chicago Departments of Otolaryngology, Medical Oncology, and Radiation Oncology protocols spanning from 1990 to 2007 (Supplementary Table 1). The presence of HNSCC was confirmed in all cases by review of histology records. Tumor stages were comparable among subgroups (Supplementary Fig. 1). Disease free and overall survival times were calculated as the difference between the date on which the patient joined the study and the last date of follow-up or date of death. Survival times were estimated using the method of Kaplan–Meier, and groups were compared using the log-rank test stratified by study protocol and treatment arm. The influence of covariates (anatomical location, tumor stage, tobacco/alcohol use, age) on survival was analyzed using Cox regression analyses.

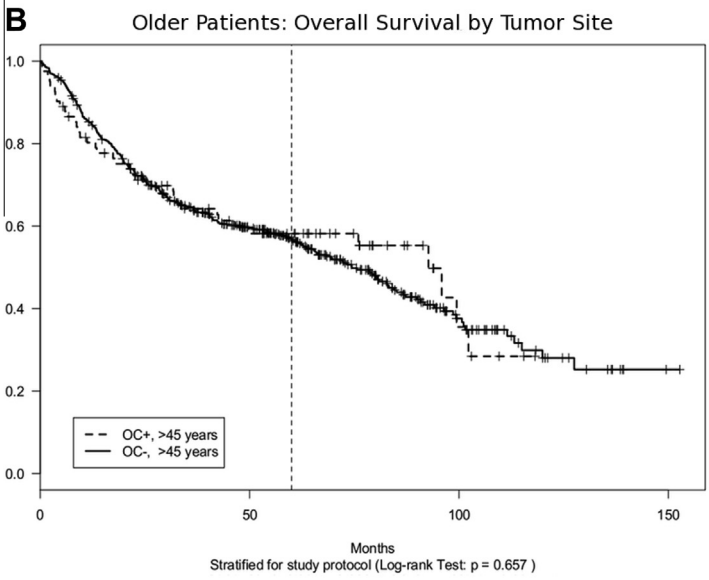
Samples

Seven frozen oral tongue tumors were selected from the University of Chicago Head and Neck Cancer tissue bank (IRB approved protocol UCCCC#8980) based on age, low or absent use of alcohol (social: $\leqslant$one glass of wine or equivalent per day) and non-use of tobacco (Table 1). Areas of tumor involvement were marked and a punch biopsy performed to achieve tumor content of $\geqslant$70%. Known HPV16(+), virus genome integrated UM-SCC47 cells were obtained from Dr. Thomas Carey (University of Michigan) and harvested at 70–80% confluence. A mouse lymphoma containing the EBV LMP2A transgene was used as a second control.[29] As a third control, publicly available RNA-Seq data of an HPV16-positive cervical carcinoma[18] were downloaded from the NCBI Sequence Read Archive (SRA, http://www.ncbi.nlm.nih.gov/Traces/sra). All tumors were screened for HPV status using an established nested PCR method.[30]

Table 1
Sample overview and clinical characteristics of sequenced oral tongue cancers. CRT = concurrent chemoradiation, S = surgery, RT = Radiation, OT = Oral tongue tumor sample.

	OT-1	OT-2	OT-3	OT-4	OT-5	OT-6	OT-7	UM-SCC47	EBV-LMP2A transgene lymphoma	Cervical cancer
Source	Patient	Patient	Patient	Patient	Patient	Patient	Patient	Cell line derived from patient	EBV LMP2A transgenic lymphoma	Literature derived
Gender	Male	Female	Female	Male	Male	Male	Male	Male	n/a	Female
Age	42	44	33	21	33	44	42	n/a	n/a	n/a
Tobacco	No	Minimal (<1 pack year in total)	Minimal (<5 pack years in total)	No	No	No	No	n/a	n/a	n/a
Alcohol	Minimal/social	Minimal/social	Minimal/social	Minimal/social	Minimal/social	No	Minimal/social	n/a	n/a	n/a
Tumor Stage	T4N0	Initially T1N0 with several recurrences	Recurrence post surgery and CRT for T3N0 tumor	T2N0	T4N2B	T2N0	Recurrence post surgery for T1N0 tumor	n/a	n/a	n/a
Treatment	CRT	Surgery, then CRT, then palliative chemotherapy	Surgery, RT, CRT, palliative chemotherapy	Surgery	CRT	Surgery, CRT	Surgery, induction chemotherapy, CRT	n/a	n/a	n/a
Outcome	Cancer free	Undergoing treatment for metastatic disease	Deceased from tumor	Cancer free	Cancer free	Cancer free	Cancer free	n/a	n/a	n/a





C

Multivariate Cox Regression Analysis

Variable	HR	95% CI	P value[1]
OSCC status (OC+ vs OC-)	0.92	(0.64, 1.32)	0.65
Age (≤45 years vs. >45years)	0.86	(0.64, 1.23)	0.42
Alcohol Use (none, low, moderate, heavy)	1.12	(1.01, 1.24)	**0.037**
Tobacco Use (Never, <20py, 20-40py, >40py)	1.32	(1.19, 1.45)	**8.8e-08**
Interaction: Young patient * OC+	2.47	(1.19, 5.10)	**0.015**
Tumor stage (II, III, IV)	1.31	(0.89, 1.94)	0.17

HR=Hazard Ratio, CI=Confidence Interval
[1] P values from Wald test of multivariate Cox regression model after adjusting for covariates

Figure 1 Survival analysis on the clinical data of 748 HNC patients estimated using Kaplan–Meier analysis. (A) Analysis for patients < 45 years of age. Patients with tumors of the oral cavity (OC+, dashed line) show a significantly decreased overall survival rate (Log-rank $p = 0.008$) when compared to younger patients with tumors at other locations of the HN region (OC-, solid line). This indicates the more aggressive phenotype of OC tumors in young patients. (B) No such difference can be observed for patients > 45 years of age (Log-rank $p = 0.66$). (C) A Cox proportional hazards model shows that significant risk factors influencing survival are tobacco/alcohol use and the interaction of young age and oral cavity tumors (py = pack years, OC = oral cavity).

