

Trends in Molecular Medicine

APOBEC3 genes: retroviral restriction factors to cancer drivers

--Manuscript Draft--

Manuscript Number:	TRMOME-D-14-00197R2
Article Type:	Review
Corresponding Author:	Tim Fenton UCL Cancer Institute London, UNITED KINGDOM
First Author:	Stephen Henderson, PhD
Order of Authors:	Stephen Henderson, PhD Tim Fenton
Abstract:	<p>The APOBEC3 cytosine deaminases play key roles in innate immunity through their ability to mutagenize viral DNA and restrict viral replication. Now recent advances in cancer genomics together with biochemical characterization of the APOBEC3 enzymes have implicated at least two family members in somatic mutagenesis during tumor development. Here we review the evidence linking these enzymes to carcinogenesis and highlight key questions, including the potential mechanisms that misdirect APOBEC3 activity to the host genome, the links to viral infection, and the association between a common APOBEC3 polymorphism and cancer risk.</p>

Henderson and Fenton: highlights

1. APOBEC3 (A3) enzymes deaminate single-stranded DNA, mutating viral genomes
2. Tumors harbor mutations likely due to off-target A3 deamination of chromosomal DNA
3. A common APOBEC3 deletion polymorphism increases breast cancer risk
4. A growing body of evidence points to a driver role for APOBEC3s in tumorigenesis

APOBEC3 genes: retroviral restriction factors to cancer drivers

Stephen Henderson¹ and Tim Fenton²

1 Bill Lyons Informatics Centre, UCL Cancer Institute, London, UK

2 Department of Oncology, UCL Cancer Institute, London, UK

Corresponding author: Tim Fenton (t.fenton@ucl.ac.uk)

Keywords: APOBEC, cancer, mutagenesis

Abstract

The APOBEC3 cytosine deaminases play key roles in innate immunity through their ability to mutagenize viral DNA and restrict viral replication. Now recent advances in cancer genomics, together with biochemical characterization of the APOBEC3 enzymes, have implicated at least two family members in somatic mutagenesis during tumor development. Here we review the evidence linking these enzymes to carcinogenesis and highlight key questions, including the potential mechanisms that misdirect APOBEC3 activity to the host genome, the links to viral infection, and the association between a common APOBEC3 polymorphism and cancer risk.

Deoxycytidine deamination in innate immunity and somatic mutagenesis

The human apolipoprotein B mRNA-editing enzyme catalytic polypeptide-like (APOBEC) gene family comprises 11 genes: the Activation-Induced Cytidine Deaminase (*AICDA*), *APOBEC1* (*A1*) and seven *A3s* encode (deoxy)cytidine deaminases capable of targeting RNA and/or DNA; *A2* does not appear to possess catalytic activity and *A4* remains uncharacterized [1]. The founding member, *A1* was cloned over 20 years ago and shown to be responsible for the deamination (see Glossary) of cytidine 6666 in the ApoB mRNA, converting it to uridine and generating a stop codon that creates a shorter ApoB⁴⁸ protein [2, 3]. *AICDA* encodes Activation-induced Deaminase (AID) which deaminates deoxycytidine (dC) in single stranded (ss) DNA, generating deoxyuridine (dU), an activity that underlies somatic hypermutation (SHM) and class switch recombination (CSR) to drive antibody diversification in B-lymphocytes (reviewed in [4, 5]). The discovery that *A1* and several *A3s* are also able to deaminate dC (in some cases with greater potency than AID) in ssDNA [6] led rapidly to an understanding that this DNA editing activity forms a key part of the innate immune response to viral infection, the best-characterized example of which is the human immunodeficiency virus (HIV) restriction factor A3G [7, 8]. Like all the *A3* enzymes, A3G can only target ssDNA, in this case the nascent proviral HIV cDNA during reverse transcription [8]. The direct deamination action is dC>dU in the minus strand reverse transcript, particularly at CC dinucleotides. However, the edited mutations are subsequent G>A transitions in the coding positive strand, inactivating the provirus [7, 9]. The HIV protein Vif (viral infectivity factor) allows the virus to overcome this response by hijacking a cellular E3 ubiquitin ligase (CUL5) to mark A3G for proteasomal degradation [10, 11].

While A3G resides in the cytoplasm, where it becomes incorporated into retroviral virions, A3 enzymes such as A3A, A3B and A3C enter the nucleus, where they are able to target other pathogens including human papillomaviruses (HPV) and also retrotransposons; mobile elements within the genome which appear to be targeted by multiple APOBEC enzymes via a variety of mechanisms [12-19]. With nuclear localization however, comes the risk of off-target activity against host genomic DNA and the potential for mutagenesis. Normally, genomic dU resulting either from dC deamination or mis-incorporation of deoxyuridine monophosphate (dUMP) during replication is efficiently removed by Uracil-DNA Glycosylase (UNG) and lesions are resolved by the Base Excision Repair (BER) pathway. Failure to replace dU with dC before the next round of replication leads to mutations, the precise nature of which depend upon the point at which replication interrupts repair and also on the polymerases involved.

The mechanistic details of these events following deamination by A3s remain uncharacterized, but working models have been proposed based upon what is known about AID-dependent SHM [5, 20-22]. As summarized in Figure 1, replication prior to removal of dU can result in C>T transitions due to incorporation of dA into the daughter strand through Watson-Crick base pairing. Alternatively, following removal of dU by UNG, translesion synthesis (TLS) polymerases can replicate over the resulting abasic site, most commonly inserting dA in the daughter strand (the 'A-rule'), again resulting in C>T transitions. Other TLS polymerases insert different deoxynucleotides opposite abasic sites; Rev1 for instance, uses dC, resulting in C>G

transversions at deaminated sites in yeast [23, 24]. If deamination were to occur in close proximity on both strands, double strand breaks (DSBs) could result, with the overhangs either chewed back by exonuclease, or filled in by polymerase (Fig 1). Here we discuss recent evidence linking off-target activity of A3s to somatic mutagenesis during tumor development. We highlight important questions regarding how and why A3s turn against chromosomal DNA, and how a polymorphism affecting A3A and A3B has come to vary so widely across the human diaspora.

APOBECs as mutators of genomic DNA: the cancer connection

Early evidence for a link between APOBECs and cancer came from transgenic animals, with overexpression of APOBEC1 (A1) in mice giving rise to hepatocellular carcinomas [25]. The finding that AID and A1 could cause mutations in DNA when expressed in *E. coli* provided not only a clue to how AID drives antibody diversification but also a mechanism by which APOBECs might promote tumor formation [6, 26]. Soon afterwards, AID was shown to cause lymphomas with mutations in c-MYC in transgenic mice and also to mediate the IgH-MYC translocations seen in Burkitt's lymphomas [27, 28]. In another key study, Beale and colleagues examined the substrate specificity of AID, A3G and A1, and noticed an enrichment for C>T mutations within *TP53* and *APC*; tumor suppressor genes that had been heavily sequenced in cancer samples, even at that time [29]. Later, it was shown that infection of gastric epithelial cells with *Helicobacter pylori* induces AID expression, leading to *TP53* mutations; one mechanism by which *H. pylori* infection may cause gastric cancer [30].

The recent application of next generation sequencing to model systems and cancer samples has provided support for A3 involvement in somatic mutagenesis, and has also begun to shed light upon the mechanisms by which this process occurs. Initially working in yeast, Roberts and colleagues demonstrated that chronic exposure to the DNA alkylating agent, methyl methanesulfonate (MMS) resulted in clustered mutations; primarily point mutations at C:G base pairs [31]. Importantly, these mutations were strand-coordinated (i.e. 'long series of either mutated cytosines or mutated guanines in the sequenced strand'), indicating that: (i) that they had arisen from a specific lesion affecting multiple cytosines or guanines in a long stretch of ssDNA (based on the mechanism of action of MMS, they postulated this to be N3-methyl cytosine; a ssDNA-specific lesion); and (ii) that they had occurred simultaneously (chronocoordinate), or at least within one cell cycle. Also apparent was strand switching, in which runs of mutated cytosines were followed by runs of mutated guanines, or vice versa. This suggested that the mutations could be occurring in stretches of ssDNA that became exposed during bidirectional 5' to 3' resection either side of DSBs. The use of mutants with replication fork dysfunction suggested that in addition to DSBs, stalled replication forks were also sources of ssDNA that gave rise to strand coordinated clusters. Next they looked at sequence data from three cancer types (head and neck (HNSC), prostate and multiple myeloma) and observed similar clusters of C- or G-coordinated clusters, often occurring close to chromosomal rearrangement breakpoints [31].

