UCSF
UC San Francisco Previously Published Works

Title

APOBEC-induced mutations and their cancer effect size in head and neck squamous cell carcinoma

Permalink

https://escholarship.org/uc/item/3343p8fh

Journal

Oncogene, 38(18)

ISSN

0950-9232

Authors

Cannataro, Vincent L
Gaffney, Stephen G
Sasaki, Tomoaki
et al.

Publication Date

2019-05-01

DOI

10.1038/s41388-018-0657-6

Copyright Information

Peer reviewed

ARTICLE

APOBEC-induced mutations and their cancer effect size in head and neck squamous cell carcinoma

Vincent L. Cannataro [1] · Stephen G. Gaffney [1] · Tomoaki Sasaki[2] · Natalia Issaeva[3,4] · Nicholas K. S. Grewal[5] · Jennifer R. Grandis[6] · Wendell G. Yarbrough[3,4,7] · Barbara Burtness[3,8] · Karen S. Anderson[2,3,9] · Jeffrey P. Townsend [1,5,10]

Abstract

Recent studies have revealed the mutational signatures underlying the somatic evolution of cancer, and the prevalences of associated somatic genetic variants. Here we estimate the intensity of positive selection that drives mutations to high frequency in tumors, yielding higher prevalences than expected on the basis of mutation and neutral drift alone. We apply this approach to a sample of 525 head and neck squamous cell carcinoma exomes, producing a rank-ordered list of gene variants by selection intensity. Our results illustrate the complementarity of calculating the intensity of selection on mutations along with tallying the prevalence of individual substitutions in cancer: while many of the most prevalently-altered genes were heavily selected, their relative importance to the cancer phenotype differs from their prevalence and from their P value, with some infrequent variants exhibiting evidence of strong positive selection. Furthermore, we extend our analysis of effect size by quantifying the degree to which mutational processes (such as APOBEC mutagenesis) contributes mutations that are highly selected, driving head and neck squamous cell carcinoma. We calculate the substitutions caused by APOBEC mutagenesis that make the greatest contribution to cancer phenotype among patients. Lastly, we demonstrate via in vitro biochemical experiments that the APOBEC3B protein can deaminate the cytosine bases at two sites whose mutant states are subject to high net realized selection intensities—PIK3CA E545K and E542K. By quantifying the effects of mutations, we deepen the molecular understanding of carcinogenesis in head and neck squamous cell carcinoma.

Introduction

Cancer has long been known to have a basis in somatic mutations that alter a diversity of cellular functions resulting in sustained proliferative signaling, evasion of growth suppressors, and genome instability [1]. Recent high-throughput sequencing of tumors has inspired genomics-driven oncology [2] and enabled the deconvolution of

---

✉ Jeffrey P. Townsend
Jeffrey.Townsend@Yale.edu

1   Department of Biostatistics, Yale School of Public Health, New Haven, CT, USA

2   Department of Pharmacology, Yale University, New Haven, CT, USA

3   Yale Cancer Center, Yale University, New Haven, CT, USA

4   Division of Otolaryngology, Department of Surgery, Yale School of Medicine, New Haven, CT, USA

5   Department of Ecology and Evolutionary Biology, Yale University, New Haven, CT, USA

6   Department of Otolaryngology-Head and Neck Surgery, University of California San Francisco, San Francisco, CA, USA

7   Department of Pathology, Yale School of Medicine, Yale University, New Haven, CT, USA

8   Department of Medicine, Yale School of Medicine, Yale University, New Haven, CT, USA

9   Department of Molecular Biophysics & Biochemistry, Yale University, New Haven, CT, USA

10 Program in Computational Biology and Bioinformatics, Yale University, New Haven, CT, USA

mutations to obtain signatures of their underlying rate [3, 4] and cause [5–7]. However, knowing the source of mutations and the rate at which they occur provides only the first step to understanding the genetic origins of substitutions that drive tumorigenesis and cancer development. After mutations occur, the mutations that are important to cancer cell survival and propagation become prevalent in tumors because of their oncogenic effect on cancer cell lineages. Quantifying the strength of selection of each mutation provides insight into the relative importance of mutations in driver genes [4], while examination of the relative contributions of different mutational processes can reveal the relative importance of each process to tumorigenesis. Independent quantification of mutational processes and selective effects is essential to knowledge-based identification of high value therapeutic targets.

Mutational processes are of particular importance in head and neck squamous cell carcinoma (HNSCC) because of the high proportion of mutations that are attributed to APOBEC (apolipoprotein B mRNA editing catalytic polypeptide-like) enzymes in HNSCC tumors [8, 9]. APOBECs are a family of enzymes that catalyze the deamination of cytosine bases, APOBEC3B expression is higher in cell lines [10] and HNSCC tumors [8] with viral infection, APOBEC activity has been found to be positively correlated with the upregulation of immune signaling pathways [9], and promoter elements and transcription factors linking HPV16 infection and APOBEC3B expression have been identified [11, 12]. Consequently, APOBEC mutagenesis is believed to play a role in the genesis of human papillomavirus-associated (HPV+) HNSCC. Some of the earliest insights into the biology of cancer arose as a result of its association with viral infection [13, 14]. Worldwide, an estimated 15 percent of cancers are attributable to viral infection [15], including Epstein-Barr virus, kaposi's sarcoma-associated herpesvirus, human T-lymphotropic virus type 1, and human papillomavirus (HPV), among others. HPV is responsible for the largest number of viral-related cancers worldwide through its contribution to cervical cancer, anal cancer, vaginal, vulvar, penile cancers, and, increasingly, HNSCC [16].

HPV-associated (HPV+) and HPV-unassociated (HPV−) HNSCC—related but clearly distinct types of cancer [17, 18]—provide an interesting pairing for comparative analysis of mutational processes and selective effects. Genomic profiling of head and neck squamous cell carcinoma has provided confirmation that, in addition to differences in etiology, natural history and treatment responsiveness [19, 20], HPV+ and HPV− HNSCC have distinct biomolecular profiles [21]. Among HPV-associated cancers, the predominant substitutions occurring are activating mutations of *PIK3CA*, inactivating mutations of *TRAF3* or *CYLD*, and amplification of

FGFR3 and the cell cycle gene *E2F*. Among HPV-unrelated cancers, inactivating substitutions in the tumor suppressor genes *TP53* and *CDKN2A* are predominant [17, 21, 22]. These distinctions in epidemiology and molecular biology are complemented by differences in natural history and responsiveness to treatment: HPV-associated tumors are more responsive to cytotoxic and radiation therapy [20], and are more frequently cured [23]. Despite these distinctions between HNSCC caused by tobacco use and those related to HPV infection [24] and a spate of recent de-escalation therapy trials for HPV+ HNSCC [25], there are currently no differences in standard therapy specific to these distinct subtypes of HNSCC [26]. Deeper insight into underlying molecular mechanisms may be instrumental to developing more specific therapies.

Patterns within the burden of mutations in a cancer provide insight into the mechanism of mutation [27]. A number of accepted mutational signatures have been identified in HNSCC including two signatures reflective of the activity of APOBEC3 [7, 28, 29]. A previous report suggests that HPV− HNSCC is characterized by a tobacco-associated mutational signature, while HPV+ cancers display an APOBEC3 signature and APOBEC3-mediated driver mutations, including characteristic helical domain PIK3CA mutations [8]. However, it is currently unknown whether differences between the genetic architecture of HPV+ and HPV− HNSCC can be attributed to distinct underlying mutation frequencies within proto-oncogenes and tumor suppressors, or to differential selection on those mutations. Improved understanding of molecular mechanisms undergirding emergence and maintenance of HNSCC and how the mechanisms differ based on HPV status has the potential to enhance clinical decision-making and inform the development of novel therapeutics. We were interested in understanding the relationship of mutational load to the weight of APOBEC signature, as well as the selective pressure associated with the mutations arising as a consequence of this mutational process.

