Genomic characterization of clonal evolution during oropharyngeal carcinogenesis driven by human papillomavirus 16

Jeesoo Chae[1,2], Weon Seo Park[3], Min Jung Kim[2], Se Song Jang[1,2], Dongwan Hong[4], Junsun Ryu[5], Chang Hwan Ryu[5], Ji-Hyun Kim[5], Moon-Kyung Choi[3], Kwan Ho Cho[6], Sung Ho Moon[6], Tak Yun[7], Jong-Il Kim[1,8,9,]* & Yuh-Seog Jung[5,]*

[1]Department of Biomedical Science, Seoul National University Graduate School, Seoul 03080, [2]Department of Biochemistry and Molecular Biology, Seoul National University College of Medicine, Seoul 03080, [3]Department of Pathology, Center for Specific Organs Cancer, Hematologic Malignancy Branch, National Cancer Center, Goyang 10408, [4]Cancer Immunology Branch, Division of Cancer Biology, National Cancer Center, Goyang 10408, [5]Department of Otorhinolaryngology, Graduate School of Cancer Science and Policy, Department of Immunotherapeutics, National Cancer Center, Goyang 10408, [6]Center for Proton Therapy, Center for Specific Organs Cancer, National Cancer Center, Goyang 10408, [7]Hematologic Oncology Clinic, Center for Specific Organs Cancer, National Cancer Center, Goyang 10408, [8]Cancer Research Institute, Seoul National University, Seoul 03080, [9]Genomic Medicine Institute, Medical Research Center, Seoul National University, Seoul 03080, Korea

Secondary prevention via earlier detection would afford the greatest chance for a cure in premalignant lesions. We investigated the exomic profiles of non-malignant and malignant changes in head and neck squamous cell carcinoma (HNSCC) and the genomic blueprint of human papillomavirus (HPV)-driven carcinogenesis in oropharyngeal squamous cell carcinoma (OPSCC). Whole-exome (WES) and whole-genome (WGS) sequencing were performed on peripheral blood and adjacent non-tumor and tumor specimens obtained from eight Korean HNSCC patients from 2013 to 2015. Next-generation sequencing yielded an average coverage of 94.3× for WES and 35.3× for WGS. In comparative genomic analysis of non-tumor and tumor tissue pairs, we were unable to identify common cancer-associated early mutations and copy number alterations (CNA) except in one pair. Interestingly, in this case, we observed that non-tumor tonsillar crypts adjacent to HPV-positive OPSCC appeared normal under a microscope; however, this tissue also showed weak p16 expression. WGS revealed the infection and integration of high-risk type HPV16 in this tissue as well as in the matched tumor. Furthermore, WES identified shared and tumor-specific genomic alterations for this pair. Clonal analysis enabled us to infer the process by which this transitional crypt epithelium (TrCE) evolved into a tumor; this evolution was accompanied by the subsequent accumulation of genomic alterations, including an *ERBB3* mutation and large-scale CNAs, such as 3q27-qter amplification and 9p deletion. We suggest that HPV16-driven OPSCC carcinogenesis is a stepwise evolutionary process that is consistent with a multistep carcinogenesis model. Our results highlight the carcinogenic changes driven by HPV16 infection and provide a basis for the secondary prevention of OPSCC. [BMB Reports 2018; 51(11): 584-589]

*Corresponding authors. Jong-Il Kim, Tel: +82-2-740-8421; Fax: +82-2-741-5423; E-mail: jongil@snu.ac.kr; Yuh-Seog Jung, Tel: +82-31-920-1685; Fax: +82-31-920-2799; E-mail: jysorl@ncc.re.kr

INTRODUCTION

Worldwide, head and neck squamous cell carcinoma (HNSCC) affects more than 550,000 people and causes 380,000 deaths each year (1). Prolonged exposure of the upper aerodigestive tract to carcinogenic environmental factors, such as tobacco, alcohol, and human papillomavirus (HPV), can result in dysplastic or premalignant lesions in the mucosa, which eventually result in HNSCC. Recent epidemiological evidence from various western and Asian countries (2, 3) has shown that the incidence of HPV-positive oropharyngeal squamous cell carcinoma (OPSCC) is increasing and is predicted to surpass that of cervical cancer by 2020 in the USA (4); thus, this prediction necessitates a better understanding of its genomic and molecular characteristics to facilitate earlier detection and secondary prevention.