At the same time, Nik-Zainal and colleagues were extracting mutational signatures (Box 2) from breast cancer sequence data, and also noticed evidence of mutation clusters [32]. These clusters were also characterized by strand-coordinated mutations at C:G base pairs. Noting the similarity of these clusters to the multi-kilobase chronocoordinate mutation showers previously described in transgenic mice expressing a *lacI* reporter gene (and predicted to occur in cancer) [33], they termed them kataegis (after the Greek *kataegisa* meaning thunder shower, or tempest) [32]. Both studies [31, 32] noticed that the predominant mutations occurring in these clusters were C>T or C>G substitutions within the optimal substrate motif (TpCpW, where W = A or T) for a subset of 'TC-specific' A3s including A3A, A3B, A3C, A3F and A3H. Since A3s are known to be specific for ssDNA and to act processively along stretches of ssDNA [34], they were proposed in both papers as candidates for producing mutation showers in tumors [31, 32].

Independently, building upon their earlier demonstration that A3B expression is often dramatically elevated in cancer compared to normal tissues [6], Harris and colleagues found that A3B expression is correlated with the extent of cytosine mutations in breast tumors, and that deoxycytidine deaminase activity in nuclear extracts from breast cancer cell lines is abolished by A3B knockdown [35]. In the same study, A3B knockdown also reduced genomic uracil loads and mutagenesis of a thymidine kinase reporter gene. A role for A3B in generating kataegis in tumors is also supported by studies showing that it can induce similar mutation showers when overexpressed in yeast, as can A3A [24]. Subsequent pan-cancer screens by all three

groups suggest a mutagenic role for one or more of the TC-specific A3s in numerous tumor types [36-38].

Mechanisms of A3 deregulation

A3-mediated mutations, whilst reminiscent of AID-dependent SHM, are clearly an off-target activity occurring in some, but not all, cancers. We know little as yet regarding how and why A3s become deregulated in tumors, but we discuss some of the possible factors leading to changes in A3 expression or activity, availability of ssDNA substrate and mutagenesis rather than repair of deaminated dC.

A3 expression and stimulation

Although there is a bias in publicly available datasets, it does appear that A3 mutagenesis is particularly prevalent in carcinomas [36-38]. This may be because epithelia are a front-line barrier to viruses, and hence the innate immune response, including A3 viral restriction, is more primed or active in these cells. A3B is highly expressed in tumor types with enrichment of A3 mutation signatures [37, 38], and its expression correlates with prevalence of A3 signatures in breast and ovarian cancers [22, 35]. Much attention therefore, has focused on A3B as an agent for genomic mutations.

At first glance though, this is difficult to reconcile with genetic evidence that the A3A_B deletion polymorphism, where the entire A3B protein coding sequence is missing, carries an increased risk of breast cancer [39, 40]. Furthermore, a recent reanalysis of public breast cancer mutation data reported higher levels of A3

mutation signatures in patients harboring a copy of A3A_B, and more still in homozygotes (albeit there were very few of the latter), consistent with the A3A_B form (essentially A3A fused to the A3B 3' untranslated region (UTR)) being more mutagenic [41].

A possible resolution of this quandary comes from Wain-Hobson and colleagues who transfected an artificial A3A_B construct into human cells. The resulting protein product was expressed at 20-fold higher levels than an A3A construct, and showed a potent hyperediting activity on nuclear DNA [42]. A possible (but as yet unexplored inference) is that the A3A 3'UTR is normally targeted by one or more micro-RNAs that repress A3A expression, and that this repression is relieved in the A3A_B allele, in which A3A is instead fused to the A3B 3'UTR (Box 1). Regardless of the mechanistic details, it suggests that although A3B is expressed at much higher levels in some tumors (e.g. breast), in some circumstances A3A can be the more potent mutagenic enzyme. Teasing apart exactly which A3 is responsible for A3 signatures in tumors is hampered by the close similarity of the proteins. Localization of A3A and A3B expression by immunohistochemistry for example, has thus far been impossible due to a lack of suitable antibodies.

In addition to differential baseline expression of A3 enzymes, another intriguing possibility is that stimulation following viral infection could increase off-target A3 genomic mutations. The preponderance of A3 mutations in HPV driven cervical cancer, together with the observation that A3 mutations are enriched in the HPV-associated subset of head and neck squamous cell carcinoma (HNSC), suggest a

possible off-target response to the virus [37, 38, 43, 44]. Consistent with this, HPV16 infection has recently been shown to upregulate A3A and A3B mRNA expression in a keratinocyte cell line, and both are upregulated in pre-invasive cervical lesions [19, 45]. However, within cervical cancer and HNSC the enrichment of A3 signature mutations is not related to the expression of A3 genes, at least as seen in the tumor biopsy [43, 44]. Of course, it could be that when the mutations are occurring during development of these tumors, they are correlated with the expression of the A3 responsible, but that this relationship is lost following subsequent downregulation; a possibility alluded to by Roberts and Gordenin in their discussion of A3s as transient hypermutators [46]. Further, in hepatocytes expression of several A3s increases following hepatitis B virus (HBV) infection [47]. However there is no sign of off-target A3 mutations following this response, nor are they evident in HBV-associated hepatocellular carcinomas [43, 48].

Existing expression of A3 enzymes, or increased expression due to viral stimulation, may be important, but is unlikely sufficient to cause A3 deregulation. As yet there is little direct evidence of what else needs to go wrong, but perhaps an intriguing hint comes from a small study of chronic lymphocytic leukemias (CLL). CLL arise from B cells, and an A3-like kataegis signature was reported within CLL samples that had undergone SHM (IgV-Mut), but not in samples that had not (IgV-UM), whilst the expression of A3A and A3B proteins was unchanged [49, 50]. Although it remains necessary to establish whether this holds true in a larger sample set, it does suggest a common mechanism may have caused permissiveness both for AID-driven SHM and A3-mediated kataegis. Highly related nuclear localization sequences (NLS) have

been described in AID and A3B [51], raising the possibility that a common signal could stimulate the nuclear translocation of both, explaining the co-occurrence of the AID and A3 mutational signatures. Another possibility is switching to a specific error-prone pathway that processes genomic uracil in B cells undergoing SHM [52]. The finding that IgV-UM CLL express more AID than IgV-Mut CLL [53] also argues that a switch is necessary before AID becomes mutagenic, even in tumor cells. The same is possibly true of A3s.

Whilst the best-characterized activity of A3G against exogenous retroviruses requires cytoplasmic localization, all seven A3 genes have the ability to inhibit nuclear retrotransposon activity, albeit with differing potencies [12, 14, 15]. The single domain A3s (A3A, C and H), at around 23kDa are able to enter the nucleus via passive diffusion, whereas the larger A3B, D, F and G require an active import mechanism [14]. A3B contains a NLS and localizes to the nuclei of transfected cells, while A3D, F and G are cytoplasmic and do not appear to come into contact with DNA, even following breakdown of the nuclear envelope during mitosis [13, 14, 54]. It is possible that these cytoplasmic A3s inhibit retroelements in an indirect manner distinct from the deamination-based inhibition by A3A [16, 17]. Again, the close similarity between the A3s is an obstacle to study of their subcellular localization. The current picture comes largely from studies on overexpression of epitope-tagged proteins, which may behave differently from endogenous A3s. This issue has been addressed recently in the study of endogenous A3A localization in monocytes; a cell type with high endogenous A3A expression. In monocytes it appears there is a mechanism absent from other cell types that keeps A3A in the cytoplasm and

prevents it from damaging the genome [55]. Different polymorphisms in A3H have been shown to alter its subcellular localization, which suggests the affected amino acids play a role in blocking the passive diffusion of this enzyme between cytoplasm and nucleus [56].

Finally, in a cohort of 115 cervical cancers, missense mutations in A3 family genes (A3B, A3F and A3G) were found in four samples [44]. In each of these samples TpC mutations accounted for at least 70% of the total mutational load, and two had the highest nonsilent mutation rate amongst the cohort. Indeed, the two most heavily mutated samples harbored phenylalanine substitutions at a conserved serine in both A3B and A3F. Although functional analysis of the mutants was not presented, this observation suggests that occasionally gain of function mutations in A3s may play a driver role in cancer development. Irrespective of the mechanism, it seems likely that the evolutionary differentiation of A3 enzymes in humans, which are under selective pressure, has arisen to target specific viral insults in specific cells whilst carefully regulating their potential for self-harm.