Next-generation tumor exome sequencing has implicated a number of molecular pathways in HNSCC tumorigenesis. This approach confirmed that known tumor suppressor genes and oncogenes (e.g., *PIK3CA, EGFR, CDKN2A, NOTCH1, TP53, FBXW7*) play a role in HNSCC, and pathway analyses found that—amongst others—proliferation, differentiation, and PI3K pathways were frequently affected [30–32]. Sequencing-based analyses have expanded our molecular understanding of the disease, and have equated the prevalence of a somatic substitution in tumors of the affected population with its importance in perturbing signaling pathways and tumor development [21, 30, 31]. Identification of targets via prevalence of substitutions assures that the altered gene will be relevant for the affected

population, but gives little characterization of the efficacy of targeting the altered gene product at the level of the individual patient. In contrast, quantification of the effects that mutations confer upon the cancer phenotype should assist with prioritization of future precision-targeted therapies [33–35], as well as identify rare substitutions that are significant drivers of tumorigenesis and tumor maintenance.

To characterize these effects, we analyzed a sample of 508 HNSCC tumors from the National Cancer Institute Genomic Data Commons [36] and 17 HNSCC tumors from Hedberg et al [37]. We evaluated the mutational signatures characteristic of HPV$^-$ and HPV$^+$ HNSCC, quantifying the role that APOBEC plays as a mutagenic factor in HNSCC. We then examined individual substitutions to evaluate the characteristic mutational signature of APOBEC-induced mutations in HNSCC and, noting in particular whether the APOBEC signature is characterized by T$\underline{C}$W transition and transversion mutations (wherein the mutated cytosine is underlined), confirmed informatic analysis with in vitro experiment. We analyzed whether APOBEC participates in both HPV$^+$ and HPV$^-$ mutagenesis and whether an APOBEC signal is correlated with the total mutational load in HPV$^-$ and HPV$^+$ HNSCC. We then applied an evolutionary biology-based approach to quantify the selection pressure upon specific mutations in a population of HNSCC tumors [4]. Our analysis confirmed well-known HNSCC tumorigenic targets, but also highlights low-prevalence mutations that are major drivers in a minority of patients, and implicates novel targets as potential drivers in HNSCC development. We quantified the importance of APOBEC mutagenic processes, identifying the underlying mutation frequencies of molecular variants and the selective effects on them that lead to the development of HNSCC.
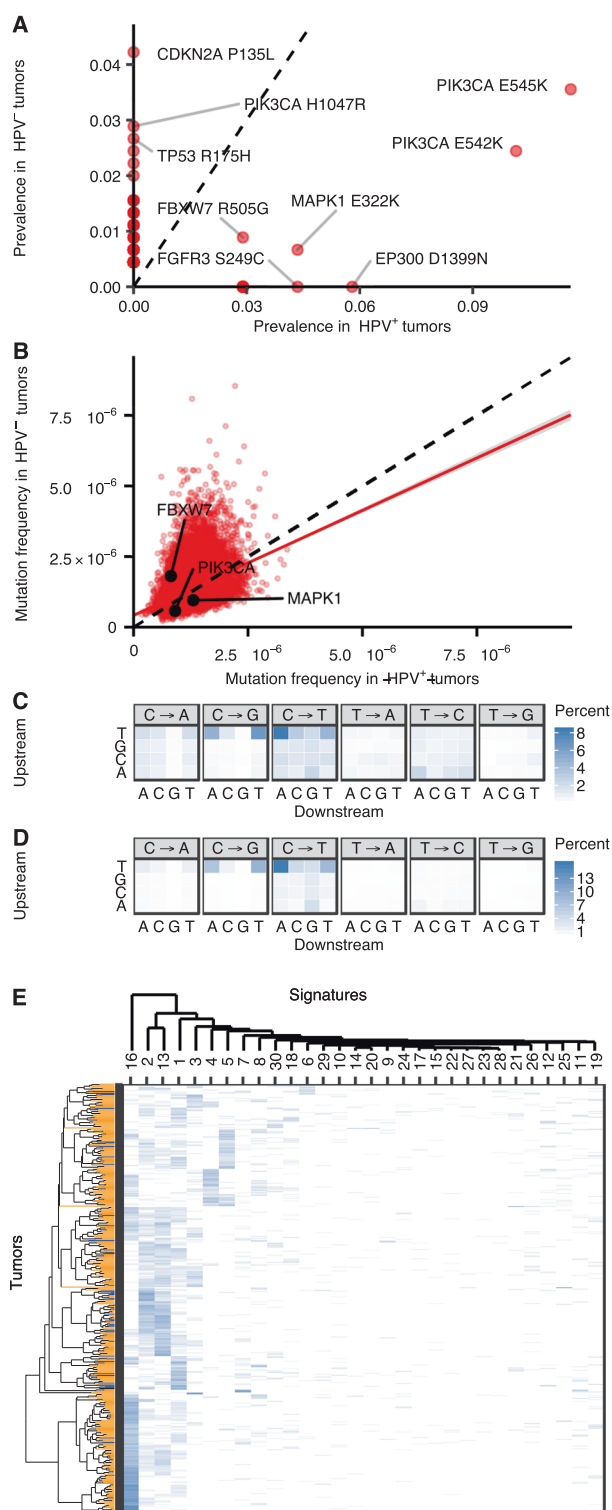
Results

HPV$^-$ and HPV$^+$ classification and HPV viral load

Our analysis incorporated 69 HPV$^+$ and 451 HPV$^-$ cases (Supplementary Table S1) retrieved from National Cancer Institute Genomic Data Commons [36] and from tumor-normal pairs from the University of Pittsburgh head and neck tissue bank, reported in Hedberg et al. [37], and sequenced at the Yale Center for Genome Analysis. Of the 69 HPV$^+$ HNSCC, one tumor carried a *TP53* substitution (missense, in this case)—a tumor that was also classified as HPV$^+$ in the comprehensive genomic characterization of HNSCC conducted by TCGA (TCGA-CR-7638) [21]. This rarity is consistent with a previously described model wherein HPV-driven cancer is associated with high HPV viral load, wild type p53, and overexpression of p16 as a consequence of HPV oncoproteins that inactivate p53 and

Rb. Non-HPV-driven cancer frequently exhibited *TP53* substitutions and an absence of p16 expression, although in some cases HPV was present at low viral load [38]. Our HPV designations differ from the TCGA biospecimen supplement data designations that are based on PCR-based MassArray—a method that has been deemed too sensitive to accurately determine whether HPV infection has played a causal role in tumorigenesis [21]. Indeed, 93 out of 508 TCGA HNSCC tumors had detectable portions of the HPV genome using this method, and 24 of these 93 tumors had substitutions in *TP53*, meaning they were unlikely to be driven by viral oncogenesis due to the criteria outlined above; thus, we applied an alternate method of detecting viral sequences in RNA-Seq data that has been relied upon in previous studies [21, 39]. Among TCGA samples with available HPV status and available mRNA expression data (429 HPV$^-$ and 66 HPV$^+$ tumors, mRNA expression data obtained from cBioPortal [40, 41]), mRNA expression was higher for APOBEC3B, APOBEC3C, APOBEC3D, APOBEC3F, APOBEC3G, and APOBEC3H in HPV$^+$ vs HPV$^-$ tumors ($P < 0.001$, one-sided Welch's t tests), and APOBEC3A mRNA expression levels were not significantly different between the two HPV classifications ($P = 0.38$, one-sided Welch's t test; Figure S1), as previously reported on a smaller subset of HNSC tumors [8]. Within the germline TCGA data, the distribution of the global APOBEC3H haplotypes [42] are 27.6% haplotype 1, 17.7% haplotype 2, 11.8% haplotype 3, 42.3% haplotype 4, and 0.6% other [43].