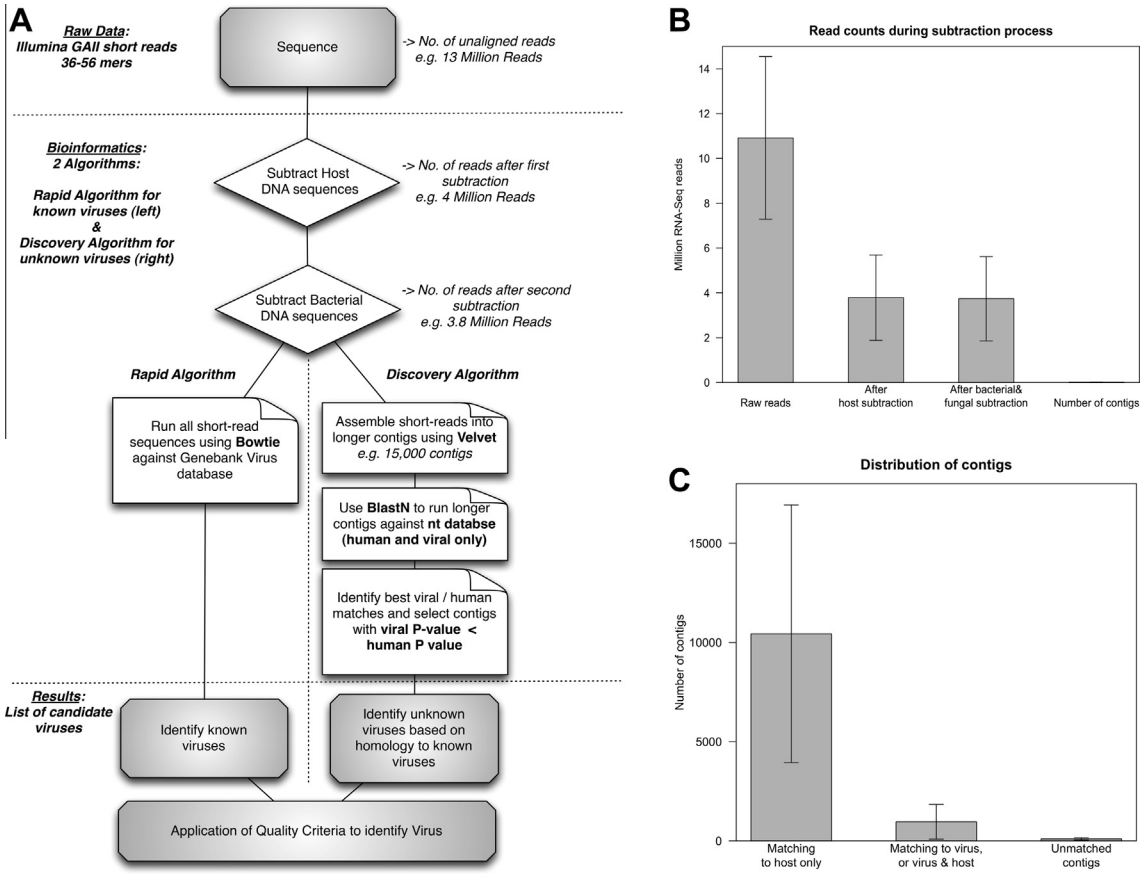


Figure 2 (A) Flowchart visualizing the different steps during the rapid and the discovery algorithm. (B) The amount of data is efficiently reduced by computational subtraction of host and non-viral microbial reads. (C) A further reduction is achieved by assembling the remaining reads into longer sequences (contigs) of which ~90% are of host origin. Only a minor fraction of contigs matches to a virus or to both a viral and a host sequence and will be evaluated subsequently. (Height of the bars correspond to the mean of all samples, error bars indicate standard deviation.)

RNA-Seq library preparation

RNA was extracted using Qiagen RNeasy columns (Qiagen, CA) and polyA purified (Ambion Micro Poly(A) purist), using between 1.5 and 10 μg RNA. cDNA libraries were prepared as described in the Supplementary methods.

Sequencing

Short read sequences were collected from an Illumina Genome Analyzer II (Illumina, CA) with a standard protocol for 36 bp or 56 bp read-lengths on one lane per sample in the GAII flowcell (Illumina, CA) at Argonne National Laboratories (Bolingbrook, IL). Expression of CDKN2A-transcripts was quantified as described in the Supplementary methods.

Virus algorithms

An overview of both algorithms is provided in Fig. 2A. A detailed explanation of the algorithms, the databases used, and the criteria to accept positive viral calls can be found in the Supplementary methods and Supplementary Tables 2 and 3.

All analyses were run on a desktop computer (Apple iMAC, 2.4 Ghz, 8 Gb RAM, Apple, CA).

In-silico simulation

To evaluate the ability to detect unknown viruses, HPV16 and EBV sequences were deleted from all databases, and both virus algorithms were repeated using the UM-SCC47/EBV-LMP2A-transgene lymphoma short read sequences.

