APOBEC3 subcellular localization and genomic editing

A DISSERTATION
SUBMITTED TO THE FACULTY OF
UNIVERSITY OF MINNESOTA
BY

Lela Lynn Lackey

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Advisor, Reuben S. Harris

October 2012

Acknowledgements

My husband, Hugh Heldenbrand, has been patient and supportive as I've worked on my thesis. I could not have done it without him. My son, Harold Ronald Heldenbrand, has give me motivation every day and made my life brighter over the last few years.

My family has supported me throughout graduate school. I would like to thank my dad and mom, Donald Harold Lackey, Jr. and Mary Edythe Lackey, and my brothers Hershel Hayes Lackey, Elijah Austin Lackey and Joseph Harold Lackey.

I would like to thank my advisor, Reuben Harris. His support and input allowed me to pursue the work I was most excited about and he has worked hard to create a challenging environment full of people committed to scientific research.

I could not have finished my thesis without the advice and assistance of the talented members of the Harris Lab. Specifically, members of the lethal mutators club – Michael Burns, Allison Land, Anurag Rathore, Brandon Leonard and Michael Carpenter. I would also like to thank the lab members with whom I have worked including – Emily Law, Judd Hultquist, Elizabeth Luengas, Zach Demorest, Donna MacDuff and Bill Brown.

I would like to thank the faculty members at the University of Minnesota who have helped me during graduate school. My committee, Howard Towle, Brian Van Ness, Perry Hackett and David Zarkower, have given me advice and support over the last few years. I would also like to thank Sean Conner and Hiroshi Matsuo, who opened their labs for me as a rotation student. I would like to thank Alex Lange, who heads the biochemistry lab course, for providing me with teaching experience.

I would like to thank the attendees of Friday Lab meeting and IMV journal club for their discussion and advice.

I would like to acknowledge my funding as a graduate student, most recently from the Institute of Molecular Virology training fellowship as well as from the National Science Foundation graduate student fellowship and 3M Science and Technology graduate fellowship.

I would like to thank the people of Venda. I learned so many things while living with them, including persevering through difficulties and the importance of education. Ro livhuwa vhatu vha Venda. Ndi pfi Doctor Tshifhiwa Nyadzanga!

I worked in several labs as an undergraduate and I would like to thank my advisors, Stu Maxwell and Rebecca Alexander. I first learned about science while working in their labs and Stu Maxwell has continued to provide me with valuable advice during graduate school.

Dedication

This thesis is dedicated to my mother - Mary Edythe Hayes Lackey.

Abstract

The APOBEC proteins are DNA cytosine deaminases with roles in immunity, including retroviral restriction and antibody maturation. Their activity theoretically makes them a danger to genomic DNA. The subfamily of *APOBEC3* genes has expanded to included seven different genes in primates. Based on their subcellular localization, only a subset of these APOBEC3 proteins have access to genomic DNA, and may potentially deaminate genomic DNA. Although the nuclear envelope breaks down during mitosis, I demonstrate that none of the APOBEC3s gain access to genomic DNA during cell division. However, APOBEC3B and other APOBEC3 proteins have access to genomic DNA during interphase. I also show that APOBEC3B is actively imported into the nuclear compartment. In general, APOBEC3 nuclear localization and deaminase activity correlate with ability to affect cell cycle progression, implicating these APOBEC3s in deamination of genomic DNA. In support of these conclusions, I observed cell death, activation of the DNA damage response and DNA mutations after ectopic expression of APOBEC3A and APOBEC3B. Moreover, endogenous APOBEC3B is demonstrably nuclear and active in breast cancer cell lines where it causes genomic deamination and mutations. Endogenous *APOBEC3B* is highly expressed in more than half of human breast cancers compared to normal breast tissues. In addition, sequences from tumors with higher levels of *APOBEC3B* have more mutations, and these mutations match APOBEC3B's deamination signature. My thesis work further defines the subcellular localization of the APOBEC3 family and provides the first evidence that APOBEC3B is involved in a human cancer type.

Table of Contents

List of Tables

List of Figures

Abbreviations

A, adenosine; AAV, Adeno-associated virus; ACF, APOBEC1 complementation factor; ADAT2, adenosine deaminase tRNA-specific; AID, activation induced cytosine deaminase; *apoB* mRNA, *apolipoprotein B* mRNA; APOBEC, apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like; A3, APOBEC3; A3A-H, APOBEC3A-H; C, cytosine; CDA, cytidine deaminase; CTD, C-terminal domain; CTNNBL1, catenin B-like 1; DSB, double-strand breaks; DNA, deoxyribonucleic acid; eGFP, enhanced green fluorescent protein; D, aspartic acid; G, guanosine; GANP, germinal center-associated nuclear protein; H2B, histone 2B; HERV-K, human endogenous retrovirus K; HBV, hepatitis B virus; HIV-1, human immunodeficiency virus, type 1; HTLV-1, Human T-cell leukemia virus type-1; I, inosine; kDa, kilodalton; lepB, leptomycin B; L1, LINE1, long interspersed elements; mRNA, messenger RNA; NHEJ, non-homologous end-joining; NTD, N-terminal domain; NES, nuclear export signal; NLS, nuclear localization signal; mRNA, messenger RNA; RNA, ribonucleic acid; SPT5, suppressor of Ty5 homolog; T, thymine; TadA, tRNA-specific adenosine deaminase; TRIB3, human homolog of *Drosophila* Tribbles 3; V, val, valine; Vif, viral infectivity factor

Chapter 1

Introduction

The zinc-dependent deaminases

Genetic information provides the necessary data for life and is encoded by all organisms in the form of deoxyribonucleic acid (DNA). However, information is conveyed in the form of ribonucleic acid (RNA). DNA and RNA are both composed of four nucleotides: adenosine, cytidine, guanosine and thymidine in DNA and adenosine, cytidine, guanosine and uridine in RNA. In order to grow and reproduce, organisms must have sufficient nucleotide pools to make RNA and replicate their DNA. The zinc-dependent cytidine deaminases (CDAs) control nucleotide pools by deaminating free cytidine to create uracil, the basic scaffold for further uridine and thymidine nucleotide synthesis as well as a substrate for degradation[1-3] (**Fig. 1-1A and B**). CDAs are present across a wide variety of species, consistent with the broad requirement for DNA and RNA turnover in all organisms[2,4].

However, organisms have evolved additional related deaminases that deaminate cytosine or adenosine within a polynucleotide context, altering the information encoded by these nucleic acids. Some of the earliest of these enzymes belong to the TadA/ADAT2 family and catalyze the deamination of adenosine within polynucleotide RNA[4,5] (**Fig. 1-1C**). These proteins target and deaminate adenosines with the codon recognition region of transfer RNAs, expanding the recognition potential of these molecules[4,6,7] (**Fig. 1-1D**). The *apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like* (*APOBEC)* genes are thought to be derived from *CDA* and *TadA*/*ADAT2* genes[2,4,5]. However, unlike CDAs and TadA/ADAT2 enzymes, APOBEC proteins do not deaminate free cytidine or adenosine in a polynucleotide context. Instead, in a combination of these functions, the APOBECs deaminate cytosine only within a polynucleotide (DNA or RNA) context

resulting in a change in informational content[8-10]. The rest of this introduction will focus on the APOBEC family members.

AID: Founder of the APOBEC family and adaptive immunity

The original *APOBEC* gene, a*ctivation induced cytosine deaminase* (*AID*), can be traced to the ancestor of humans and cartilaginous fish nearly 550 million years ago, corresponding with an expanded antigen response[11-16] (**Fig 1-2**). In all species, AID's deaminase activity helps solve the problem of a limited number of immune recognition molecules in a world of nearly unlimited immunogens[12,15]. AID allows cells to change their antibody locus, which combined with clonal expansion, selects for highly specific antigen-binding immune molecules[12,15,17].

Studies of human AID have shown that it targets both the variable and constant regions of the antibody locus for deamination, leading to somatic hypermutation and class switch recombination[18-20]. In somatic hypermutation, diversity at the variable region is generated by unrepaired or improperly repaired AID-induced uracils, resulting in significant C-to-T transition mutations, as well as other repair-mediated transitions and transversions[17,21] **(Fig. 1-3A)**. In class switch recombination, uracils are converted to nicks and double-strand breaks, recombining through non-homologous end-joining (NHEJ) to merge a new constant region to the rest of the antibody gene[17,21] **(Fig. 1-3B)**. Individuals with defects in AID have severe immune disorders and do not undergo antibody maturation[20,22,23].

While AID is clearly important for antibody maturation, its ability to deaminate genomic DNA may also lead to point mutations and translocations that contribute to cancer[24-26]. AID's oncogenic activity is discussed further in Chapters 3-5.

AID is regulated at both the mRNA and protein levels

To prevent undesired genomic deamination, *AID* expression is carefully regulated. As expected for an adaptive immune gene, *AID* is primarily expressed in B-cells and within the spleen[19,27,28]. Within B-cell lineages, *AID* mRNA is present only at low levels in immature B-cells, but induced through NFκB signaling in activated B-cells[29,30]. Post-transcriptionally, the stability of *AID* mRNA is regulated by microRNA-155 and microRNA-181[31-34].

AID is also regulated post-translationally through protein-protein interactions[35]. For example, it interacts with SPT5 (suppressor of Ty5 homolog), a protein associated with stalled RNA polymerase II, to guide it to the antibody locus and to off-target genes prone to deamination[36-39]. AID is also regulated by subcellular localization[40] (see Chapter 2). It contains a non-canonical active nuclear localization signal and enters the nucleus in a process that may be mediated by two additional interacting proteins – CTNNBL1 and GANP[40,41]. Interestingly, AID also has a nuclear export signal that functions through the CRM1 pathway and causes it to be cytoplasmic[42-44] **(Fig 1-3C)**. The functional importance of AID's ability to shuttle between the cytoplasm and nucleus is unclear, but must be balanced between its role in antibody maturation and undesirable genomic deamination[45].

APOBEC2 is conserved, but has unknown function

APOBEC2 has been identified in bony fish, suggesting that *APOBEC2* arose after *AID* but is still an early and conserved *APOBEC* gene **(Fig. 1-2)**. Unlike AID, APOBEC2 does not appear to have DNA cytosine deaminase activity[10,46]. It is, however, conserved and under purifying selection and appears to be involved in muscle biology, specifically in heart development, implying that is performs an important biological function[47-52]. Presently, the most important contribution of APOBEC2 to our understanding of the APOBEC family is the fact that it is relatively soluble and was the first APOBEC protein available for structural and oligomeric studies[53,54].

