APOBEC-Mediated Cytosine Deamination Links *PIK3CA* Helical Domain Mutations to Human Papillomavirus-Driven Tumor Development

Stephen Henderson,[1,2,4] Ankur Chakravarthy,[1,4] Xiaoping Su,[3] Chris Boshoff,[1,5] and Tim Robert Fenton[1,*]
[1]Department of Oncology, UCL Cancer Institute, University College London, London WC1E 6BT, UK
[2]Bill Lyons Informatics Centre, UCL Cancer Institute, University College London, London WC1E 6BT, UK
[3]Department of Bioinformatics and Computational Biology, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA
[4]Co-first author
[5]Present address: Pfizer Oncology, 10555 Science Center Drive, La Jolla, CA 92121, USA
*Correspondence: t.fenton@ucl.ac.uk
http://dx.doi.org/10.1016/j.celrep.2014.05.012

SUMMARY

APOBEC3B cytosine deaminase activity has recently emerged as a significant mutagenic factor in human cancer. APOBEC activity is induced in virally infected cells, and APOBEC signature mutations occur at high frequency in cervical cancers (CESC), over 99% of which are caused by human papillomavirus (HPV). We tested whether APOBEC-mediated mutagenesis is particularly important in HPV-associated tumors by comparing the exomes of HPV+ and HPV− head and neck squamous cell carcinomas (HNSCCs) sequenced by The Cancer Genome Atlas project. As expected, HPV− HNSCC displays a smoking-associated mutational signature, whereas our data suggest that reduced exposure to exogenous carcinogens in HPV+ HNSCC creates a selective pressure that favors emergence of tumors with APOBEC-mediated driver mutations. Finally, we provide evidence that APOBEC activity is responsible for the generation of helical domain hot spot mutations in the *PIK3CA* gene across multiple cancers. Our findings implicate APOBEC activity as a key driver of *PIK3CA* mutagenesis and HPV-induced transformation.

INTRODUCTION

The *apo*lipoprotein *B* mRNA *e*diting enzyme *c*atalytic polypeptide-like (APOBEC) family of enzymes catalyze the deamination of cytosine bases in nucleic acids, resulting in conversion of the target cytosine to uracil and thus an alteration of the substrate DNA or RNA sequence (Conticello, 2008). Depending upon the repair pathway used to remove and replace the resulting uracil base, cytosine deamination in DNA most commonly results in mutations to thymine or guanine (Roberts et al., 2012; Simonelli et al., 2005). One APOBEC family member, the

activation-induced deaminase (AID), is expressed in B lymphocytes, wherein it participates in the processes of somatic hypermutation and class switch recombination key to antibody generation (Pham et al., 2005), whereas another, APOBEC3G, participates in the response to infection by HIV and other retroviruses, acting on viral cDNA to elicit mutagenesis of the viral genome (Harris et al., 2003; Mangeat et al., 2003; Zhang et al., 2003). APOBEC3 enzymes also participate in the response to DNA viruses, including human papillomavirus (HPV) and hepatitis virus B (HBV) (Vartanian et al., 2008, 2010). A number of studies have suggested a role for APOBECs in carcinogenesis: both AID and APOBEC1 promote tumorigenesis in transgenic mouse models, and AID upregulation by *Helicobacter pylori* has also been shown to generate *TP53* mutations in gastric epithelial cells (Matsumoto et al., 2007; Okazaki et al., 2003; Yamanaka et al., 1995). Analysis of sequence data from human tumors revealed clusters of strand-specific cytosine mutations suggestive of APOBEC editing of cellular DNA, a process termed kataegis (Nik-Zainal et al., 2012), likely resulting from APOBEC-mediated deamination of single-stranded DNA generated during repair of double-strand DNA breaks (Nik-Zainal et al., 2012; Roberts et al., 2012).

Many of the cytosine mutations characteristic of kataegis fit the consensus target site for a subset of APOBECs, collectively termed "TC-specific APOBECs," that display preference for thymine immediately 5′ to the target cytosine (Roberts et al., 2012). Studies on deamination of retroviral cDNA and on recombinant APOBEC3B indicate a preference for adenine, thymine, or guanine immediately 3′ to the target cytosine (Bishop et al., 2004; Burns et al., 2013b). However, C > T mutations also occur via the spontaneous deamination of methylated CpG sites, a mutational process that has been linked to aging (Alexandrov et al., 2013a). The use of "TCW mutations" (C > T and C > G mutations at T-C-A/T [TCW] trinucleotides) to evaluate APOBEC-mediated mutagenesis avoids this confounding factor (Roberts et al., 2013). Furthermore, TCW mutations are enriched within kataegis clusters and near to rearrangement breakpoints—both indicative of APOBEC involvement (Drier et al., 2013; Nik-Zainal et al., 2012; Roberts et al., 2012). Two recent analyses of such APOBEC signature mutations in whole-exome

CrossMark

(all known coding regions of the genome) sequence data from The Cancer Genome Atlas (TCGA) suggest cytosine deamination catalyzed by APOBEC enzymes is a mutagenic mechanism in multiple cancer types (Burns et al., 2013b; Roberts et al., 2013). Of the TC-specific APOBECs, APOBEC3B has been posited to play the prominent role in this process, due to its localization to the nuclei of interphase cells and correlation with cytosine mutations in breast and ovarian cancer (Burns et al., 2013a, 2013b; Leonard et al., 2013). Among those cancers with the highest *APOBEC3B* mRNA levels and greatest fraction of TCW mutations are cervical and head and neck squamous cell carcinomas (CESC and HNSCC, respectively). CESC and HNSCC were also among 16 cancer types reported by Stratton and colleagues to display the putative APOBEC mutational signature (Alexandrov et al., 2013a). Over 99% of CESCs and a subset of HNSCCs are caused by high-risk variants of HPV (Chaturvedi et al., 2011; Ramqvist and Dalianis, 2010; Walboomers et al., 1999). The majority of HNSCCs, meanwhile, are unrelated to HPV and typically occur in heavy smokers. Of relevance to our findings, sequencing of CESC and HNSCC has revealed that the *PIK3CA* proto-oncogene is mutated at high frequency in these tumors. Indeed, in some HPV+ HNSCCs, it was the only mutation observed among a large panel of known cancer genes (Lechner et al., 2013b; Lui et al., 2013; Ojesina et al., 2014). The established role of APOBEC enzymes in antiviral responses, together with the finding of such high levels of APOBEC signature mutations in CESC has led to the proposal that APOBEC-mediated mutagenesis may play a particular role in the development of virally driven cancers (Alexandrov et al., 2013a; Kuong and Loeb, 2013; Ojesina et al., 2014; Roberts et al., 2013). Analyzing HNSCC enabled direct comparison of exomes from 40 HPV+ and 258 HPV− tumors arising at similar sites from a single study. We were therefore able to test our hypothesis that it is the HPV+ tumors in which APOBEC activity accounts for the greatest proportion of mutations and to establish a clear link between HPV-driven tumorigenesis and APOBEC-mediated mutagenesis.

