APOBEC Enzymes: Mutagenic Fuel for Cancer Evolution and Heterogeneity

Charles Swanton1,2, Nicholas McGranahan1,3, Gabriel J. Starrett4, and Reuben S. Harris4

1The Francis Crick Institute, London, United Kingdom. 2UCL Cancer Institute, CRUK Lung Cancer Centre of Excellence, London, United Kingdom. 3Centre for Mathematics & Physics in the Life Sciences & Experimental Biology (CoMPLEX), University College London, London, United Kingdom. 4Department of Biochemistry, Molecular Biology and Biophysics, Institute for Molecular Virology, and Masonic Cancer Center, University of Minnesota, Minneapolis, Minnesota.

Corresponding Authors: Charles Swanton, The Francis Crick Institute, 44 Lincoln’s Inn Fields, London, WC2A 3LY, United Kingdom. Phone: 44-207-269-3463; Fax: 44-20-7269-3094; E-mail: charles.swanton@crick.ac.uk; and Reuben S. Harris, Department of Biochemistry, Molecular Biology and Biophysics, Institute for Molecular Virology, and Masonic Cancer Center, University of Minnesota, Minneapolis, MN 55455. Phone: 612-624-0457; E-mail: rsh@umn.edu
doi: 10.1158/2159-8290.CD-15-0344
© 2015 American Association for Cancer Research.

THE IMPORTANCE OF CANCER DIVERSITY

The majority of human cancers display one or more patterns of genomic instability, ranging from large-scale changes in chromosome architecture and tumor karyotype through to multitudes of single-nucleotide mutations. The diversity created by these seemingly chaotic processes provides the substrate for selection within tumors. The relative importance of selection and elevated mutation rates (which when used in its broadest sense covers both large-scale chromosomal and single-nucleotide changes) has been a subject of much debate (reviewed in ref. 1). Although an elevated mutation rate is not necessarily a requirement for the initiation of a sporadic tumor, it will likely precipitate more rapid tumor evolution and adaptation (2–4).

Considerable evidence is now accumulating that specific patterns of diversity within tumors are associated with worse clinical outcomes. Work over the last two decades has established that chromosomal instability (CIN), a driver of intercellular heterogeneity, is associated with poorer clinical outcomes across a range of cancer types (for review, see ref. 5). This work is underpinned by elegant mouse models that have shown that forcing transient CIN in lung cancer causes more frequent tumor recurrence (6).

Maley and colleagues (7) demonstrated in a prospective study of Barrett’s esophagus that clonal diversity, assessed using ecologic indices, predicts progression to invasive adenocarcinoma. The advent of next-generation sequencing has enabled deeper resolution of cancer genomic instability and its potential association with clinical outcome. It is increasingly appreciated that driver events do not always occur early in tumor evolution and can arise in subclones and contribute to tumor maintenance and subclonal expansions later in tumor evolution (8–10). Consistent with this, Landau and colleagues (11) have revealed that the presence of a subclonal driver event is associated with worse outcome in chronic lymphocytic leukemia. Schwarz and colleagues (12) have shown in high-grade serous ovarian cancer that measures of heterogeneity correlate with progression-free survival times. Increasing evidence is emerging that subclonal populations that are either present at low frequency or undetectable at diagnosis may dominate the disease at recurrence (12, 13) and, in cases of low-grade glioma, be catalyzed by temozolomide therapy (14).

However, evidence is strengthening for the “just right” hypothesis for tumor diversity (15, 16); too little diversity and
cancers fail to adapt sufficiently robustly in the face of selection pressures; too much genomic instability (in its many forms) results in cell-autonomous lethality, as genomic chaos may be too severe to ensure sufficiently faithful propagation of genomic information to daughter cells. Thus, an intermediate or “just right” rate of tumor genomic instability may exist, ensuring sufficient diversity required to overcome selection barriers, while retaining propagation of genomic information sufficient for viability of daughter cells. An analysis of CIN quartiles in non–small cell lung cancer (NSCLC) and gastric, estrogen receptor (ER)–negative breast, and ovarian cancers has revealed that such intermediate thresholds of CIN appear to exist and correlate with the poorest clinical outcomes, in contrast to excessive or minimal CIN in cases where outcomes are better (17, 18).