Availability of ssDNA substrate

As A3s are only able to target ssDNA, the availability of this substrate is likely often a limiting factor for A3-mediated mutagenesis in tumor cells. ssDNA exposed during the resection stage of DSB repair is vulnerable to mutations, and kataegis is often associated with rearrangement breakpoints in human tumors [31, 32]. In addition to rearrangements, segmental amplifications or deletions (CNVs) also involve DSBs. Among breast cancers, the *HER2*-amplified subset displays the greatest enrichment

for A3 signature mutations [38] and also displays elevated markers of genomic instability, harboring twice the number of CNVs seen in other breast tumors [57, 58]. Breaks associated with CNVs may be an important source of ssDNA substrate in tumors such as these, although they are not correlated with A3 signature enrichment across multiple tumor types, at least as measured using exome sequencing data [38]. In yeast models A3-induced kataegis was potentiated by, and occurred close to, targeted DSBs introduced using a restriction endonuclease [24]. In the absence of endonuclease-induced DSBs, mutation showers induced by heterologous expression of cytosine deaminases are dependent upon UNG, which is required for the creation of DSBs following removal of uracil from opposing strands (as occurs during AID-induced class-switch recombination; Figure 1), [24, 59]. A3-mediated kataegis may therefore be either self-seeding or associated with DSBs arising via other mechanisms. Since they are associated with chromothripsis, both spatially and temporally (at least in breast cancer) [32, 60], this raises the question of whether kataegis is triggered by chromothripsis, or possibly vice-versa?

Another source of kataegis could be stretches of ssDNA exposed at stalled or collapsed replication forks. Yeast mutants with increased fork stalling display increased mutation showers upon MMS exposure [31] and fork stalling is a common feature of transformed cells; a consequence of replication stress caused by unregulated S-phase entry [61]. Further work in yeast has revealed that break-induced replication (BIR), a highly conserved mechanism for the repair of DSBs occurring at collapsed replication forks or eroded telomeres produces long stretches of ssDNA, resulting in kataegis-like mutation clusters [62]. The authors suggest that

BIR arising from defects in replication, rather than strand resection during the repair of DSBs, could account for the longer stretches of strand-coordinated editing seen in some kataegis events. Replication stress and A3-induced mutagenesis have recently been linked to loss of a chromosomal fragile site gene, FHIT [63]. FHIT is frequently lost very early in tumor development, causing replication stress due to deoxythymidine triphosphate (dTTP) depletion [64]. When genomic sequences from lung adenocarcinomas were stratified by A3B and FHIT expression, those with high A3B and FHIT loss showed significantly higher A3 signature mutation loads than high A3B expressers with normal FHIT levels, thus the mutagenic potential of A3B may be unleashed in the absence of FHIT [63].

In addition to BIR, mismatch repair (MMR) has also been implicated in A3-mediated mutagenesis. When Furano and colleagues introduced episomes with specific mismatches into human cell lines they observed that while the mismatches were repaired, the flanking sequences frequently incurred T_pC mutations characteristic of A3 activity [52]. Knockdown experiments showed that these A3-like mutations were dependent upon A3 expression and BER and MMR proteins. They proposed a model in which A3 mutations occur when the BER pathway is “hijacked” by a non-canonical MMR mechanism: the same pathway deployed during AID-dependent immunoglobulin SHM [52]. Indeed, the activation of such a pathway might explain the co-existence of AID- and A3-dependent mutation clusters in CLL [50].

Another possible exposure of ssDNA to A3s is in the transcription bubble. Transcription bubbles, structures in which ssDNA is exposed due to helicase activity

of the RNA polymerase complex, are known targets of AID [65-68] and represent a likely source of mutations in coding regions if replication occurs before repair of dU. In bladder cancer, A3 mutations are enriched in highly transcribed genes, particularly on the sense strand [69], suggesting that A3s also act on ssDNA during transcription. It is possible that a transcription-coupled repair pathway may suppress mutagenesis on the coding strand, although the evidence linking repair of deaminated dC to transcription remains controversial [70]. Alternatively, this strand bias is reminiscent of transcription-coupled AID mutagenesis, which acts upon structures called R-loops, in which the non-coding strand is selectively looped out and deaminated while the coding strand is hybridized to the primary transcript and thus protected [71]. Several mechanisms that have been posited to expose stretches of ssDNA to A3 activity are depicted in Figure 2.

Failure of DNA repair pathways

In HNSC and breast cancers a linear relationship between A3 signature mutations and other point mutations across cancers with widely varying mutational loads is observed [35, 43], consistent with a model in which various mutagenic processes are kept in check by DNA repair mechanisms and only manifest when these pathways fail. In B cells, for instance, the mutagenic activity of AID is restricted outside of SHM by a combination of BER and MMR [72]. Since mutagenesis at dU likely occurs when BER is interrupted by TLS (Figure 1), it is possible the unregulated progression through S-phase that occurs following loss of cell cycle checkpoint control during transformation acts to increase the mutagenic effect of A3 activity. Specific defects in DNA repair pathways have not been linked to the appearance of the A3 signature

in tumors however, and it may be the case that the repair pathways remain functional but unable to cope with increased rate of mutation following A3B upregulation [35], aberrant A3 localization or increased exposure of ssDNA.

In summary, many mechanisms may expose DNA to A3 deamination and a number of *in vitro* models have replicated kataegis-like phenomena. However even within and between different breast cancer samples a mixed presentation of A3-like mutations within large rearrangement-associated kataegis macroclusters, or sporadically throughout the genome, or with varying proportions of transition or transversion can be observed [32]. Indeed, a recent analysis of ovarian cancer has linked C>A transversions to A3B activity, with the implication that a different TLS polymerase may act on abasic sites caused by cytosine deamination in these tumors [22] (Figure 1). This suggests that many of the proposed models of A3-mediated mutagenesis maybe valid, and that together with increased A3A or A3B expression (as occurs during HPV infection, for example) there are multiple points of failure in DSB repair, replication or transcription that might potentiate A3 off-target activity, even within cancers arising in the same tissue.

A3 mutations: passengers or drivers?

An important question arising from the finding of widespread A3 mutations in many tumor types is whether this is a bystander phenomenon occurring as a result of a chaotic tumor genome, or whether the A3 mutational process contributes to tumor development. Transgenic expression of AID and A1 are tumorigenic in mice, and C>T mutations suggestive of AID/APOBEC involvement were previously noted in *TP53*,

MYC and *APC* [25, 27, 29, 30]. In exome data, enrichment of A3 mutations amongst putative cancer driver genes as defined by either the catalogue of somatic mutations in cancer (COSMIC) database [73] and/or the MutSig collection [74] of recurrently mutated genes [38, 43] have been observed. Upon inspection of the genes most frequently A3-mutated in HPV-associated cancers, the *PIK3CA* proto-oncogene is almost exclusively mutated at two helical domain hotspots (E542K and E545K) [43]. These are always G>A substitutions within the ApGpT context, thus the result of TpCpA transitions on the noncoding strand. Indeed, *PIK3CA* helical domain mutations are strongly correlated with A3 mutation enrichment across multiple tumor types, while an equally oncogenic and prevalent hotspot mutation (H1047R), which is not caused by an A3 signature mutation is largely confined to tumor types with reduced enrichment for this signature across the exome [43]. These observations, together with the finding that *PIK3CA* helical domain mutation coincides with appearance of the A3 signature in lung cancer subclones [75] strongly implicate A3 activity in generating these driver mutations.

Accumulating evidence suggests that the stage of tumor development at which the A3 signature appears varies considerably between tumor types. Analysis of metachronous paired samples representing early superficial noninvasive (Ta) and subsequent invasive (T1/T2) bladder tumors both displayed the A3 signature equally [69]. In breast cancers, A3 mutations begin to accumulate early but make a greater contribution to the mutational load (particularly C>G transversions) of later subclones, suggesting an increasing fraction of A3 mutagenesis as the tumors evolve [60]. The same study revealed evidence of distinct kataegis events occurring at

separate points during the development of a given tumor. Similarly, in non-small cell lung adenocarcinomas a recent evolutionary analysis of dissected subclones showed the A3 signature is weaker early in the evolution of disease (when tobacco-associated mutations predominate) but appears strongly in subclones [75]. This pattern was also apparent in lung adenocarcinomas but not in lung squamous carcinomas. In summary, there is no simple pan-cancer story of A3 mutagenesis. There is variation even between different histological subtypes arising in a common tissue, however it appears that A3 mutagenesis contributes to the early development of certain tumors, and can still fuel metastasis and/or drug-resistance at advanced stages by driving the evolution of tumor subclones.