RESULTS

Mutations Characteristic of APOBEC Activity Are Enriched in HPV+ HNSCC

Recent work on the detection of viral gene expression in TCGA samples using RNA sequencing (RNA-seq) data permitted definitive identification of HPV status for 299 samples in the TCGA HNSCC cohort (Chen et al., 2013; Tang et al., 2013). We retrieved RNA-seq and somatic mutation data (see Experimental Procedures) and analyzed the mutational signatures present in these samples. A separate HNSCC cohort was previously shown to harbor four mutational signatures, implicating APOBEC activity, age, tobacco smoking, and exposure to ultraviolet light as the factors responsible for the majority of mutations in these tumors (Alexandrov et al., 2013a). The TCGA HNSCC data set that we analyzed contains only one surface-exposed tumor, a lip squamous cell carcinoma, that we removed from our analysis to eliminate UV as a potential confounding source of C > T mutations (Pfeifer et al., 2005). Using the method of Alexandrov and colleagues (Alexandrov et al.,

2013b), we extracted three stable mutational signatures (Figures 1A and S1A), which closely resemble the APOBEC, age, and smoking-associated signatures previously defined (Alexandrov et al., 2013a). Consistent with our hypothesis that APOBEC-mediated mutagenesis is particularly important in HPV-driven tumors, the application of binomial regression showed that the candidate fractional APOBEC signature per sample was associated with HPV status, but importantly not with age or smoking (Figure 1B). This effect was magnified when analysis was restricted to the more conservative TCW > TTW/TGW (hereafter TCW > TKW) mutation signature used by Roberts and colleagues (Roberts et al., 2013). As expected, smoking levels were higher among the HPV- HNSCCs (Figure S1B), as was the fraction of mutations attributable to the smoking signature (data not shown). We also observed the age-associated mutational signature characterized by C > T mutations at CpG sites. Despite a lower median age in the HPV+ patients (Figure S1B), these tumors actually harbored a higher fraction of these mutations. This could be attributable to increased spontaneous deamination due to the hypermethylation of CpG sites previously noted in HPV+ HNSCC (Lechner et al., 2013a). The TCW mutation signature excludes the TCG component of the APOBEC signature, thus eliminating this potential confounder. We further sought to establish the enrichment of TCW mutation signature by calculating a per-sample enrichment score (ES) (see Experimental Procedures) of C > K mutations at TCW sites within defined sample fractions of the exome as previously reported (Roberts et al., 2013). Consistent with our other estimates of the APOBEC signature, the mean ES is higher in HPV+ HNSCC (Figure 1C; 95% bootstrapped confidence interval [CI]; permutation p value 0.0045). The stronger TCW signature in HPV+ HNSCC is clear when visualizing the sequence context surrounding C > K mutations (Figure 1D). APOBEC hyperediting of viral DNA has been reported in both HPV- and HBV-infected cells (Vartanian et al., 2008, 2010). To establish whether the APOBEC mutagenic signature may be a general feature of virally driven cancers, we analyzed 213 exome sequences from the International Cancer Genome Consortium (ICGC) analysis of hepatitis virus (HBV or hepatitis C virus [HCV])-associated hepatocellular carcinoma (HCC) (http://dcc.icgc.org; Project [LINC-JP] Liver Cancer - NCC, JP). In contrast to HNSCC, we saw no enrichment for the TCW motif at sites of C > K mutations and the overall frequency of TCW mutations was significantly lower than that seen in either HPV+ or HPV− HNSCC (Figure 1D; data not shown).

APOBEC3B Expression Is Elevated in HPV+ HNSCC but Is Weakly Correlated with the Fraction of TCW Mutations in These Tumors

Using TCGA gene expression data generated by RNA-seq, we examined the expression of *APOBEC3B* mRNA in the same samples and found significantly higher expression in HPV+ HNSCC than in HPV− HNSCC (p = 3.66×10^{-11} by Wilcoxon; Figure 2A). Whereas the HPV+ tumors have higher *APOBEC3B* expression and a higher frequency of TCW mutations, expression of *APOBEC3B* is only weakly, albeit significantly, correlated with TCW mutations across all HNSCC samples (p = 0.0127; by Spearman's rho; r = 0.144; Figure 2B). It is clear
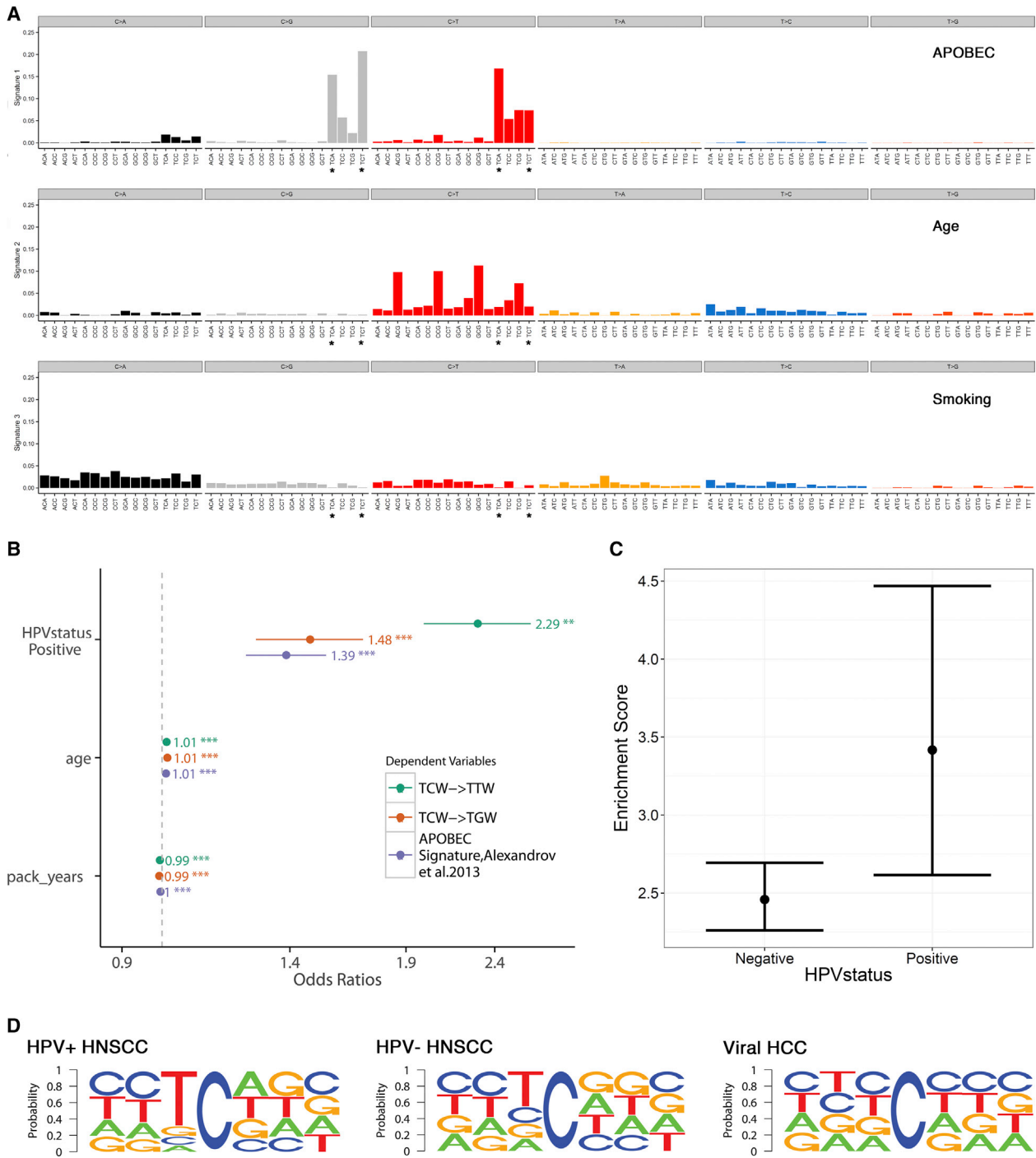
Figure 1. HPV+ HNSCC Harbors a Strong APOBEC-Associated Mutational Signature