With recent advances in next-generation sequencing (NGS) technology, large-scale genomic analyses of HNSCC have revealed recurrently mutated molecular pathways, such as p53, RTK/PI(3)K and various differentiation pathways (5). However, the carcinogenetic effects, including the effects of long-term HPV infection, remain elusive. HPV has been identified in circular extrachromosomal elements that are known as episomes during initial HPV infection, and viral-host

integration that induces E6 and E7 viral oncogene expression is known to be an important step in the progression of HNSCC and genital cancer (6, 7). However, intensive efforts to detect active HPV16 in tissues surrounding such tumors have provided scant evidence of this process (8). In addition, studies examining the prevalence of HPV infection in non-tumor tonsillar specimens have been performed (9-11) but have failed to find or have found only a low (1%-3.1%) rate of high-risk HPV infection. Thus, HPV-infected non-tumor tissues are rare; and no comparative analyses of HPV-infected non-tumor and corresponding tumor tissues, which might elucidate the precise serial action of HPV, have yet been reported.

Here, we present a comparative genomic analysis of adjacent non-tumor and tumor tissue pairs that highlights the stepwise accumulation of genomic alterations and clonal evolution during high-risk HPV16-driven OPSCC development. We suggest that HPV-driven OPSCC carcinogenesis is a step-by-step evolutionary process that is consistent with the multistep carcinogenesis model.

RESULTS

Clonal populations of tumors emerge separately in adjacent non-tumor mucosa

To understand the genomic features of non-tumor mucosa in close proximity to tumors and to find early mutations shared with the tumor that might provide a growth advantage, we performed comparative WES analysis for eight pairs of adjacent non-tumor and HNSCC tissues (Supplementary Table S1 and Fig. S1). Despite rigorous pathologic review by a specialized pathologist (W.S., P.), no gross or microscopic findings suggested any relevant features of dysplasia for carcinoma in situ (CIS) using conventional hematoxylin and eosin (H&E) staining in all eight non-tumor samples. Genome-wide copy number alterations (CNA) patterns revealed that most non-tumor mucosa maintained genomic stability, except for one sample, HNT-009N (Fig. 1A, Supplementary Fig. S2). In contrast, the tumors showed a significantly increased number of somatic CNAs (SCNAs) (total number of SCNAs: 74 in non-tumors and 2,748 in tumors; Mann-Whitney P-value = 0.001; Supplementary Table S2).

Non-tumor mucosa samples had few somatic coding mutations, regardless of the HPV status of the matched tumor (median: 5.5 mutations per sample, ranging from 0 to 51; median allele frequency [AF] [excluding HNT-009N] = 0.073; AF including HNT-009N = 0.102; Fig. 1B, C and Supplementary Table S3). Moreover, tumors showed significantly more somatic mutations than did non-tumors (median: 59.5 mutations, ranging from 3 to 201; median AF = 0.235; Mann-Whitney P-value = 0.027). Common mutations associated with HNSCC development were not observed in each pair, except for HNT-009, even during the manual re-confirmation of mutations that showed the possibility of having a small fraction of progenitors (Fig. 1B, Supplementary Fig. S3).
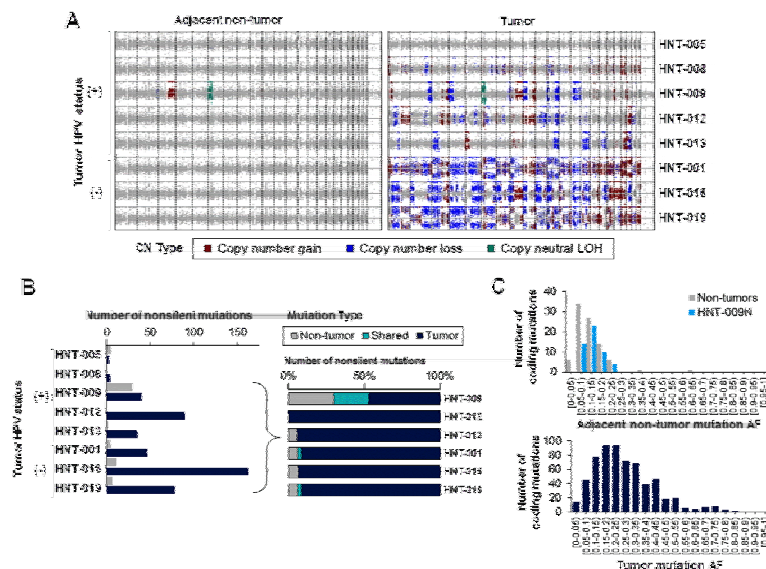