Between HPV$^+$ and HPV$^-$ tumor tissues, signatures of mutagenesis are not distinct, but frequencies of mutation among genes are distinct

To probe into the relative importance of the substitutions driving these two types of HNSCC, we examined the prevalence of substitutions among the cancers, the frequency we expected to observe substitutions in the absence of selection, and ultimately the intensity of selection that drives any difference between the observed and expected fluxes. The prevalences of recurrent somatic nucleotide variants in HPV$^+$ and HPV$^-$ tumor tissue were markedly different: only four substitutions were recurrently mutated in the assembled data sets for these cancers (Fig. 1a, Supplementary Table S2). Some of these differences in prevalence arise because of differences in gene-level mutation frequencies (Fig. 1b, Supplementary Table S3) that reflecting underlying differences between the covariates governing gene-level mutation events—such as gene expression levels—within HPV$^+$ and HPV$^-$ cells [21]. In contrast, differences in prevalence between HPV$^+$ and HPV$^-$ tumors are unlikely to arise because of differing trinucleotide mutation frequencies, because HPV$^+$ and HPV$^-$

spontaneous deamination of 5-methylcytosine and correlated with aging (Signature 1 of 30) [6].

APOBEC signatures are not equally frequent in HPV$^-$ and HPV$^+$ HNSCC

All 30 COSMIC mutational signatures were represented in one or more of the 461 HNSCC primary tumors examined that exhibited greater than 50 mutations (Fig. 1e, Supplementary Table S4). The two most common signatures were those attributed to APOBEC activity (COSMIC signatures #2 and #13). After APOBEC, signature #1, a typical somatic mutation signature attributed to aging, was the most prevalent signature in HNSCC—higher than tobacco (#4). Mean weights of signatures #13 and #2 were 0.19 and 0.30 in HPV$^+$ tumors and 0.13 and 0.11 in HPV$^-$ tumors, respectively.

APOBEC signatures were not equally frequent in HPV$^-$ and HPV$^+$ HNSCC (Fisher's exact test, $P = 0.0001$). Among tumors with a known HPV status and enough SNVs to calculate mutational signatures [5], 98% of HPV$^+$ HNSCC tumors exhibited an APOBEC signature, while 76% of HPV$^-$ tumors exhibited an APOBEC signature. Consistent with expectation, HPV$^+$ status in tumor tissue was more likely among those tumors exhibiting a high total APOBEC weight (Fig. 2; logistic regression, $P = 1 \times 10^{-8}$).

HPV$^+$ and HPV$^-$ tumors with a signature of APOBEC-associated mutagenesis do not differ in mutation load, but do differ in the proportion of TCW→TKW substitutions

In HPV$^-$ tumors, there was no significant difference between the typical SNV count of tumors with and without an APOBEC signature ($P = 0.49$, two-sided Wilcoxon rank-sum test). Only one tumor was both HPV$^+$ and had no

tumor tissues exhibited similar trinucleotide mutation profiles (Figs. 1c, d). The dominant signatures in these trinucleotide mutation profiles include APOBEC-related C → T and C → G mutations in the TCW motif context, and C → T mutations in a 5′-CG-3′ context that are likely reflective of

detectable signature of APOBEC activity (PY-16T). Among samples with an APOBEC signal, typical mutation load was not significantly different between HPV$^-$ and HPV$^+$ HNSCC ($P = 0.39$, two-sided Wilcoxon rank-sum test). Among tumors with an APOBEC signature, HPV$^+$ samples typically exhibited a higher proportion of the TCW to TKW mutations that are commonly attributed to APOBEC ($P = 0.008$, two-sided Wilcoxon rank-sum test). Nevertheless, the majority of TCW to TKW mutations among tumors with an APOBEC signal occurred within HPV$^-$ tumors, as a supermajority of tumors were HPV$^-$. Thus, when analyzing the substitutions that likely occurred through APOBEC mutagenesis, we consider all HNSCC samples regardless of HPV status.

The selective pressures within HPV$^+$ and HPV$^-$ HNSCC tumors largely differ

Much of the difference in the frequencies of substitutions between HPV$^+$ and HPV$^-$ HNSCC can be attributed to
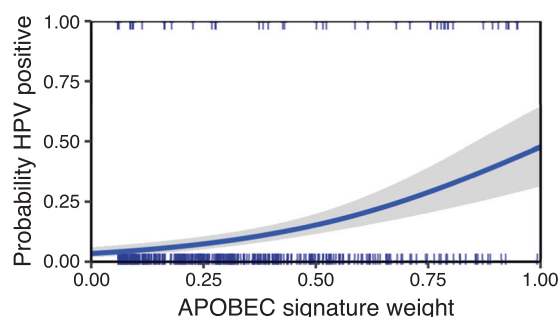
Fig. 2 Tumors with a high total APOBEC weight were more likely to be HPV positive

different selective intensities imposed by mutations occurring in the HPV$^+$ vs. HPV$^-$ context. Comparison of observed frequencies of substitutions and the expected frequencies of mutations yielded a ranked list of single nucleotide variants based on their selection intensity (Fig. 3, Supplementary Table S2), excluding variants observed fewer than two times in our sample set. These recurrent variants represent the point mutations that are the strongest drivers of the neoplastic HNSCC disease phenotype. Among the recurrent SNV with the largest 25 selection intensities in HPV$^-$ HNSCC and all recurrent SNV in HPV$^+$, only FBXW7 R505G was shared. Substitutions in TP53 and HRAS are frequent in HPV$^-$, but are almost never found in HPV$^+$. Among all variants, four recurrent substitutions are shared (MAPK1 E322K, PIK3CA E542K and E545K, and FBXW7 R505G) and have similar selection intensities within the two tumor types (Fig. 3c).

Within this set of the strongest drivers, many align with existing understanding of HNSCC tumorigenesis. Several of the most-selected variants occurred in genes—*TP53, HRAS, PIK3CA, MAPK1*—previously associated with the landscape of substitutions observed in HNSCC via genomic screening [17, 31]. Interestingly, one of the most strongly selected variants among HPV$^-$ HNSCC (and the strongest selected variant in HPV$^+$ HNSCCs) is characterized by a low expected mutation frequency and a low prevalence of oncogenic substitution, occurring in the E3 ubiquitin ligase-encoding gene *FBXW7*, a known tumor suppressor which has been molecularly characterized as a regulator of common HNSCC oncogenes [44, 45]. In HPV$^+$ tumors, fibroblast growth factor receptor genes such as FGFR3 have previously been found to be enriched for mutations [17], but
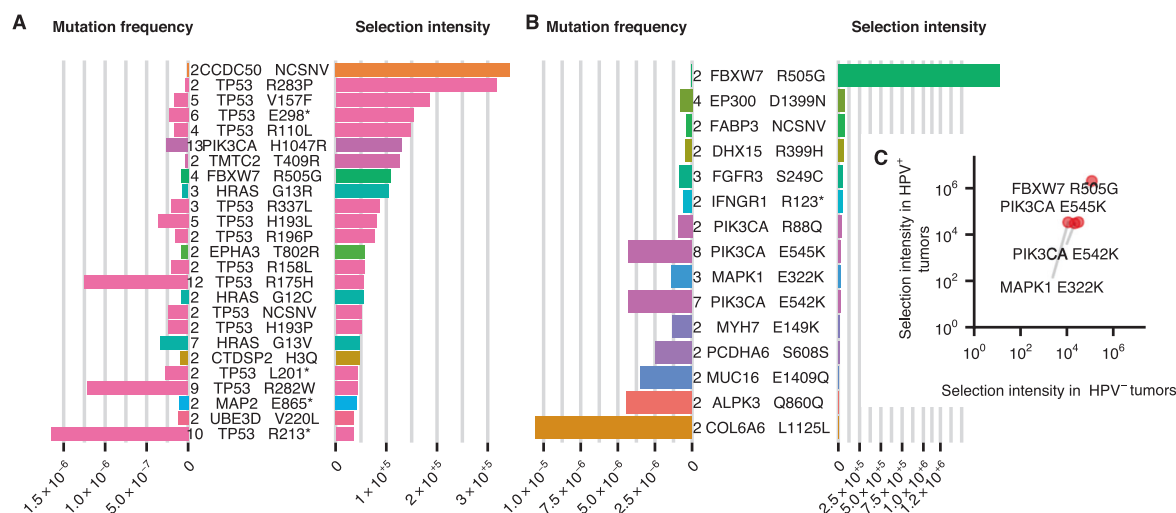
Fig. 3 Expected mutation frequency (left-hand bar), prevalence (columnated numbers, along with HUGO gene name and the amino acid change), and selection intensity (right-hand bar) associated with recurrently observed mutations (**a**) for the top 25 selection intensities of point mutations among 451 HPV$^-$ HNSCC tumor tissues, and (**b**) for all recurrent substitutions among 69 HPV$^+$ HNSCC tumor tissues. Inset, (**c**) the selection intensities on the recurrently observed mutations within HPV$^-$ HNSCCs and within HPV$^+$ HNSCCs

the high intensity of their effect on the proliferation and survival of cancer cell lineages has not previously been quantified relative to other somatic nucleotide mutations.