Taken together, although some diversity ensures more rapid tumor evolution, excessive genomic instability may be incompatible with cell viability and result in the requirement for a buffering of such processes in order to achieve the “just right threshold.” Determining the molecular basis for such buffering mechanisms might prove fruitful to devise new approaches to sensitize tumor cells to the underlying threat of genome instability. Here, we review mutational processes that foster diversity, focusing on the APOBEC family of enzymes that are emerging as major sources of mutation for subclonal heterogeneity.

**CHROMOSOME-LEVEL DRIVERS OF TUMOR DIVERSITY**

The fields of genomic instability, intratumor heterogeneity, and cancer evolution have collided over recent years as next-generation sequencing technologies have begun to reveal genomic signatures of DNA damage and errors in DNA repair processes, and longitudinal studies have begun to shed light on both endogenous and exogenous influences shaping the cancer genome. Mechanisms of genomic instability within cancer genomes can be considered at both the chromosome (and can be considered macroevolutionary sources of diversity) and single-nucleotide levels (21).

CIN refers to an increased rate of change of chromosome number or structure. Errors in chromosome segregation can be classified as premitotic aberrations that may arise due to defective DNA replication or repair (22) or defects in telomere maintenance, generating structurally unstable acentric or dicentric chromosomes (for review, see ref. 21). Mitotic errors in chromosome segregation may arise through transient aberrant spindle geometry (23), defects in sister chromatid cohesion (24), spindle assembly checkpoint dysfunction (25), or aberrant microtubule–kinetochore attachments generating merotelic chromosomes (where one kinetochore of a chromosome is attached by microtubules to both spindle poles), generating numerically unstable chromosomes (26).

Next-generation sequencing studies are revealing additional large-scale chromosomal aberrations that contribute to genome complexity, such as chromothripsis (27), chromoplexy (28), and genome doubling events (29, 30). In tandem, live cell imaging experiments provide insights into how such macroevolutionary events arise in tumors (31) and their consequences for cancer genome evolution (30). It is now clear that structural CIN and numerical CIN cannot be considered separately, and that one may lead to another. Janssen and colleagues (32) demonstrated how cytokinesis errors entrap chromosomes in the cytokinetic furrow, resulting in DNA double-strand breaks and chromosome translocations. Crasta and colleagues (31) demonstrated how lagging chromosomes in mitosis form micronuclei that are subject to asynchronous DNA replication triggering DNA damage and chromosomal fragmentation, shedding light on how chromothripsis may originate in cancers. Similarly, premitotic errors in DNA replication resulting in structurally unstable chromosomes can initiate numerical CIN. DNA replication stress results in replication fork arrest or collapse and single-stranded DNA (ssDNA) exposure, initiating a DNA-damage response. Loss of three candidate CIN-suppressor genes encoded on 18q in colorectal cancer was found to drive structural CIN, precipitating numerical CIN, manifested as deviation in the modal centromeric number, a phenotype that could be rescued by the addition of nucleosides (22).

Whole-genome doubling events are a common and likely important mechanism for enabling additional genomic complexity and heterogeneity. Genome doubling events are estimated to occur in 11% to 64% of solid tumors (33) and have been linked with progression from Barrett’s esophagus to esophageal adenocarcinoma (34). We have shown that the occurrence of genome doubling events in vitro is associated with an increase in chromosome instability due to a doubling in the rate of chromosome segregation errors (30). In addition, unlike diploid cells, which die or enter a prolonged interphase arrest following a chromosome segregation error, tetraploid cancer cells tolerate chromosome segregation errors and more effectively propagate aneuploid progeny than diploid cells (30).