Fig. 1. Genomic characteristics of adjacent non-tumor and tumor tissues from eight HNSCC cases. (A) B allele frequency (BAF) and SCNAs according to WES analysis. Non-tumor mucosa did not show CNAs except for one p16-positive non-tumor, HNT-009N. LOH: Loss of heterozygosity; CN: Copy number. (B) The number of non-silent mutations (left panel) and the proportion of shared non-silent mutations from each pair of non-tumor and tumor tissues (right panel). HNT-005 and -008 were omitted because of the limited number of mutations. (C) A histogram of the AF distribution of somatic coding mutations.
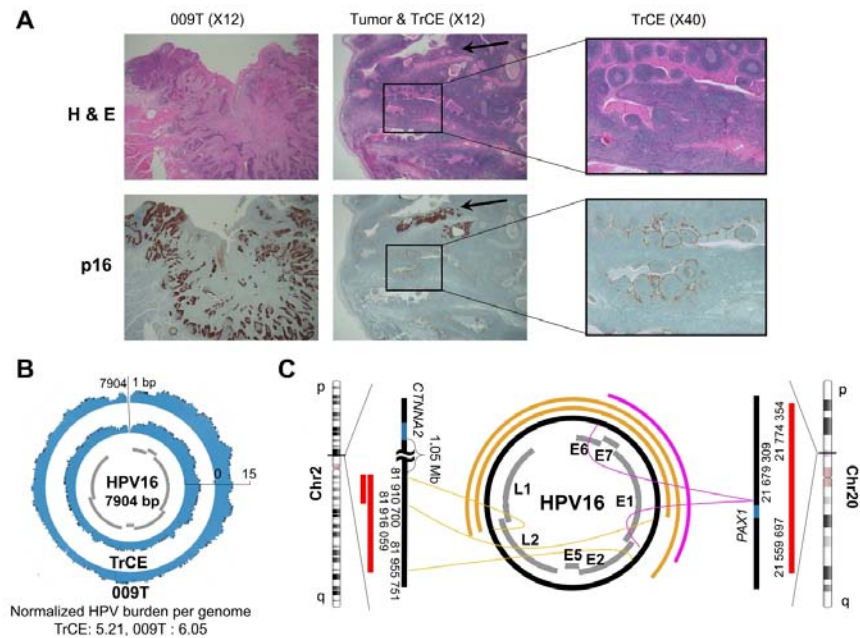
Fig. 2. Description of the HNT-009 case with HPV16-positive(+) transitional crypt epithelium (TrCE). (A) Histopathological confirmation of TrCE in HNT-009. TrCE exhibited weaker immunoreactivity (H-score: 12) to p16 than did 009T (H-score: 26). This epithelium was confined within a grossly normal epithelial layer within the tonsillar crypt epithelium (box), in contrast to the adjacent tumor tissue (arrow). (B) Per-genome HPV16 viral load in the TrCE and 009T. (C) HPV16 integration into the human genome. Two integration sites accompanying CNA were observed in both the TrCE and 009T.

p16 expression and HPV16 integration in normal-appearing tonsillar crypt epithelium

Notably, we observed that non-tumor crypt epithelial tissue adjacent to the HPV(+) tumor of HNT-009 case exhibited numerous somatic alterations when compared to seven other non-tumor samples, such as the amplification of 3q and the copy-neutral loss of heterozygosity (cnLOH) of 6p and somatic mutations (Fig. 1). The HNT-009 case was a 56-year-old female who neither smoked nor drank (Supplementary Table S4). H&E staining showed that the squamous epithelium of the tonsillar crypt was histologically indistinguishable from normal crypt epithelium. Interestingly, however, the tissue exhibited weak p16 immunoreactivity (H-score: 12), although it clearly differed from the neighboring established OPSCC based on its higher p16 expression (H-score: 26, Fig. 2A). Because this epithelium showed common genomic alterations in 009T (tumor from HNT-009) and p16 expression, we hereafter refer to this tissue as transitional crypt epithelium (TrCE).