APOBEC1 is a RNA and DNA editor

APOBEC1 arose in the common ancestor of reptiles and mammals around 300 million years ago[55,56] **(Fig. 1-2)**. Although thought to have originated from *AID* and appearing after *APOBEC2*, *APOBEC1* was the first APOBEC to be scientifically discovered. It was identified as the editor of *apolipoprotein B* (*apoB*) mRNA, where it specifically targets cytosine 6666 to uracil, producing a premature stop codon and resulting in a different protein product with an impact on lipid metabolism[8,9,57,58] **(Fig 1-4)**. APOBEC1 is targeted specifically to this residue in this mRNA through an interacting protein, ACF (APOBEC1 complementation factor), which contacts an anchoring sequence on the mRNA[59-61]. APOBEC1 is also responsible for additional RNA editing events, although the majority of these are non-coding[62,63]. Mice null for APOBEC1 are

missing *apoB* mRNA and other mRNA editing events and have altered subtly lipid metabolism[63-65].

Interestingly, APOBEC1 is a strong, single-strand, DNA cytosine deaminase[10,66]. In fact, reptilian APOBEC1 does not appear to have a role in editing mRNA and entirely lack *apoB* mRNA, implying that RNA-editing activity may have been a later acquisition and DNA-editing is the more ancestral trait[55]. Several mammalian APOBEC1s, but not human APOBEC1, can deaminate and restrict some retroelements, including both retrotransposons and retroviruses[67-69]. APOBEC1 is a versatile enzyme with roles in both RNA and DNA deamination.

APOBEC1 regulation

Like *AID,* expression of *APOBEC1* is controlled transcriptionally. Human *APOBEC1* mRNA is primarily expressed in the small intestine where the majority of *apoB* mRNA editing occurs[8,70,71]. In rodents, *APOBEC1* is expressed in both the small intestine and the liver, where three distinct promoters control developmental, hormonal and nutritional expression levels[72-74]. Protein interactions govern the targeting of APOBEC1, as it requires ACF to recognize and bind *apoB* mRNA[59,60]. Similar to AID, the subcellular localization of APOBEC1 is also regulated. APOBEC1 functions in the nucleus, where it edits spliced *apoB* mRNA transcripts[74,75], but APOBEC1 is actively exported out of the nucleus and into the cytoplasm[76-78]. APOBEC1 is actively imported into the nuclear compartment, but it is not clear whether APOBEC1 or ACF controls localization[61,78]. The role of ACF or other APOBEC1-interactors in DNA deamination and retroelement restriction is unclear.

APOBEC3s restrict lentiviruses

The dramatic expansion of *APOBEC3* genes began in early placental mammals nearly 100 million years ago with the duplication of an ancestral *AID* gene, resulting in primates with seven distinct *APOBEC3s*[2,5,56,79,80] **(Fig. 1-2)**. While APOBEC1, and to a minor extent AID, have been implicated in retroelement restriction[67,68,81,82], evidence suggests that the APOBEC3 proteins truly excel in this arena[83,84]. For example, human APOBEC3F (A3F) and APOBEC3G (A3G) can inhibit human immunodeficiency virus type-1 (HIV-1) infection by targeting the HIV-1 genome for cytosine deamination during reverse transcription[85-89] **(Fig. 1-5A)**. Deaminase-independent inhibitory mechanisms may play a role in A3G and A3F restriction of HIV-1, although these effects are relatively minor in comparison to the level of restriction by the fully active enzymes[83,90,91]. HIV-1 encodes the viral infectivity factor (Vif), which targets many of the APOBEC3 proteins, including A3F and A3G, for proteasome degradation[86-89,92-95]. Other human APOBEC3 proteins have been implicated in HIV-1 restriction, including APOBEC3D (A3D) and APOBEC3H (A3H), but there is less consensus on their role in human infections[83,96,97]. APOBEC3 proteins in other species also inhibit the replication of their respective lentiviruses, with some cross inhibition between species[98] (*e.g.*, human A3G and cat A3Z2-Z3 both restrict feline leukemia virus[99,100]).

APOBEC3s affect other viruses including HBV, HTLV-1 and AAV

Human APOBEC3s restrict additional DNA-based viruses outside of the lentiviral family. Human T-cell leukemia virus type-1 (HTLV-1) is a retrovirus with a lifecycle similar to HIV-1, where its RNA genome must undergo reverse transcription into DNA for subsequent production of viral mRNAs and progeny RNA genomes. This similarity to HIV-1 makes restriction by APOBEC3 proteins unsurprising. APOBEC3A (A3A), A3B and A3G (and potentially other APOBEC proteins) have been shown to restrict HTLV-1 in cell culture infectivity assays and APOBEC3 mutational signatures have been identified in patient HTLV-1 isolates[101-103]. More surprising is the restriction of hepatitis B virus (HBV) by the APOBEC3 proteins. HBV and other hepadnaviruses have a partially double-stranded DNA genome that replicates through an RNA intermediate. This dependency on reverse transcription, as well as reports of hypermutated HBV isolates, led to experiments showing that A3B, APOBEC3C (A3C), A3F and A3G can inhibit HBV through both deaminase dependent and independent mechanisms[104,105]. Adeno-associated virus (AAV) belongs to the parvovirus family, which does not pass through a RNA intermediate. Instead AAV has a single-stranded DNA genome making it a potential substrate for single-stranded DNA deaminating enzymes. A3A has been shown to inhibit the replication of AAV, although no other APOBEC3 proteins appear to have an effect on AAV replication[106,107]. Various members of the APOBEC3 family have been reported to restrict other viruses as well, including human papillomavirus, Epstein-Barr virus, respiratory syncytial virus and others[84,108].

APOBEC3s have a role in retrotranposon inhibition

Retrotransposons are essentially retroviruses that are integrated into the host genome. Rather than replicate and infect other cells, they re-infect the host cell. Proportionally, the human genome is composed of nearly 50% active or degraded retrotransposons[109]. Retrotransposons produce their genomes as RNA and undergo reverse transcription, providing an opportunity for APOBEC3 restriction **(Fig. 1-5A)**. There have been many reports demonstrating that the majority of APOBEC3s can inhibit endogenous retroviruses, including human endogenous retrovirus K (HERV-K), murine IAP and yeast Ty1, as well as the non-LTR retrotransposons, including LINE and SINE elements[67,106,110-125].

The conclusion that the APOBEC3s inhibit retrotransposition comes with the caveat that studying such repression in mammalian cell lines requires transient and abnormal expression of retrotransposon genomes because retrotransposition is a stochastic and slow process (an estimated 1-2 transpositions every generation)[108,126,127]. Several APOBEC3 proteins are able to restrict foreign DNA that enters the cell, including the plasmids used for the majority of these retrotransposition assays[128]. Nevertheless, simple plasmid restriction cannot explain A3F's ability to inhibit LINE1 retrotransposition in a deaminase independent manner[114,128]. Furthermore, the positive selection of the APOBEC family is evident long before the arrival of lentiviruses, implying a formative role by an unknown retroelement[47,79,129,130]. Thus, a role for the APOBEC3 proteins in retrotransposon restriction is likely, although it is not presently well-defined[84].

The mechanism that targets the APOBEC3 proteins to viral or retrotransposon sequences for DNA deamination is not known. APOBEC3s could theoretically deaminate genomic DNA just as easily, although prior to this thesis work only A3A had demonstrated this ability, and only with ectopic expression[131-133]. AID is different story and is discussed in detail in Chapter 5. The ability of other APOBEC3 proteins to deaminate genomic DNA and the link between deamination and cancer is discussed further in Chapters 3 and 4.

Transcriptional regulation of the *APOBEC3* genes

In humans, mRNA expression profiles of the *APOBEC* genes vary by gene and tissue. For example, *A3G* and *A3F* mRNAs are present in a broad variety of tissues, but are evident in $CD4^+$ T cells and the thymus and spleen, consistent with their role in HIV-1 restriction, while A3H is poorly detected in most tissues with the exception of the lung and thymus[134,135]. In addition to tissue specific regulation the APOBEC3s can be induced by a variety of cellular signals, such as interferon. For instance, expression of *A3A* changes dramatically in myleloid cells, where mRNA levels change by more than 1000-fold in a matter of hours[97,128,136]. The expression of several *APOBEC3* genes cannot be clearly linked to a particular tissue type, including *A3B* and *A3D*[134]. However, as demonstrated by *A3A*, the *APOBEC3* genes have the capacity for sudden and specific induction, making it possible that we have not looked yet at the exact cell type under the appropriate circumstances. Even with these limitations, tissue specific expression can be very helpful. For example, one of our first clues that *A3B* might be involved in breast cancer genomic mutations was its high expression in breast cancer lines[71] (Chapter 4).

Interestingly, several studies suggest high levels of *APOBEC3* expression in the germline tissues, the ovaries and testis, suggesting a role for APOBEC3 there[115,137-139].

APOBEC3 protein interactors

The targeting of AID and APOBEC1 to specific substrates is affected by protein interactors[36-39,59,60]. In contrast, no proteins important for APOBEC3 function have been identified. Instead, identified interactors include a plethora of RNA-binding proteins, including hnRNP proteins[140-147]. For example, hnRNP K interacts with A3B and may have a general effect on promoter expression[140]. In addition, A3G and A3F reproducibly interact with MOV10, AGO2 and many others, which may indicate an unknown role for the APOBEC3s in RNA biology or serve as a means to sequester APOBEC3 activity[141-147]. Thus far, none of these protein interactors have been shown to impact APOBEC3 deaminase activity or yielded a clear functional result. However, the well-known ACF (an APOBEC1-interactor) is an hnRNP R protein, lending hope that these candidate proteins are important interactors in the proper context[74]. There are also examples of APOBEC3 interactors outside of the RNA-binding proteins. Recently, TRIB3 (human homolog of *Drosophila* Tribbles 3) was identified as an A3A interactor, and it may have a role in protecting the genome from A3A-mediated deamination and subsequent DNA damage[148].

Subcellular localization of the APOBEC proteins

Unlike AID and APOBEC1, none of the APOBEC3 proteins have demonstrated shuttling activity[149,150]. Instead, the APOBEC3 proteins can be divided into cell-wide,

nuclear and cytoplasmic categories. The cell-wide proteins include the smaller deaminases A3A and A3C[96,106,111,115,119,120,151-153] **(Fig. 1-5B)**. Also technically in this group is A3H, however, to complicate this matter there are at least seven human isoforms of A3H[111,119,122,139,154]. A3H haplotype II is one of the more stable and active isoforms of A3H and is present in the cytoplasm and the nucleoli[96,122,139,154] **(Fig. 1-5B, arrow)**. A3B is the only prominently nuclear APOBEC3 protein, an observation important for understanding its role in genomic deamination[96,111,115,117,119,150,151,153,155,156] **(Fig. 1-5B, arrow)**. The other APOBEC3 proteins (A3D, A3G and A3F) are cytoplasmic and, interestingly, the APOBEC family members favored as most important for physiological HIV-1 restriction[87,96,97,111,115,117,119,120,143,144,149-151,153,155,157] **(Fig. 1-5B)**.