(A) Deconvolution of the HNSCC point mutations via nonnegative matrix factorization identifies three signatures closely resembling those previously associated with APOBEC activity, age, and smoking. Asterisks denote TCW mutations.

(B) Multiple binomial regressions of the fractional APOBEC signatures extracted in (A) as well as the fraction of TCW mutations (TCW > TTW and TCW > TGW) show a strong association with HPV status, but not with age or smoking history.

(C) Per-sample enrichment analysis of C > K mutations at TCW sites in the genomic neighborhood relative to non-TCW C > K mutations confirms significant enrichment in HPV+ HNSCCs. The 95% CI was calculated from 10,000 bootstraps of the mean.

(D) Sequence logos illustrate the enrichment of the TCW sequence context at C > K mutations in HPV+ HNSCCs relative to HPV− HNSCCs, whereas no such pattern is evident in viral HCC.
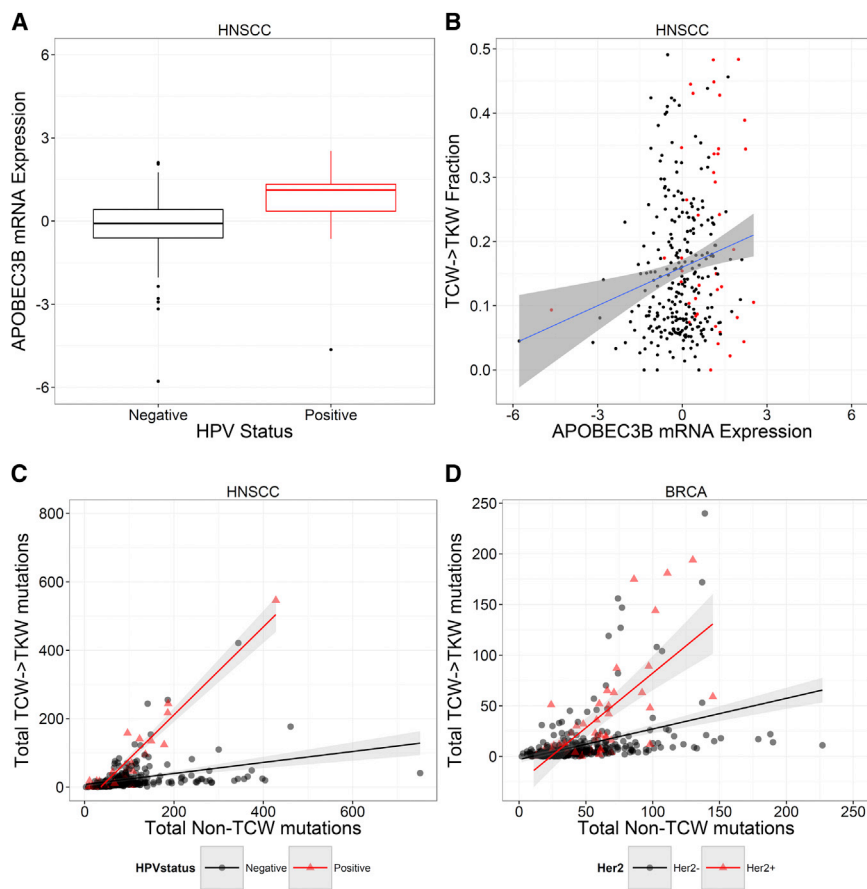
See also Figure S1 and Table S1.

Figure 2. TCW Mutations Are Correlated Weakly with *APOBEC3B* mRNA Levels but Strongly with Overall Mutational Load in HNSCC and BRCA

(A) Box plot displaying *APOBEC3B* mRNA levels in HPV− and HPV+ HNSCC.

(B–D) Scatterplots showing the relationship between (B) *APOBEC3B* mRNA levels and TCW mutations, (C) TCW and non-TCW mutations in HNSCC (HPV+ samples, red triangles; HPV− samples, gray diamonds), and (D) TCW and non-TCW mutations in *HER2*+ (red triangles) and *HER2*− (gray circles) BRCA. Shading around fitted lines represents 95% CI.

See also Figures S2 and S3 and Table S1.

checked for relationships between total mutation load and TCW mutations. In certain HNSCC samples, particularly in the HPV+ subset, TCW mutations represent a substantial fraction of total point mutations, so we examined the relationship between the number of TCW mutations and all other point mutations (non-TCW mutations). The data are best fit (adjusted R^2 = 0.51) by a linear regression model with a positive slope for HPV− HNSCC (slope 0.24 TCW/non-TCW) and a separate, steeper line of fit for HPV+ HNSCC (slope 0.44 TCW/non-TCW), reflecting the higher fraction of TCW mutations seen in the HPV+ tumors (Figure 2C). To establish whether the link between TCW mutations and non-TCW mutations occurs across different tumor types, we analyzed TCGA data for breast adenocarcinoma (BRCA) (n = 458; Parker et al., 2009; Cancer Genome Atlas Network, 2012). In BRCA, it is the *HER2*-amplified subset that displays the highest level of TCW mutations (Roberts et al., 2013). Like HNSCC, we find that, in BRCA, TCW mutations are closely correlated with non-TCW mutations and the relationship is best fit (adjusted R^2 = 0.34) with a separate steeper line of fit for *HER2*+ than *HER2*− tumors (*HER2*− = 0.2; *HER2*+ = 0.49; Figure 2D).