Furthermore, WGS confirmed HPV16 infection in both the TrCE and 009T (Supplementary Table S5-7). The normalized per-genome HPV16 viral burden was 5.2 in the TrCE and was slightly higher at 6.1 in 009T (Fig. 2B). Although HPV16 in the early infectious stage is expected to exist mainly in the episomal form, two integration regions on chromosomes 2 and 20 were identified in the TrCE as well as in 009T. One

integration site was located 7 kb upstream of *PAX1* on chromosome 20, and the other was found 1.05 Mb downstream of *CTNNA2* on chromosome 2 (Fig. 2C). The HPV16-integrated loci were accompanied by structural alterations, such as amplification of the region between the HPV16 insertional breakpoints and inverted duplication across the integration breakpoint (Supplementary Fig. S4).

Stepwise genomic alterations during HPV16-driven carcinogenesis

The genomic alterations found in the TrCE were more evident in the matched tumor, 009T. Of the coding sequence (CDS) mutations in the TrCE, 53% (27 of 51) showed significant increase of AF in 009T (t-test P-value = 2.1E-10), although the function of these mutations in cancer is unknown (Fig. 3A). In addition to these mutations, copy-neutral LOH of 6p21.1-25 was also more prominent in 009T (Fig. 3B).

Several tumor-specific features were also present. Mutational signature analysis (12) revealed the presence of two common signatures, signatures 1 and 2 (Fig. 3C), which were correlated with the age at cancer diagnosis and the APOBEC enzymatic activity, respectively. In addition, a new signature, 6, was identified in 009T and was associated with defects in the DNA mismatch repair (dMMR) system; however, mutations in dMMR-related genes such as *MLH1, MSH2/6* and *PMS2* were
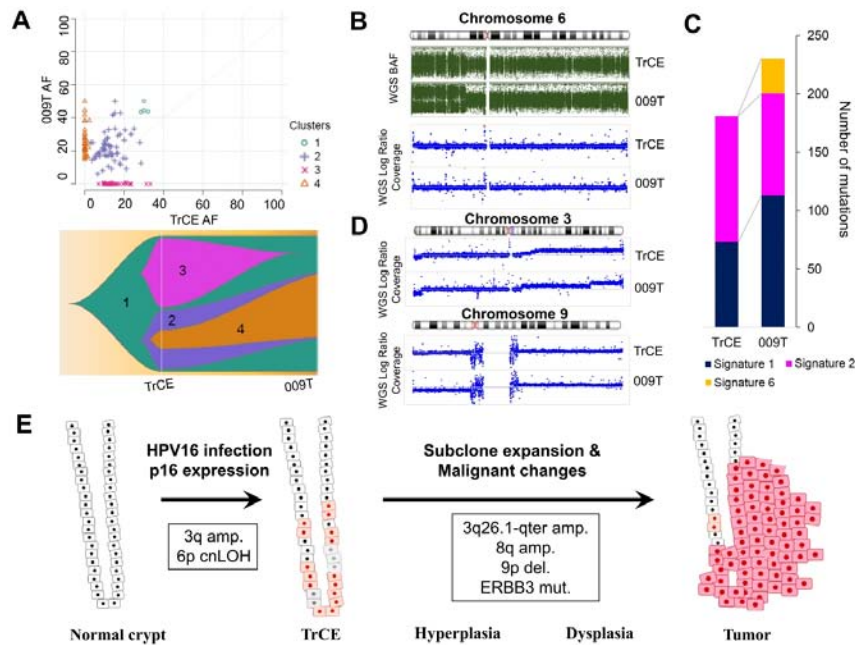
Fig. 3. Comparative genomic features of the TrCE and matched tumor 009T and the proposed model of HPV-driven carcinogenesis. (A) Inferred subclonal architecture of the TrCE and 009T from HNT-009. Cluster 4 involved mutations acquired in the tumor, including an *ERBB3* mutation. The AF of the shared mutations in clusters 1 and 2 were significantly increased in 009T (P = 2.1E-10). (B) Copy-neutral LOH of 6p. (C) Mutational signature analysis; Signature 1: age at diagnosis; Signature 2: APOBEC enzymatic activity; Signature 6: defective mismatch repair. (D) A series of SCNAs from the TrCE and the tumor. The Y axis shows the range from −2 to 2. (E) The proposed model of stepwise HPV16-driven oropharyngeal carcinogenesis.