Concluding remarks and future perspectives

The discovery of A3 signature mutations, and more broadly the dissection of recurring mutational signatures across cancers, has been a highlight of the effort to collect and “data-mine” the tumor databases. Based upon the current literature, we can speculate on the causes and consequences of A3-mediated mutagenesis in tumor development (Figure 3), but it is essential that these observational studies are followed by further experimentation before we can confidently link these signature mutations to A3 activity. To date, of the 11 APOBEC genes in humans only A1 and AID have been shown to cause tumors when overexpressed in mice [25, 27]. Conversely, A3 loss-of-function experiments in models of tumor development are also currently lacking, as are studies testing the models of A3-mediated mutagenesis that have been proposed based upon AID function in lymphocytes. Indeed, whilst a diversity of A3 signatures in tumors is observed (from none to many, from sporadic

to clustered kataegis, and from early beginning to late developing), specific patterns in specific tumors *in vivo* have not been associated with the putative mechanisms discovered *in vitro*. This would be an important step forward, as the mutational signature of a particular tumor could serve as a genomic readout of the underlying physiological problems occurring in DNA repair pathways of that tumor. This could open up new avenues of data-mining in the tumor databases as we might relate types of A3 signature (rather than just their quantity) to different polymorphisms, to expression signatures, or mutations of specific genes and pathways. DNA repair pathways have become drug targets either to prevent therapeutic resistance or kill cancer cells dependent upon particular pathways [76]. A genomic readout indicating which DNA repair pathways are dysregulated would provide useful guidance for this work.

Further observational work on A3 mutational signatures is likely to follow the general trend of cancer genomics, which is to recognize cancer as an evolving clonal disease. Early indications are that A3 signatures arise early in some tumors whilst in others they appear more strongly in later subclones. These analyses will continue as we seek to understand whether A3 signatures may point to preventive therapies or possibly could be biomarkers of early disease.

Finally, the association between A3 polymorphisms, mutagenic potential and cancer risk is particularly intriguing. The dramatic effect that underlying genetics and/or environmental factors can have on mutational signatures was recently highlighted by a global analysis of hepatocellular carcinoma (HCC), in which cancers of Japanese,

European and US-Asian patients were characterized by different mutational signatures [77]. Given the association of the A3 chimeric polymorphism with breast cancer, one is left wondering why it has increasingly gained prevalence outside of African populations. Cancer genome studies from more diverse populations may help us untangle the competing selective pressures behind the evolution of the A3 family.

Acknowledgments

This work was supported by funding from Rosetrees Trust and Cancer Research UK. We thank Dr Elza De Bruin and Dr Jenni Nikkila for critical reading of the manuscript. The authors have no conflicts of interest to declare.

References

1. Harris RS, Liddament MT: **Retroviral restriction by APOBEC proteins.** *Nature reviews Immunology* 2004, **4**(11):868-877.
2. Teng B, Burant CF, Davidson NO: **Molecular cloning of an apolipoprotein B messenger RNA editing protein.** *Science* 1993, **260**(5115):1816-1819.
3. Powell LM, Wallis SC, Pease RJ, Edwards YH, Knott TJ, Scott J: **A novel form of tissue-specific RNA processing produces apolipoprotein-B48 in intestine.** *Cell* 1987, **50**(6):831-840.
4. Conticello SG, Langlois M-AA, Yang Z, Neuberger MS: **DNA deamination in immunity: AID in the context of its APOBEC relatives.** *Advances in immunology* 2007, **94**:37-73.
5. Di Noia JM, Neuberger MS: **Molecular mechanisms of antibody somatic hypermutation.** *Annual review of biochemistry* 2007, **76**:1-22.
6. Harris RS, Petersen-Mahrt SK, Neuberger MS: **RNA editing enzyme APOBEC1 and some of its homologs can act as DNA mutators.** *Molecular Cell* 2002, **10**(5):1247-1253.
7. Harris RS, Bishop KN, Sheehy AM, Craig HM, Petersen-Mahrt SK, Watt IN, Neuberger MS, Malim MH: **DNA Deamination Mediates Innate Immunity to Retroviral Infection.** *Cell* 2003, **113**(6):803-809.
8. Mangeat B, Turelli P, Caron G, Friedli M, Perrin L, Trono D: **Broad antiretroviral defence by human APOBEC3G through lethal editing of nascent reverse transcripts.** *Nature* 2003, **424**(6944):99-103.
9. Bishop KN: **APOBEC-Mediated Editing of Viral RNA.** *Science* 2004.
10. Sheehy AM, Gaddis NC, Malim MH: **The antiretroviral enzyme APOBEC3G is degraded by the proteasome in response to HIV-1 Vif.** *Nature medicine* 2003, **9**(11):1404-1407.
11. Stopak K, de Noronha C, Yonemoto W, Greene WC: **HIV-1 Vif blocks the antiviral activity of APOBEC3G by impairing both its translation and intracellular stability.** *Molecular Cell* 2003, **12**(3):591-601.
12. Bogerd HP, Wiegand HL, Doehle BP, Lueders KK, Cullen BR: **APOBEC3A and APOBEC3B are potent inhibitors of LTR-retrotransposon function in human cells.** *Nucleic Acids Research* 2006, **34**(1):89-95.
13. Bogerd HP, Wiegand HL, Hulme AE, Garcia-Perez JL, O'Shea KS, Moran JV, Cullen BR: **Cellular inhibitors of long interspersed element 1 and Alu retrotransposition.** *Proceedings of the National Academy of Sciences* 2006, **103**(23):8780-8785.
14. Kinomoto M, Kanno T, Shimura M, Ishizaka Y, Kojima A, Kurata T, Sata T, Tokunaga K: **All APOBEC3 family proteins differentially inhibit LINE-1 retrotransposition.** *Nucleic Acids Research* 2007, **35**(9):2955-2964.
15. Muckenfuss H, Hamdorf M, Held U, Perkovic M, Lower J, Cichutek K, Flory E, Schumann GG, Munk C: **APOBEC3 Proteins Inhibit Human LINE-1 Retrotransposition.** *Journal of Biological Chemistry* 2006, **281**(31):22161-22172.
16. Richardson SR, Narvaiza I, Planegger RA, Weitzman MD, Moran JV: **APOBEC3A deaminates transiently exposed single-strand DNA during LINE-1 retrotransposition.** *eLife* 2014, **3**.

17. Stenglein MD, Burns MB, Li M, Lengyel J, Harris RS: **APOBEC3 proteins mediate the clearance of foreign DNA from human cells.** *Nature structural & molecular biology* 2010, **17**(2):222-229.
18. Vartanian J-P, Guetard D, Henry M, Wain-Hobson S: **Evidence for Editing of Human Papillomavirus DNA by APOBEC3 in Benign and Precancerous Lesions.** *Science* 2008, **320**(5873):230-233.
19. Warren CJ, Xu T, Guo K, Griffin LM, Westrich JA, Lee D, Lambert PF, Santiago ML, Pyeon D: **APOBEC3A Functions as a Restriction Factor of Human Papillomavirus.** *Journal of virology* 2014.
20. Chaudhuri J, Alt FW: **Class-switch recombination: interplay of transcription, DNA deamination and DNA repair.** *Nature reviews Immunology* 2004, **4**(7):541-552.
21. Krokan HE, Saetrom P, Aas PA, Pettersen HS, Kavli B, Slupphaug G: **Error-free versus mutagenic processing of genomic uracil--relevance to cancer.** *DNA Repair (Amst)* 2014, **19**:38-47.
22. Leonard B, Hart SN, Burns MB, Carpenter MA, Temiz NA, Rathore A, Vogel RI, Nikas JB, Law EK, Brown WL *et al*: **APOBEC3B upregulation and genomic mutation patterns in serous ovarian carcinoma.** *Cancer Research* 2013, **73**(24):7222-7231.
23. Chan K, Resnick MA, Gordenin DA: **The choice of nucleotide inserted opposite abasic sites formed within chromosomal DNA reveals the polymerase activities participating in translesion DNA synthesis.** *DNA Repair (Amst)* 2013, **12**(11):878-889.
24. Taylor BJ, Nik-Zainal S, Wu YL, Stebbings LA, Raine K, Campbell PJ, Rada C, Stratton MR, Neuberger MS: **DNA deaminases induce break-associated mutation showers with implication of APOBEC3B and 3A in breast cancer kataegis.** *eLife* 2013, **2**.
25. Yamanaka S, Balestra ME, Ferrell LD, Fan J, Arnold KS, Taylor S, Taylor JM, Innerarity TL: **Apolipoprotein B mRNA-editing protein induces hepatocellular carcinoma and dysplasia in transgenic animals.** *Proc Natl Acad Sci U S A* 1995, **92**(18):8483-8487.
26. Petersen-Mahrt SK, Harris RS, Neuberger MS: **AID mutates E. coli suggesting a DNA deamination mechanism for antibody diversification.** *Nature* 2002, **418**(6893):99-103.
27. Okazaki IM, Hiai H, Kakazu N, Yamada S, Muramatsu M, Kinoshita K, Honjo T: **Constitutive expression of AID leads to tumorigenesis.** *The Journal of experimental medicine* 2003, **197**(9):1173-1181.
28. Ramiro AR, Jankovic M, Eisenreich T, Difilippantonio S, Chen-Kiang S, Muramatsu M, Honjo T, Nussenzweig A, Nussenzweig MC: **AID is required for c-myc/IgH chromosome translocations in vivo.** *Cell* 2004, **118**(4):431-438.
29. Beale RC, Petersen-Mahrt SK, Watt IN, Harris RS, Rada C, Neuberger MS: **Comparison of the differential context-dependence of DNA deamination by APOBEC enzymes: correlation with mutation spectra in vivo.** *Journal of molecular biology* 2004, **337**(3):585-596.
30. Matsumoto Y, Marusawa H, Kinoshita K, Endo Y, Kou T, Morisawa T, Azuma T, Okazaki IM, Honjo T, Chiba T: **Helicobacter pylori infection triggers aberrant expression of activation-induced cytidine deaminase in gastric epithelium.** *Nature medicine* 2007, **13**(4):470-476.