APOBEC Activity Accounts for a High Fraction of Mutations in Putative Cancer-Driver Genes in HPV+ Tumors

The E6 and E7 oncoproteins from high-risk HPV strains disable p53 and pRb tumor suppressor function, causing immortalization and cell cycle re-entry, but are not sufficient to affect cellular transformation. Tumor development in HPV-infected epithelia therefore requires the acquisition of mutations in cellular genes, reflected in the considerable latency between HPV infection and the emergence of carcinoma (McLaughlin-Drubin and Münger, 2009; Psyrri and DiMaio, 2008). To investigate whether APOBEC activity directly contributes to tumorigenesis in HPV+ tumors, we initially focused our analysis on likely driver mutations in
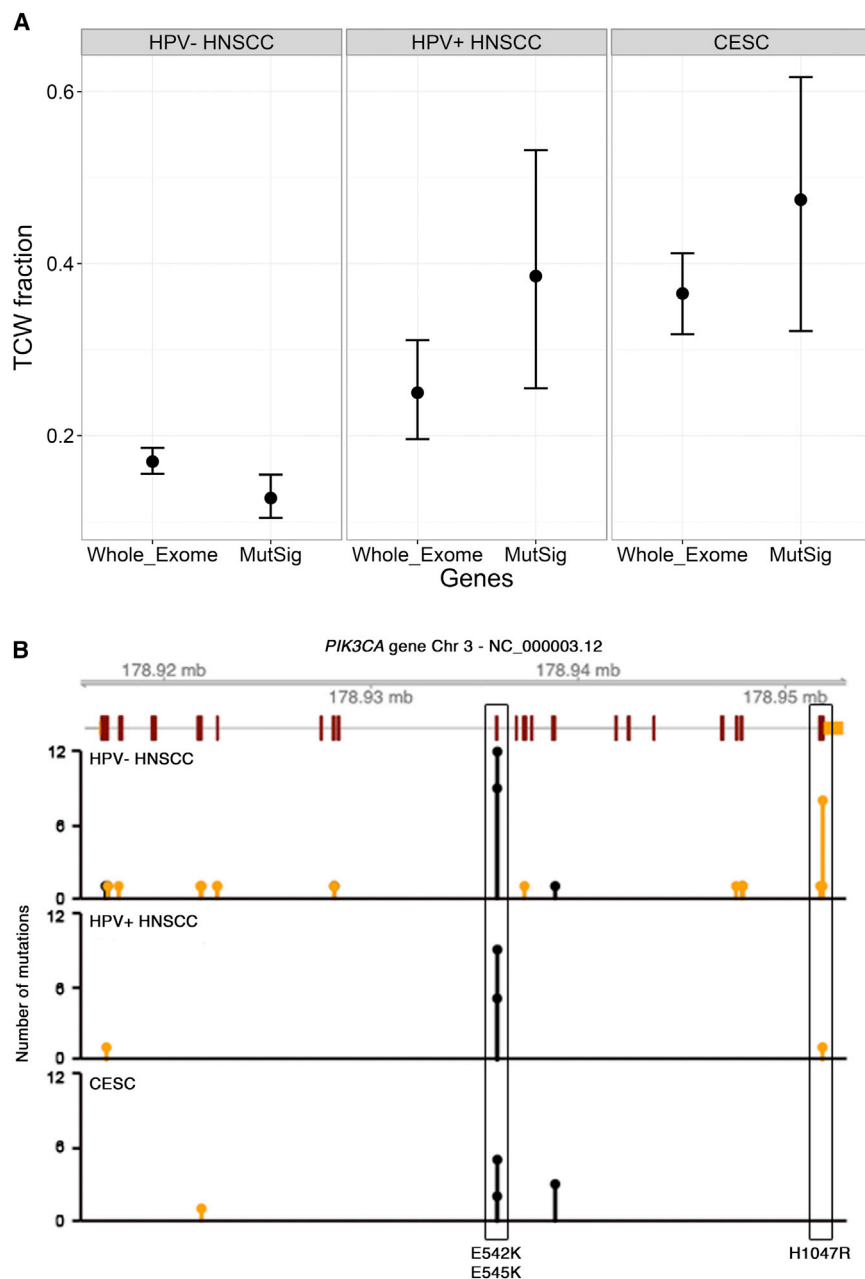
from this plotted relationship that the observed level of *APOBEC3B* can only partially explain the higher fraction of TCW mutations in HPV+ HNSCC. *APOBEC3B* expression is not correlated with the number of background (non-TCW) mutations in HNSCC (data not shown). We also checked the expression of the six other *APOBEC3* family members that reside together with *APOBEC3B* in a gene cluster on chromosome 22. With the exception of *APOBEC3A*, all are significantly upregulated at the mRNA level in the HPV+ samples (Figure S2A), but again, expression is only weakly correlated with TCW mutations (Figure S2B). Furthermore, using viral gene expression data available for 32 of the 40 HPV+ samples (Table S1), we did not observe correlations between levels of the bicistronic transcript encoding the major HPV oncogenes, *E6* and *E7*, and either *APOBEC3B* expression or TCW mutations (Figures S3A and S3B).

TCW Mutations Are Correlated with the Background Mutational Load

From our analysis of *APOBEC3B* and *E6/E7* mRNA levels, it appears possible that another factor plays the dominant role in determining the fraction of APOBEC-mediated mutagenesis in HNSCC. Because HPV+ HNSCC has a significantly lower number of mutations per exome than HPV− HNSCC (Figure S1C; Agrawal et al., 2011; Stransky et al., 2011), we

A

Figure 3. *PIK3CA* Is Mutated at Helical Domain Hot Spot TCW sites in HPV+ HNSCC and CESC

(A) 95% CI of the TCW mutation fractions in HNSCC and CESC among MutSig genes.

(B) Location and TCW status of mutations in *PIK3CA* among HNSCC and CESC. Black lollipops, TCW mutations; yellow lollipops, non-TCW mutations.