**NUCLEOTIDE-LEVEL DRIVERS OF TUMOR DIVERSITY**

To contrast with the chromosome-level events described above, this review will use the phrase “nucleotide-level drivers” to refer to mutational sources that cause predominantly single-nucleotide changes. Accepted and conventional endogenous sources of DNA damage and mutation include DNA replication errors, oxidative stress, hydrolytic deamination, and alkylation damage (reviewed in refs. 35, 36). Obvious exogenous sources include UVA and UVB radiation in sunlight, reactive compounds in tobacco products (reviewed in ref. 37), and emerging dietary carcinogens such as aristolochic acid (38). The importance of nucleotide-level drivers is evidenced by defective post-replicative mismatch repair in hereditary and sporadic colorectal cancers, defective excision repair and higher incidences of skin cancer in xeroderma pigmentosum patients, and other cancer predispositions caused...
by defective DNA repair processes (reviewed in refs. 37, 39, 40). Moreover, many chemotherapeutics, such as platinum-based drugs and nucleotide analogs, are effectively DNA-damaging agents that preferentially kill replicating cancer cells. Recent deep sequencing efforts have confirmed the established links between these and other sources of mutation and various cancer types, but they have also helped to shine light on several additional sources of DNA damage and mutation (41, 42).

**APOBEC ENZYMES: MAJOR CONTRIBUTORS TO CANCER MUTAGENESIS**

The largest newly defined source of mutation in many different tumor types is enzymatic DNA cytosine deamination by several members of the APOBEC family (reviewed recently in refs. 35, 43–45). These enzymes were identified independently in 2002 as DNA mutators and as antiviral factors (46–48). Apart from activation-induced deaminase (AID), which deaminates expressed antibody genes, all of the family members are named after the founder, APOBEC1, which edits both ssDNA cytosines and cellular mRNA cytosines (as implied by its rarely used full name: apolipoprotein B mRNA editing enzyme catalytic polypeptide 1). Importantly, the sheer existence of multiple DNA-mutating enzymes in human cells strongly suggested that at least one could become deregulated and promote mutagenesis in cancer (46). However, dominant roles in antiviral innate immunity and particularly in HIV-1 restriction took center stage for most of the following decade as technologies and experimental systems have had to be developed to test the cancer hypothesis (49).

It is now clear that human cells have the capacity to express a total of 11 distinct APOBEC family members (Fig. 1A). The seven APOBEC3 proteins are encoded by a tandem array of genes located on chromosome 22, AID and APOBEC1 are encoded by genes on chromosome 6, and APOBEC2 and APOBEC4 by more distantly related genes on different chromosomes. All seven APOBEC3 proteins as well as AID and APOBEC1 have demonstrated ssDNA cytosine to uracil (C-to-U) deaminase activity in a wide variety of biochemical assays and cell-based experimental systems (reviewed in ref. 49). Demonstrated physiologic functions include AID-catalyzed deamination of expressed antibody gene sequences, which drives antibody diversification by somatic hypermutation and isotype switch recombination, and APOBEC3D/F/G/H-catalyzed deamination of HIV-1 cDNA replication intermediates during reverse transcription (reviewed, respectively, in ref. 50 and refs. 49, 51). DNA cytosine deamination is therefore established as the hallmark activity of this protein family (Fig. 1B). As elaborated below, cytosine to thymine (C-to-T) transition mutations are common outcomes of APOBEC-mediated DNA deamination events, although ensuing DNA repair intermediates, such as abasic sites and DNA breaks, can also lead to cytosine to guanine (C-to-G) transversions and other mutational outcomes (for recent mechanistic reviews, see refs. 44, 45).

APOBEC family members function as single or double domain enzymes (Fig. 1A). APOBEC3A/C/H, AID, and APOBEC1 are single domain enzymes because each is made up of one conserved zinc-coordinating domain (Fig. 1C). APOBEC3B/D/F/G are double domain enzymes composed of two conserved zinc-coordinating domains and thought to be connected by a short flexible region. Each domain has an overall globular fold with a hydrophobic β-sheet at the core, surrounded by six α-helices (Fig. 1C). The α2 and α3 helices coordinate zinc through conserved H-X-E and C-P-X2,4-C motifs and define the active site, with the glutamate serving as a proton shuttle during DNA cytosine deamination (Fig. 1B).