Other variants with high inferred intensity of selection implicate novel candidate genes in HNSCC tumorigenesis or cancer development also deserve further attention to validate their role. The coiled-coil domain-containing protein 50 (CCDC50) gene contained a recurrent splice-site substitution within two HPV⁻ tumors: this mutation was found to have the highest selection intensity of all recurrent somatic nucleotide variants within HPV⁻ tumors, as it is subject to an extremely low expected frequency of mutation. The four lowest selection intensities among recurrently observed mutations in HPV⁺ tumors were one substitution within the second longest human protein MUC16 (the *MUC16* gene is subject to frequent mutation in cancers and

is likely spuriously associated with tumorigenesis [3]), and three synonymous substitutions within *ALPK3*, *PCDHA6*, and *COL6A6*.

Substitutions from the putative APOBEC binding site have a lower average selection intensity

Among the 314 recurrent amino acid substitutions in the HPV⁺ and HPV⁻ datasets, 62 were the product of a TCW to TKW nucleotide mutation, and thus can be putatively associated with APOBEC mutagenesis. Substitutions that could putatively be attributed to APOBEC mutagenesis typically exhibited a lower selection intensity than substitutions that are not in a TCW to TKW trinucleotide context (Fig. 4; Supplementary Table S2, $P = 0.002$, one-sided Welch two-sample *t*-test).

PIK3CA, FBXW7, and FGFR3 are the genes whose APOBEC-based tumor burden contributes the most to population-level cancer

Next, we quantified tumor burden arising from APOBEC processes by evaluating which mutations have the largest effect on cancer growth and the highest prevalences among tumors, and distinguished whether these mutations most likely result from APOBEC mutagenesis. Among mutations that were likely APOBEC-independent, FBXW7 R505G, EP300 D1399N, PIK3CA H1047R, and several mutations in TP53 had the highest net realized selection intensity. Among mutations that were likely APOBEC induced, FBXW7 R505G, PIK3CA E545K, FGFR3 S249C, PIK3CA E542K, and DHX15 R399H had the highest net realized selection intensities. (Fig. 5; Supplementary Table S5).
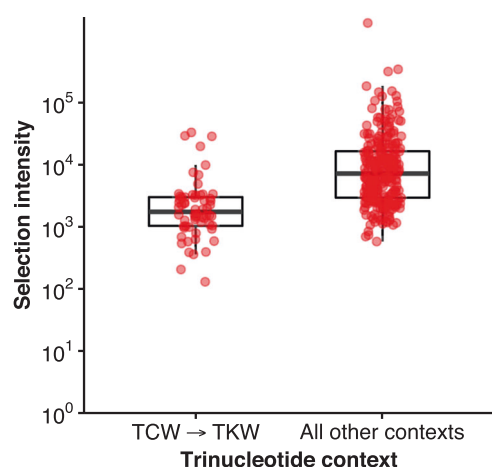

Fig. 4 The selection intensities of recurrent, amino acid replacement TCW→TKW nucleotide mutations compared to other recurrent, amino acid replacement mutations (All other contexts)

Fig. 5 Net realized selection intensity within cancer patients for recurrently observed mutations attributable to signatures 1, 2, 13, 4, 16, and the remaining 25 signatures. PABP3 NCSNV is a non-coding single-nucleotide variant on chromosome one at position 31838696
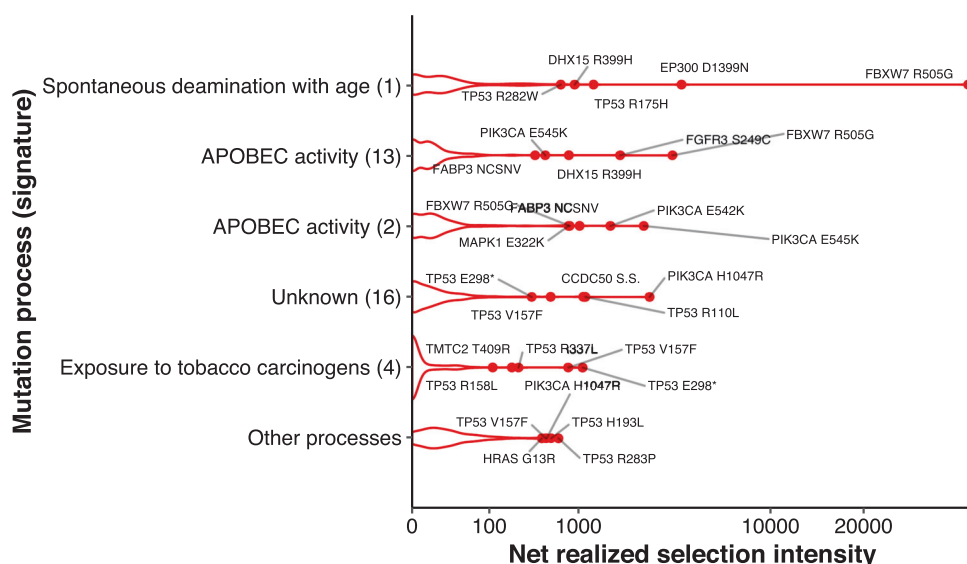
APOBEC3B efficiently deaminates PIK3CA DNA in vitro

To validate whether these bioinformatically identified targets are true APOBEC targets, we investigated deamination of sites in *PIK3CA* using a traditional uracil DNA glycosylase (UDG) deaminase assay. Using a recombinantly expressed and purified preparation of full-length APOBEC3B—the APOBEC3 member most closely associated to human cancers—we tested its activity on two 25-mer substrates mimicking the DNA sequence surrounding the E542 and E545 sites. APOBEC3 deaminated the TC motif in the center of these two sequences more than it deaminated a centered TC motif within an AT-rich 25-mer nonspecific oligonucleotide control. Upon determination of catalytic rates by fitting of the deamination time course to a single exponential curve, we found that both *PIK3CA*-derived oligonucleotides were deaminated at a higher efficiency than the benchmark 25-mer, which we adapted from a published 43-mer [46], reflecting published literature including a systematic biochemical analyses of all 16 cytosine-flanking trinucleotide studies and structural studies that captured thymine at the -1 position relative to the target cytosine [47] (Fig. 6). This experiment shows that A3B can deaminate single-stranded DNA matching these cancer mutation hotspots.

Discussion

Here we have shown that the COSMIC mutation signatures of HPV$^+$ and HPV$^-$ head and neck squamous cell carcinomas do not cluster into distinct groups of self-similar mutation signatures when considering all 30 COSMIC signatures. Nonetheless, all but one (98%) HPV$^+$ tumor had
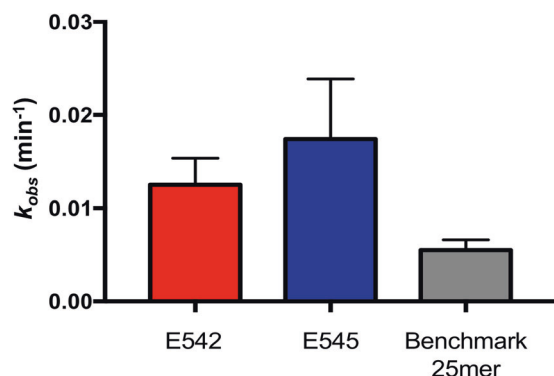