not identified. Thirty-four CDS mutations in genes including *ERBB3, SYNE1,* and *WDR90* were identified only in 009T. A missense mutation in *ERBB3*, resulting in the substitution of amino acid D922 in the tyrosine kinase domain with N, showed an AF of 21.7% and was predicted to impair protein function according to several in-silico tools. Several tumor-specific CNAs, such as the amplification of 3q26.1-qter and 8q21-q24 and the deletion of 9p13-p22, were also observed (Fig. 3D).

Based on the observation of serial genomic changes in the TrCE and 009T, subclonal structures and evolutionary patterns were inferred (Fig. 3A bottom panel; Supplementary Fig. S5 and Supplementary Table S8). These results suggested that, after the emergence of the ancestral clone with mutation cluster 1, the TrCE evolved into two subclones with mutation clusters 2 and 3, respectively. The subclone with cluster 2 subsequently progressed into a tumor by accumulating additional genomic changes (cluster 4), and the other subclone underwent negative clonal selection. A proposed model of HPV16-driven OPSCC carcinogenesis was developed by combining the comprehensive genomic features of this case as shown in Fig. 3E.

DISCUSSION

Here, we provide a novel glimpse of the landscape of oropharyngeal carcinogenesis driven by longstanding HPV16 infection; this process occurs via the stepwise accumulation of genomic alterations and clonal evolution.

We screened genomic alterations in eight pairs of HNSCC and adjacent non-tumor tissues. Tumors exhibited somatic alterations that were distinct from those in non-tumor mucosa, indicating the generation of malignant clones during tumor progression. Although *TP53* mutations were frequently reported in the tumor-free margin of HPV(−) HNSCC (13, 14), genetic aberrations, including *TP53* mutations, were only observed in a tumor-specific manner in three of the HPV(−) cases.

Regarding the important role of HPV in OPSCC, several studies have investigated the presence of HPV infection in the non-malignant tonsil tissues of cancer-free individuals or in adjacent tumors (8-11) as well as in the genomic profiles of premalignant lesions prior to oral cancer (15). Attempts to detect HPV infection in almost 4,000 non-tumor tonsil tissue specimens in the UK did not find evidence of an association with HPV (10), and two other studies conducted in Finland and in the US found that only 1.0% and 3.1% of patients, respectively, had high-risk HPV infections (in 477 and 5,579

non-malignant tonsil tissues, respectively) (9, 11). Nevertheless, in-depth genomic characterization of the normal to premalignant to malignant transition process driven by persistent HPV16 infection in OPSCC has not been performed. Considering that the TrCE tissue was obtained from a 56-year-old female who neither smoked nor drank alcohol, we postulate that the molecular features of the TrCE might reflect a carcinogenetic process that is driven exclusively by HPV16 during the development of OPSCC.

p16 immunohistochemistry (IHC) has been reported as an adjunctive method for assessing dysplasia in genital skin lesions and the head and neck (16-18). However, in this study, despite the presence of microscopically normal tonsillar crypt epithelium without dysplasia, the TrCE showed weak p16 immunoreactivity, and HPV16 infection and integration was also identified. Ilmarinen *et al.* reported an episomal form of HPV in HPV-infected non-malignant tonsils using fluorescence in situ hybridization (FISH); however, since WGS offers sensitive detection of viral integration, it is possible that HPV integration is able to occur even prior to dysplasia. Because of the sponge- and fissure-like structures of the tonsillar crypts, it is difficult to detect early premalignant legions. However, as therapeutic vaccines effectively treat high-grade precancerous cervical lesions caused by HPV (19), epithelium with TrCE-like features might also be a target for this type of immunotherapy and for surgical resection, if the lesions are localized.

In addition, the TrCE exhibited several other genomic features. Although Ilmarinen *et al.* demonstrated the presence of several sequence mutations in genes such as *PIK3CA* in HPV-infected chronic tonsillitis specimens using the targeted sequencing of 578 genes, our TrCE contained large-scale CNAs such as 3q amplification and 6p cnLOH as well as mutations. Amplification of 3q has been reported repeatedly in HNSCC, and 6p cnLOH has also been reported in 2 of 19 HNSCCs (20). Recurrent cnLOH of specific regions might indicate the presence of important genes with a role in disease pathogenesis.