Results

Survival analysis

Survival of 748 HNC patients treated at the University of Chicago with chemoradiotherapy was analyzed. A lower rate of survival was evident in younger patients with oral cavity tumors compared to those with non-oral cavity tumors: among the young patients (Age < 45 years, $n = 135$), patients with primary neoplasms of the oral cavity ($n = 24$) had 5-year overall survival of 42%, compared to 65% 5-year survival in 111 patients with non-oral cavity neoplasms [$p = 0.008$, HR = 3.25, (95%CI: 1.50–7.05.)] (Fig. 1A). Progression free survival was also different with 35.7% of oral cavity patients versus 64% in non-oral cavity patients progression-free at 5 years [$p = 0.002$; HR = 2.96, (95%CI: 1.42–6.17)]. This difference is by and large not driven by good prognosis oropharynx tumors (majority HPV(+) based on prior data[31–33]), with young oral cavity tumor patients showing worse overall survival than both oropharynx, and non-oropharynx/non-oral cavity tumors ($p = 0.023$, log-rank test comparing three groups; Supplementary Fig. 2). Within the cohort of young patients with oral cavity tumors there was no difference in clinical outcome between oral tongue (50% of young oral cavity patients, Supplementary Fig. 3) and all other oral cavity primary sites ($p = 0.70$, Supplementary Fig. 4). In the age group of patients >45 years there was no significant difference in overall survival for oral cavity vs.

non-oral cavity tumors [Fig. 1B, $p = 0.66$, HR = 1.08 (95%CI 0.76–1.55)]. A multivariate Cox regression analysis on data of all patients (Fig. 1C) showed a significant increase in risk for young patients with oral cavity cancer ($p = 0.015$, HR = 2.47).

This difference in survival underlines the more aggressive clinical phenotype of tumors originating in the oral cavity and the oral tongue in younger patients compared to HNSCC at other locations. To date no known risk factor has been shown to contribute for this difference, making additional influences such as an infective agent more likely.

Oral tongue cancer samples

Seven oral tongue cancers from young patients (average age 37 years) with no or minimal tobacco and alcohol use were selected from the University of Chicago tissue bank (Table 1). Patients were treated depending on clinical stage with surgery, chemoradiotherapy or combined modality approaches. Five patients were disease free at the time of manuscript preparation, one patient was deceased, and one patient was undergoing palliative treatment for metastatic disease. Of note, three patients experienced recurrences after initial treatment despite adequate therapy results (43%) and one patient progressed on induction chemotherapy (14%). All tumors were confirmed to be HPV-negative using an established PCR method.[30]

Algorithm results

An overview of the two-algorithm approach and its performance is provided in Fig. 2 and Supplementary Table 2. After applying positive selection criteria (Supplementary Table 3), the rapid algorithm clearly identified HPV16 and EBV in the positive control samples (Fig. 3A and Supplementary Fig. 5A).

In contrast, none of the oral tongue tumors showed significant matches to known viruses. To further investigate the reliability of the virus call criteria, and to foster the distinction of false positive calls by the rapid algorithm, the proportion of reads determined as being of viral origin divided by all reads obtained per sample was calculated (Fig. 4A). Even though some reads aligned to viral genomes for the oral tongue samples, the number of reads aligning to a virus normalized for all input reads is significantly higher for the positive controls (Kruskal–Wallis $p = 0.017$). Moreover, in the control samples a mean of 95.7% (95%CI of the true expected mean 86.4%-100%) of all viral reads aligned to a single best virus (HPV16/EBV, Fig. 4B). This shows, that true viral reads dominate the total virus count over false positive hits. Similarly high values cannot be observed for the OT samples. The amounts of viral reads in the OT samples rather resemble the amount and distribution of viral hits identified in UM-SCC47 that do not match to HPV16, and that can serve as a baseline of expected false positive hits (Figs. 3A and 4*B*).

Sequences that were aligned to viruses by the rapid algorithm, but did not fulfill the positive call criteria, were subjected to a manual BLAST search. For the majority of those sequences, either sequence homology to human genes or low complexity/repetitive reads were reported, thus supporting the notion of being false positives. The large number of viral hits that also match human sequence may relate to the homology of human and viral ORFs. Up to 8% of the human genome are reported to have originated from endogenous retroviruses[34] and 13% of herpesvirus proteins have clear sequence similarity to products of the human genome.[35]

Furthermore, using the discovery algorithm, HPV16 and EBV (HHV4) were also unequivocally identified in all positive control samples (Fig. 3B and Supplementary Fig. 5B).

Despite the presence of only one single transgenic viral protein (LMP2A) in one of the control samples, both algorithms identified the presence of EBV, suggesting a high degree of sensitivity.

To simulate the detection of yet unknown viruses related to known viruses, HPV16 and EBV sequences were deleted from the databases and the algorithm was rerun for UM-SCC47 and the EBV-LMP2A transgene lymphoma. The rapid algorithm produced identical results with the singular change of HPV16 and EBV matches missing (*not shown*). The discovery algorithm identified related E6 transcripts in UM-SCC47 derived sequences (Fig. 3C). In addition, a multitude of HPV viruses from different HPV types and species (that by definition share 60–70% sequence similarity[36])