See also Figure S4 and Tables S2 and S3.

tumor development. Upon ranking all genes by their TCW mutation content, as expected, we observed a bias toward large genes within the top ten; however, we were struck by the presence of the *PIK3CA* proto-oncogene, which encodes the class 1A phosphoinositide-3-kinase p110α catalytic subunit, in all three lists (Table S3). This was of particular interest to us, because the major oncogenic *PIK3CA* mutations have been well characterized and occur at high frequency in both HPV+ HNSCC and CESC (Lechner et al., 2013b; Lui et al., 2013; Ojesina et al., 2014; Vogt et al., 2007). The most-common and well-studied mutations in *PIK3CA* lie in two "hot spots": E542K and E545K in the helical domain and H1047R in the kinase domain. Upon inspection of our mutation data, we realized that both E542K (c.1624G > A) and E545K (c.1633G > A) are caused by TCW mutations on the opposing strand, whereas H1047 (typically c.3140A > G) mutations are not of this type (Figure S4). We found that, in HPV+ HNSCC and CESC, the *PIK3CA* mutations are almost exclusively (22 of 25) of the TCW type, with the majority occurring at the helical domain hot spots, whereas no H1047R mutations were seen. In HPV− HNSCC, only 22 of 48 *PIK3CA* mutations were TCW and we observed eight H1047R mutations (Figure 3B). Thus, a significant skew toward TCW mutations in *PIK3CA* is evident in the HPV+ tumors (chi-square test p = 0.001).

HNSCC and CESC as determined using MutSig, an algorithm developed to detect recurrently mutated genes that accounts for various factors that affect how often a particular gene may be mutated by chance (Table S2; Broad Institute TCGA Genome Data Analysis Center, 2013; Lawrence et al., 2013). In HPV+ HNSCC and CESC, the high exome-wide fraction of TCW mutations is also seen among candidate driver genes, whereas in HPV− HNSCC, the fraction of TCW mutations among MutSig genes is again much lower than that seen in the HPV+ tumors (Figure 3A).

This observation prompted us to look in further detail at the genes that might be mutated by APOBEC and contributing to

PIK3CA Helical Mutations Are Associated with the APOBEC Signature in Multiple Cancer Types

Our observations in HPV+ HNSCC and CESC suggest that APOBEC activity may cause helical domain mutations in *PIK3CA*. To test this hypothesis, we investigated the *PIK3CA* mutation spectrum in multiple cancers by comparing the incidence of helical (c.1624G > A; c.1633G > A) versus kinase (H1047) hot spot mutations. Consistent with our proposition that APOBEC mediates helical mutations, the tumor
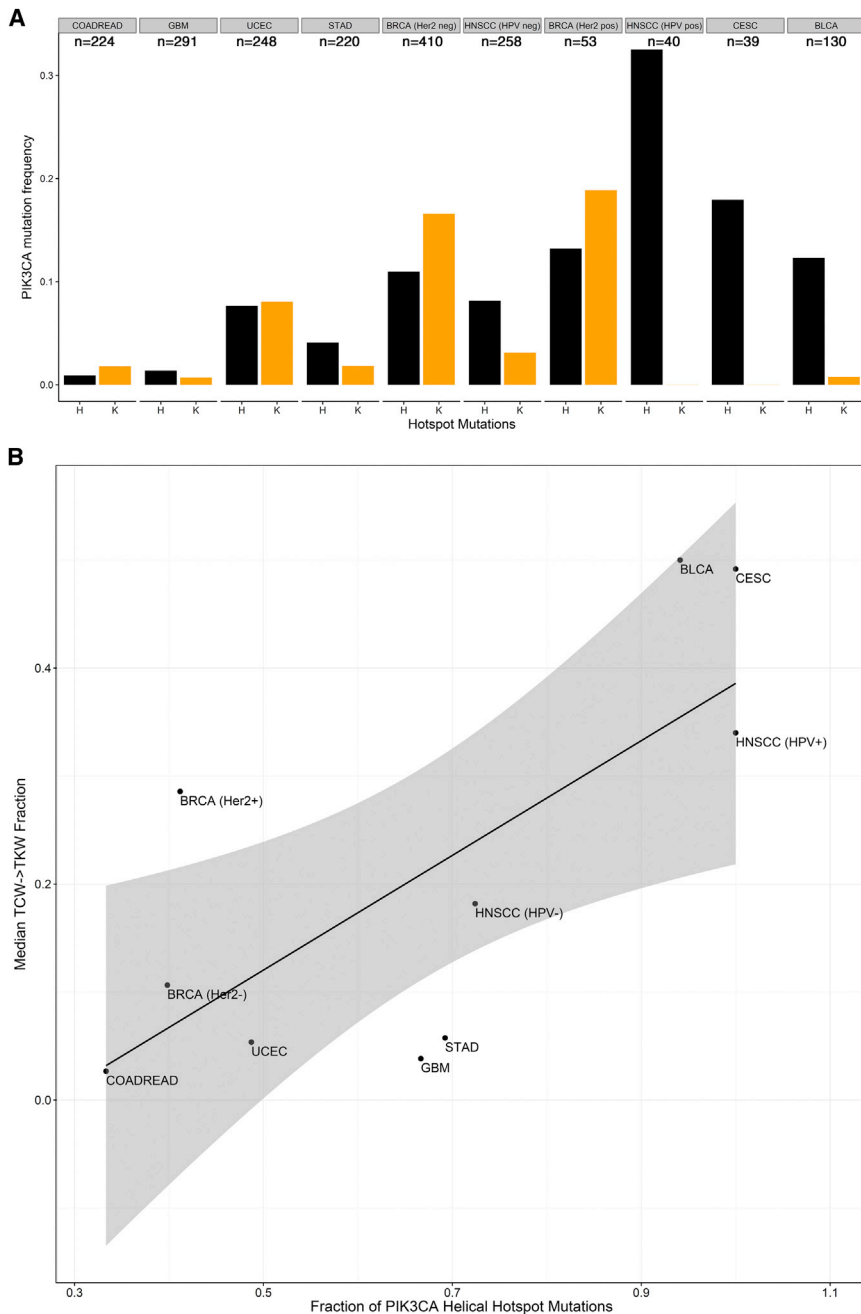
Figure 4. Tumors with Enrichment for APOBEC Signature Mutations Harbor a High Proportion of *PIK3CA* Helical Domain Mutations

(A) Relative levels of *PIK3CA* helical and kinase domain hot spot mutations (as a fraction of the total number of samples in each cancer type) across multiple cancer types. BLCA, bladder carcinoma; COADREAD, colorectal adenocarcinoma; GBM, glioblastoma multiforme; H, helical; K, kinase hot spot mutations; STAD, stomach adenocarcinoma; UCEC, uterine corpus endometrioid carcinoma.