31. Roberts SA, Sterling J, Thompson C, Harris S, Mav D, Shah R, Klimczak LJ, Kryukov GV, Malc E, Mieczkowski PA *et al*: **Clustered mutations in yeast and in human cancers can arise from damaged long single-strand DNA regions.** *Molecular cell* 2012, **46**(4):424-435.
32. Nik-Zainal S, Alexandrov LB, Wedge DC, Van Loo P, Greenman CD, Raine K, Jones D, Hinton J, Marshall J, Stebbings LA *et al*: **Mutational processes molding the genomes of 21 breast cancers.** *Cell* 2012, **149**(5):979-993.
33. Wang J, Gonzalez KD, Scaringe WA, Tsai K, Liu N, Gu D, Li W, Hill KA, Sommer SS: **Evidence for mutation showers.** *Proc Natl Acad Sci U S A* 2007, **104**(20):8403-8408.
34. Chelico L, Pham P, Calabrese P, Goodman MF: **APOBEC3G DNA deaminase acts processively 3[prime] [rarr] 5[prime] on single-stranded DNA.** *Nat Struct Mol Biol* 2006, **13**(5):392-399.
35. Burns MB, Lackey L, Carpenter MA, Rathore A, Land AM, Leonard B, Refsland EW, Kotandeniya D, Tretyakova N, Nikas JB *et al*: **APOBEC3B is an enzymatic source of mutation in breast cancer.** *Nature* 2013, **494**(7437):366-370.
36. Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio SA, Behjati S, Biankin AV, Bignell GR, Bolli N, Borg A, Brackenbury AP *et al*: **Signatures of mutational processes in human cancer.** *Nature* 2013, **500**(7463):415-421.
37. Burns MB, Temiz NA, Harris RS: **Evidence for APOBEC3B mutagenesis in multiple human cancers.** *Nat Genet* 2013, **45**(9):977-983.
38. Roberts SA, Lawrence MS, Klimczak LJ, Grimm SA, Fargo D, Stojanov P, Kiezun A, Kryukov GV, Carter SL, Saksena G *et al*: **An APOBEC cytidine deaminase mutagenesis pattern is widespread in human cancers.** *Nat Genet* 2013, **45**(9):970-976.
39. Long J, Delahanty RJ, Li G, Gao Y-TT, Lu W, Cai Q, Xiang Y-BB, Li C, Ji B-TT, Zheng Y *et al*: **A common deletion in the APOBEC3 genes and breast cancer risk.** *Journal of the National Cancer Institute* 2013, **105**(8):573-579.
40. Xuan D, Li G, Cai Q, Deming-Halverson S, Shrubsole MJ, Shu X-OO, Kelley MC, Zheng W, Long J: **APOBEC3 deletion polymorphism is associated with breast cancer risk among women of European ancestry.** *Carcinogenesis* 2013, **34**(10):2240-2243.
41. Nik-Zainal S, Wedge DC, Alexandrov LB, Petljak M, Butler AP, Bolli N, Davies HR, Knappskog S, Martin S, Papaemmanuil E *et al*: **Association of a germline copy number polymorphism of APOBEC3A and APOBEC3B with burden of putative APOBEC-dependent mutations in breast cancer.** *Nature genetics* 2014, **46**(5):487-491.
42. Caval V, Suspene R, Shapira M, Vartanian J-PP, Wain-Hobson S: **A prevalent cancer susceptibility APOBEC3A hybrid allele bearing APOBEC3B 3'UTR enhances chromosomal DNA damage.** *Nature communications* 2014, **5**:5129.
43. 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 reports* 2014, **7**(6):1833-1841.

44. Ojesina AI, Lichtenstein L, Freeman SS, Peadarallu CS, Imaz-Rosshandler I, Pugh TJ, Cherniack AD, Ambrogio L, Cibulskis K, Bertelsen B *et al*: **Landscape of genomic alterations in cervical carcinomas.** *Nature* 2014, **506**(7488):371-375.
45. Vieira VC, Leonard B, White EA, Starrett GJ, Temiz NA, Lorenz LD, Lee D, Soares MA, Lambert PF, Howley PM *et al*: **Human Papillomavirus E6 Triggers Upregulation of the Antiviral and Cancer Genomic DNA Deaminase APOBEC3B.** *mBio* 2014, **5**(6).
46. Roberts SA, Gordenin DA: **Clustered and genome-wide transient mutagenesis in human cancers: Hypermutation without permanent mutators or loss of fitness.** *BioEssays : news and reviews in molecular, cellular and developmental biology* 2014.
47. Bonvin M, Achermann F, Greeve I, Stroka D, Keogh A, Inderbitzin D, Candinas D, Sommer P, Wain-Hobson S, Vartanian JP *et al*: **Interferon-inducible expression of APOBEC3 editing enzymes in human hepatocytes and inhibition of hepatitis B virus replication.** *Hepatology* 2006, **43**(6):1364-1374.
48. Lucifora J, Xia Y, Reisinger F, Zhang K, Stadler D, Cheng X, Sprinzl MF, Koppensteiner H, Makowska Z, Volz T *et al*: **Specific and nonhepatotoxic degradation of nuclear hepatitis B virus cccDNA.** *Science (New York, NY)* 2014, **343**(6176):1221-1228.
49. Puente XS, Pinyol M, Quesada Vc, Conde L, Ord $\sqrt{\geq\sqrt{\pm ez GR, Villamor N, Escaramis G, Jares P, Be\sqrt{\dagger Sl, Gonz\sqrt{\circ lez-D\sqrt{\neq az M et al$: **Whole-genome sequencing identifies recurrent mutations in chronic lymphocytic leukaemia.** *Nature* 2011, **475**(7354):101-105.
50. Rebhandl S, Huemer M, Gassner FJ, Zaborsky N, Hebenstreit D, Catakovic K, Grossinger EM, Greil R, Geisberger R: **APOBEC3 signature mutations in chronic lymphocytic leukemia.** *Leukemia* 2014.
51. Lackey L, Demorest ZL, Land AM, Hultquist JF, Brown WL, Harris RS: **APOBEC3B and AID have similar nuclear import mechanisms.** *Journal of molecular biology* 2012, **419**(5):301-314.
52. Chen J, Miller BF, Furano AV: **Repair of naturally occurring mismatches can induce mutations in flanking DNA.** *eLife* 2014, **3**.
53. Heintel D, Kroemer E, Kienle D, Schwarzinger I, Gleiss A, Schwarzmeier J, Marculescu R, Le T, Mannhalter C, Gaiger A *et al*: **High expression of activation-induced cytidine deaminase (AID) mRNA is associated with unmutated IGVH gene status and unfavourable cytogenetic aberrations in patients with chronic lymphocytic leukaemia.** *Leukemia* 2004, **18**(4):756-762.
54. Lackey L, Law EK, Brown WL, Harris RS: **Subcellular localization of the APOBEC3 proteins during mitosis and implications for genomic DNA deamination.** *Cell Cycle* 2013, **12**(5):762-772.
55. Land AM, Law EK, Carpenter MA, Lackey L, Brown WL, Harris RS: **Endogenous APOBEC3A DNA cytosine deaminase is cytoplasmic and nongenotoxic.** *J Biol Chem* 2013, **288**(24):17253-17260.
56. Li MM, Emerman M: **Polymorphism in human APOBEC3H affects a phenotype dominant for subcellular localization and antiviral activity.** *Journal of virology* 2011, **85**(16):8197-8207.