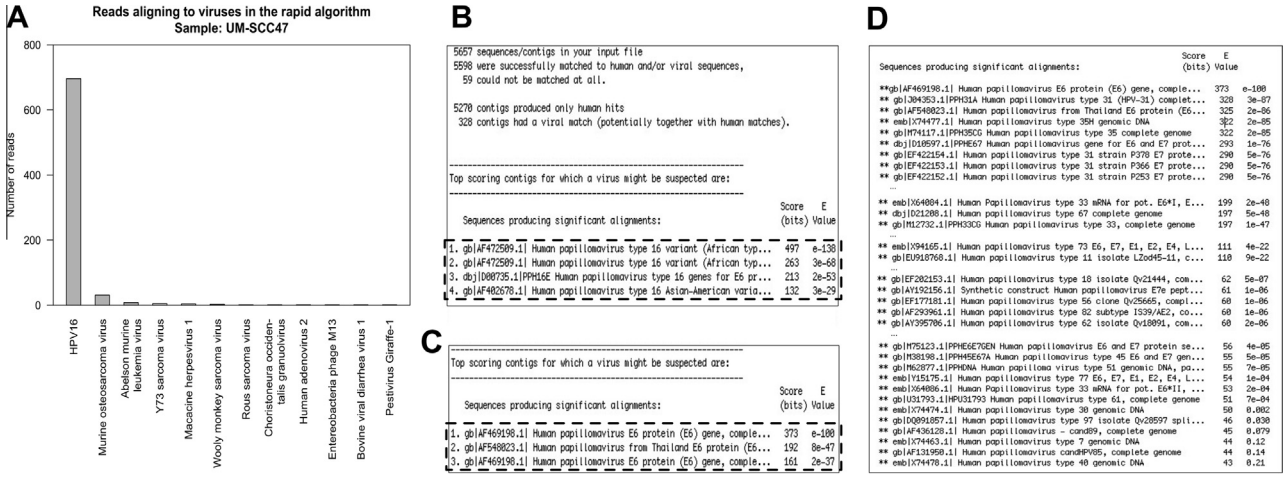


Figure 3 Algorithm results for UM-SCC47 (A) Quantification of all raw reads aligning to a virus in the rapid algorithm. (B) Summary results of the discovery algorithm matching against a subset of human and viral sequences derived from the BLAST nt database. All contigs aligning to a virus and meeting the candidate criteria (see Supplementary Table 3) are reported in a concise format to provide a prompt overview. (C) Summary results of the discovery algorithm aligning to the viral and human nt database after HPV16 was deleted from it. Less contigs matching to a virus are reported, but several related viruses are successfully identified. (D) Example of the more detailed output of the discovery algorithm using the viral and human nt database after deleting all HPV16 sequences. The list of successful alignments of the top-scoring contig is shown in Fig. 3C. Multiple HPV viruses and species are reported demonstrating the algorithm's ability to identify related viruses with a sequence homology of down to 60–70%. (**indicates a match to a virus family known to be involved in cancer; Structure of BLAST results: original source of the sequence aligned to |Accession number| Description of the organism/gene aligned to; Score (bits): score calculated by the number of concordant basepairs, substitutions and gaps in the alignment, higher scores indicate better agreement; *E*-value: likelihood that a sequence with a similar score will occur in the database by chance, the smaller the *E*-value, the more significant).