Additionally, the subsequent acquisition of alterations in the TrCE is probably crucial for cancer progression due to their activation of cellular proto-oncogenes and preventative role against tumor suppressor genes. From the subclone analysis, we inferred that the TrCE diverged into two subclones from an initial precursor, and one subclone predominated in the progressed tumor with the acquisition of alterations such as an *ERBB3* mutation and the amplification of 3q and deletion of 9p. The amplification of 3q and deletion of 9p, recurrently reported in HNSCCs, may be involved in the HPV-related progression of tonsillar carcinoma (21). Although *ERBB3* mutations appear in approximately 3% of HNSCCs (5), a recent TCGA study of cervical cancer highlighted a significant recurrence of *ERBB3* mutations (22), suggesting that this gene could be an HPV-related cancer driver. The observed increase in the AF of the *ERBB3* mutation from 0.028 in the TrCE to 0.22 in 009T might reflect the enhanced fitness of the subclone.

Finally, we attempted to find more cases exhibiting TrCE-like features in the pathologic archives of HPV(+) HNSCCs (n = 93) among patients who had surgery-based treatment; however, we could not find any additional cases with TrCE features. Therefore, we assume that TrCE is very rare, consistent with previous observations, and that this study is meaningful as the first paper to report the detailed genomic profile of this lesion in matched tumors. Nevertheless, this study has several limitations: (1) common premalignant mutations were not identified, perhaps because of the small sample size and sampling bias; (2) changes in gene expression or in the epigenome of adjacent non-tumor tissues were not fully evaluated; (3) the expression of p16 in adjacent non-tumor tissue in one case and the derived carcinogenesis model should be further verified. Multi-region sequencing or single cell sequencing of the TrCE-like tissues would enable us to expand our understanding of clonal evolution during HPV-associated OPSCC carcinogenesis. This would enable the findings to be generalized and provide a basis for the screening and secondary prevention of OPSCC.

MATERIALS AND METHODS

Detailed materials and experimental procedures are available in Supplementary Data.

ACKNOWLEDGEMENTS

CONFLICTS OF INTEREST

The authors declare no conflict of interest.

REFERENCES

1. Global Burden of Disease Cancer, Fitzmaurice C and Allen C (2016) 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 3, 524-548

2. Gooi Z, Chan JY and Fakhry C (2016) The epidemiology of the human papillomavirus related to oropharyngeal head and neck cancer. Laryngoscope 126, 894-900

3. Chaturvedi K, Engels A and Pfeiffer M (2011) Human papillomavirus and rising oropharyngeal cancer incidence in the United States. J Clin Oncol 29, 4294-4301

4. Chaturvedi K, Anderson F and Lortet-Tieulent J (2013) Worldwide trends in incidence rates for oral cavity and oropharyngeal cancers. J Clin Oncol 31, 4550-4559

5. Cancer Genome Atlas Network (2015) Comprehensive genomic characterization of head and neck squamous cell carcinomas. Nature 517, 576-582

6. Wentzensen N, Vinokurova S and von Knebel Doeberitz M (2004) Systematic review of genomic integration sites of human papillomavirus genomes in epithelial dysplasia and invasive cancer of the female lower genital tract. Cancer Res 64, 3878-3884

7. Vogelstein B and Kinzler W (1993) The multistep nature of cancer. Trends Genet 9, 138-141

8. Rietbergen M, Braakhuis J and Moukhtari N (2014) No evidence for active human papillomavirus (HPV) in fields surrounding HPV-positive oropharyngeal tumors. J Oral Pathol Med 43, 137-142

9. Ilmarinen T, Munne P and Hagstrom J (2017) Prevalence of high-risk human papillomavirus infection and cancer gene mutations in nonmalignant tonsils. Oral Oncol 73, 77-82

10. Palmer E, Newcombe G and Green C (2014) Human papillomavirus infection is rare in nonmalignant tonsil tissue in the UK: implications for tonsil cancer precursor lesions. Int J Cancer 135, 2437-2443

11. Gillison L, Broutian T and Pickard K (2012) Prevalence of oral HPV infection in the United States, 2009-2010. JAMA 307, 693-703