57. Ellsworth RE, Ellsworth DL, Patney HL, Deyarmin B, Love B, Hooke JA, Shriver CD: **Amplification of HER2 is a marker for global genomic instability.** *BMC Cancer* 2008, **8**:297.
58. Isola J, Chu L, DeVries S, Matsumura K, Chew K, Ljung BM, Waldman FM: **Genetic alterations in ERBB2-amplified breast carcinomas.** *Clin Cancer Res* 1999, **5**(12):4140-4145.
59. Lada AG, Dhar A, Boissy RJ, Hirano M, Rubel AA, Rogozin IB, Pavlov YI: **AID/APOBEC cytosine deaminase induces genome-wide kataegis.** *Biol Direct* 2012, **7**:47; discussion 47.
60. Nik-Zainal S, Van Loo P, Wedge DC, Alexandrov LB, Greenman CD, Lau KW, Raine K, Jones D, Marshall J, Ramakrishna M *et al*: **The life history of 21 breast cancers.** *Cell* 2012, **149**(5):994-1007.
61. Zeman MK, Cimprich KA: **Causes and consequences of replication stress.** *Nat Cell Biol* 2014, **16**(1):2-9.
62. Sakofsky CJ, Roberts SA, Malc E, Mieczkowski PA, Resnick MA, Gordenin DA, Malkova A: **Break-induced replication is a source of mutation clusters underlying kataegis.** *Cell reports* 2014, **7**(5):1640-1648.
63. Waters CE, Saldivar JC, Amin ZA, Schrock MS, Huebner K: **FHIT loss-induced DNA damage creates optimal APOBEC substrates: Insights into APOBEC-mediated mutagenesis.** *Oncotarget* 2014.
64. Saldivar JC, Miuma S, Bene J, Hosseini SA, Shibata H, Sun J, Wheeler LJ, Mathews CK, Huebner K: **Initiation of genome instability and preneoplastic processes through loss of Fhit expression.** *PLoS Genetics* 2012, **8**(11):e1003077.
65. Chaudhuri J, Tian M, Khuong C, Chua K, Pinaud E, Alt FW: **Transcription-targeted DNA deamination by the AID antibody diversification enzyme.** *Nature* 2003, **422**(6933):726-730.
66. Larson ED, Maizels N: **Transcription-coupled mutagenesis by the DNA deaminase AID.** *Genome Biology* 2004, **5**(3):211.
67. Ramiro AR, Stavropoulos P, Jankovic M, Nussenzweig MC: **Transcription enhances AID-mediated cytidine deamination by exposing single-stranded DNA on the nontemplate strand.** *Nature immunology* 2003, **4**(5):452-456.
68. Sohail A, Klapacz J, Samaranayake M, Ullah A, Bhagwat AS: **Human activation induced cytidine deaminase causes transcription dependent, strand biased C to U deaminations.** *Nucleic Acids Research* 2003, **31**(12):2990-2994.
69. Nordentoft I, Lamy P, Birkenkamp-Demtroder K, Shumansky K, Vang S, Hornshoj H, Juul M, Villesen P, Hedegaard J, Roth A *et al*: **Mutational context and diverse clonal development in early and late bladder cancer.** *Cell reports* 2014, **7**(5):1649-1663.
70. Hanawalt PC, Spivak G: **Transcription-coupled DNA repair: two decades of progress and surprises.** *Nat Rev Mol Cell Biol* 2008, **9**(12):958-970.
71. Vaidyanathan B, Yen WF, Pucella JN, Chaudhuri J: **AIDing Chromatin and Transcription-Coupled Orchestration of Immunoglobulin Class-Switch Recombination.** *Front Immunol* 2014, **5**:120.

72. Liu M, Duke JL, Richter DJ, Vinuesa CG, Goodnow CC, Kleinstein SH, Schatz DG: **Two levels of protection for the B cell genome during somatic hypermutation.** *Nature* 2008, **451**(7180):841-845.
73. Futreal PA, Coin L, Marshall M, Down T, Hubbard T, Wooster R, Rahman N, Stratton MR: **A census of human cancer genes.** *Nature reviews Cancer* 2004, **4**(3):177-183.
74. Lawrence MS, Stojanov P, Polak P, Kryukov GV, Cibulskis K, Sivachenko A, Carter SL, Stewart C, Mermel CH, Roberts SA *et al*: **Mutational heterogeneity in cancer and the search for new cancer-associated genes.** *Nature* 2013, **499**(7457):214-218.
75. de Bruin EC, McGranahan N, Mitter R, Salm M, Wedge DC, Yates L, Jamal-Hanjani M, Shafi S, Murugaesu N, Rowan AJ *et al*: **Spatial and temporal diversity in genomic instability processes defines lung cancer evolution.** *Science (New York, NY)* 2014, **346**(6206):251-256.
76. Rehman FL, Lord CJ, Ashworth A: **Synthetic lethal approaches to breast cancer therapy.** *Nat Rev Clin Oncol* 2010, **7**(12):718-724.
77. Totoki Y, Tatsuno K, Covington KR, Ueda H, Creighton CJ, Kato M, Tsuji S, Donehower LA, Slagle BL, Nakamura H *et al*: **Trans-ancestry mutational landscape of hepatocellular carcinoma genomes.** *Nature genetics* 2014.
78. LaRue RS, Jonsson SR, Silverstein KA, Lajoie M, Bertrand D, El-Mabrouk N, Hotzel I, Andresdottir V, Smith TP, Harris RS: **The artiodactyl APOBEC3 innate immune repertoire shows evidence for a multi-functional domain organization that existed in the ancestor of placental mammals.** *BMC Mol Biol* 2008, **9**:104.
79. Munk C, Beck T, Zielonka J, Hotz-Wagenblatt A, Chareza S, Battenberg M, Thielebein J, Cichutek K, Bravo IG, O'Brien SJ *et al*: **Functions, structure, and read-through alternative splicing of feline APOBEC3 genes.** *Genome Biology* 2008, **9**(3):R48.
80. Munk C, Willemsen A, Bravo IG: **An ancient history of gene duplications, fusions and losses in the evolution of APOBEC3 mutators in mammals.** *BMC evolutionary biology* 2012, **12**:71.
81. Kidd JM, Newman TL, Tuzun E, Kaul R, Eichler EE: **Population stratification of a common APOBEC gene deletion polymorphism.** *PLoS genetics* 2007, **3**(4).
82. Harari A, Ooms M, Mulder LC, Simon V: **Polymorphisms and splice variants influence the antiretroviral activity of human APOBEC3H.** *Journal of virology* 2009, **83**(1):295-303.
83. OhAinle M, Kerns JA, Li MM, Malik HS, Emerman M: **Antiretroelement activity of APOBEC3H was lost twice in recent human evolution.** *Cell Host & Microbe* 2008, **4**(3):249-259.
84. Wang X, Abudu A, Son S, Dang Y, Venta PJ, Zheng YH: **Analysis of human APOBEC3H haplotypes and anti-human immunodeficiency virus type 1 activity.** *Journal of virology* 2011, **85**(7):3142-3152.
85. Helleday T, Eshtad S, Nik-Zainal S: **Mechanisms underlying mutational signatures in human cancers.** *Nature reviews Genetics* 2014, **15**(9):585-598.
86. Dulak AM, Stojanov P, Peng S, Lawrence MS, Fox C, Stewart C, Bandla S, Imamura Y, Schumacher SE, Shefler E *et al*: **Exome and whole-genome**

- sequencing of esophageal adenocarcinoma identifies recurrent driver events and mutational complexity.** *Nature genetics* 2013, **45**(5):478-486.
87. Roberts SA, Gordenin DA: **Hypermutation in human cancer genomes: footprints and mechanisms.** *Nature reviews Cancer* 2014, **14**(12):786-800.
88. Curtin NJ: **DNA repair dysregulation from cancer driver to therapeutic target.** *Nature reviews Cancer* 2012, **12**(12):801-817.

Box 1: Evolution and population variation within the A3 gene family

The A3 genes have evolved and diversified extensively throughout mammalian evolution, and still show signs of ongoing selective pressure in humans. The ancestral A3 enzyme evolved from the ancestral AID cytidine deaminase that is widespread throughout vertebrates. The mouse genome has a single A3 gene, the dog has two, but there are four in cats, six in horses and seven in primates, with the human A3 family all residing within a single 100 kb region on chr22 (A3A, A3B, A3C, A3D, A3F, A3G, A3H). This complexity has arisen from a series of duplications and fusions amongst three ancient clades of A3 genes (A3Z1-A3Z3) established subsequent to the evolution of placental mammals some 100 million years ago (A3 homologs are not found in monotremes or marsupials) [78-80]. For instance, the human A3A gene arose from A3Z1, and the human A3B gene arose from fusion of A3Z1 and A3Z2 genes (Figure 1). The close relatedness of A3 genes presents a challenge to researchers studying the expression and functions of the different family members.

Adding further diversity, a common deletion polymorphism between A3A exon 5 and A3B exon 8 produces an A3A_B fusion transcript, completely removing the A3A 3'UTR and A3B CDS. This allele, in which the A3A CDS is linked to the A3B 3'UTR displays a remarkable geographical gradient across the human population; rare in Africans (~1%), but found in 6% of Europeans, 37% in East Asians, 58% in Amerindians, and 93% of Oceanians [81]. This suggests that A3B may be under competing selective pressures, speculatively with the full A3B protective against a particular African pathogen but progressively lost as humans migrated further out of Africa. Other A3 genes also display a variety of haplotypes, including A3H, in which

SNPs common in the African population confer increased stability, altered subcellular localization and enhanced anti-retroviral function to the resulting proteins [56, 82-84].