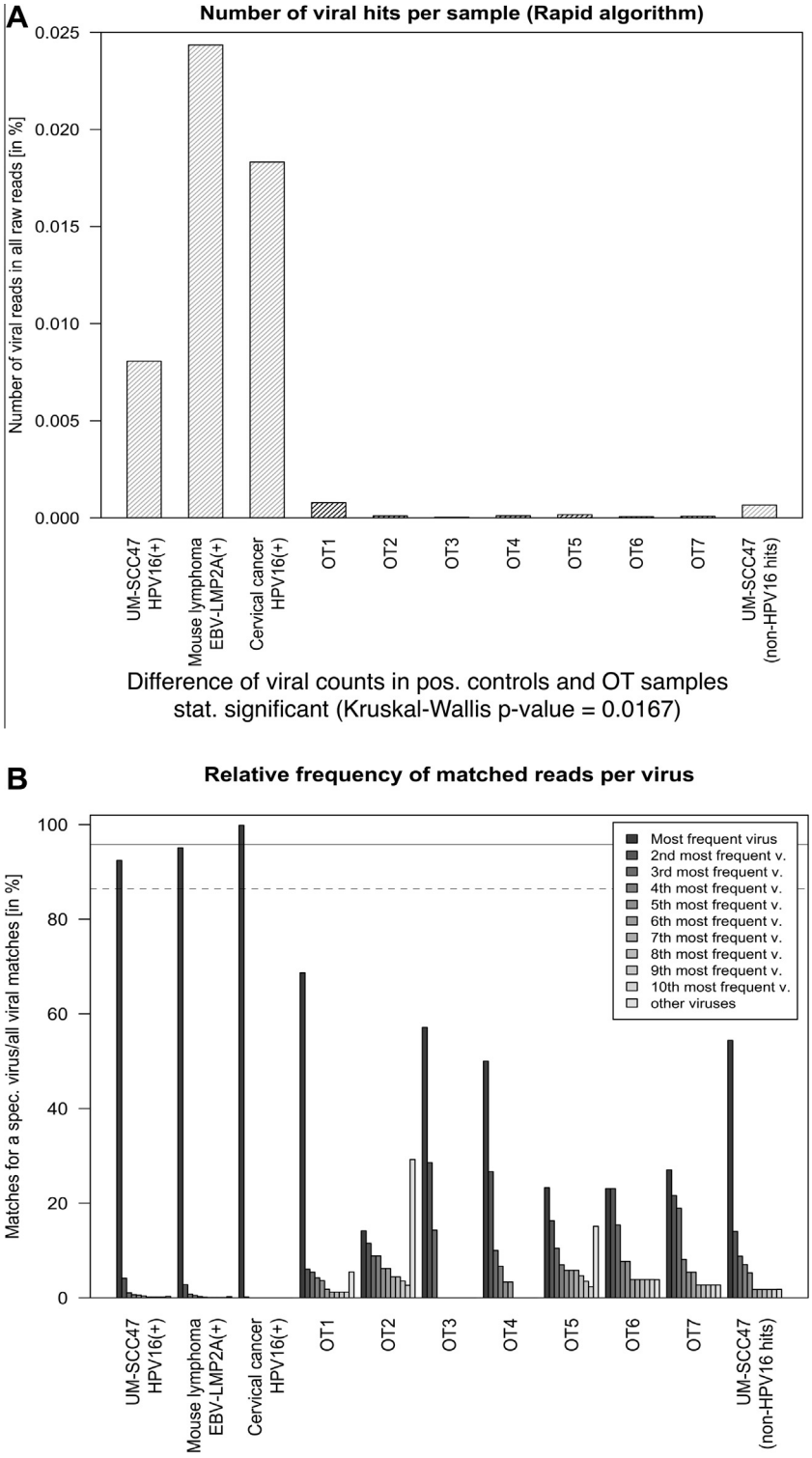


Figure 4 Quantification of viral reads (A) The positive control samples show a significantly higher number of reads aligning to a virus than the oral tongue (OT) samples when normalized for the amount of input reads (Kruskal–Wallis $p = 0.017$). The viral counts in the OT samples appear more similar to UM-SCC47 when only considering viral reads not originating from HPV16 (on the right; likely false positives). (B) In the control samples on average 95.7% of all viral hits (solid line indicates mean, dashed line the lower limit of the 95% CI of the mean at 86.4%) stem from a single virus. For the OT samples it is less distinct. Instead the viral reads are spread out among a number of different viruses similar to the viral reads not matching to HPV16 for UM-SCC47 (on the right). This provides a second indicator that viral hits observed in the OT samples likely are false positive matches.

are identified (Fig. 3D), demonstrating a high degree of sensitivity for related viruses.

For the EBV-LMP2A transgene lymphoma the discovery algorithm identified a single candidate hit (Herpesvirus papio latent membrane protein 2 (Supplementary Fig. 5C), an alpha-herpesvirus occurring in baboons[37]). Since there is only one single EBV protein in the LMP2A transgenic lymphoma, as expected fewer herpesvirus hits are identified. Other hits e.g. Kaposi's sarcoma

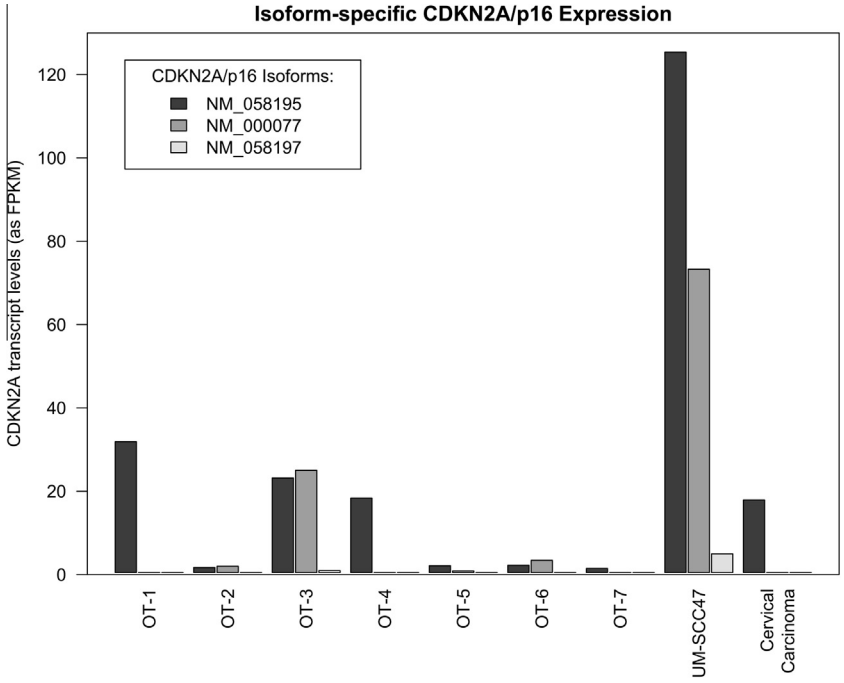


Figure 5 Isoform-specific CDKN2A (p16) expression. The expression levels of CDKN2A/p16 mRNA isoforms in the samples were quantified by RNA-Seq as CDKN2A/p16 reads normalized for transcript length and total reads per sample (FPKM, see Supplementary methods). Total CDKN2A/p16 expression was significantly greater in high-expression samples (OT-1, OT-3, OT-4) compared to the low-expression samples OT-2, OT-5, OT-6, and OT-7 ($p = 0.037$, one-sided Welch t-test). Bars represent transcript levels of different isoforms per sample, corresponding RefSeq-IDs are given in the legend.

virus (HHV8), and Herpes simplex virus 2 (HHV2) are also found, but did not meet the candidate criteria (Supplementary Fig. 5D).

In three of the seven oral tongue tumors, possible viral candidates meeting the positive virus call criteria were reported, but verification of the respective contigs by a manual web-based BLAST query readily identified them as most likely being false positive hits (Supplementary Table 4); *detailed output for each sample is available in the* Supplementary materials). None of the seven oral tongue tissue samples showed convincing evidence of viral transcripts in the rapid or the discovery algorithm underlining the absence of viral transcripts belonging to either a known virus or a novel virus related to a known virus.

p16/CDKN2A expression

p16 is commonly used as a surrogate marker for HPV in oropharyngeal tumors. Interestingly CDKN2A mRNA (p16) was significantly expressed in three of the seven samples ($p = 0.037$, one-sided Welch t-test; Fig. 5). This is consistent with recent reports suggesting that p16 is expressed in OCSCC in the absence of HPV, making it an unreliable HPV surrogate marker when evaluating oral cavity tumors.[7,10]

Discussion

In this study, we analyzed differences in survival and clinical outcome of a prospectively treated cohort of 748 HNC patients (Supplementary Table 1). We found a significantly inferior outcome for young patients (<45 years) with cancer of the oral cavity (OCSCC) compared to young patients with HNC in other anatomic locations. To our knowledge, this is the largest cohort on which a study of this kind has been done. Our study highlights the influence of tumors at different anatomical locations on the overall survival rate of young HNC patients. Our data show that there are prognostic differences between younger and older oral cavity patients, in contrast to the majority of recent reports claiming a comparable stage by stage prognosis for younger and older HNSCC patients.[38–41] To date no known risk factor could be shown to convey the decreased prognosis in young OCSCC patients. In fact young OCSCC patients have a lower exposure to known risk factors: never-smokers represent 20.7% of the young non-oral cavity patients, but 37.5% of the young oral cavity patients in our study cohort. In addition in our dataset a non-significant trend towards lower stages was present for young OCSCC patients.