(B) Scatterplot of TCW mutation fraction versus the fraction of hot spot mutations accounted for by helical mutations. Shading around fitted lines represents 95% CI.

suggests that, rather than occurring by chance, these are indeed mediated by APOBEC.

DISCUSSION

Our comparison of HPV+ and HPV− HNSCC suggests a strong link between HPV and APOBEC-mediated mutagenesis of cellular genes. Furthermore, the lack of an association between HBV/HCV and TCW mutations suggests the high level of these mutations seen in HPV-associated tumors is not simply a vestige of APOBEC enzyme activity associated with persistent viral infection. The relatively high fraction of APOBEC signature mutations in HNSCC, irrespective of HPV status, suggests that APOBEC-mediated mutagenesis is a feature of certain tumor types and that this activity is either potentiated or selected for in the presence of HPV.

We were surprised that increased expression of several APOBEC3 genes including *APOBEC3B* in HPV+ HNSCC could only partially explain the increased level of TCW mutations in the HPV+ tumors. This contrasts somewhat with BRCA, in which *APOBEC3B* expression is more strongly correlated with C > T mutations and overall mutational loads, but is consistent with findings in CESC (Burns et al., 2013a; Ojesina et al., 2014). Nonetheless, the observed fraction of TCW mutations in tumor samples is a late snapshot in tumor progression. Whereas we do not see evidence that viral gene expression (*E6/E7*) or APOBEC levels at this late point are related to the accumulated TCW fraction, we cannot rule out that we are observing these fluctuating variables past the crucial phase during which the mutations are occurring. This scenario is consistent with a transient hypermutator model recently

types with the highest APOBEC signature and highest *APOBEC3B* expression levels (bladder carcinoma [BLCA], CESC, and HPV+ HNSCC; Burns et al., 2013b; Roberts et al., 2013) showed the highest ratio of helical-to-kinase mutations (Figure 4A). We next related the *PIK3CA* mutation spectrum to the fraction of TCW mutations in each exome and, again consistent with our model, observed a significant enrichment for TCW mutations in samples harboring helical domain mutations (Figure 4B; Spearman's rho 0.71; p 0.01). Our finding that *PIK3CA* mutations occur more frequently at the TCW helical sites in TCW-mutation-enriched tumors strongly

proposed for APOBEC involvement in somatic mutagenesis, in which the authors note that sustained mutagenic activity is incompatible with cell viability and would thus be selected against following the accumulation of key driver mutations (Roberts and Gordenin, 2014). Alternatively, rather than an increased rate of APOBEC-mediated mutagenesis in HPV+ versus HPV− cells, the elevated fraction of TCW mutations in HPV+ HNSCC may result from a reduced mutational load of other kinds. HPV− HNSCC typically occurs in heavy smokers, whereas HPV+ HNSCC is often seen in nonsmokers and harbors less mutations per tumor (Agrawal et al., 2011; Goon et al., 2009; Stransky et al., 2011); thus in HPV+ HNSCC, the lack of exogenous mutagens from tobacco smoke may confer a stronger selection pressure for APOBEC-mediated mutagenesis.

Regardless of mechanism, APOBEC-mediated mutagenesis is not just a curious passenger mutation phenomenon; rather, our finding that it mediates driver mutations, including known oncogenic *PIK3CA* mutations, suggests it is a vital factor in HPV-driven tumor development. Furthermore, our data support a model in which APOBEC activity is the major source of *PIK3CA* mutations in multiple cancers, whereas in tumors with low APOBEC activity, *PIK3CA* is equally likely to be mutated at the kinase domain hot spot and the helical domain mutations seen may be due to other mutational processes that cause $C > T$ mutations. Whereas APOBECs have previously been implicated in mutagenesis of putative driver gene sets and both AID and APOBEC3B cause mutations in *TP53* in cultured cells (Burns et al., 2013a; Matsumoto et al., 2007; Roberts et al., 2013), this report links their activity to specific oncogenic mutations in tumors. Our data indicate that whereas mutations in the helical and kinase domains of *PIK3CA* cause activation via different mechanisms (Zhao and Vogt, 2008), the mutational processes at play in the tumor may in fact be the major determinant of which hot spot is mutated. In summary, we have uncovered a critical role for APOBEC-mediated mutagenesis in HPV-driven tumor development and identified a specific oncogenic event: the generation of *PIK3CA* helical domain mutations, through which this is mediated, both in HPV+ tumors and in other cancers that show evidence of high APOBEC activity.

EXPERIMENTAL PROCEDURES

Data Retrieval
BRCA and HCC sequence data were from the Broad Institute's Genome Data Analysis Center (GDAC) portal (https://confluence.broadinstitute.org/display/GDAC/MAF+Dashboard) and the ICGC data portal (ftp://data.dcc.icgc.org/), respectively. Mutation data for all other tumor types (and RNA-seqV2 data for HNSCC) were retrieved from the TCGA data portal (https://tcga-data.nci.nih.gov). HPV status of HNSCC samples as determined by RNA-seq was obtained from Tang et al. (2013). *PIK3CA* mutation data were from cBioPortal for Cancer Genomics (http://www.cbioportal.org/). Duplicate entries were removed, and data were reduced to point mutations before further processing.

PAM50 annotations for the BRCA samples were retrieved from TCGA portal at https://tcga-data.nci.nih.gov/docs/publications/brca_2012/. Analysis was restricted to 458 samples for which the PAM50 subtype was cancer-associated (Parker et al., 2009; Cancer Genome Atlas Network, 2012).

Point mutation data (including hg19 coordinates) for samples were retrieved from the maf files, and the sequence context 3 bp up and downstream of each mutation was retrieved using the Bioconductor BSgenome.Hsapiens. UCSC.hg19 annotation package.

Mutational Signatures
Mutational signatures were deconvoluted using the Wellcome Trust Sanger Institute Mutational Signature Framework (Alexandrov et al., 2013b), a method based on nonnegative matrix factorization. The optimal number of mutational signatures to extract was determined by searching for maximum stability and minimum Frobenius reconstruction errors across two to eight candidate signatures using 400 iterations each, whereas the optimal three signatures shown were subsequently extracted using 1,200 iterations of the algorithm.

Per-Sample Enrichment
ESs were calculated as previously described (Roberts et al., 2013) on a per-sample basis by comparing the fraction of $C > K$ mutations in and out of the TCW sequence context to the ratio of cytosines and guanines in and out of the TCW/WGA sequence context in sample fractions determined using 41 nucleotide windows centered on mutant bases. As the ES score was strongly skewed, the 95% confidence intervals were defined using 10,000 iterations of bootstrapping, and p values for enrichment in HPV+ HNSCCs were calculated using 10,000 permutations (NB t tests on mean difference of ES were also significant).