Box 2: Extracting mutation signatures

The “mutational portraits” of different cancers are composed from differing quotients of many distinct mutational processes, and the snapshot at the point a tumor sample is sequenced is an accumulation of the history of these processes and a combination across the heterogeneity of the tumor sample [85]. Extensive sequencing across many samples is required so that recurrent and overlapping mutational processes can be teased apart into their constituent signatures. Recently a breakthrough in the study of point mutation processes in cancer came with the analysis of a large collection of many thousands of tumor samples using the Non-Negative Matrix Factorization (NMF) algorithm [36]. For the cancer genome study it was used to separate up to 20 distinct mutational signatures from the noisy overlapping collection of signatures, including two signatures dominated by C>T and C>G mutations at TpC sites, indicative of A3 activity. Exogenous damage processes underlie other signatures, including UV radiation on dipyrimidines (CC>TT) strongly apparent in melanoma, whilst (G>T) transversions associated with smoking proliferate in lung cancers, and spontaneous deamination at methyl-CpG sites leading to transitions (C>T) are a common age-associated phenomenon throughout many types of cancer [36]. The separation of these mutational signals can give insight into the driving forces behind tumorigenesis. For instance, a study of

esophageal cancers has revealed a quite unique signature of (AA>AC) transversions making up to 29% of all substitutions in this tumor [86]. It is likely that this reveals a particular mutagenic effect associated with gastroesophageal reflux, raising possibilities for esophageal cancer prevention or surveillance. When one is interested in the extent to which a specific mutational signature might be enriched in a particular tumor or subset of tumors, methods such as NMF can be complemented by calculating the observed frequency at which the signature mutations occur in an exome or genome versus that expected by chance, allowing statistical comparison between individual samples. This approach has been used to identify enrichment for A3 signature mutations in HER2-amplified breast cancer and HPV-associated HNSC, for instance [38, 43, 87]. In addition to the signatures arising from external DNA damaging processes, endogenous error in the replication or repair process may introduce errors independently, or if defective, fail to properly repair the normal physiological range of DNA damage properly. How these repair and replication processes fail in different cancers may also modulate the mutation signature, and is a fertile new field in mutation research [88].

Box 3: Outstanding questions

1. How does the A3 innate immune system become deregulated in cancer? Is it an excess of DNA damage creating substrate, or failure of the downstream correction pathways? Is increased A3B expression sufficient? Does loss of a regulatory switch misdirect A3 activity to the genome?
2. Is A3B uniquely harmful but protective against an African pathogen, hence its retention there and loss elsewhere? Why is it harmful?

3. The A3A_B polymorphism predisposes to breast cancer but does it affect risk of developing other cancers, in particular those such as bladder and HPV-related cancers, in which A3s appear to play a prominent role in tumorigenesis?
4. Which A3 is primarily responsible for somatic mutagenesis during cancer development and does the A3 family member involved vary from one cancer to another?
5. Could viral infections, even those subsequently cleared, cause potentially cancer-initiating mutations in our DNA by triggering an A3 response?
6. Does A3 activity contribute significant numbers of DSBs in cancer cells? Could A3s be responsible for a portion of the insertion/deletions, rearrangements and even chromothripsis observed in certain tumors?

Glossary

Base excision repair (BER): The major pathway by which deaminated cytosine is repaired, involving removal of the damaged base by a DNA glycosylase (UNG) to create an apyrimidinic (AP) site, nicking of the phosphodiester backbone on the affected strand by an AP endonuclease (e.g. APE1) and then replacement of the cytosine and ligation to complete repair. Problems occur when this process is interrupted by replication via the action of TLS polymerases.

Chromothripsis: From the Greek meaning chromosome shattering, in which many rearrangements are observed in a chromosome region suggesting fragmentation followed by error-prone DSB repair.

Break-Induced replication: A specialized variant of replication that is used for the repair of single-ended DSBs such as those occurring at collapsed replication forks and eroded telomeres. BIR creates very long stretches of ssDNA, presumably a prerequisite for A3-mediated kataegis and is highly conserved throughout eukaryotes although so far, evidence linking it to kataegis is restricted to yeast models.

Deamination: Here refers to the removal of an NH₂ group from a base in RNA or DNA (commonly cytidine/deoxycytidine, in

which case the product is uridine/deoxyuridine respectively).

Exome: Refers to the coding portion of the genome. This comprises only around 1% of the genome and is thus cheaper to sequence and the data generated are more straightforward to analyze.

Human papillomaviruses (HPV): Small double stranded DNA viruses with a tropism for either mucosal or cutaneous epithelia. High-risk mucosal variants (e.g. HPV16) cause almost all cervical cancers and a subset of head and neck squamous carcinomas.

Hypermutator: The term used to describe certain cancer samples that are statistical outliers, displaying many more mutations than other tumors of a similar type. In hypermutated breast cancers, many of these mutations are of the A3 type and these are the samples in which kataegis events are commonly seen.

Mismatch repair (MMR): A mode of post-replication repair in which the parental strand is used as a template from which to recognize and repair mismatches in the daughter strand that have occurred during replication.

Replication fork: The structure formed by the replication complex, at which point helicases have separated the two DNA strands to allow replication.

- Retrotransposons:** Retrotransposons are mobile elements (essentially endogenous retroviruses) that make up around 40% of human genomic DNA. They proliferate via transcription to RNA, followed by reverse transcription and re-integration. They are divided into two broad groups: the viral long terminal repeat (LTR) transposons, and the long (LINE-1/L1) and short (SINE) interspersed elements. Retrotransposons cause inherited disorders and cancer through insertional mutagenesis.
- Transition mutation:** A mutation in which one purine is replaced by another, and hence one pyrimidine is replaced by another on the complementary strand (and vice-versa).
- Translesion synthesis (TLS):** The mode of replication in which specialized DNA polymerases mediate strand synthesis in the presence of lesions (e.g. abasic sites) on the parental strand. This allows replication to be completed in a timely manner but at the expense of generating mutations.
- Transversion mutation:** A mutation in which a purine is replaced by a pyrimidine (and vice versa).

Figure Legends

Figure 1. Error-prone processing of genomic uracil downstream of A3-mediated cytosine deamination results in transitions, transversions or double strand breaks (DSBs).

A3 mutation is a multistep process that begins with the exposure of ssDNA (shown here to occur either when the complementary strand is absent such as following resection at DSBs, or in transiently unwound DNA such as in transcription bubbles) to the activity of an A3 enzyme. Multiple deoxycytidines (usually within TpCpA/T/G motifs) may then be deaminated to deoxythymidines. The normal process of BER would then be removal of the base by UNG to yield an apyrimidinic (AP) site, 'nicking' by endonuclease APE1 and then repair by XRCC1-recruited polymerases and ligases. Interruption of this process by replication prior to UNG-catalyzed base excision results in C>T transitions (left). More commonly, base excision occurs efficiently but the resulting AP site can be replicated over by translesion synthesis (TLS) polymerases, leading generally either to transitions or to transversions (centre), depending upon the polymerase involved. Deamination of bases on opposing strands could lead to DSBs, as occurs during AID-mediated CSR (right). Figure based on models presented in [5, 20-23].

Figure 2. Potential mechanisms exposing ssDNA to A3 activity

Four potential mechanisms are shown by which ssDNA could become exposed to A3 activity, resulting in mutagenesis, shown here when error-prone translesion synthesis (TLS) polymerases mediate replication over the abasic sites generated by UNG. Any or all of these mechanisms may account for both single mutations and for kataegis, represented here on rainfall plots (center) of intermutation distance

against the genomic locations of mutations for a breast cancer (BRCA_PD4107A), originally featured in the seminal kataegis paper of Nik-Zainal *et al.* [32]. The plots show a large multi Mb region (lower panel) of kataegis (over a region of chromothripsis). Zooming in (upper panel) we can see evidence of strand-switching: multi-kb showers of cytosine mutations on one strand (red dots) and then the other (blue dots). (1) BIR induced by a one-ended DSB occurring as a result of replication fork collapse. This mode of replication is conservative; all newly synthesized DNA is incorporated into the recipient chromosome (R), using the donor (D) as the template. Note that ssDNA is exposed on the recipient strand both during initial resection and during asynchronous synthesis as the replication bubble migrates often very large distances along the chromosome. (2) Bidirectional 5'-3' resection during repair of double sided breaks creates ssDNA on opposing strands either side of the DSB. This could explain the strand-switching observed both in yeast models of kataegis [31] and in tumor samples such the example shown here. (3) Stretches of ssDNA can result from problems during replication, including polymerase blockage at replication forks as shown here. (4) R-loops are structures in which GC-rich coding strand DNA hybridizes with nascent mRNA, exposing the non-coding strand to deamination by AID. Figure based on models presented in [87] and references therein.