This thesis is predicated on the hypothesis that only deaminases with access to genomic DNA would have the potential to cause mutations in genomic DNA. Here I describe the nuclear localization and active nuclear import activity of A3B (Chapter 2) as well as the ability of the APOBEC3 proteins to access genomic DNA during nuclear envelope breakdown during cell division (Chapter 3). I also present data correlating the localization of the APOBEC3 proteins with their ability to deaminate genomic DNA (Chapter 3). I then describe how A3B is capable of deaminating and mutating the human genome in breast cancer (Chapter 4). Finally, I relate this information back to its broader significance for our understanding of APOBEC3 activity and regulation as well as broader human health relating in genomic mutations and cancer (Chapter 5).

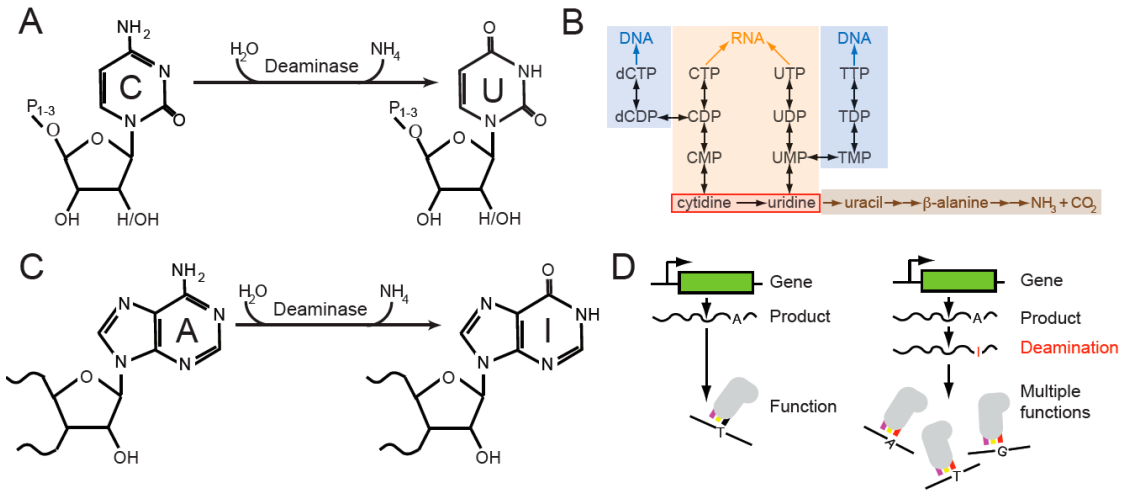

Figure 1-1. Reaction and functions catalyzed by the cytidine deaminase and TadA/ADAT2 families. **(A)** Cytidine deaminases (CDAs) catalyze the deamination of free cytidine (C) in RNA or DNA to produce uridine (U). **(B)** Selected aspects of cytidine and uridine in human biology. Uridine is an important precursor for DNA and RNA pyrimidines as well as a substrate for degradation (based on information from[3]). **(C)** TadA/ADAT2 enzymes catalyze the deamination of adenosine (A) within RNA into inosine (I). **(D)** Deamination of adenosine changes the base recognition of tRNA molecules. In addition to thymidine (T), inosine also recognizes guanosine (G) and adenosine (A), expanding the specificity of this tRNA to multiple codons.

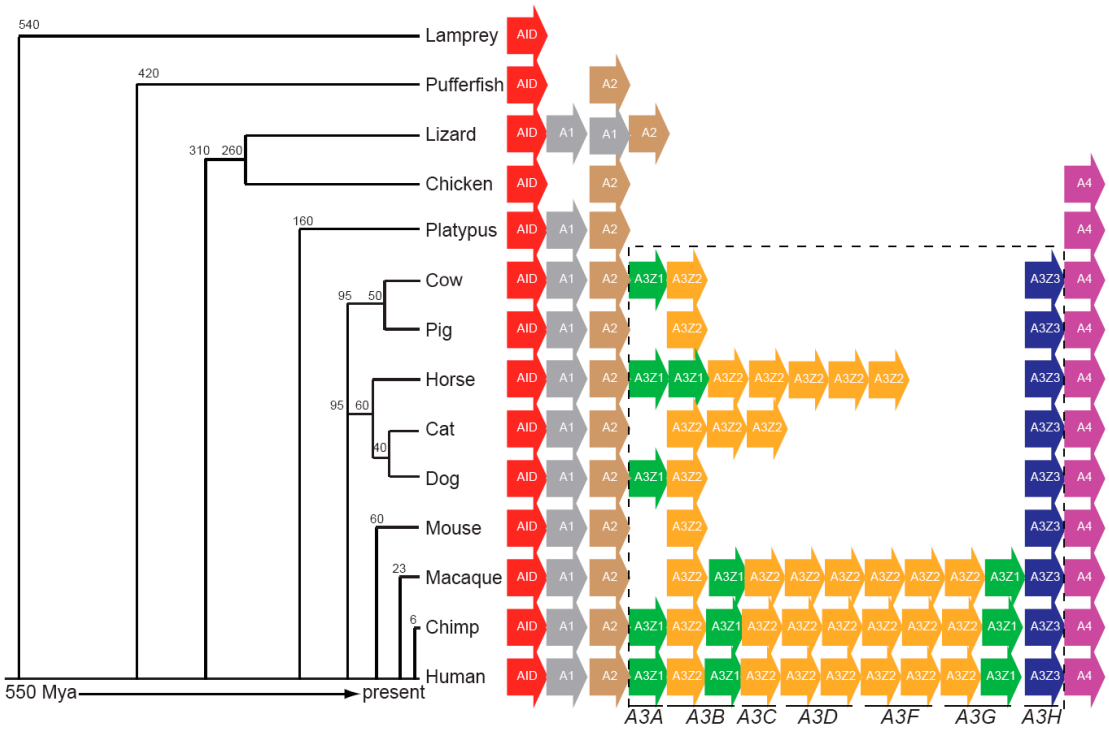

Figure 1-2. Evolutionary tree with the APOBEC repertoire indicated for each organism. *AID* has been sequenced from the lamprey, implying that *AID* existed in common ancestor of early vertebrates around 540 million years ago. *APOBEC2* can be identified in pufferfish and other bony fish that arose 420 million years ago, as well as all species to date. *APOBEC1* has been sequenced in lizards and birds and mammals, although not in chickens. *APOBEC4* appears in the same general time-frame and can be sequenced from chickens and mammals. The *APOBEC3* genes are specific to placental mammals and have undergone complex expansions and deletions in different species. Dates for species divergence are based on[16,56,158] and sequences for the *APOBEC* genes are based on[2,4,5,11-15,55,79,80,99,159-161].

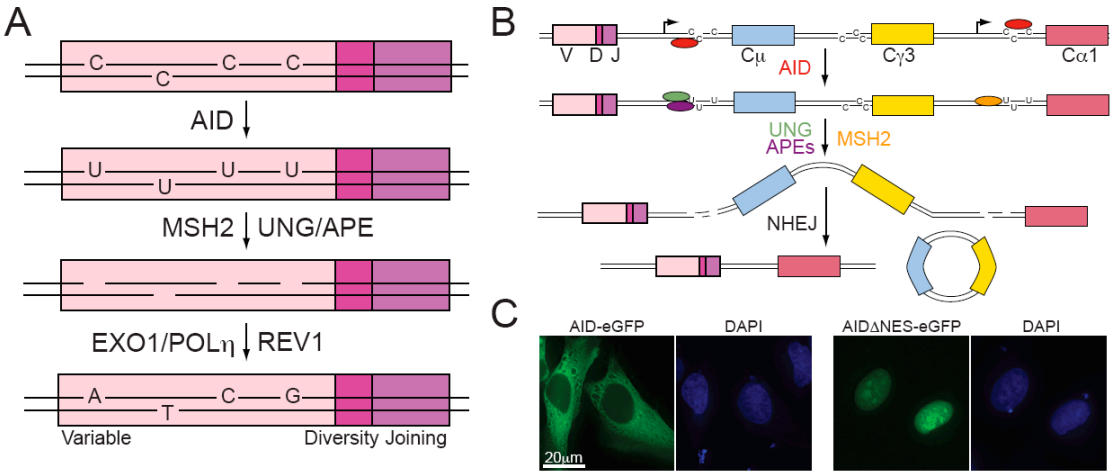

Figure 1-3. Activity and localization of AID. (A) AID deaminates cytosine in the variable region, creating uracils within DNA. U:G mismatches are recognized by mismatch and base excision repair proteins (MSH2 and UNG are shown). In combination with other repair enzymes (like EXO1, APE1 and APE2) as well as error-prone or translesion polymerases (such as POLη and REV1), the original cytosines and the surrounding bases can be mutated to create diversity in the variable region. Additional details are available in Di Noia, *et al.*[17]. **(B)** AID deaminates cytosines within the switch (S) regions upstream of joint constant (C) regions in a transcription dependent manner. The high proportion of AID motifs within the S regions produces uracils on both strands in close proximity. Repair of these uracils leads to double-strand DNA breaks (DSBs). These DSBs can be repaired through classical or alternative NHEJ, bringing a new constant region close to the V(D)J region and resulting in an antibody with different attributes. Additional details are available in Xu, *et al.*[162]. In this simplified representation the antibody locus rearranges to replace the constant region encoding IgM to the region

encoding IgA. **(C)** The C-terminus of AID contains a nuclear export signal and is essential for nuclear export and class switch recombination. Representative images of HeLa cells expressing AID-eGFP or AIDΔNES-eGFP, which has a deletion removing the nuclear export signal, are shown with the nucleus indicated by Hoechst staining (DAPI). Similar nuclear accumulation is seen in AID expressing leptomycin B treated cells, implicating the CRM1 export pathway[40].

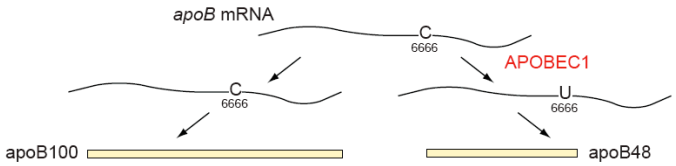

Figure 1-4. RNA-editing activity of APOBEC1. APOBEC1 deaminates cytosine 6666 in *apoB* mRNA for cytosine deamination. The resulting uracil creates a pre-mature stop codon, which translates to a shorter protein product (apoB48 compared to apoB100). These protein products have different functions in lipid metabolism. Additional details available in[163].