Fig. 6 Deamination of *PIK3CA* mimic substrates. UDG deamination experiments were conducted with 25-mer substrates corresponding to the *PIK3CA* E542 site, E545 site, and the benchmark 25-mer. The bar graph shows a comparison of the observed catalytic rate k_{obs} with each oligonucleotide substrate. Error bars represent the standard deviation of the exponential fit to the data

a detectable APOBEC signature, whereas a considerably lower percentage (76%) of HPV$^-$ tumors had a detectable APOBEC signature. Our analysis revealed significant differences in the genetic architecture of HPV$^+$ and HPV$^-$ HNSCC tumors. Evaluating each mutation based on its expected frequency in the absence of positive selection and the actual frequency at which it was observed, we identified several well-known culprits: *PIK3CA, TP53*, and *HRAS* [21]. Our analysis of the effect sizes of mutations underlying HPV$^+$ and HPV$^-$ HNSCC tumors not only substantiates the difference in genetic architecture between these tumors based on HPV status, but also identifies some rare oncogenic mutations that contribute strongly to tumorigenesis in both HPV$^+$ and HPV$^-$ HNSCC tumors (e.g., FBXW7 R505G), as well as oncogenic mutations that are somatically selected for in either HPV$^+$ (e.g., EP300 D1399N, FGFR3 F249C) or HPV$^-$ tumors (e.g., PIK3CA H1047R, oncogenic mutations in HRAS and TP53). Our results primarily agree with the existing literature on the strongest driver mutations in HNSCC, but segregate the contributions of oncogenes in particular as major contributors to one or the other of HPV$^+$ and HPV$^-$ HNSCC tumors.

It has been previously remarked that mutations arising from APOBEC-related processes could be a double-edged sword: [48] while APOBEC-related processes provide innate immunity against viruses by targeting them for mutation [49], they also are a source of numerous mutations driving cancer [7, 28, 50–52]. We have shown that the inner edge of that sword is somewhat blunt: APOBEC mutations typically have lower selection intensities than mutations caused by non-APOBEC processes. Our analysis implies that it would on average be preferable during tumorigenesis for a person to exchange an unknown, non-APOBEC neoplastic mutation for an unknown, putatively APOBEC neoplastic mutation. Regardless, it remains highly plausible that prevention of APOBEC mutagenesis during tumorigenesis would diminish the rate of substitution and decelerate cancer progression and even the evolution of therapeutic resistance [53]. Indeed, some of the variants with the highest net realized selection intensities among patients are attributable to APOBEC activity. The net realized selection intensity of mutations quantifies the amount of cancer phenotype at the population level that is attributable to their cause, incorporating both the prevalence of each substitution and the degree to which it contributes to the cancer phenotype in the patients with that substitution. Accordingly, it quantifies the population-wide amount of cancer phenotype that could be eliminated by universal patient prescription of a perfect therapeutic that abrogates oncogenic function of each mutation, corresponding to potential population health benefits of a perfect mutation-targeted therapeutic. One caveat of our analysis is that this

inference applies to a population as though our data were a representative sample of the population. HPV infection prevalence varies with socioeconomic factors [54], racial disparities [35, 55] and genetic polymorphisms [56, 57] associated with outcomes in head and neck cancer so greater attention to representative sampling would provide more precise estimates of the frequencies of mutations within populations [58] and the net realized selection intensities in each population.

The mutations in *PIK3CA* that are highly selected in HPV$^+$ HNSCC—E542K and E545K—lie in its helical domain [59]. Mutated sites in the helical domain have been proposed as targets of APOBEC-mediated cytosine deamination [8]. The Henderson et al. (2014) analysis reported these helical domain substitutions to be HPV$^+$ HNSCC-specific based on correlations between APOBEC3B expression, and HPV status and the prevalence of TCW→TKW substitutions, yet it appears that the cancer effect size of these mutations is substantial in both HPV$^-$ HNSCC and HPV$^+$ HNSCC. The four shared recurrently observed mutations among HPV$^-$ and HPV$^+$ HNSCC are similarly ordered by selection intensity within their tumor type.

Our biochemical experimental data show that APOBEC3B deaminates the two PIK3CA mimic oligonucleotides at a higher efficiency compared to a benchmark oligomer of a similar size. Importantly, this analysis is the first quantitative assessment of deamination of two bioinformatically identified putative target sites relative to a benchmark using a full-length APOBEC enzyme. This observation is particularly interesting given that we are beginning to understand the molecular features of APOBEC-driven mutagenesis in vivo. For instance, APOBEC mutation patterns are present at a significant levels in six distinct cancer types [28] and APOBEC proteins specifically target regions of the human genome containing exposed single stranded DNA, including chromosome rearrangement breakpoints [51] and replication intermediates [60]. This specificity of targeting might explain observations of a temporal dependence within the cell cycle, wherein APOBEC specifically induces mutations in early replicating regions in chromosomal DNA that are highly active in gene expression [52]. In light of these bioinformatic discoveries, our biochemical result demonstrating context-specific targeting of APOBEC3B provides an additional layer of validation of APOBEC mutagenesis in cancer, and opens new questions in the field—including questions regarding the molecular mechanisms by which APOBEC enzymes target these sites in specific genes.

A key difference in our methodology from the gene-level focus of other recent exome sequencing studies is the application of inference at the level of individual mutations of amino acid residues in our analysis. Our analysis is

therefore restricted to substitution mutations, whereas comprehensive genomic characterizations have investigated deletions, copy number alterations, and other rare genomic events. For instance, Hajek et al [22] flagged deletions in the gene *TRAF3*, demonstrating the importance of nonsense mutations in this gene to the genetic architecture of HPV$^+$ HNSCC. Because our approach enables the estimation of expected frequencies of nucleotide mutations, but not expected frequencies of deletions, the lack of recurrent *TRAF3* substitutions precluded it from any estimation of selective pressure using our current methodology.

In addition to refining the interpretation of genes exhibiting higher-than-expected mutation loads, our analysis suggests some novel avenues for investigation. This methodology highlights gene variants that have been missed by gene-based surveys of somatic substitution prevalence. One of the somatic variants with very high effect sizes in both HPV$^+$ and HPV$^-$ tumors, FBXW7 R505G, occurs in an E3 ubiquitin ligase and modifies E3 ligase binding to the substrates [61, 62]. *FBXW7* has been characterized as a tumor suppressor due to its ability to target several known oncoproteins including MYC, NOTCH, and Cyclin E [44]. FBXW7 is also a negative regulator of Aurora A and B, and thus regulates a mitotic checkpoint [63, 64]. *FBXW7* mutation has been demonstrated to associate with NOTCH1 activation in T-cell acute lymphoblastic leukemia; [65] however, not until *NOTCH1*'s implication in HNSCC tumorigenesis was *FBXW7* linked to this disease [30, 45].

These results underscore one important difference of our evolutionary approach from earlier approaches toward the identification of the molecular basis of tumorigenesis and cancer development from exome sequence data: its utility at ranking oncogenic mutations by importance, making use of tissue- and nucleotide- specific mutation frequencies. Here, we make use of bulk sequencing of primary tumors at the time of resection, and thus our estimates are based on average mutation frequencies and correspond to selection intensities that are averaged over the time course of tumorigenesis to resection. However, processes of mutagenesis vary over time [66], and finer resolution measurements of the expected frequencies of mutations [67] and corresponding selection intensities are possible, especially with multiple samples gathered metachronously or even synchronously. Our approach has been enabled by extensive previous research on the intrinsic mutation rates of genes and their covariates [3, 68, 69], and relative mutation rates of single nucleotide variants and their putative mutagenic source [7]. Nonetheless, comparisons of selection intensities on somatic single nucleotide variants only paint part of the complex picture of tumorigenesis: copy number variation, epigenetic alterations, tumor heterogeneity and microenvironment, germline variants, and other factors also contribute to the cancer phenotype [58, 70–72]. Methods to

calculate measures of selection intensities that incorporate these factors and their epistatic effects would enable a holistic model—and a more holistic understanding—of tumorigenesis.