Mechanistic studies in breast cancer, model organism work, and breast cancer genome sequencing efforts independently converged on APOBEC3 enzymes as major sources of mutation in cancer (52–54). We used a panel of specific quantitative PCR assays to identify upregulation of APOBEC3B in breast cancer cell lines and primary tumor samples (53). Other DNA deaminase family members were either not expressed or expressed at similar levels in normal and patient-matched tumor samples. Knockdown of endogenous APOBEC3B levels in several breast cancer cell lines ablated all measurable DNA cytosine deamination activity from nuclear protein extracts and caused proportional declines in steady-state genomic uracil levels, C-to-T mutations, and reporter gene mutation frequencies (53). Importantly, the biochemical preference of APOBEC3B in vitro for chemically defined ssDNA substrates was shown to closely resemble the actual cytosine mutation signature for chemically defined ssDNA substrates was shown
in two publicly accessible cohorts of breast cancer genomic DNA sequences, with both biased toward 5′-TCA and 5′-TCG motifs. A subset of the cytotoxic mutations within 5′-TCG motifs may also be due to hydrolytic deamination of methyl-cytosine, which inevitably occurs over time (aging), but there is no reason that this should preferentially occur at cytosome nucleobases with thymine on the 5′ side. Taken together, with positive correlations between APOBEC3B expression levels and C-to-T mutation loads and overall mutation loads, we concluded that APOBEC3B is a major source of mutation in a dominant fraction of breast cancers.

Roberts and colleagues (54) have shown that a chemical mutagen could generate clusters of cytosome mutations in yeast and that analogous clusters are evident in publicly available cancer genomic DNA sequences. They used a genetic approach to generate extended tracks of ssDNA, followed by a selection for mutations that simultaneously inactivated adjacent reporter genes. This elegant system enabled the identification of strand-coordinated clusters of mutations that were much more frequent in the presence of the chemical mutagen methyl-methanesulfonate (MMS). They predicted that similar ssDNA mutagenic processes may be contributing to human cancer and found strand-coordinated clusters of mutations at cytosome (or complementary guanines) in the genomes of multiple myelomas, head/neck cancers, and prostate tumors. These mutations were enriched within 5′-TCA and 5′-TCT motifs previously deduced for several APOBEC enzymes, thus implicating this protein family in cancer mutagenesis. Although these were not cause-and-effect experiments as above for knockdown studies in cancer cell lines, other potential mutation-generating mechanisms, such as chemicals, were considered less likely because they are not known to be preferentially biased toward trinucleotide motifs.

Nik-Zainal and colleagues (52) generated full genome sequences of 21 breast cancers in order to deduce mutation landscapes. One of the most striking findings from this study was the uncovering of extensive clusters of strand-coordinated cytosome mutations. These clusters were termed kataegis due to an analogy with localized thundershowers (kataegis is an ancient Greek word for thunder). These cytosome-biased mutations occurred almost exclusively within 5′-TCG dinucleotide motifs characteristic of several APOBEC family members. Kataegis cytosome mutations are characterized as strand-coordinated C-to-T transitions and C-to-G transversions (52, 54). The most likely mechanism is clustered deamination of cytosome within extended regions of ssDNA (during replication or recombination repair), followed by uracil excision to create abasic sites, and error-prone DNA synthesis of the complementary DNA strand resulting in the insertion of adenine (transition outcome) or guanine (transversion outcome) nucleobases opposite each nonstructural abasic lesion (for recent mechanistic review, see ref. 44).