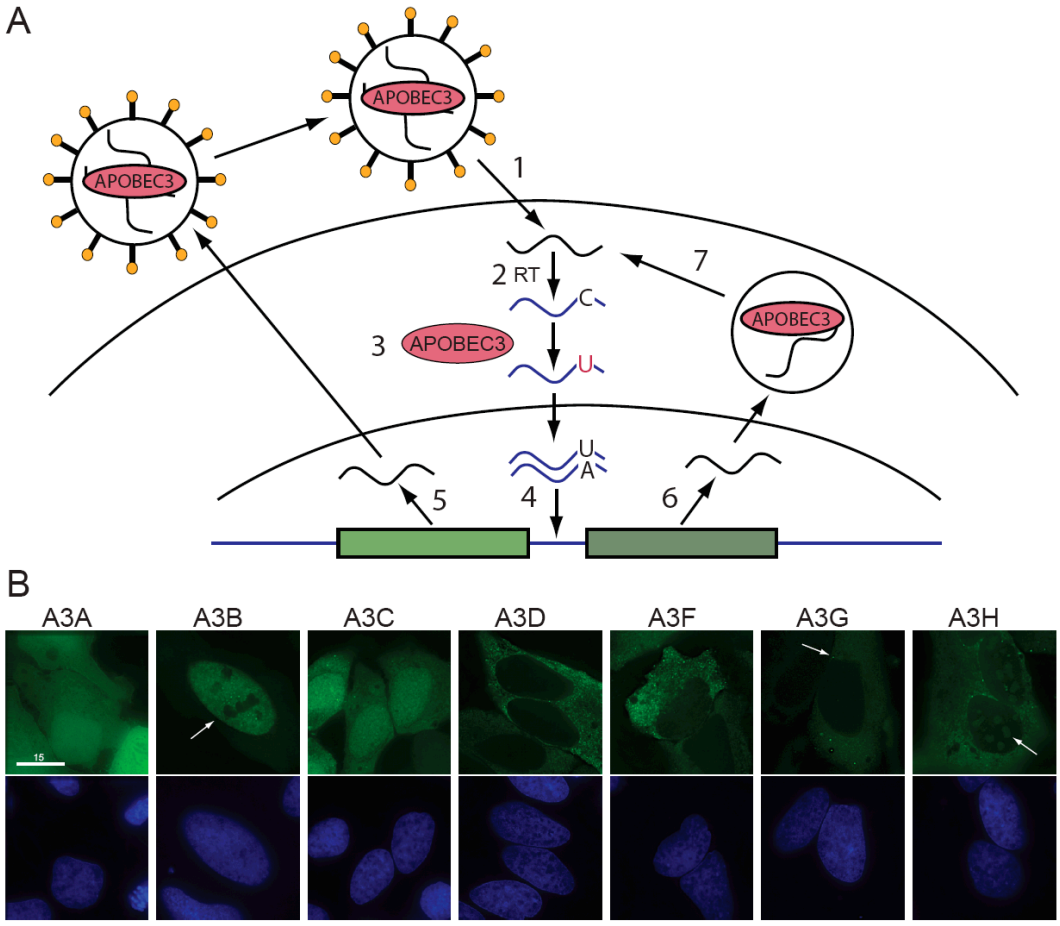

Figure 1-5. Activity and localization of the APOBEC3 proteins. (A) Retroelement replication showing an exogenous virus entering the cell (1) and using reverse transcriptase (RT) to convert its RNA genome into single-stranded DNA (2). This is a substrate for APOBEC3, which deaminates cytosine to uracil (3). These uracils can be replicated to create C/T to G/A mutations or can cause the degradation of the viral genome. Hypermutated genomes can be integrated (4) and may produce new viral particles and continue the infectious cycle (5) or may be inactive due to APOBEC3 mutagenesis. Endogenous retrotransposons replicate through an RNA intermediate (6),

but are thought to form viral-like particles in the cytoplasm (7) and proceed through similar RT and APOBEC3 susceptible stages (2-4). This schematic is based on retroelement restriction in MacDuff, *et al.*[164] **(B)** Images from HeLa cells expressing each of the seven members of the APOBEC3-eGFP proteins (unpublished data). These enzymes are cell-wide (A3A, A3C, A3H), nuclear (A3B, white arrow) or cytoplasmic (A3D, A3F, A3G). P-bodies are characteristic of A3F and A3G expression (A3G, white arrow). A3H haplotype II appears in the cytoplasmic and nucleolar compartments (A3H, white arrow). The nuclei are indicated below with Hoechst staining.

Chapter 2

APOBEC3B and AID have similar nuclear import mechanisms

This chapter is reprinted with permission from Lackey, L., Demorest, Z. L., Land, A. M., Hultquist, J. F., Brown, W. L. & Harris, R. S. (2012). *J Mol Biol* **419**, 301-14.

Foreword

While several of the smaller members of the APOBEC3 subfamily are cell-wide and have access to the nuclear compartment, only one protein is specifically nuclear by fluorescence microscopy – A3B[96,111,115,117,119,150,151,153,155,156]. Prior work reasoned that A3B must be actively imported because it approaches the limit for passive diffusion [115]. However, since A3B is so close to this limit (45.9 vs 50 kDa limit) and shape plays an important role in passage through the nuclear pore, we felt that this assumption should be tested[165]. In addition, prior to this study, AID had been shown to have active nuclear import[41]. We were intrigued by the idea that A3B might be actively imported in the same way as AID, since these proteins are evolutionarily related[2,4,5]. Here we demonstrate active nuclear import of A3B and differential functions of A3B and AID.

Contributions

I imaged and analyzed AID and A3B in HEK293T, HeLa and other cell lines. I performed the digitonin permeabilitization experiments and time-lapse microscopy that showed active nuclear import of A3B. ZLD and I performed the co-immunoprecipitation with adaptor importin proteins. ZLD identified the mislocalizing AID point mutant and I identified the A3B point mutant and imaged and analyzed both. JFH performed the single-cycle HIV-1 infection assay and western blots. I imaged the HIV-1 infected and APOBEC3 expressing cells. I performed the LINE1 inhibition assay. RSH contributed to experimental design and editing of the manuscript.

Summary

Members of the APOBEC protein family catalyze DNA cytosine deamination and underpin a variety of immune defenses. For instance, several family members, including APOBEC3B, elicit strong retrotransposon and retrovirus restriction activities. However, unlike the other proteins, APOBEC3B is the only family member with steady-state nuclear localization. Here we show that APOBEC3B nuclear import is an active process requiring at least one amino acid (Val54) within an N-terminal motif analogous to the nuclear localization determinant of the antibody gene diversification enzyme AID. Mechanistic conservation with AID is further suggested by APOBEC3B's capacity to interact with the same subset of importin proteins. Despite these mechanistic similarities, enforced APOBEC3B expression cannot substitute for AID-dependent antibody gene diversification by class switch recombination. Regulatory differences between APOBEC3B and AID are also visible during cell cycle progression. Our studies suggest that the present day APOBEC3B enzyme retained the nuclear import mechanism of an ancestral AID protein during the expansion of the *APOBEC3* locus in primates. Our studies also highlight the likelihood that, after nuclear import, specialized mechanisms exist to guide these enzymes to their respective physiological substrates and prevent gratuitous chromosomal DNA damage.

Introduction

Human cells can express up to eleven distinct polynucleotide cytosine deaminases (reviewed in[5,161]). AID, APOBEC1, A3A, A3B, A3C, A3D, A3F, A3G, and A3H have demonstrated DNA deaminase activity in one or more biochemical or biological assays (reviewed in[5]). Two other family members, APOBEC2 and APOBEC4, do not appear to have DNA editing activity[48,51,159]. Genomic cytosine to uracil deamination is potentially dangerous because uracil lesions can template the insertion of adenines and cause C/G-to-T/A transition mutations and, depending on DNA repair fidelity, other types of base substitutions, DNA breaks, and larger-scale chromosome aberrations (reviewed in[25,26,166]).

AID is arguably the most ancient family member because all vertebrates use it for adaptive immunity through antibody gene diversification (reviewed in[5,17,167]) (**Fig. 2-1A**). *APOBEC1* appears to have evolved more recently, most likely from the duplication of an ancestral *AID* gene prior to the split of species such as birds and lizards from the vertebrate phylogenetic tree[2,55] (**Fig. 2-1A**). In most species, the *AID* and *APOBEC1* genes are located adjacent to one another on the same chromosome. Current models posit that these genes provided the substrate to root the *APOBEC3* locus prior to the radiation of placental mammals, most likely by another ancestral gene duplication event[79,168] (**Fig. 2-1A**). Thus, despite this considerable evolutionary gap, their common origin suggests that some present day APOBEC3 protein activities and/or regulatory programs may still be conserved with those of AID (in addition to their shared DNA deaminase activity). Indeed, we and others have seen that localization through CRM1 mediated nuclear export is a conserved property of AID from a variety of different species[14,42,169-171] (**Fig. 2-1B**).

The physiological functions of AID in antibody diversification and APOBEC1 in *APOB* mRNA editing necessitate the transport of these enzymes to the nuclear compartment (reviewed in[5,163]). Despite <30 kDa protein sizes, which should allow for passive diffusion through nuclear pores, both enzymes are imported actively[41,77,172]. At least for AID, import is accomplished through a non-canonical nuclear localization signal (NLS) interacting with the adaptor importins α1, α3, and α5 in concert with the import factor β1[41]. AID import may also be affected by interactions with other proteins such as GANP and CTNNBL1 (reviewed in[35]). Similarly, APOBEC1 binds with at least one other cellular protein, ACF, to enter the nuclear compartment and access *APOB* mRNA substrates[8,9,78,173]. However, at steady state, both AID and APOBEC1 appear cytoplasmic due to a strong C-terminal leucine-rich nuclear export signal (NES)[42,44,169]. This NES directs export through the CRM1 pathway, which can be blocked by incubating cells with leptomycin B (lepB) (**Fig. 2-1B**)[174].

Curiously, A3B is the only human DNA deaminase family member with an apparent steady-state nuclear localization[111,115]. This property is also evident in rhesus macaque A3B[96]. Although human A3B has a putative positively charged NLS spanning residues 206-212, mutagenesis studies showed that these residues are dispensable[150]. However, individual domain as well as A3B/G chimera analyses were able to implicate an N-terminal region spanning residues 1-60 in nuclear import[115,150,156].