Methods

Determining HPV status

Curated sequencing data for 508 cases of HNSCC was sourced from National Cancer Institute Genomic Data Commons (NCI GDC) [36], which provides information on the limited diversity of ethnicity and race in the sample. 477 of these tumors had HPV RNA viral transcript loads as assessed by the VirusScan software in Cao et al [39]. Out of the 31 tumors with no previous viral load data, 24 had the RNA-Seq data required for VirusScan. HPV infection status for these data was determined by counting HPV16 and HPV33 reads within RNA-Seq data, as attributed by VirusScan [39]. Tumors were designated as HPV$^+$ if they contained greater than 100 HPV RNA viral transcript reads per hundred million, and tumors were designated HPV$^-$ if they contained fewer than 100 HPV RNA viral transcript reads per hundred million. Of the 501 HNSCC tumors analyzed with VirusScan, 66 tumors were deemed HPV$^+$, and 435 HNSCCs that did not match that criterion were classified HPV$^-$. Out of the remaining seven tumors sourced from the NCI GDC with no viral load data, two tumors had consistent clinical results from p16 and ISH tests (one HPV$^+$ and one HPV$^-$). Additionally, we included 17 HNSCCs that were whole-exome sequenced at Yale University [37] and one (PY-16T) was also determined to be HPV-associated by HPV in situ hybridization and p16 immunohistochemistry [73], and the remaining 16 cases were determined to be HPV$^-$.

Bioinformatic and statistical analysis of HNSCC data

Before analysis, all SNV data downloaded from the NCI were converted to human genome 19 reference coordinates using the R package rtracklayer [74] to make the data compatible with the various software packages used in our analysis. Somatic variants from the Hedberg et al [37]. dataset were obtained from published supplementary material, and underwent the following processing: (1) variants with labels other than "Transmitted" and "Primary Only" were removed to obtain the set of primary tumor variants, (2) neighboring SNVs were redesignated as DNVs and extraneous alleles (denoted with lower case) were removed, (3) variants were re-annotated using the software vcf2maf v1.6.13 [75, 76], (4) those variants annotated as "common" or classified as intronic and flanking were

removed, and (5) shared variants for five tumors that were identified as duplicated in the NCI dataset (PY-19T, PY-1T, PY-14T, PY-13T, PY-7T) were removed. For the selection intensity analysis, we removed all neighboring variants to ensure that DNVs did not influence our analyses of the frequencies of single nucleotide mutations. We performed hierarchical clustering on the distance matrix of mutational signatures for all tumors that had available data on HPV status and mutational signatures.

Calculating selection intensity

The selection intensity for point mutations was calculated by an updated approach based on that of Cannataro et al [4]. Briefly, the expected frequency μ that nucleotide mutations occur before they are acted on by selection over the average amount of time elapsed throughout the evolutionary process driving tumorigenesis (from initialization to resection) was determined by calculating the expected frequency that silent mutations occur at the gene level using dndscv [91], then scaling each possible nucleotide mutation by a coefficient corresponding to the relative expected frequency within its trinucleotide context. The effect of the trinucleotide context was quantified as the average distribution of trinucleotide contexts for that tissue calculated analyzing all HNSCC tumors with greater than 50 mutations with deconstructSigs [5]. We then defined the frequency of substitution, λ, as the frequency at which genetic variants were observed within sequence data. We corrected for the fact that one can only observe one substitution per site, even though a flux of mutations at a given rate will generate a Poisson distributed number of substitutions [4]. This observed frequency of substitutions was divided by the expected frequency of mutations to determine $\frac{\lambda}{\mu} = \gamma$, the selection intensity on a point mutation during the intratumoral fixation process.

Calculating net realized selection intensity per putative mutation process

We calculated the effect sizes γ_i of somatic amino acid substitution i within a growing tumor. Effect sizes were calculated using the prevalence $p_{i,h}$ of that substitution among tumors, where $p_{i,h} = \frac{n_{i,h}}{N_h}$, $n_{i,h}$ is the number of tumors with variant i detected in tumors with HPV status h, N_h is the total number of tumors with HPV status $h \in \{0, 1\}$, where $h = 0$ indicates HPV$^-$ and $h = 1$ indicates HPV$^+$). The net realized selection intensity was then $\frac{p_{i,h_1}\left(\alpha_{i,j}\gamma_i p_{i,h_1}\right) + p_{i,h_2}\left(\alpha_{i,j}\gamma_i p_{i,h_2}\right)}{p_{i,h_1} + p_{i,h_2}}$ for all recurrent amino acid substitutions i in our dataset, where the median proportional contribution of the individual COSMIC signatures to the specific single nucleotide substitution j resulting in amino acid substitution i was $\alpha_{i,j}$.

Code availability

The scripts used to process data and perform these analyses are available online at https://github.com/Townsend-Lab-Yale/HNSCC_APOBEC.

Assessing APOBEC3B deamination of PIK3CA helical domain sites

Recombinant full-length APOBEC3B with a N-terminal maltose binding protein tag was sequenced to ensure that the plasmid DNA did not contain any insertions, deletions, or mutations. The plasmid was expressed and purified from *E. coli* (the Overexpress C43(DE3) cell line purchased from Lucigen) as described previously [77] and authenticated using antibiotic selection markers. For deamination assays, a traditional UDG deaminase assay was employed in which 1 μM of APOBEC3B was incubated with 40 nM of 5′-radiolabeled oligonucleotide in a reaction buffer consisting of 50 mM Tris pH 8.0, 100 mM NaCl, 0.1% Triton X-100, and 1 mM DTT. Upon quenching timepoints, the reactions were post-processed with 5 units of UDG and 0.2 N NaOH. The resulting solution was then run on a 20% denaturing PAGE gel and exposed to a phosphor screen overnight. Gel densities were analyzed with QuantityOne software to calculate % product at each timepoint. A representative gel image from which the densitometry analysis has been performed is available as a supplement to the manuscript (Figure S2). Catalytic efficiency of the purified full length APOBEC3B protein has been reported in a number of previous publications [77, 78]. While differences exist in constructs including expression system, presence of affinity/solubility tags, length of oligonucleotide substrate, and assay conditions used, the APOBEC3B activity in the present study reflects expected values based on previously reported kinetics. The benchmark oligonucleotide tested in the current study is a shortened version (25-mer) generic oligonucleotide substrate based on early characterization of APOBEC3B [46]. Our studies also confirm that catalytic efficiency increases proportionally to substrate length (data not shown) similar to that previously shown with APOBEC3G [79].

All DNA oligonucleotides were ordered HPLC-purified from Integrative DNA Technologies. The E542 and E545 sites are located very close to one another in PI3KCA; thus, when we constructed the 25-mer PI3KCA test oligos, the sequences had significant overlap. To ensure that our analysis focused on the deamination of each respective cytosine, we constructed oligos for each site that were composed with a thymine (another pyrimidine) at the other site. The test oligonucleotide sequences,

E542: 5′-TGCTTAGTGATTTCAGAGAGAGGAT-3′, and

E545: 5′-TGCTCAGTGATTTTAGAGAGAGGAT-3′, were based on the cDNA sequence of PIK3CA surrounding the test sites. The benchmark 25-mer oligonucleotide sequence

control: 5′-ATTATTATGGATCAATTATTTATTA-3′,

was a modification of a substrate oligonucleotide used in an earlier report using this assay to investigate activity of APOBEC3B [46].

For calculation of the observed catalytic rate k_{obs}, each enzymatic time course was fit to a single exponential curve defined by

$$\%\text{product} = A(1 - e^{-k_{obs}t})$$

in which A is the maximum % product, and time t was measured in minutes.