Additional work from several groups, as well as ours, have combined to implicate APOBEC mutagenesis in approximately half of all human cancers, with breast, head/neck, lung, bladder, and cervical cancers bearing the largest APOBEC-attributable mutation burdens (41, 43, 55–65; and reviewed recently in refs. 35, 43–45, 63–65). APOBEC3B is upregulated in over half of all cancers, with 5% to 10% of tumors at the lower end of the spectrum showing upregulation and nearly 90% to 95% of tumors at the higher end showing upregulation (Fig. 2). Other APOBEC family members do not appear to be similarly upregulated, and other correlating factors have yet to emerge. In the largest study to date, Alexandrov and colleagues (41) used a computational algorithm to rank the APOBEC mutational signature second only to “aging” in cancer mutagenesis. Aging-associated mutations are largely C-to-T transitions within CpG dinucleotide motifs. Most of these mutations are likely due to spontaneous hydrolytic deamination of methyl-C-to-T, because this modified base is more vulnerable to deamination than normal-C. However, some APOBEC3 family members, such as APOBEC3A (and by homology, APOBEC3B), can accommodate bulkier nucleobase substrates, such as methyl-C, in ssDNA, and therefore “aging” may be a mix of both spontaneous and enzyme-catalyzed deamination (66–68).

Altogether, a compelling case can be made for APOBEC3B as a major mutational source in breast, head/neck, lung, bladder, cervical, and ovarian cancers, and it is also likely to be a more modest but still significant contributor to many other cancers. However, it is equally clear that APOBEC-like 5′-TC″-biased cytosome mutations can still occur in some breast tumors in the complete absence of APOBEC3B due to a deletion allele (69). This observation strongly suggests that at least one other APOBEC family member is also contributing to the overall mutational burden in cancer. One candidate is APOBEC3A due to deregulation caused by the APOBEC3B deletion allele, which fuses the coding sequence of APOBEC3A to the 3′-untranslated region of APOBEC3B.
A CG-to-ATG mutation in EGFR that causes the with the EGFR inhibitors gefitinib or erlotinib through a half of EGFR-activating mutant NSCLCs fail treatment′ level 5-ACT to ATT transition mutation. Moreover, nearly As an example, MEK2 Leu46 to Phe is due to a nucleotide genes potentially mediated by APOBEC3 deamination (72).

tions in extreme morphologic heterogeneity for many tumor types, helical domain mutations (75). It implies that APOBEC3B is the predominant source of the rare APOBEC3-catalyzed event into a commonly observed 5′-TC motif, selective pressure may conceivably transform a cytosine mutation does not occur in the preferred APOBEC3 important consideration because, although this particular distribution in the genesis and not viral infection (61, 74). This skewed mutation implying that the common denominator is APOBEC muta-

tion resulting in a kinase domain His1047 to Arg substitution. (complementary strand 5′-TCA to TTA) that convert both helical domain mutations in HPV-negative tumors (61). Specifically, HPV-positive tumors have 5′-TGA to TAA transitions (complementary strand 5′-TCA to TTA) that convert both helical domain Glu542 and Glu545 to Lys, whereas HPV-negative tumors often have a 5′-CAT to CGT transition mutation resulting in a kinase domain His1047 to Arg substitution. Similar helical domain biases have also been reported for PIK3CA mutations in other APOBEC signature tumor types, implying that the common denominator is APOBEC mutagenesis and not viral infection (61, 74). This skewed mutation distribution in the PIK3CA oncogene is significant because it implies that APOBEC3B is the predominant source of the helical domain mutations (75).

Importantly, for many decades, pathologists have reported extreme morphologic heterogeneity for many tumor types, including breast and lung cancers (76). It is likely that an underlying genetic heterogeneity exerts influence on this morphologic heterogeneity. Single sample ultradepth sequencing and SNP array studies have detected mutations at different frequencies depending upon when they arose during tumorigenesis and outgrowth (52, 77, 78). Multiregion deep sequencing of primary tumors has revealed strikingly different results and clear evidence for the existence of multiple subclonal lineages (59, 79–82). Evidence for ongoing mutations comes from comparisons of deep sequences of primary tumors and metastatic outgrowths or local tumor recurrences following therapy (14, 59, 83–87). In all of these cases, a proportion of mutations are common to both the primary tumor and the metastasis, and a proportion of mutations are unique to the primary tumor or the metastasis/recurrence. We have recently found evidence that APOBEC-attributable C-to-T and C-to-G mutations in 5′-TC motifs may be early as well as later events in tumor evolution, occurring in the trunk and branches of the tumor’s evolutionary tree (74). Truncal mutations occurring early in tumor evolution would be expected to be propagated following the metastatic process.