Here we find that A3B actively imports into the nucleus through a non-canonical N-terminal NLS within the same region as the non-canonical NLS of AID. Additionally, both A3B and AID interact with adaptor importin proteins. Despite previous work showing that these enzymes can both inhibit retrotransposon replication[111,114], other

aspects of the regulation and function of A3B and AID have clearly bifurcated. For instance, we show that A3B and AID localize differently during mitosis. In addition, over-expression of A3B in primary B-cells does not confer an ability to perform class switch recombination, and, unlike A3B, AID demonstrates little or no ability to restrict HIV-1 replication in single-cycle infectivity assays. Since A3B appears to have constitutive access to genomic DNA, and it is actively shuttled to the nuclear compartment, understanding additional levels of A3B regulation will be important for determining whether in certain circumstances it is able, like AID, to mutate genomic DNA and contribute to carcinogenesis.

Results

Relationship between A3B and AID

Consistent with a prior report[41], we noticed that AID-eGFP has an import delay in HEK293T cells compared to HeLa cells after lepB treatment (see 15 minutes time-point in **Fig. 2-1C**). Similarly, we noticed less nuclear A3B-eGFP in HEK293T in comparison to HeLa cells, an average of 60% vs 70% of measurable fluorescent signal, respectively (**Fig. 2-1D**). A3B-eGFP displayed a HeLa-like nuclear concentration in a wide variety of other cell types (*e.g.*, **Fig 2-1E** and additional data not shown). This activity was independent of the epitope tag and imaging procedure, as similar results were obtained for A3B-HA in fixed cells (**Fig. 2-S1**). These observations suggested that HEK293T cells have an import defect and provided correlative support for the hypothesis that A3B and AID share a common import mechanism.

A3B is actively imported into the nucleus of plasma membrane-permeabilized cells

The nuclear pore complex (NPC) allows objects smaller than 50 kDa to enter into the nucleus via passive diffusion, but passage is also influenced strongly by protein shape[175,176]. Native A3B is approximately 45.9 kDa, but the GFP tag brings the total mass to approximately 70 kDa. Since we see A3B-eGFP in the nuclear compartment, we would predict A3B to enter the nucleus through active import. However, occasionally proteins larger than 50 kDa are able to enter the nucleus through passive transport[177]. To address this possibility, we used a permeabilization assay in which the detergent digitonin permeabilizes the plasma membrane while leaving the nuclear membrane and pores intact[177,178]. Then, the original cytoplasm is washed away and the cellular import machinery can be added back with the addition of any cargos, such as a fluorescently tagged protein of interest.

Using this assay, we found that full-length A3B-eGFP is imported within 20-30 minutes into the nuclear compartment, as is the N-terminal half of the protein (A3B NTD-eGFP) (**Fig. 2-2A,** quantified in **2-2B**). In contrast, APOBEC3A-eGFP and eGFP, which are cell-wide[106,115], were only found in nuclei at very low levels (barely distinguishable from non-fluorescent HeLa lysate) consistent with a passive diffusion mechanism (**Fig. 2-2A and B**). As a positive control, fusing the SV40 NLS[178,179] to APOBEC3A-eGFP allowed the protein to concentrate in the nucleus (**Fig. 2-2A and B**). These data demonstrate an active nuclear import mechanism for A3B-eGFP.

A3B and AID localize differently during the cell cycle

Human cells undergo open mitosis with nuclear envelope breakdown during prophase and reformation after telophase. We hypothesized that both A3B-mCherry and AID-mCherry would become cell wide upon breakdown of the nuclear envelope and return to their respective subcellular locations after telophase. Cell-wide proteins with no regulatory signals, such as mCherry, are cell-wide throughout the cell cycle, but excluded from condensed chromosomes. In comparison, histone2B-eGFP (H2B-eGFP) elicits the expected strong but broad nuclear localization during G1 and a concentrated chromosomal distribution through mitosis (**Fig. 2-3A** and **Movies 1-4**).

Interestingly, during mitosis, although A3B-mCherry appears to shift as expected to a more cell-wide localization, it also seems to be excluded from metaphase chromosomes marked by H2B-eGFP. However, this localization phenotype only persists until ~10 minutes after telophase when A3B-mCherry rapidly (~20-30 minutes) returns to its characteristic nuclear localization (**Fig. 2-3B** and **Movies 5-8**). This rapid re-entry is further consistent with an active import mechanism. These data also demonstrate that A3B is not becoming trapped (and retained) with nuclear constituents as the nuclear membrane reforms after mitosis. AID-mCherry also shifts as expected to a more cell-wide localization at the early stages of mitosis. However, in contrast to A3B-mCherry, it appears to co-localize with H2B-eGFP labeled DNA during anaphase and telophase, and it remains concentrated there for ~20 minutes before rapidly shuttling to the cytoplasm (compare 24-32 minute images in **Fig. 2-3C** and **Movies 9-12**). These data support the conclusion that A3B is imported actively, and they also provide new visual evidence for differential regulation of A3B and AID during the cell cycle.

APOBEC3B and AID depend on the same predicted β2 region for nuclear import

Active nuclear import implies the existence of an NLS that directly or indirectly binds an importin. AID, APOBEC1, and A3B have putative classical localization signals that have been shown to be irrelevant for localization (although the putative NLS in AID is more controversial), implying that the localization determinant for each of these enzymes is non-classical or indirect[41,42,150,169,172,180]. Since the initial 60 residues are responsible for the nuclear localization of A3B[150], we mutated residues in this region to those of the related, but cytoplasmic, A3G protein. However, no single mutational swap affected the localization of A3B NTD-eGFP (**Fig. 2-S2**).

Our second strategy for identifying residues involved in A3B nuclear import was through comparison with AID. Amino acids 40-53 of AID have been implicated in nuclear import[41]. Focusing on the homologous region in A3B, we discovered that the single amino acid substitution mutant, valine (V) 54 to aspartic acid (D), caused full-length A3B-eGFP to become cell-wide (**Fig. 2-4A**). Mutation of A3B V54 to D also diminished the nuclear import of full-length A3B-eGFP in the digitonin import assay (**Fig. 2-S3**). The homologous residue in AID, tyrosine 48, when mutated to histidine, modestly decreased the efficiency of nuclear import (**Fig. 2-4B**). This region of AID was predicted to form a β-strand (β2) when modeled using the recent A3G catalytic domain structure (pdb 3IR2)[181-183] (**Fig. 2-4C**). Modeling of A3B using the same methodology predicted a similar β2 region on the external surface of the protein (**Fig. 2-4C**). The residue identified in A3B is consistent with previous reports on the importance of the N-terminal region for nuclear localization of A3B[115,150] and complements what we know

about the localization of AID. We conclude that the same structured region of AID and A3B, despite differences in primary amino acid sequence, is involved in nuclear import.

APOBEC3B can interact with the adaptor importins α1, α3 and α5

The classical pathway for nuclear import is driven by adaptor import proteins (reviewed in[175,184-186]). These proteins interact with the NLS as well as with karyopherin (importin β1) and the entire complex crosses through the nuclear pore using a GTP energy gradient. Because AID has been reported to interact with three of the adaptor importins, we tested these known interactors for interaction with A3B, with the hypothesis that A3B might have the same binding preferences. Indeed, pull-down experiments with GST-tagged adaptor importins indicate that A3B can also interact with importin α1, α3 and α5 (**Fig. 2-5**). As expected based on the localization defect described above, A3B V54D weakened these interactions (**Fig. 2-5**). As previously reported[41], wild-type AID bound these same importins, but we were surprised to find that AID Y48H bound more strongly (**Fig. 2-5**). It is possible that the differential impact of the single amino acid substitution mutations may relate to interaction strength, with an optimal intermediate affinity being required for import and higher or lower affinities compromising this process. Nevertheless, despite this incongruence, these results show that both AID and A3B can form complexes with adaptor importins α1, α3 and/or α5.

AID and A3B are functionally divergent for class switch recombination and HIV restriction

Since AID performs its predominant physiological function in the nucleus, we asked whether A3B could substitute for AID in promoting class switch recombination. Specifically, we tested whether AID-deficient mouse B-cells, which are unable to switch antibody type, could be supplemented with A3B and convert from IgM to IgG[18]. In this assay, only the positive control, catalytically active AID, was able to affect the switch from IgM to IgG (**Fig. 2-6A**). Catalytically inactive AID (T27E), wild-type A3B and catalytically inactive A3B (E255Q) were all unable to promote isotype switching (**Fig. 2-6A**). We conclude that A3B's ability to enter the nucleus, even via a similar pathway as AID, does not enable it to perform class switch recombination.

To do the reciprocal experiment, we asked whether AID has the capacity to restrict HIV-1. AID has been previously tested for the capacity to restrict HIV[187,188], but not in a dose-response study in which expression levels range. In comparison to A3B and A3G, which strongly restrict Vif-deficient HIV-1 in single-cycle infection assays, AID showed little activity despite having the capacity to package into viral particles (**Fig. 2-6B and C**). At low expression levels the catalytic glutamate of A3G (E259) was required, consistent with prior work[189] (reviewed in[190]), whereas the analogous residue in A3B (E255) was only partly required (consistent with the latter enzyme having two active sites[191]). We also found that AID was a poor LINE1 (L1) inhibitor in comparison to A3B (**Fig. 2-S4**). We conclude that AID does not possess a strong capacity to restrict Vif-deficient HIV-1 or L1, but note that this does not exclude it from possible roles in restricting the replication of other types of retroelement substrates, as suggested previously[81,164,192].

One prior report indicated that HIV-1 infection, and specifically expression of the

viral protein Vif, caused re-localization of several APOBEC3 proteins, including A3B[151]. In a reconstruction experiment, we infected A3B- or A3G-mCherry expressing cells with replication competent HIV_{LAI} (Nef replaced by eGFP). No changes in localization were observed for mCherry, A3G-mCherry, or A3B-mCherry in HIV-1 infected (GFP-positive) cells (**Fig. 2-6D** and data not shown). Thus, HIV-1 does not appear to grossly alter the localization phenotypes described here, and we cannot readily explain why our studies diverge from the prior report that tried to address this question.

Discussion

AID is a well-studied enzyme with important functions in the adaptive immune response. A3B has not been as extensively studied because it is specific to primates (there are no exact homologs of any specific APOBEC3 in mice[161]), and it is more difficult to work with (it is lethal to *E. coli*[117]). To address the hypothesis that A3B and AID are regulated in the same way, we confirmed that A3B is generally nuclear, and we extended present knowledge by showing that A3B is actively imported into the nucleus through a predicted β2 region and that it has the capacity to interact with importins α1, α3, and α5. Why A3B is targeted specifically and constitutively to the nucleus is an avenue of current research.