The worse outcome in young OCSCC patients led us to investigate the hypothesis that tumors of the mobile tongue, which are a majority of oral cavity cancers, may have a viral etiology[9] specifically implying a so far unknown virus. We therefore developed a novel viral screening algorithm using massively parallel short-read sequencing and were able to demonstrate the absence of viral transcripts in our collection of oral tongue cancers.

Our method focuses on mRNA transcripts, an approach that has proven successful for viral detection in other tumors,[42,43] as viral protein expression is essential for viral oncogenesis in all known human oncoviruses. While we believe that our approach provides a high level of sensitivity for detection of all active oncoviruses, there are possible scenarios which would lead the algorithms to be falsely negative: (1) a virus that does not produce mRNA transcripts or does not polyadenylate them.[44] It has been hypothesized that viruses cause cancer by a "hit-and-run" mechanism[45–47] or by viral oncomodulation.[48,49] Nevertheless, there are currently no known examples of viral oncogenesis of this type in humans. (2) Oncogenesis by an unknown virus without sufficient homology to any known viruses. Such lack of homology would lead to the inability of BLAST to detect a new virus. Given the marked partial homology between distinct viral families, we think this is a remote possibility. We estimate that the discovery algorithm (based on BLAST identification of distant HPV-species, Fig. 3C/D) will identify viruses down to 60–70% sequence homology. Furthermore homol-

ogy within a small part of the viral genome would be sufficient for detection. (3) A virus may use sequences that closely resemble human sequences[34,35,37] leading the automated algorithm to discard such "human" information. This again seems unlikely since all known oncoviruses have distinct viral proteins, that are essential to proper viral function, and are therefore reliably identifiable. (4) Involvement of non-viral microbial factors (e.g. bacteria/fungi): Because there is a substantial overlap in genetic material between viruses and bacteria and the presence of bacterial/fungal DNA contamination of samples from the oral cavity subtraction of non-viral microbial sequences was necessary. (5) Finally, our small sample set of seven tumors may have led us to randomly only select non-virus associated tumors. We specifically selected representative, phenotypically ideal cases (i.e. young age, non-smokers, non-drinkers). We believe that our cohort is optimal at representing the clinical phenotype, but larger sample sets, that will undoubtedly become available in the future, can further substantiate our results.

In oropharyngeal cancer, measurement of p16 protein expression is a reliable surrogate marker for HPV infection and is used widely.[50,51] In oral cavity tumors however, two recent reports demonstrate a high false positive rate when evaluating p16 for OCSCC.[7,10] In our study a high abundance of CDKN2A/p16 mRNA-transcripts in 3 out of 7 (43%) HPV-negative OT samples is observed. Bearing in mind, that mRNA-expression levels do not perfectly reflect protein levels, this finding nonetheless validates previous reports questioning p16 as a single HPV-surrogate marker in OCSCC.

In summary, our algorithm successfully identifies viruses in three positive control tumors, even in the presence of just a single viral gene, and when simulating the discovery of a distantly related but potentially unknown virus. No viruses were found in the oral tongue cancer samples. We thereby show for the first time that oral tongue tumors in non-smokers/non-drinkers do not contain transcriptionally active oncoviruses, and are therefore unlikely to be virus-associated. As such we propose refocusing the etiological investigations of OCSCC on non-viral environmental and genetic factors. Our report also has broader implications for the field of virus discovery, making it feasible for smaller laboratories to perform such analyses using a rapid and sensitive dual algorithm for known and unknown viruses.

Funding

1. T.S. is supported by a Young Clinical Scientist Award from the Flight Attendant Medical Research Institute (FAMRI).
2. R.L. is supported by the Public Health Service grants CA73507 from the National Cancer Institute. K.B. is supported by the Carcinogenesis Training Program (T32CA009560).
3. The authors acknowledge the Univ. of Chicago CTSA pilot grant funding, as well as the Gleason family.
4. E.V. is supported by an American Society of Clinical Oncology (ASCO) Translational Research Professorship
5. M.K. is supported bya Ruth L. Kirchstein National Research Service Award Short-Term Institutional Research Training Grant (5T35DK062719-24).

Conflict of interest statement

None declared.

Acknowledgement

We would like to acknowledge Thomas Carey for his kind contribution of the UM-SCC47 cell line.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.oraloncology.2013.02.003.