Some of the strongest evidence for the ongoing potency of APOBEC mutagenic processes in cancer evolution has come from recent studies that provide temporal resolution of mutagenic processes during cancer evolution, reviewed below.

EVIDENCE FOR ONGOING APOBEC MUTAGENESIS IN TUMOR EVOLUTION

It is important to emphasize that no single study is definitive with respect to ongoing APOBEC mutation in cancer, but aggregate findings from many studies make a very strong case. First, as described above, mutation is ongoing and dependent upon APOBEC3B in breast and ovarian cancer cell lines, suggesting that it may have been similarly ongoing in the tumors from which they were originally derived (53, 55). Second, high levels of APOBEC3B correlate with poor outcomes for ER-positive breast cancer patients (71). Significantly different clinical outcomes are unlikely to be due to transient or brief exposures but more likely to ongoing and compounded accumulations of APOBEC mutations that continually provide the underlying substrate for tumor diversification and adaptation under selective pressures.

Third, a number of drug resistance mutations occur at cytosine bases, and a subset of these occur at APOBEC3-preferred 5′-TC dinucleotide motifs. For instance, a proportion of BRAFV600E-mutant melanomas fail BRAF inhibitor treatment (vermurafenib or dabrafenib) due to cytosine mutations in MEK1, MEK2, or other signal transduction pathway genes potentially mediated by APOBEC3 deamination (72). As an example, MEK2 Leu46 to Phe is due to a nucleotide level 5′-ACT to ATT transition mutation. Moreover, nearly half of EGFR-activating mutant NSCLCs fail treatment with the EGFR inhibitors gefitinib or erlotinib through a single 5′-ACG-to-ATG mutation in EGFR that causes the Thr790Met “gatekeeper” substitution (73). This raises an important consideration because, although this particular cytosine mutation does not occur in the preferred APOBEC3 5′-TG motif, selective pressure may conceivably transform a rare APOBEC3-catalyzed event into a commonly observed clinical outcome.

Fourth, the distribution of PIK3CA-activating mutations is different in head/neck cancers, with exclusively helical domain C-to-T transitions observed in human papillomavirus (HPV)-positive tumors and a combination of helical domain and kinase domain mutations in HPV-negative tumors (61). Specifically, HPV-positive tumors have 5′-TGA to TAA transitions (complementary strand 5′-TCA to TTA) that convert both helical domain Glu542 and Glu545 to Lys, whereas HPV-negative tumors often have a 5′-CAT to CGT transition mutation resulting in a kinase domain His1047 to Arg substitution. Similar helical domain biases have also been reported for PIK3CA mutations in other APOBEC signature tumor types, implying that the common denominator is APOBEC mutagenesis and not viral infection (61, 74). This skewed mutation distribution in the PIK3CA oncogene is significant because it implies that APOBEC3B is the predominant source of the helical domain mutations (75).

ORDERING MUTATIONAL PROCESSES DURING CANCER EVOLUTION

The heterogeneity observed within a tumor can shed light on the evolutionary history of the cancer genome and the temporal order of genetic events acquired during the disease course. Clonal mutations that are found in all tumor cells reflect somatic events that occurred relatively “early” in a tumor’s evolutionary history, before or during the latest clonal sweep. These events represent a mixture of driver aberrations implicated in tumorigenesis, and passenger mutations, which can vastly outnumber the drivers, but provide important contextual and mechanistic information. Conversely, mutations present in only a subset of tumor cells, by definition, occurred after the latest clonal sweep, and thereby reflect somatic events that may be involved in driving tumor progression and subclonal diversification. In the case of genome doubling or chromosomal amplification events, the temporal order of somatic events can be further refined. For example, a mutation occurring before a genome doubling will likely be present in at least two copies, whereas a mutation that occurred after a doubling event will be present only at one copy (Fig. 3A). Using such approaches, distinct mutational processes are becoming apparent that contribute to subclonal expansions and intratumor heterogeneity.