While human *AID* mutations result in severe immune deficiencies [20,22,23], *A3B* is deleted in certain global populations with no obvious phenotypes (*e.g.*, ~93% of Polynesians harbor homozygous *A3B* deletion alleles[193]). It is not surprising to find that A3B was unable to perform class switch recombination in AID deficient murine B-cells. However, this was somewhat disappointing because we could envisage that forced

expression of A3B in B-cells might have resulted in some lesions in the antibody locus that allow for class switch recombination. The inability of A3B to perform class switch recombination echoes the failure of APOBEC1 to perform class switch recombination in AID-deficient B-cells[194,195] and it is unlikely that any family member, apart from AID itself, will have this capability.

We also found evidence for differential regulation of A3B and AID during the cell cycle. While A3B does not associate with H2B-labeled DNA during mitosis (particularly in late telophase), AID is cell-wide during division and appears to concentrate transiently with H2B-labeled DNA shortly after telophase and before being re-exported. We propose that A3B and AID interact with different proteins during mitosis. Several interactors of AID[7,41,196-198] and A3B[140,199] have been reported and one or more these proteins may contribute to the new localization phenotypes described here (**Fig. 2-3** and **Movies 1-12)**. In conclusion, our studies have demonstrated several similarities between A3B and AID with respect to nuclear import, but they have also highlighted several differences in regulation and function. Understanding the regulation of A3B, including its subcellular localization, is important for knowing how the cell regulates A3B's cytosine deaminase activity - either for its benefit or detriment.

Materials and Methods

Constructs

Several constructs used in this paper have been reported previously including A3B-eGFP, N-terminal A3B-eGFP, AID-eGFP, zebrafish AID-eGFP, and MLV AID viral constructs [81,96,150,200]. The pmCherry vectors used for time-lapse microscopy and the HIV localization experiment are direct derivatives of these constructs. The eGFP-tagged histone2B construct was a gift from J. Mueller. The GST-tagged importin α1, α3 and α5 constructs were gifts from J. Di Noia lab[41]. The HIV_{LAI} nef::eGFP construct was a gift from M. Stevenson. The A3A NLS construct was made with an NLS encoding primer set that annealed to create a BsrGI site. This double-stranded oligo was inserted into the BsrGI site at the 3' end the GFP gene of A3A-GFP (Fwd 5'-GTACAAGGATCCAAAAAAGAAGAGAAAGGTAGATCCAAAAAAGAAGAGAA AGGTAGATCCAAAAAAGAAGAGAAAGGTACT-3' and Rev 5'-GTACAGTACCTTTCTCTTCTTTTTTGGATCTACCTTTCTCTTCTTTTTTGGATCT ACCTTTCTCTTCTTTTTTGGATCCTT-3'). Site-directed mutagenesis was used to create Y48H in AID and V54D in A3B. (A3B V54D Fwd 5'-TGGGACACAGGGGACTTTCGAGGCCAG-3' and Rev 5'-CTGGCCTCGAAAGTCCCCTGTGTCCCA-3') (AID Y48H Fwd 5'-CTGGACTTTGGTCATCTTCGCAATAAG-3' and Rev 5'-CTTATTGCGAAGATGACCAAAGTCCAG-3'). The A3B transducing virus was inserted into the AID MLV transducing virus backbone containing an IRES GFP and MLV long terminal repeats, but because A3B restricts production of its own virus, its coding sequence was disrupted with an inverted β-globin intron just before the second catalytic site (Fwd 5'- NNNNGTCGACATGAATCCACAGATCAGAA-3' and Rev 5'-CCTTCTTCTCTTTCCTACAGGACCCAGGTGCCATTGTCC -3', Fwd 5'-

CTGTAGGAAAGAGAAGAA-3', Rev 5'-GCCCATGTGCTGGTCCATCAGGTGAGTTTGGGGACCCTTG-3' and Fwd 5'-GGGCTTTCTATGCAACGAGGCTAAGAATCTTCTCTG -3', Rev 5'-NNNNACGCGTCTACTTGTACAGCTCGTCCAT-3'). The catalytic mutant was made by site-directed mutagenesis (Fwd 5'- GCCGCCATGCGCAGCTGCGCTTC-3', Rev 5'-GAAGCGCAGCTGCGCATGGCGGC-3').

Cell culture

All cells except the MCF10A cell-line were grown in DMEM (Invitrogen) supplemented with 10% fetal bovine serum (HyClone). MCF10A cells were grown in MEBM base media with 100 ng/mL cholera toxin and MEGM additives except for gentamicin (Lonza). HeLa, HEK293T and MCF10A cells were obtained from the ATCC. U2OS cells were a kind gift from J.Mueller and the JSQ3 and TR146 lines were a gift from M. Herzberg.

Live cell fluorescence microscopy experiments

Microscopy experiments were performed as described[150] with minor modifications. Approximately 20,000 cells of HeLa, HEK293T, U2OS, TR146, JSQ3 or MCF10A cells were plated on LabTek chambered cover glasses (Nunc), grown for 24 hours, transfected (Trans-IT LT1; Mirus) with 200 ng of DNA and incubated overnight. For live cell imaging, cells were incubated with clear DMEM containing 0.1% Hoechst dye to stain the nuclei and 10 mM HEPES, pH7.4 and kept at 37 degrees Celsius. Time lapse microscopy for cell cycle imaging was set up to take pictures at one-minute intervals

using sealed chambered cover glasses with normal growth media. For nuclear import analysis of AID, AID-eGFP expressing live cells were treated with lepB (40 ng/ml) and followed over time[200]. Quantification of cells for lepB was performed by fixing the cells at the indicated time-points in 4% paraformaldehyde in PBS for 20 minutes before treating with 0.1% Hoechst dye in PBS. A DeltaVision deconvolution microscope (Applied Precision) at 60x or 40x magnification was used to collect all the images, and deconvolution was performed using DeltaVision softWoRx software (Applied Precision).

Quantification of imaging

Images were quantified for comparison with CellProfiler (Broad Institute)[201,202]. To quantify the nuclear localization of A3B-eGFP in HEK293T and HeLa cells, A3B-eGFP and A3B V54D-eGFP in HeLa cells, and AID-eGFP and AID Y48H-eGFP in HeLa cells, pipelines were set-up in CellProfiler to identify the nuclear compartment (based on Hoescht staining) and the cell, based on the GFP fluorescence. The intensity was measured in the defined nuclear and cellular compartments and the nuclear intensity divided by the total intensity. Data for analysis was collected from CellProfiler and assembled and statistics calculated with Prism 5.0 (GraphPad Software, Inc). The student t-test was used to calculate p-values where indicated. To quantify the import of eGFP tagged lysate into digitonin permeabilized cells we used CellProfiler to identify the nuclear compartment (based on Hoescht staining) and measured the intensity of fluorescence in the nucleus for each construct. This direct measurement was graphed with Prism 5.0 (GraphPad Software, Inc) and the Mann-Whitney test used to calculate statistics for A3B-eGFP and A3B V54D-eGFP.

Digitonin permeabilization import assays

Digitonin permeabilization has been previously described[178] and modified as below. HeLa cells were plated onto 30 window slides at 50-150 cells/window and allowed to settle overnight (#63434-02; Electron Microscopy Sciences). The slides were incubated for five minutes on ice, washed five minutes in cold transport buffer (20 mM HEPES, pH 7.3, 110 mM potassium acetate, 2 mM magnesium acetate), and then permeabilized in 70 ug/mL digitonin (EMD Chemicals, Inc) in transport buffer for five minutes. The digitonin solution was removed by washing with cold transport buffer then the import mixture was applied individually to each window. To make the lysates ~2.5 million HeLa cells were transfected with 5 ug of DNA (Trans-IT LT1; Mirus) and allowed to express protein for 48 hours before harvesting in 100 uL of transport buffer with protease inhibitors (Complete Protease Inhibitor Cocktail Tablets; Roche). To check for protein expression in the lysates, samples were run on acrylimide gels (BioRad), then transferred to PVDF membranes. After blocking for 1 hour (3% casein in PBS with 0.1% Tween), mouse anti-GFP (JL-8, BD Clontech 632 381) was used at a 1:2000 dilution. The membranes were incubated overnight 4 degrees C. The membranes were washed four times with PBS Tween before incubating with 1:5000 goat anti-mouse HRP (BioRad) secondary for 1 hour at room temperature. The total import mixture contained cell lysate, 0.5 mM GTP and 1 mM ATP. Each set of lysates was done in triplicate. The import mixture incubated on the slides for 20 minutes at room temperature and was washed away with cold transport buffer. The cells were fixed for 20 minutes with 4% paraformaldehyde on ice, washed with PBS, incubated for five minutes with PBS containing 0.1% Hoechst dye to

stain the nuclei, and a coverslip was mounted with a 50% glycerol:PBS solution. A DeltaVision deconvolution microscope (Applied Precision) at 40x magnification was used to collect the images, using the same shutter exposure (1 second) for GFP fluorescence.

Protein modeling and alignments

The homology modeling server Swiss Model (http://swissmodel.expasy.org/) was used to generate structures of A3B and AID based off the crystal structure of the C-terminal domain of APOBEC3G (3IR2 Chain A)[181-183]. For A3B the QMEAN4 score was 0.52 and for AID the QMEAN4 score was 0.57 (range from 0-1, higher numbers characterize more accurate models)[203]. Models were loaded into the molecular visualization program, VMD[204], and the selected residues highlighted. The chosen views were rendered for display. For alignment A3B sequence NCBI NP_004891.3, A3G sequence NCBI NP_068594.1, A3F sequence NCBI NP_660341.2 and AID sequence NCBI NP_065712.1 were used in a ClustalW alignment.

GST pulldown experiments

GST alone and GST-tagged importin α1, α3 and α5[41] were grown in BL21DE3 *E. coli* to OD600 0.6, induced with Isopropyl β-D-1-thiogalactopyranoside (IPTG) at 0.4 mM and expressed at 14 degrees for 16 hours. The cell pellet was resuspended in GST lysis buffer (25 mM HEPES pH 7.4, 10% glycerol, 150 mM NaCl, 0.5% Triton X-100, 1 mM EDTA, 1 mM $MgCl_2$, 1 mM $ZnCl_2$) with protease inhibitors. The suspension was sonicated (Misonix XL-2000) three times for 2 minutes total (15 second pulses) with rests on ice.