Driver Mutations
Lists of genes identified as significantly mutated using MutSig CV0.9 for HNSCC (https://www.broadinstitute.org/cancer/cga/mutsig; Broad Institute TCGA Genome Data Analysis Center, 2013; Lawrence et al., 2013) and MutSig vS2N for CESC were obtained from the relevant Firehose reports on the GDAC portal. Significantly mutated genes were defined as those with a q value of 0.1 or less.

The list of genes in which missense or nonsense mutations have been causally implicated in cancer progression was downloaded from the Cancer Gene Census website (http://www.sanger.ac.uk/research/projects/cancergenome/census.html).

For code used to analyze all data, see Supplemental Experimental Procedures.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, four figures, and three tables and can be found with this article online at http://dx.doi.org/10.1016/j.celrep.2014.05.012.

AUTHOR CONTRIBUTIONS

S.H., A.C., C.B., and T.R.F. designed the research. S.H., A.C., and T.R.F. analyzed the data and wrote the paper. S.H. and A.C. performed the experiments, and X.S. contributed data.

REFERENCES

Agrawal, N., Frederick, M.J., Pickering, C.R., Bettegowda, C., Chang, K., Li, R.J., Fakhry, C., Xie, T.-X., Zhang, J., Wang, J., et al. (2011). Exome

sequencing of head and neck squamous cell carcinoma reveals inactivating mutations in NOTCH1. Science 333, 1154–1157.

Alexandrov, L.B., Nik-Zainal, S., Wedge, D.C., Aparicio, S.A.J.R., Behjati, S., Biankin, A.V., Bignell, G.R., Bolli, N., Borg, A., Børresen-Dale, A.-L., et al.; Australian Pancreatic Cancer Genome Initiative; ICGC Breast Cancer Consortium; ICGC MMML-Seq Consortium; ICGC PedBrain (2013a). Signatures of mutational processes in human cancer. Nature 500, 415–421.

Alexandrov, L.B., Nik-Zainal, S., Wedge, D.C., Campbell, P.J., and Stratton, M.R. (2013b). Deciphering signatures of mutational processes operative in human cancer. Cell Reports 3, 246–259.

Bishop, K.N., Holmes, R.K., Sheehy, A.M., Davidson, N.O., Cho, S.-J., and Malim, M.H. (2004). Cytidine deamination of retroviral DNA by diverse APOBEC proteins. Curr. Biol. 14, 1392–1396.

Broad Institute TCGA Genome Data Analysis Center B.-T.-G. (2013). Mutation Analysis (MutSigCV v0.9). http://gdac.broadinstitute.org/runs/analyses__2013_05_23/reports/cancer/HNSC/MutSigNozzleReportCV/nozzle.html.

Burns, M.B., Lackey, L., Carpenter, M.A., Rathore, A., Land, A.M., Leonard, B., Refsland, E.W., Kotandeniya, D., Tretyakova, N., Nikas, J.B., et al. (2013a). APOBEC3B is an enzymatic source of mutation in breast cancer. Nature 494, 366–370.

Burns, M.B., Temiz, N.A., and Harris, R.S. (2013b). Evidence for APOBEC3B mutagenesis in multiple human cancers. Nat. Genet. 45, 977–983.

Cancer Genome Atlas Network (2012). Comprehensive molecular portraits of human breast tumours. Nature 490, 61–70.

Chaturvedi, A.K., Engels, E.A., Pfeiffer, R.M., Hernandez, B.Y., Xiao, W., Kim, E., Jiang, B., Goodman, M.T., Sibug-Saber, M., Cozen, W., et al. (2011). Human papillomavirus and rising oropharyngeal cancer incidence in the United States. J. Clin. Oncol. 29, 4294–4301.

Chen, Y., Yao, H., Thompson, E.J., Tannir, N.M., Weinstein, J.N., and Su, X. (2013). VirusSeq: software to identify viruses and their integration sites using next-generation sequencing of human cancer tissue. Bioinformatics 29, 266–267.

Conticello, S.G. (2008). The AID/APOBEC family of nucleic acid mutators. Genome Biol. 9, 229.

Drier, Y., Lawrence, M.S., Carter, S.L., Stewart, C., Gabriel, S.B., Lander, E.S., Meyerson, M., Beroukhim, R., and Getz, G. (2013). Somatic rearrangements across cancer reveal classes of samples with distinct patterns of DNA breakage and rearrangement-induced hypermutability. Genome Res. 23, 228–235.

Goon, P.K., Stanley, M.A., Ebmeyer, J., Steinsträsser, L., Upile, T., Jerjes, W., Bernal-Sprekelsen, M., Görner, M., and Sudhoff, H.H. (2009). HPV & head and neck cancer: a descriptive update. Head Neck Oncol. 1, 36.

Harris, R.S., Bishop, K.N., Sheehy, A.M., Craig, H.M., Petersen-Mahrt, S.K., Watt, I.N., Neuberger, M.S., and Malim, M.H. (2003). DNA deamination mediates innate immunity to retroviral infection. Cell 113, 803–809.

Kuong, K.J., and Loeb, L.A. (2013). APOBEC3B mutagenesis in cancer. Nat. Genet. 45, 964–965.

Lawrence, M.S., Stojanov, P., Polak, P., Kryukov, G.V., Cibulskis, K., Sivachenko, A., Carter, S.L., Stewart, C., Mermel, C.H., Roberts, S.A., et al. (2013). Mutational heterogeneity in cancer and the search for new cancer-associated genes. Nature 499, 214–218.

Lechner, M., Fenton, T., West, J., Wilson, G., Feber, A., Henderson, S., Thirlwell, C., Dibra, H.K., Jay, A., Butcher, L., et al. (2013a). Identification and functional validation of HPV-mediated hypermethylation in head and neck squamous cell carcinoma. Genome Med. 5, 15.

Lechner, M., Frampton, G.M., Fenton, T., Feber, A., Palmer, G., Jay, A., Pillay, N., Forster, M., Cronin, M.T., Lipson, D., et al. (2013b). Targeted next-generation sequencing of head and neck squamous cell carcinoma identifies novel genetic alterations in HPV+ and HPV- tumors. Genome Med. 5, 49.

Leonard, B., Hart, S.N., Burns, M.B., Carpenter, M.A., Temiz, N.A., Rathore, A., Vogel, R.I., Nikas, J.B., Law, E.K., Brown, W.L., et al. (2013). APOBEC3B upregulation and genomic mutation patterns in serous ovarian carcinoma. Cancer Res. 73, 7222–7231.

Lui, V.W.Y., Hedberg, M.L., Li, H., Vangara, B.S., Pendleton, K., Zeng, Y., Lu, Y., Zhang, Q., Du, Y., Gilbert, B.R., et al. (2013). Frequent mutation of the PI3K pathway in head and neck cancer defines predictive biomarkers. Cancer Discov. 3, 761–769.