References

1. Parkin DM, Bray F, Ferlay J, Pisani P. Global cancer statistics, 2002. *CA Cancer J Clin* 2005;**55**(2):74–108.
2. Chung CH, Gillison ML. Human papillomavirus in head and neck cancer: its role in pathogenesis and clinical implications. *Clin Cancer Res* 2009;**15**(22):6758–62 [November 15].
3. Haddad RI, Shin DM. Recent advances in head and neck cancer. *N Engl J Med* 2008;**359**(11):1143–54 [September 11].
4. D'Souza G, Kreimer AR, Viscidi R, Pawlita M, Fakhry C, Koch WM, et al. Case-control study of human papillomavirus and oropharyngeal cancer. *N Engl J Med* 2007;**356**(19):1944–56 [May 10].
5. Näsman A, Attner P, Hammarstedt L, Du J, Eriksson M, Giraud G, et al. Incidence of human papillomavirus (HPV) positive tonsillar carcinoma in Stockholm, Sweden: an epidemic of viral-induced carcinoma? *Int J Cancer* 2009;**125**(2):362–6 [July 15].
6. 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 [March 19].
7. Lingen MW, Xiao W, Schmidt A, Jiang B, Pickard R, Kreinbrink P, et al. Low etiologic fraction for high-risk human papillomavirus in oral cavity squamous cell carcinomas. *Oral Oncol* 2012 [July 27].
8. Siebers TJH, Merkx MAW, Slootweg PJ, Melchers WJG, Cleef P, Wilde PCM. No high-risk HPV detected in SCC of the oral tongue in the absolute absence of tobacco and alcohol—a case study of seven patients. *Oral Maxillofac Surg* 2008;**12**(4):185–8 [December 2].
9. Salem A. Dismissing links between HPV and aggressive tongue cancer in young patients. *Ann Oncol* 2010;**21**(1):13–7.
10. Kabeya M, Furuta R, Kawabata K, Takahashi S, Ishikawa Y. Prevalence of human papillomavirus in mobile tongue cancer with particular reference to young patients. *Cancer Sci* 2011;**103**(2):161–8 [December 15].
11. Liang X-H, Lewis J, Foote R, Smith D, Kademani D. Prevalence and significance of human papillomavirus in oral tongue cancer: the Mayo Clinic experience. *J Oral Maxillofac Surg* 2008;**66**(9):1875–80 [September 1].
12. Rautava J, Luukkaa M, Heikinheimo K, Alin J, Grenman R, Happonen R-P. Squamous cell carcinomas arising from different types of oral epithelia differ in their tumor and patient characteristics and survival. *Oral Oncol* 2007;**43**(9):911–9 [October 1].
13. Lingen MW, Chang KW, McMurray SJ, Solt DB, Kies MS, Mittal BB, et al. Overexpression of p53 in squamous cell carcinoma of the tongue in young patients with no known risk factors is not associated with mutations in exons 5–9. *Head Neck* 2000;**22**(4):328–35.
14. Chitapanarux I, Lorvidhaya V, Sittitrai P, Pattarasakulchai T, Tharavichitkul E, Sriuthaisiriwong P, et al. Oral cavity cancers at a young age: analysis of patient, tumor and treatment characteristics in Chiang Mai University Hospital. *Oral Oncology* 2006;**42**(1):83–8.
15. Hasegawa W, Pond GR, Rifkind JT, Messner HA, Lau A, Daly AS, et al. Long-term follow-up of secondary malignancies in adults after allogeneic bone marrow transplantation. *Bone Marrow Transplantation* 2004;**35**(1):51–5 [November 1].
16. Curtis RE, Rowlings PA, Deeg HJ, Shriner DA, Socíe G, Travis LB, et al. Solid cancers after bone marrow transplantation. *N Engl J Med* 1997;**336**(13):897–904 [March 27].
17. Poreba E, Broniarczyk JK, Gozdzicka-Jozefiak A. Epigenetic mechanisms in virus-induced tumorigenesis. *Clin Epigenet* 2011;**2**(2):233–47 [March 23].
18. Arron ST, Ruby JG, Dybbro E, Ganem D, Derisi JL. Transcriptome sequencing demonstrates that human papillomavirus is not active in cutaneous squamous cell carcinoma. *J Invest Dermatol* 2011;**131**(8):1745–53 [August 1].
19. Braakhuis BJM, Snijders PJF, Keune W-JH, Meijer CJLM, Ruijter-Schippers HJ, Leemans CR, et al. Genetic patterns in head and neck cancers that contain or lack transcriptionally active human papillomavirus. *J Natl Cancer Inst* 2004;**96**(13):998–1006 [July 7].
20. Wiest T, Schwarz E, Enders C, Flechtenmacher C, Bosch FX. Involvement of intact HPV16 E6/E7 gene expression in head and neck cancers with unaltered p53 status and perturbed pRb cell cycle control. *Oncogene* 2002;**21**(10):1510–7 [February 28].
21. Jung AC, Briolat J, Millon R, de Reyniès A, Rickman D, Thomas E, et al. Biological and clinical relevance of transcriptionally active human papillomavirus (HPV) infection in oropharynx squamous cell carcinoma. *Int J Cancer* 2010;**126**(8):1882–94 [April 15].
22. Attner P, Du J, Näsman A, Hammarstedt L, Ramqvist T, Lindholm J, et al. The role of human papillomavirus in the increased incidence of base of tongue cancer. *Int J Cancer* 2010;**126**(12):2879–84 [June 15].
23. Rota PA, Oberste MS, Monroe SS, Nix WA, Campagnoli R, Icenogle JP, et al. Characterization of a novel coronavirus associated with severe acute respiratory syndrome. *Science* 2003;**300**(5624):1394–9 [May 30].