Debris was pelleted from the solution at 14,000 rpm and the supernatent was incubated with glutathione beads (Glutathione Sepharose 4 fast flow; GE Healthcare) overnight. The beads were washed four times with GST lysis buffer. HEK293T cells were transfected with 5 ug of DNA (Trans-IT LT1; Mirus) and allowed to express the GFP tagged proteins for 48 hours. These cells were harvested, lyzed and the insoluble fraction pelleted by centrifugation. Lysates were added to the glutathione beads and incubated overnight. The beads were washed six times with lysis buffer and the bound proteins eluted by boiling with SDS loading dye. Samples were run on acrylimide gels (BioRad), then transferred to PVDF membranes. After blocking for 1 hour (3% casein in PBS with 0.1% Tween), mouse anti-GFP (JL-8, BD Clontech 632 381) was used at a 1:2000 dilution. The membranes were incubated overnight 4 degrees C. The membranes were washed four times with PBS Tween before incubating with 1:5000 goat anti-mouse HRP (BioRad) secondary for 1 hour at room temperature. The blots were developed with chemiluminescent reagents A and B and film (Denville HyGlo A and B, HyBlot).

Class switch recombination assays

Both the class switch recombination assay and the AID deficient mice used to generate the B-cells have been previously described[18,200]. All experiments followed the guidelines set forth by the University of Minnesota Animal Care and Use Committee. Briefly, B-cells were purified from the mouse spleen using magnetic sorting (Miltenyi Biotech) and cultured in RPMI supplemented with 10% FBS, 50 ng/mL IL-4 and 50 ug/mL LPS. After 48 hours the cells were infected with GFP, AID, A3B and point mutant viruses, resuspended in fresh media and cultured for 4 days. Conversion of IgM to IgG was

measured by flow cytometry (FACSCanto II, BD Biosciences) after staining with anti-IgG1-PE (BD Biosciences).

HIV single-cycle experiments

The single-cycle HIV assay and specific antibodies used for analysis have been described[96] with the only modification being the use of HeLa cells. Briefly, either HEK293T or HeLa cells were transfected with 1 ug of the HIV1IIIB A200C proviral construct and 0, 25, 50 or 100 ng of HA-tagged APOBEC3 construct. Twice as much AID-HA (50, 100, 200 ng) was required for similar expression levels. The supernatents from the transfected cells were added to the reporter cell line, CEM-GFP, and these cells were subsequently analyzed by flow cytometry to quantify viral titers. Cell and viral particle lysates were prepared and immunoblotted as previously reported[96].

HIV Infection and fluorescent microscopy studies

HeLa cells transfected with mCherry, A3G-mCherry and A3B-mCherry were seeded into 6 well plates on coverslips and allowed to adhere overnight. HIV_{LAI} nef::eGFP was added to half the samples and both treated and untreated cells were fixed with 4% PFA for 20 minutes in PBS at 0, 24 and 48 hours time-points. The nuclei were stained with Hoescht dye (0.1% in PBS). The coverslips were mounted in 50% glycerol in PBS and imaged on a DeltaVision deconvolution microscope (Applied Precision) at 40x magnification. Deconvolution was performed using DeltaVision softWoRx software (Applied Precision).

Immunofluorescence experiments

Approximately 20,000 cells of HeLa, HEK293T, U2OS, TR146, JSQ3 or MCF10A cells were plated on LabTek chambered cover glasses (Nunc) and transfected with A3B-HA and incubated overnight. Cells were washed with PBS, then fixed with 4% paraformaldehyde in PBS for 20 minutes at room temperature. Cells were washed with PBS, incubated in blocking buffer for 1 hour (1% BSA, 1% goat serum and 0.2% Triton-X 100 in PBS) then stained with 1:200 mouse anti-HA 11 (Covance MMS-101P) overnight. Cells were washed with PBS five times then incubated with 1:200 goat anti-mouse FITC (Jackson 115095146) for 1 hour. After washing four times cells were incubated in PBS containing 0.1% Hoechst dye to stain the nuclei. A DeltaVision deconvolution microscope (Applied Precision) at 60x magnification was used to collect the images, and deconvolution was performed using DeltaVision softWoRx software (Applied Precision).

LINE1 studies

L1 assays were carried as previous described[117]. Briefly, a LINE1 construct containing eGFP with an inverted intron was transfected into HEK293T cells along with HA-tagged constructs of interest. All constructs were tested in triplicate. After 48 hours the cells were tested for eGFP expression by flow cytometry and then selected with 0.8 ug/mL of puromycin. Flow cytometry was performed at the indicated time-points. Western blots were performed from pooled cell lysates collected at 48 hours and probed with anti-HA and anti-tubulin antibodies.

Specific Contributions

We thank J. Di Noia, M. Herzberg, J. Mueller and M. Stevenson for reagents, R. LaRue for the phylogenetic schematic, and J. Lee for technical assistance. This work was supported by NIH R01 AI064046 and P01 GM091743. L. Lackey was supported in part by an NSF Pre-doctoral Fellowship and subsequently by a position on the Institute for Molecular Virology Training Grant NIH T32 AI083196. Z.L. Demorest was supported in part by NIH T32 AI007313. A.M. Land was supported by a CIHR Postdoctoral Fellowship. J.F. Hultquist was supported by an NSF Predoctoral Fellowship.

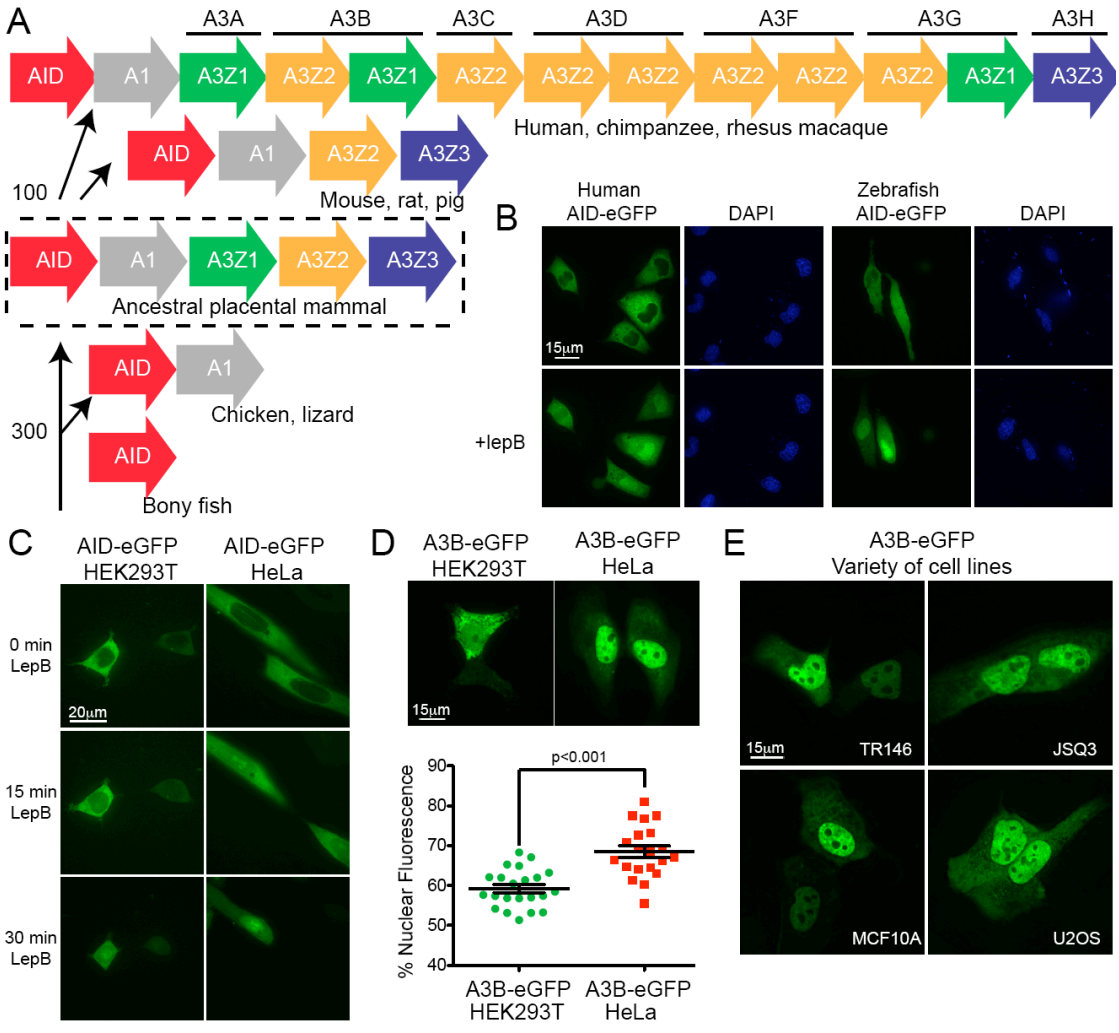

Figure 2-1. Relationships between AID and A3B. (A) Phylogenetic tree depicting the *APOBEC* loci in the indicated species. The split between fish and birds (~300 million years ago) and the divergence of the original placental mammal (~100 million years ago) are shown[2,5,55]. **(B)** Representative images of HeLa cells transfected with human or zebrafish AID-eGFP and treated with lepB or ethanol as a vehicle control. **(C)** Representative images of HEK293T or HeLa cells expressing human AID-eGFP after treatment for the indicated times with lepB. **(D)** Representative images of HEK293T or

HeLa cells expressing A3B-eGFP (quantified below; mean and SD shown for >20 individual cell measurements). **(E)** Representative images of A3B-eGFP expressed in the buccal tumor epithelial line TR146, the squamous cell carcinoma line JSQ3, the breast epithelial line MCF10A, and the osteosarcoma line U2OS.