Mangeat, B., Turelli, P., Caron, G., Friedli, M., Perrin, L., and Trono, D. (2003). Broad antiretroviral defence by human APOBEC3G through lethal editing of nascent reverse transcripts. Nature 424, 99–103.

Matsumoto, Y., Marusawa, H., Kinoshita, K., Endo, Y., Kou, T., Morisawa, T., Azuma, T., Okazaki, I.-M., Honjo, T., and Chiba, T. (2007). Helicobacter pylori infection triggers aberrant expression of activation-induced cytidine deaminase in gastric epithelium. Nat. Med. 13, 470–476.

McLaughlin-Drubin, M.E., and Münger, K. (2009). Oncogenic activities of human papillomaviruses. Virus Res. 143, 195–208.

Nik-Zainal, S., Alexandrov, L.B., Wedge, D.C., Van Loo, P., Greenman, C.D., Raine, K., Jones, D., Hinton, J., Marshall, J., Stebbings, L.A., et al.; Breast Cancer Working Group of the International Cancer Genome Consortium (2012). Mutational processes molding the genomes of 21 breast cancers. Cell 149, 979–993.

Ojesina, A.I., Lichtenstein, L., Freeman, S.S., Pedamallu, C.S., Imaz-Rosshandler, I., Pugh, T.J., Cherniack, A.D., Ambrogio, L., Cibulskis, K., Bertelsen, B., et al. (2014). Landscape of genomic alterations in cervical carcinomas. Nature 506, 371–375.

Okazaki, I.M., Hiai, H., Kakazu, N., Yamada, S., Muramatsu, M., Kinoshita, K., and Honjo, T. (2003). Constitutive expression of AID leads to tumorigenesis. J. Exp. Med. 197, 1173–1181.

Parker, J.S., Mullins, M., Cheang, M.C.U., Leung, S., Voduc, D., Vickery, T., Davies, S., Fauron, C., He, X., Hu, Z., et al. (2009). Supervised risk predictor of breast cancer based on intrinsic subtypes. J. Clin. Oncol. 27, 1160–1167.

Pfeifer, G.P., You, Y.-H., and Besaratinia, A. (2005). Mutations induced by ultraviolet light. Mutat. Res. 571, 19–31.

Pham, P., Bransteitter, R., and Goodman, M.F. (2005). Reward versus risk: DNA cytidine deaminases triggering immunity and disease. Biochemistry 44, 2703–2715.

Psyrri, A., and DiMaio, D. (2008). Human papillomavirus in cervical and head-and-neck cancer. Nat. Clin. Pract. Oncol. 5, 24–31.

Ramqvist, T., and Dalianis, T. (2010). Oropharyngeal cancer epidemic and human papillomavirus. Emerg. Infect. Dis. 16, 1671–1677.

Roberts, S.A., and Gordenin, D.A. (2014). Clustered and genome-wide transient mutagenesis in human cancers: Hypermutation without permanent mutators or loss of fitness. BioEssays, Published online February 26, 2014. http://dx.doi.org/10.1002/bies.201300140.

Roberts, S.A., Sterling, J., Thompson, C., Harris, S., Mav, D., Shah, R., Klimczak, L.J., Kryukov, G.V., Malc, E., Mieczkowski, P.A., et al. (2012). Clustered mutations in yeast and in human cancers can arise from damaged long single-strand DNA regions. Mol. Cell 46, 424–435.

Roberts, S.A., Lawrence, M.S., Klimczak, L.J., Grimm, S.A., Fargo, D., Stojanov, P., Kiezun, A., Kryukov, G.V., Carter, S.L., Saksena, G., et al. (2013). An APOBEC cytidine deaminase mutagenesis pattern is widespread in human cancers. Nat. Genet. 45, 970–976.

Simonelli, V., Narciso, L., Dogliotti, E., and Fortini, P. (2005). Base excision repair intermediates are mutagenic in mammalian cells. Nucleic Acids Res. 33, 4404–4411.

Stransky, N., Egloff, A.M., Tward, A.D., Kostic, A.D., Cibulskis, K., Sivachenko, A., Kryukov, G.V., Lawrence, M.S., Sougnez, C., McKenna, A., et al. (2011). The mutational landscape of head and neck squamous cell carcinoma. Science 333, 1157–1160.

Tang, K.W., Alaei-Mahabadi, B., Samuelsson, T., Lindh, M., and Larsson, E. (2013). The landscape of viral expression and host gene fusion and adaptation in human cancer. Nat. Commun. 4, 2513.

Vartanian, J.-P., Guétard, D., Henry, M., and Wain-Hobson, S. (2008). Evidence for editing of human papillomavirus DNA by APOBEC3 in benign and precancerous lesions. Science 320, 230–233.

Vartanian, J.-P., Henry, M., Marchio, A., Suspène, R., Aynaud, M.-M., Guétard, D., Cervantes-Gonzalez, M., Battiston, C., Mazzaferro, V., Pineau, P., et al. (2010). Massive APOBEC3 editing of hepatitis B viral DNA in cirrhosis. PLoS Pathog. 6, e1000928.

Vogt, P.K., Kang, S., Elsliger, M.-A., and Gymnopoulos, M. (2007). Cancer-specific mutations in phosphatidylinositol 3-kinase. Trends Biochem. Sci. 32, 342–349.

Walboomers, J.M.M., Jacobs, M.V., Manos, M.M., Bosch, F.X., Kummer, J.A., Shah, K.V., Snijders, P.J.F., Peto, J., Meijer, C.J.L.M., and Muñoz, N. (1999). Human papillomavirus is a necessary cause of invasive cervical cancer world-wide. J. Pathol. 189, 12–19.

Yamanaka, S., Balestra, M.E., Ferrell, L.D., Fan, J., Arnold, K.S., Taylor, S., Taylor, J.M., and Innerarity, T.L. (1995). Apolipoprotein B mRNA-editing protein induces hepatocellular carcinoma and dysplasia in transgenic animals. Proc. Natl. Acad. Sci. USA 92, 8483–8487.

Zhang, H., Yang, B., Pomerantz, R.J., Zhang, C., Arunachalam, S.C., and Gao, L. (2003). The cytidine deaminase CEM15 induces hypermutation in newly syn-thesized HIV-1 DNA. Nature 424, 94–98.

Zhao, L., and Vogt, P.K. (2008). Helical domain and kinase domain mutations in p110alpha of phosphatidylinositol 3-kinase induce gain of function by different mechanisms. Proc. Natl. Acad. Sci. USA 105, 2652–2657.