24. Palacios G, Quan P-L, Jabado OJ, Conlan S, Hirschberg DL, Liu Y, et al. Panmicrobial oligonucleotide array for diagnosis of infectious diseases. *Emerging Infect Dis* 2007;**13**(1):73–81.
25. Wang D, Coscoy L, Zylberberg M, Avila PC, Boushey HA, Ganem D, et al. Microarray-based detection and genotyping of viral pathogens. *Proc Natl Acad Sci USA* 2002;**99**(24):15687–92 [November 26].
26. Lipkin WI. Microbe hunting. *Microbiol Molec Biol Rev* 2010;**74**(3):363–77 [August 30].
27. Feldhahn M, Menzel M, Weide B, Bauer P, Meckbach D, Garbe C, et al. No evidence of viral genomes in whole-transcriptome sequencing of three melanoma metastases. *Exp Dermatol* 2011;**20**(9):766–8 [June 14].
28. Kostic AD, Ojesina AI, Pedamallu CS, Jung J, Verhaak RGW, Getz G, et al. PathSeq: software to identify or discover microbes by deep sequencing of human tissue. *Nat Biotechnol* 2011;**29**(5):393–6 [May 1].
29. Bieging KT, Amick AC, Longnecker R. Epstein-Barr virus LMP2A bypasses p53 inactivation in a MYC model of lymphomagenesis. *Proc Natl Acad Sci USA* 2009;**106**(42):17945–50 [October 20].
30. Sotlar K, Diemer D, Dethleffs A, Hack Y, Stubner A, Vollmer N, et al. Detection and typing of human papillomavirus by e6 nested multiplex PCR. *J Clin Microbiol* 2004;**42**(7):3176–84 [July 1].
31. Seiwert TY, Haraf DJ, Cohen EE, Blair EA, Stenson K, Salama JK, et al. A randomized phase II trial of cetuximab-based induction chemotherapy followed by concurrent cetuximab, 5-FU, hydroxyurea, and hyperfractionated radiation (CetuxFHX), or cetuximab, cisplatin, and accelerated radiation with concomitant boost (CetuxPX) in patients with locoregionally advanced head and neck cancer (HNC). *J Clin Oncol* 2011 [September 27 suppl-abstr5519].
32. 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 [November 8].
33. 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 [July 1].
34. Griffiths DJ. Endogenous retroviruses in the human genome sequence. *Genome Biol* 2001;**2**(6) [REVIEWS1017].
35. Holzerlandt R, Orengo C, Kellam P, Albà MM. Identification of new herpesvirus gene homologs in the human genome. *Genome Res* 2002;**12**(11):1739–48 [November 1].
36. de Villiers E-M, Fauquet C, Broker TR, Bernard H-U. Hausen zur H. Classification of papillomaviruses. *Virology* 2004;**324**(1):17–27 [June 20].
37. Tyler SD, Severini A. The complete genome sequence of herpesvirus papio 2 (Cercopithecine herpesvirus 16) shows evidence of recombination events among various progenitor herpesviruses. *J Virol* 2006;**80**(3):1214–21 [February 1].
38. Sasaki T, Moles DR, Imai Y, Speight PM. Clinico-pathological features of squamous cell carcinoma of the oral cavity in patients. *J Oral Pathol Med* 2005;**34**(3):129–33.
39. Sargeran K, Murtomaa H, Safavi SMR, Vehkalahti MM, Teronen O. Survival after diagnosis of cancer of the oral cavity. *Br J Oral Maxillofac Surg* 2008;**46**(3):187–91.
40. Pytynia KB, Grant JR, Etzel CJ, Roberts D, Wei Q, Sturgis EM. Matched analysis of survival in patients with squamous cell carcinoma of the head and neck diagnosed before and after 40 years of age. *Arch Otolaryngol Head Neck Surg* 2004;**130**(7):869–73.
41. Veness MJ, Morgan GJ, Sathiyaseelan Y, Gebski V. Anterior tongue cancer and the incidence of cervical lymph node metastases with increasing tumour thickness: should elective treatment to the neck be standard practice in all patients? *ANZ J Surg* 2005;**75**(3):101–5.
42. Feng H, Taylor JL, Benos PV, Newton R, Waddell K, Lucas SB, et al. Human transcriptome subtraction by using short sequence tags to search for tumor viruses in conjunctival carcinoma. *J Virol* 2007;**81**(20):11332–40 [October 1].
43. Feng H, Shuda M, Chang Y, Moore PS. Clonal integration of a polyomavirus in human Merkel cell carcinoma. *Science* 2008;**319**(5866):1096–100 [February 22].
44. Polacek C, Friebe P, Harris E. Poly(A)-binding protein binds to the non-polyadenylated 3′ untranslated region of dengue virus and modulates translation efficiency. *J Gener Virol* 2009;**90**(3):687–92 [March 1].
45. Nevels M, Täuber B, Spruss T, Wolf H, Dobner T. "Hit-and-run" transformation by adenovirus oncogenes. *J Virol* 2001;**75**(7):3089–94 [April 1].
46. Stevenson PG, May JS, Connor V, Efstathiou S. Vaccination against a hit-and-run viral cancer. *J Gener Virol* 2010;**91**(Pt 9):2176–85 [September 1].
47. Niller HH, Wolf H, Minarovits J. Viral hit and run-oncogenesis: genetic and epigenetic scenarios. *Cancer Lett* 2011;**305**(2):200–17 [June 28].
48. Baryawno N, Rahbar A, Wolmer-Solberg N, Taher C, Odeberg J, Darabi A, et al. Detection of human cytomegalovirus in medulloblastomas reveals a potential therapeutic target. *J Clin Invest* 2011;**121**(10):4043–55 [October 3].
49. Michaelis M, Doerr HW, Cinatl J. The story of human cytomegalovirus and cancer: increasing evidence and open questions. *Neoplasia* 2009;**11**(1):1–9.
50. Robinson M, Sloan P, Shaw R. Refining the diagnosis of oropharyngeal squamous cell carcinoma using human papillomavirus testing. *Oral Oncology* 2010;**46**(7):492–6.
51. Jordan RC, Lingen MW, Perez-Ordonez B, He X, Pickard R, Koluder M, et al. Validation of methods for oropharyngeal cancer HPV status determination in US cooperative group trials. *Am J Surg Pathol* 2012;**36**(7):945–54.