In a recent analysis of intratumor heterogeneity in early-stage NSCLC from our laboratory, multiregion sampling allowed the timing of mutational processes during tumor evolution to be deciphered, establishing the temporal dynamics of APOBEC mutational processes (59, 79). Enrichment of the APOBEC signature was seen in the branches of tumor evolutionary trees relative to the early clonal truncal mutations, particularly prominent in adenocarcinomas of the lung (Fig. 3B; refs. 59, 79). Consistent with this endogenous
mutational process driving subclonal expansions, mutations within an APOBEC3B context discussed above were found in driver genes such as **PTPRD**, **PIK3CA**, **EP300**, **TGFBR1**, and **AKAP9** (59). Even in current smokers, APOBEC appeared to be the dominant mutational process in tumor subclones at disease presentation.

There was also evidence for spatial heterogeneity in APOBEC activity; in one adenosquamous tumor, the APOBEC signature appears to occur prior to subclonal diversification but after genome doubling. A mutation in **PIK3CA** occurs in an APOBEC context. Figure is adapted from ref. 59.

Figure 3. Temporal order of mutational processes during cancer evolution. **A**, a schematic of mutations accumulating during cancer. Mutations can be temporally dissected from sequencing data by determining which mutations occurred before or after genome doubling or amplification events and also identifying mutations occurring before or after subclonal diversification. GD, genome doubling. **B**, a diagram of the evolutionary history of a lung adenocarcinoma. The APOBEC signature appears to occur prior to subclonal diversification but after genome doubling. A mutation in **PIK3CA** occurs in an APOBEC context. Figure is adapted from ref. 59.

In a broader analysis of over 2,500 tumors across nine cancer types, we found evidence for a specific APOBEC mutational signature increasing in prevalence during the course of tumor evolution in four cancer types, including bladder cancer, ER-negative breast cancer, head/neck cancer and, as previously described, lung adenocarcinoma (59, 74). In contrast, an age-related signature as well as mutational processes linked to exogenous sources of DNA damage (smoking in NSCLC and UV light in melanoma) were generally found to decrease in prevalence during the evolutionary history of tumors (74).

The importance of APOBEC, particularly later in tumor evolution, is highlighted by the observation that this mutational process was found to be the major source of subclonal cancer gene mutations in bladder, breast, and head/neck squamous cancers, lung adenocarcinoma, and lung squamous cell carcinoma (74). In four of those five cancer types (bladder, head/neck, lung adenocarcinoma, and squamous carcinoma), mutations attributable to an APOBEC source were enriched within subclonal cancer gene mutations relative to clonal driver gene mutations, suggesting APOBEC is a mutagenic source, fueling cancer heterogeneity and subclonal diversification. Strikingly, in head/neck cancer, lung adenocarcinoma, and lung squamous cell carcinoma, we found evidence that over 85% of subclonal mutations in **PIK3CA** occurred in an APOBEC context (74). The vast majority of these subclonal mutations were mutations in the helical domain (E545K) that...
have previously been linked to APOBEC-mediated mutagenesis in cervical and head/neck tumors (61).

**FUTURE PERSPECTIVES**

The role of APOBEC as an ongoing mutagenic process, contributing to subclonal diversification and intratumor heterogeneity, suggests the need for therapeutically tractable approaches to inhibit these enzymes and limit further tumor adaptation. However, in parallel to embarking upon drug-discovery efforts focused on APOBEC3B inhibition, it will be important to unambiguously identify the other APOBECs that also contribute to tumor evolution and heterogeneity. An additional APOBEC may work with APOBEC3B to compound evolutionary benefits for the tumor, or it may only compensate in the absence of APOBEC3B, as suggested by APOBEC3A deregulation studies.