Compliance with ethical standards

References

1. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. Cell. 2011;144:646–74.
2. Hayes DN, Kim WY. The next steps in next-gen sequencing of cancer genomes. J Clin Invest. 2015;125:462–8.
3. Lawrence MS, Stojanov P, Polak P, Kryukov GV, Cibulskis K, Sivachenko A, et al. Mutational heterogeneity in cancer and the search for new cancer-associated genes. Nature. 2013;499:214–8.
4. Cannataro VL, Gaffney SG, Stender C, Zhao Z, Philips M, Greenstein AE, et al. Heterogeneity and mutation in KRAS and associated oncogenes: evaluating the potential for the evolution of resistance to targeting of KRAS G12C. Oncogene. 2018;37:2444–55.
5. Rosenthal R, McGranahan N, Herrero J, Taylor BS, Swanton C. DeconstructSigs: delineating mutational processes in single tumors distinguishes DNA repair deficiencies and patterns of carcinoma evolution. Genome Biol. 2016;17:31.
6. Alexandrov LB, Jones PH, Wedge DC, Sale JE, Campbell PJ, Nik-Zainal S, et al. Clock-like mutational processes in human somatic cells. Nat Genet. 2015;47:1402–7.
7. Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio SAJR, Behjati S, Biankin AV. et al. Signatures of mutational processes in human cancer. Nature. 2013;500:415–21.
8. Henderson S, Chakravarthy A, Su X, Boshoff C, Fenton TR. APOBEC-mediated cytosine deamination links PIK3CA helical domain mutations to human papillomavirus-driven tumor development. Cell Rep. 2014;7:1833–41.
9. Faden DL, Thomas S, Cantalupo PG, Agrawal N, Myers J, DeRisi J. Multi-modality analysis supports APOBEC as a major source of mutations in head and neck squamous cell carcinoma. Oral Oncol. 2017;74:8–14.
10. Vieira VC, Leonard B, White EA, Starrett GJ, Temiz NA, Lorenz LD, et al. Human papillomavirus E6 triggers upregulation of the antiviral and cancer genomic DNA deaminase APOBEC3B. MBio. 2014. https://doi.org/10.1128/mBio.02234-14
11. Mori S, Takeuchi T, Ishii Y, Kukimoto I. Identification of APOBEC3B promoter elements responsible for activation by human papillomavirus type 16 E6. Biochem Biophys Res Commun. 2015;460:555–60.
12. Mori S, Takeuchi T, Ishii Y, Yugawa T, Kiyono T, Nishina H, et al. Human Papillomavirus 16 E6 Upregulates APOBEC3B via the TEAD Transcription Factor. J Virol. 2017. https://doi.org/10.1128/JVI.02413-16
13. Weiss RA, Vogt PK. 100 years of Rous sarcoma virus. J Exp Med. 2011;208:2351–5.
14. Rous P. A sarcoma of the fowl transmissible by an agent separable from the tumor cells. J Exp Med. 1911;13:397–411.
15. zur Hausen H. Viruses in human cancers. Science. 1991;254:1167–73.
16. Parkin DM. The global health burden of infection-associated cancers in the year 2002. Int J Cancer. 2006;118:3030–44.
17. Seiwert TY, Zuo Z, Keck MK, Khattri A, Pedamallu CS, Stricker T, et al. Integrative and comparative genomic analysis of HPV-positive and HPV-negative head and neck squamous cell carcinomas. Clin Cancer Res. 2015;21:632–41.
18. Hammerman PS, Hayes DN, Grandis JR. Therapeutic insights from genomic studies of head and neck squamous cell carcinomas. Cancer Discov. 2015;5:239–44.
19. Applebaum KM, Furniss CS, Zeka A, Posner MR, Smith JF, Bryan J, et al. Lack of association of alcohol and tobacco with HPV16-associated head and neck cancer. J Natl Cancer Inst. 2007;99:1801–10.
20. 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;100:261–9.
21. Cancer Genome Atlas Network. Comprehensive genomic characterization of head and neck squamous cell carcinomas. Nature. 2015;517:576–82.
22. Hajek M, Sewell A, Kaech S, Burtness B, Yarbrough WG, Issaeva N. TRAF3/CYLD mutations identify a distinct subset of human papillomavirus-associated head and neck squamous cell carcinoma. Cancer. 2017;123:1778–90.
23. 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;363:24–35.
24. Hayes DN, Neil Hayes D, Van Waes C, Seiwert TY. Genetic landscape of human papillomavirus–associated head and neck cancer and comparison to tobacco-related tumors. J Clin Oncol. 2015;33:3227–34.
25. Mirghani H, Blanchard P. Treatment de-escalation for HPV-driven oropharyngeal cancer: Where do we stand? Clin Transl Radiat Oncol. 2018;8:4–11.
26. Bhatia A, Burtness B. Human papillomavirus–associated oropharyngeal cancer: defining risk groups and clinical trials. J Clin Orthod. 2015;33:3243–50.
27. Alexandrov LB, Stratton MR. Mutational signatures: the patterns of somatic mutations hidden in cancer genomes. Curr Opin Genet Dev. 2014;24:52–60.
28. Roberts SA, Lawrence MS, Klimczak LJ, Grimm SA, Fargo D, Stojanov P, et al. An APOBEC cytidine deaminase mutagenesis pattern is widespread in human cancers. Nat Genet. 2013;45:970–6.
29. Burns MB, Temiz NA, Harris RS. Evidence for APOBEC3B mutagenesis in multiple human cancers. Nat Genet. 2013;45:977–83.
30. Agrawal N, Frederick MJ, Pickering CR, Bettegowda C, Chang K, Li RJ, et al. Exome sequencing of head and neck squamous cell carcinoma reveals inactivating mutations in NOTCH1. Science. 2011;333:1154–7.
31. Stransky N, Egloff AM, Tward AD, Kostic AD, Cibulskis K, Sivachenko A, et al. The mutational landscape of head and neck squamous cell carcinoma. Science. 2011;333:1157–60.
32. Lui VWY, Hedberg ML, Li H, Vangara BS, Pendleton K, Zeng Y, et al. Frequent mutation of the PI3K pathway in head and neck cancer defines predictive biomarkers. Cancer Discov. 2013;3:761–9.
33. Martin D, Abba MC, Molinolo AA, Vitale-Cross L, Wang Z, Zaida M, et al. The head and neck cancer cell oncogenome: a platform for the development of precision molecular therapies. Oncotarget. 2014;5:8906–23.
34. Theurer JA, Stecho W, Yoo J, Kwan K, Wehrli B, Harry V, et al. Feasibility of targeting PIK3CA mutations in head and neck squamous cell carcinoma. Pathol Oncol Res. 2016;22:35–40.
35. Wu ES, Park JY, Zeitouni JA, Gomez CR, Reis IM, Zhao W, et al. Effect of actionable somatic mutations on racial/ethnic disparities in head and neck cancer prognosis. Head Neck. 2016;38:1234–41.
36. Grossman RL, Heath AP, Ferretti V, Varmus HE, Lowy DR, Kibbe WA, et al. Toward a shared vision for cancer genomic data. N Engl J Med. 2016;375:1109–12.
37. Hedberg ML, Goh G, Chiosea SI, Bauman JE, Freilino ML, Zeng Y, et al. Genetic landscape of metastatic and recurrent head and neck squamous cell carcinoma. J Clin Invest. 2016;126:169–80.
38. Weinberger PM, Yu Z, Haffty BG, Kowalski D, Harigopal M, Brandsma J, et al. Molecular classification identifies a subset of human papilloma virus—associated oropharyngeal cancers with favorable prognosis. J Clin Oncol. 2006;24:736–47.
39. Cao S, Wendl MC, Wyczalkowski MA, Wylie K, Ye K, Jayasinghe R, et al. Divergent viral presentation among human tumors and adjacent normal tissues. Sci Rep. 2016;6:28294.

40. Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. Sci Signal. 2013;6:l1.

41. Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. Cancer Discov. 2012;2:401–4.

42. Wang X, Abudu A, Son S, Dang Y, Venta PJ, Zheng Y-H. Analysis of human APOBEC3H haplotypes and anti-human immunodeficiency virus type 1 activity. J Virol. 2011;85:3142–52.

43. Huang K-L, Mashl RJ, Wu Y, Ritter DI, Wang J, Oh C, et al. Pathogenic germline variants in 10,389 adult cancers. Cell. 2018;173:355–70.e14.

44. Welcker M, Clurman BE. FBW7 ubiquitin ligase: a tumour suppressor at the crossroads of cell division, growth and differentiation. Nat Rev Cancer. 2008;8:83–93.

45. 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;5:49.

46. Burns MB, Lackey L, Carpenter MA, Rathore A, Land AM, Leonard B, et al. APOBEC3B is an enzymatic source of mutation in breast cancer. Nature. 2013;494:366–70.