12. Alexandrov B, Nik-Zainal S and Wedge C (2013) Signatures of mutational processes in human cancer. Nature 500, 415-421

13. Tabor P, Brakenhoff H, Ruijter-Schippers J, Kummer A, Leemans R and Braakhuis J (2004) Genetically altered fields as origin of locally recurrent head and neck cancer: a retrospective study. Clin Cancer Res 10, 3607-3613

14. Brennan A, Mao L and Hruban H (1995) Molecular assessment of histopathological staging in squamous-cell carcinoma of the head and neck. N Engl J Med 332, 429-435

15. Wood M, Conway C and Daly C (2015) The clonal relationships between pre-cancer and cancer revealed by ultra-deep sequencing. J Pathol 237, 296-306

16. Ezaldein H, Lott P, McNiff M, Hui P, Buza N and Ko J (2015) Grading of aypia in genital skin lesions: routine microscopic evaluation and use of p16 immunostaining. J Cutan Pathol 42, 519-526

17. Gultekin E, Senguven B, Klussmann P and Dienes P (2015) P16(INK 4a) and Ki-67 expression in human papilloma virus-related head and neck mucosal lesions. Invest Clin 56, 47-59

18. Tabrizi N, Tan E, von Knebel Doeberitz C, Tan J, Quinn A and Garland M (2015) Evaluation of p16INK4a immunostaining for the detection of high-grade changes in cervical cytology. Pathology 47, 314-319

19. Trimble L, Morrow P and Kraynyak A (2015) Safety, efficacy, and immunogenicity of VGX-3100, a therapeutic synthetic DNA vaccine targeting human papillomavirus 16 and 18 E6 and E7 proteins for cervical intraepithelial neoplasia 2/3: a randomised, double-blind, placebo-controlled phase 2b trial. Lancet 386, 2078-2088

20. Marescalco S, Capizzi C, Condorelli F and Barresi V (2014) Genome-wide analysis of recurrent copy-number alterations and copy-neutral loss of heterozygosity in head and neck squamous cell carcinoma. J Oral Pathol Med 43, 20-27

21. Singh B, Stoffel A and Gogineni S (2002) Amplification of the 3q26.3 locus is associated with progression to invasive cancer and is a negative prognostic factor in head and neck squamous cell carcinomas. Am J Pathol 161, 365-371

22. Cancer Genome Atlas Research (2017) Integrated genomic and molecular characterization of cervical cancer. Nature 545, 378-384

23. Holsinger C, McWhorter J, Menard M, Garcia D and Laccourreye O (2005) Transoral lateral oropharyngectomy for squamous cell carcinoma of the tonsillar region: I. Technique, complications, and functional results. Arch Otolaryngol Head Neck Surg 131, 583-591

24. Rhie A, Park WS and Choi MK (2015) Genomic copy Number Variations Characterize the Prognosis of Both P16-positive and P16-negative oropharyngeal squamous cell carcinoma after curative resection. Medicine (Baltimore) 94, e2187

25. Hirsch R, Varella-Garcia M and Bunn A (2003) Epidermal growth factor receptor in non-small-cell lung carcinomas: correlation between gene copy number and protein expression and impact on prognosis. J Clin Oncol 21, 3798-3807

26. Park WS, Ryu J and Cho KH (2012) Human papillomavirus in oropharyngeal squamous cell carcinomas in Korea: use of G1 cycle markers as new prognosticators. Head Neck 34, 1408-1417

27. Li H and Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25, 1754-1760

28. McKenna A, Hanna M and Banks E (2010) The genome analysis toolkit: a mapReduce framework for analyzing next-generation DNA sequencing data. Genome Res 20, 1297-1303

29. Cibulski K, Lawrence S and Carter L (2013) Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. Nat Biotechnol 31, 213-219

30. Krumm N, Sudmant H and Ko A (2012) Copy number variation detection and genotyping from exome sequence data. Genome Res 22, 1525-1532

31. Rosenthal R, McGranahan N, Herrero J, Taylor S and Swanton C (2016) DeconstructSigs: delineating mutational processes in single tumors distinguishes DNA repair deficiencies and patterns of carcinoma evolution. Genome Biol 17, 31

32. Larson B and Fridley L (2013) PurBayes: estimating tumor cellularity and subclonality in next-generation sequencing data. Bioinformatics 29, 1888-1889