Figure 3. Model depicting the possible causes and consequences of A3 mediated mutagenesis during tumor development. Based on the current literature we propose a speculative model in which multiple factors combine to cause off-target A3 activity and that this in turn drives further mutation and genomic instability,

fuelling tumor development. Note that *TP53* mutations, genomic instability and DSBs have been proposed as both causes and consequences of A3 activity [35, 45].

Abbreviation: TSG, tumor suppressor gene.

Box 1 Figure I. Evolution of the A3 gene family and distribution of the A3A_B polymorphism.

(A) The A3 family has evolved solely within the mammalia, probably from a duplication of the ancestral AID gene. There are 3 ancient clades (A3Z1-A3Z3) that have fused and recombined to make the seven human A3 genes (A, B, C, D, F, G, H). Figure reproduced from [80] with permission from Dr Ignacio Bravo.

(B) A deletion polymorphism produces a hybrid, A3A_B, in which the A3A CDS is fused to the A3B 3'UTR. The polymorphism is very rare in African populations but expanding to near fixation in Oceanian populations (arrows represent human migration out of Africa over the last approx. 100,000 years). Figure adapted from [81] with permission from Dr Evan Eichler.

Figure 1
[Click here to download high resolution image](#)

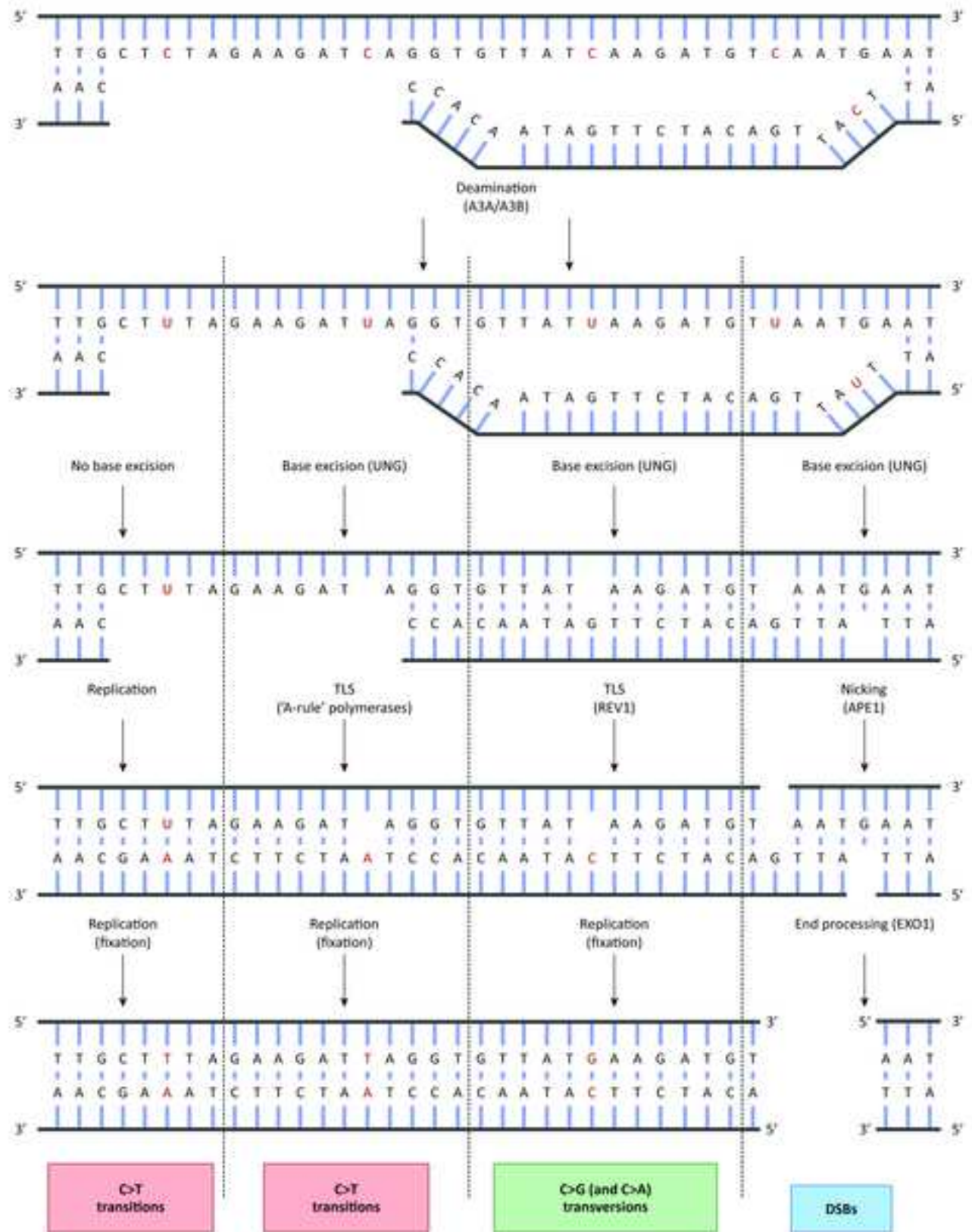


Figure 2
[Click here to download high resolution image](#)

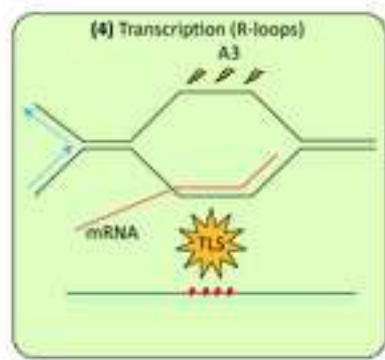
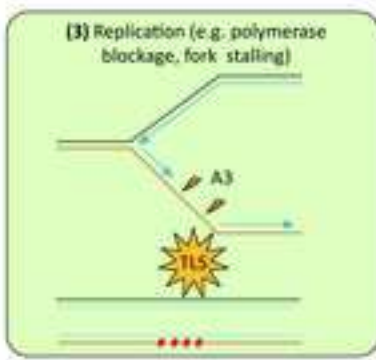
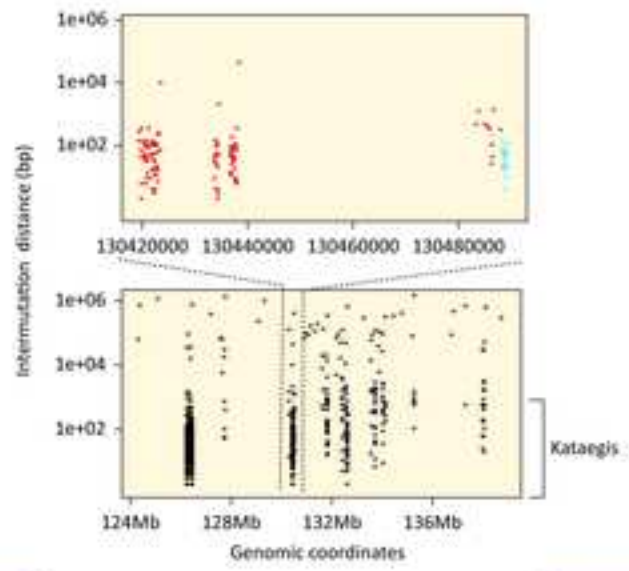
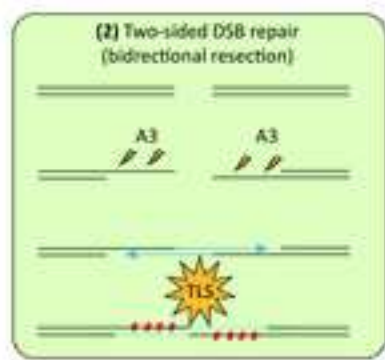
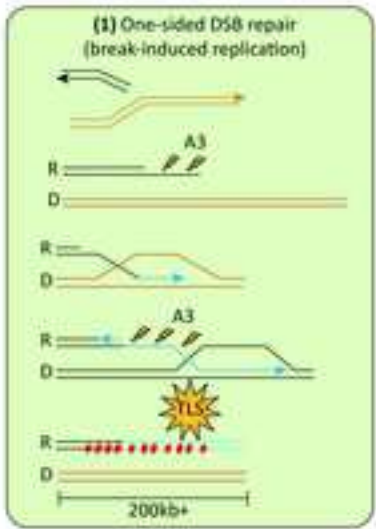


Figure 3
[Click here to download high resolution image](#)