47. Shi K, Carpenter MA, Banerjee S, Shaban NM, Kurahashi K, Salamango DJ, et al. Structural basis for targeted DNA cytosine deamination and mutagenesis by APOBEC3A and APOBEC3B. Nat Struct Mol Biol. 2017;24:131–9.

48. Glaser AP, Fantini D, Wang Y, Yu Y, Rimar KJ, Podojil JR, et al. APOBEC-mediated mutagenesis in urothelial carcinoma is associated with improved survival, mutations in DNA damage response genes, and immune response. Oncotarget. 2018;9:4537–48.

49. Vieira VC, Soares MA. The role of cytidine deaminases on innate immune responses against human viral infections. Biomed Res Int. 2013;2013:1–18.

50. Nik-Zainal S, Wedge DC, Alexandrov LB, Petljak M, Butler AP, Bolli N, et al. Association of a germline copy number polymorphism of APOBEC3A and APOBEC3B with burden of putative APOBEC-dependent mutations in breast cancer. Nat Genet. 2014;46:487–91.

51. Roberts SA, Sterling J, Thompson C, Harris S, Mav D, Shah R, et al. Clustered mutations in yeast and in human cancers can arise from damaged long single-strand DNA regions. Mol Cell. 2012;46:424–35.

52. Kazanov MD, Roberts SA, Polak P, Stamatoyannopoulos J, Klimczak LJ, Gordenin DA, et al. APOBEC-induced cancer mutations are uniquely enriched in early-replicating, gene-dense, and active chromatin regions. Cell Rep. 2015;13:1103–9.

53. Venkatesan S, Rosenthal R, Kanu N, McGranahan N, Bartek J, Quezada SA, et al. APOBEC mutagenesis in drug resistance and immune escape in HIV and cancer evolution. Ann Oncol. 2018. https://doi.org/10.1093/annonc/mdy003.

54. Krupar R, Hartl M, Wirsching K, Dietmaier W, Strutz J, Hofstaedter F. Comparison of HPV prevalence in HNSCC patients with regard to regional and socioeconomic factors. Eur Arch Otorhinolaryngol. 2014;271:1737–45.

55. Jiron J, Sethi S, Ali-Fehmi R, Franceschi S, Struijk L, van Doorn L-J, et al. Racial disparities in human papillomavirus (HPV) associated head and neck cancer. Am J Otolaryngol. 2014;35:147–53.

56. Flores-Obando RE, Gollin SM, Ragin CC. Polymorphisms in DNA damage response genes and head and neck cancer risk. Biomarkers. 2010;15:379–99.

57. Ramakodi MP, Devarajan K, Blackman E, Gibbs D, Luce D, Deloumeaux J, et al. Integrative genomic analysis identifies ancestry-related expression quantitative trait loci on DNA polymerase β and supports the association of genetic ancestry with survival disparities in head and neck squamous cell carcinoma. Cancer. 2017;123:849–60.

58. Ramakodi MP, Kulathinal RJ, Chung Y, Serebriiskii I, Liu JC, Ragin CC. Ancestral-derived effects on the mutational landscape of laryngeal cancer. Genomics. 2016;107:76–82.

59. Chaussade C, Cho K, Mawson C, Rewcastle GW, Shepherd PR. Functional differences between two classes of oncogenic mutation in the PIK3CA gene. Biochem Biophys Res Commun. 2009;381:577–81.

60. Hoopes JI, Cortez LM, Mertz TM, Malc EP, Mieczkowski PA, Roberts SA. APOBEC3A and APOBEC3B preferentially deaminate the lagging strand template during DNA replication. Cell Rep. 2016;14:1273–82.

61. Davis RJ, Welcker M, Clurman BE. Tumor suppression by the Fbw7 ubiquitin ligase: mechanisms and opportunities. Cancer Cell. 2014;26:455–64.

62. Orlicky S, Tang X, Willems A, Tyers M, Sicheri F. Structural basis for phosphodependent substrate selection and orientation by the SCFCdc4 ubiquitin ligase. Cell. 2003;112:243–56.

63. Teng C-L, Hsieh Y-C, Phan L, Shin J, Gully C, Velazquez-Torres G, et al. FBXW7 is involved in Aurora B degradation. Cell Cycle. 2012;11:4059–68.

64. Kwon Y-W, Kim I-J, Wu D, Lu J, Stock WA Jr, Liu Y, et al. Pten regulates Aurora-A and cooperates with Fbxw7 in modulating radiation-induced tumor development. Mol Cancer Res. 2012;10:834–44.

65. Kox C, Zimmermann M, Stanulla M, Leible S, Schrappe M, Ludwig W-D, et al. The favorable effect of activating NOTCH1 receptor mutations on long-term outcome in T-ALL patients treated on the ALL–BFM 2000 protocol can be separated from FBXW7 loss of function. Leukemia. 2010;24:2005–13.

66. McGranahan N, Favero F, de Bruin EC, Birkbak NJ, Szallasi Z, Swanton C. Clonal status of actionable driver events and the timing of mutational processes in cancer evolution. Sci Transl Med. 2015;7:283ra54.

67. de Bruin EC, McGranahan N, Mitter R, Salm M, Wedge DC, Yates L, et al. Spatial and temporal diversity in genomic instability processes defines lung cancer evolution. Science. 2014;346:251–6.

68. Fousteri M, Mullenders LHF. Transcription-coupled nucleotide excision repair in mammalian cells: molecular mechanisms and biological effects. Cell Res. 2008;18:73–84.

69. Stamatoyannopoulos JA, Adzhubei I, Thurman RE, Kryukov GV, Mirkin SM, Sunyaev SR. Human mutation rate associated with DNA replication timing. Nat Genet. 2009;41:393–5.

70. Montgomery ND, Selitsky SR, Patel NM, Hayes DN, Parker JS, Weck KE. Identification of germline variants in tumor genomic sequencing analysis. J Mol Diagn. 2018;20:123–5.

71. Sun W, Bunn P, Jin C, Little P, Zhabotynsky V, Perou CM, et al. The association between copy number aberration, DNA methylation and gene expression in tumor samples. Nucleic Acids Res. 2018. https://doi.org/10.1093/nar/gky131.

72. Mroz EA, Tward AD, Tward AM, Hammon RJ, Ren Y, Rocco JW. Intra-tumor genetic heterogeneity and mortality in head and neck cancer: analysis of data from the Cancer Genome Atlas. PLoS Med. 2015;12:e1001786.

73. Lewis JS Jr, Beadle B, Bishop JA, Chernock RD, Colasacco C, Lacchetti C, et al. Human papillomavirus testing in head and neck carcinomas: guideline from the college of american pathologists. Arch Pathol Lab Med [Internet]. 2017. https://doi.org/10.5858/arpa.2017-0286-CP.

74. Lawrence M, Gentleman R, Carey V. rtracklayer: an R package for interfacing with genome browsers. Bioinformatics. 2009;25:1841–2.

75. vcf2maf. Github; 2017. https://github.com/mskcc/vcf2maf/releases/tag/v1.6.13. [Accessed 18 June 2017].

76. Martincorena I, Raine KM, Gerstung M, Dawson KJ, Haase K, Van Loo P, et al. Universal patterns of selection in cancer and somatic tissues. Cell. 2017. https://doi.org/10.1016/j.cell.2017.09.042.

77. Fu Y, Ito F, Zhang G, Fernandez B, Yang H, Chen XS. DNA cytosine and methylcytosine deamination by APOBEC3B:

enhancing methylcytosine deamination by engineering APOBEC3B. Biochem J. 2015;471:25–35.

78. Adolph MB, Love RP, Feng Y, Chelico L. Enzyme cycling contributes to efficient induction of genome mutagenesis by the cytidine deaminase APOBEC3B. Nucleic Acids Res. 2017;45:11925–40.

79. Chelico L, Pham P, Calabrese P, Goodman MF. APOBEC3G DNA deaminase acts processively $3' \rightarrow 5'$ on single-stranded DNA. Nat Struct Mol Biol. 2006;13:392.