A strong link between APOBEC3B upregulation and viral infection, specifically HPV in head/neck, cervical, and perhaps some bladder tumors, is intriguing and suggests that other cancer types may have an as-yet-unknown association with a virus or DNA-based parasite. APOBEC3B upregulation may benefit the virus by facilitating viral genetic variation or by helping to transform the cell and thereby increase the virus’s chances of spreading. However, there are also likely to be non-viral mechanisms for APOBEC3B upregulation, as a microbial etiology is unlikely to have been missed in deep sequencing studies, especially in heavily studied areas such as breast cancer.

Knowledge of APOBEC3 mutagenesis in cancer may yield significant diagnostic and prognostic value as well as new therapeutic opportunities. For instance, the diagnosis of APOBEC-high tumors may encourage more aggressive multi-drug therapies to avert poor clinical outcomes, especially if the data for ER-positive breast cancer translates to additional cancer types (71, 88). APOBEC high tumors are also likely to have higher levels of DNA damage, and may be amenable to a synthetic lethal approach analogous to PARP inhibitor treatment of BRCA-mutant cancers (89). Indeed, given evidence for the “just-right” threshold of genome instability in cancers (17, 18), exacerbating genomic instability in APOBEC-high tumors may also prove fruitful.

Understanding what activates APOBEC is also of critical importance, as it could inspire methods to prevent its activation. For instance, the mutator phenotype caused by HPV infection could be averted by eliminating the virus from the system or preventing infection through vaccination. The proclivity for ssDNA substrates for these enzymes suggests the possibility that DNA replication or repair defects that enhance exposure of ssDNA may contribute to upregulation of this mutagenic process. Moreover, the fact that APOBEC3B and other family members are enzymes with defined active sites and the additional fact that APOBEC3B is nonessential should encourage the development of small-molecule inhibitors to suppress APOBEC-mediated mutagenesis. This strategy could prolong progression-free survival times with current targeted therapy approaches by decreasing the significant problems caused by the evolution of mutations conferring drug resistance. There is scope for further advances in precision cancer medicine by targeting the processes that drive intratumor heterogeneity. APOBEC represents one such process fueling subclonal expansions and represents an attractive target to limit adaptation and rapid evolution of APOBEC-high tumors.

**Disclosure of Potential Conflicts of Interest**

C. Swanton and R.S. Harris have ownership interest (including patents) in ApoGen Biotechnologies. No potential conflicts of interest were disclosed by the other authors.

**Acknowledgments**

The authors thank the members of the Swanton and Harris laboratories for helpful comments on the manuscript. The authors apologize to colleagues whose work could not be cited due to space limitations.

**Grant Support**

C. Swanton is a senior Cancer Research UK (CRUK) clinical research fellow and is funded by the CRUK, the Rosetrees Trust, EU FP7 (projects PREDICT and RESPONSIFY; ID:259303), the Prostate Cancer Foundation, the European Research Council (The-seus), and the Breast Cancer Research Foundation. The Swanton laboratory is supported by the National Institute for Health Research University College London Hospitals Biomedical Research Centre. Cancer research in the Harris laboratory is supported by grants from the Department of Defense Breast Cancer Research Program (BC121347), the Jimmy V Foundation for Cancer Research, the Norwegian Centennial Chair Program, the Minnesota Partnership for Biotechnology and Medical Genomics, and the Randy Shaver Cancer Research and Community Fund. G.J. Starrett is supported by the U.S. National Science Foundation (DGE 13488264).

Received March 20, 2015; revised May 14, 2015; accepted May 14, 2015; published OnlineFirst June 19, 2015.

**REFERENCES**


APOBEC and Cancer Heterogeneity


34. Li X, Galipeau PC, Paulson TG, Sanchez CA, Arnaudo J, Liu K, et al. Temporal and spatial evolution of somatic chromosomal altera-


51. Desiminnie BA, Delviks-Frankenberg RA, Burdick QC, Qi D, Izumi T, Pathak VK. Multiple APOBEC3 restriction factors for HIV-1 and one Vif to rule them all. J Mol Biol 2014;426:1220–45.