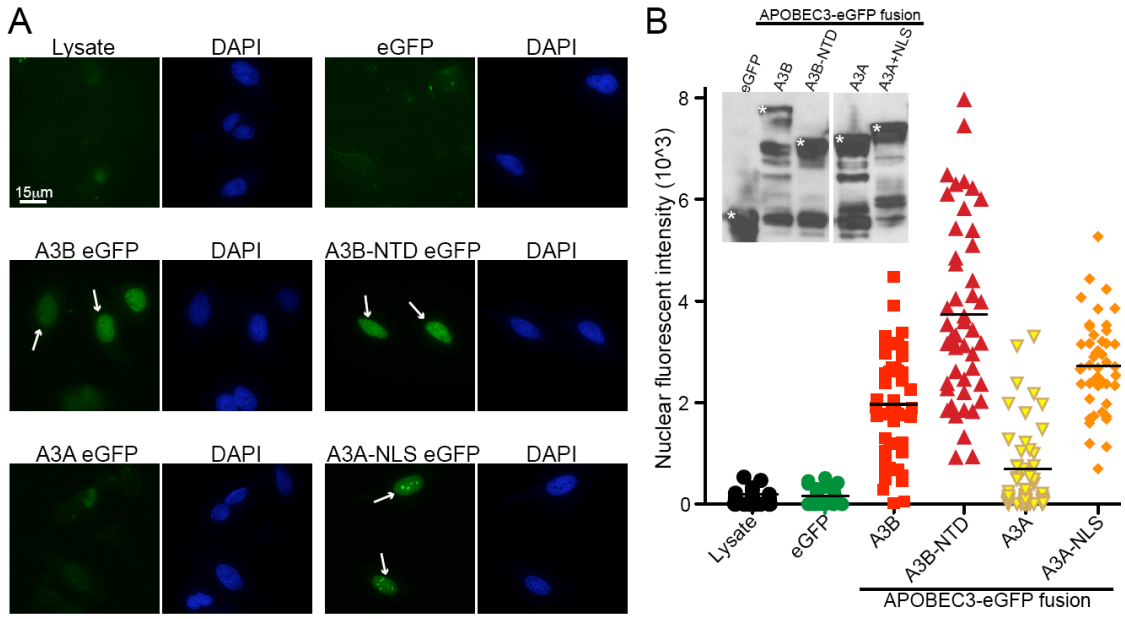

Figure 2-2. A3B is actively imported into the nucleus. (A) Representative images of digitonin treated HeLa cells incubated with lysates containing GFP, A3B-eGFP, A3B NTD-eGFP, A3A-eGFP, or A3A NLS-eGFP. White arrows highlight instances of active nuclear import. **(B)** Quantification of the results from (A) using the same exposure for all conditions (FITC=1 second; n≥30 cells were examined and the mean nuclear fluorescent intensity is indicated for each condition). The inset is an anti-GFP immunoblot of representative cell lysates with asterisks indicating the correct bands.

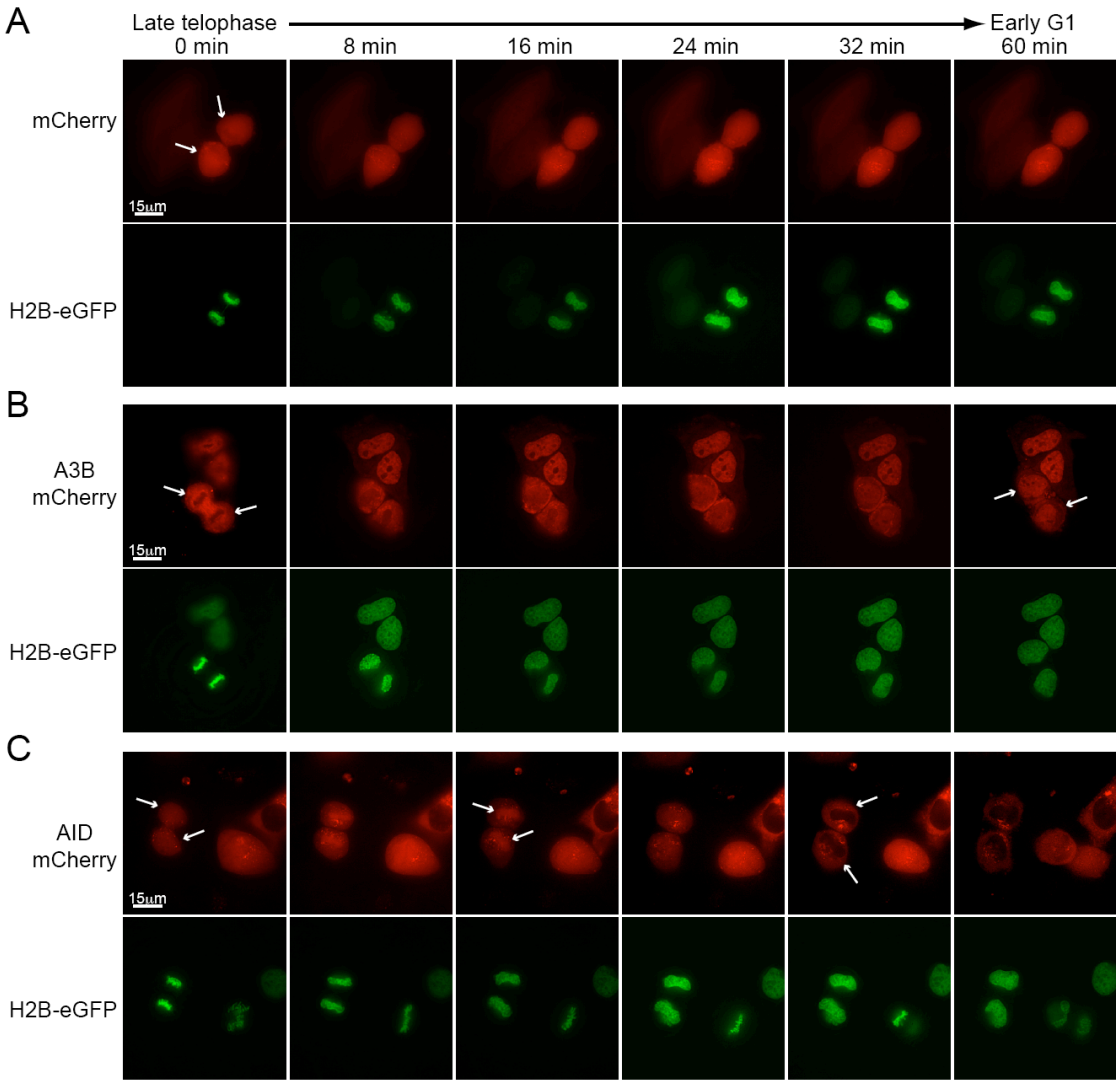

Figure 2-3. A3B and AID localize differently during the mitosis. Representative image frames of **(A)** mCherry, **(B)** A3B-mCherry, and **(C)** AID-mCherry localization during HeLa cell cycle progression from late telophase of mitosis to early interphase. H2B-eGFP images of the same cells are shown below each time series. The white arrows at 0 min highlight informative cells, and the arrows at other time-points indicate significant localization events discussed in the text.

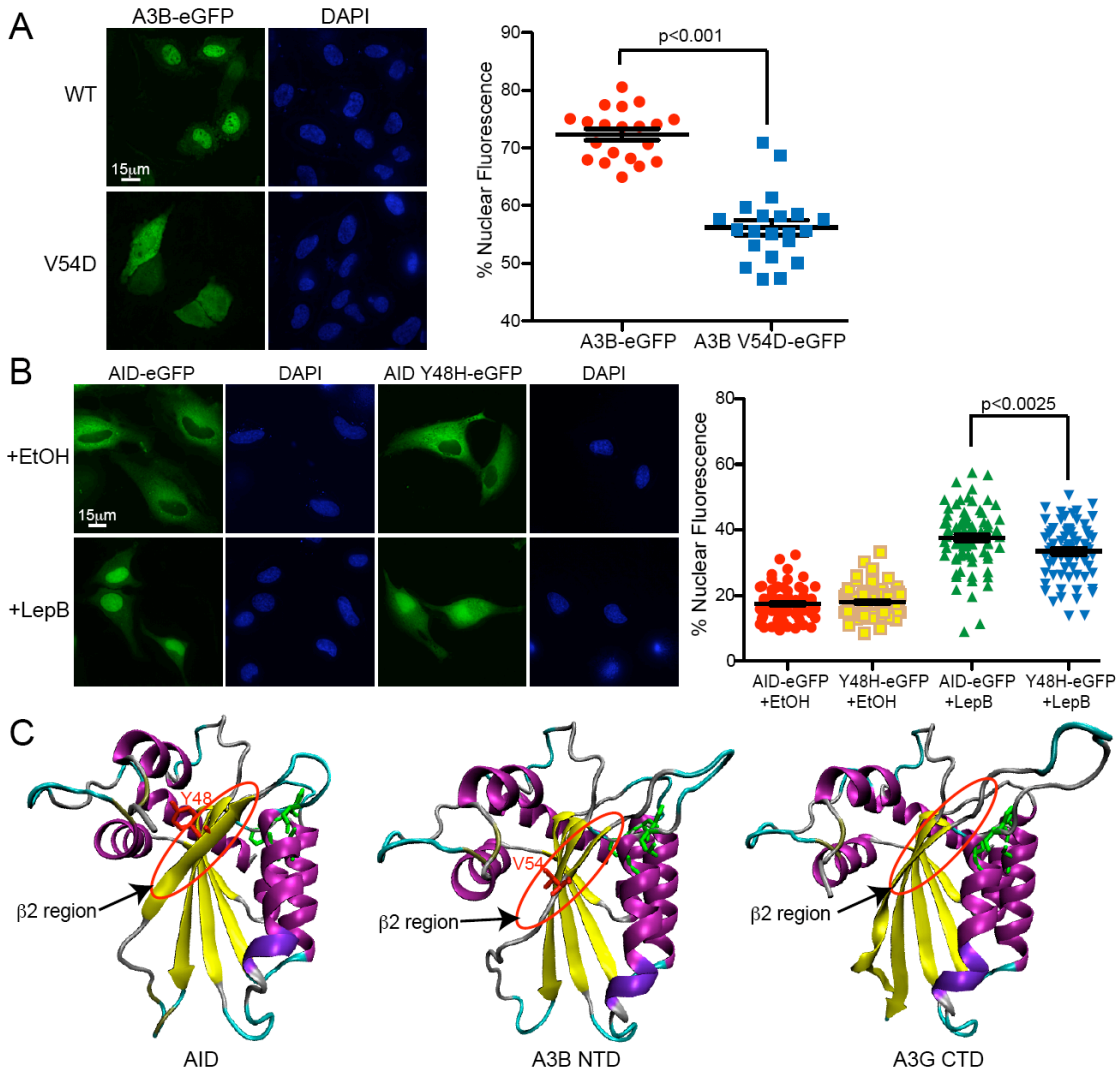

Figure 2-4. Single amino acid changes within the β2 region affects nuclear localization of A3B and AID. (A) Representative images of A3B-eGFP and A3B V54D-eGFP in HeLa cells. The adjacent dot plot reports quantification of the nuclear to total fluorescent signal (n≥25 cells were analyzed for each condition with mean and SD shown). **(B)** Representative images of AID-eGFP and AID Y48H-eGFP in HeLa cells taken 3 hrs after ethanol or lepB treatment. The adjacent dot plot reports quantification of

the nuclear to total fluorescent signal (n≥50 cells were analyzed for each condition with mean and SD shown). **(C)** Model ribbon structures of AID and A3B NTD depicted adjacent to the actual structure of A3G C-terminal domain (CTD). The β2 region and key amino acids in AID and A3B are labeled in red and circled; the side chains of zinc-coordinating residues are illustrated in green.

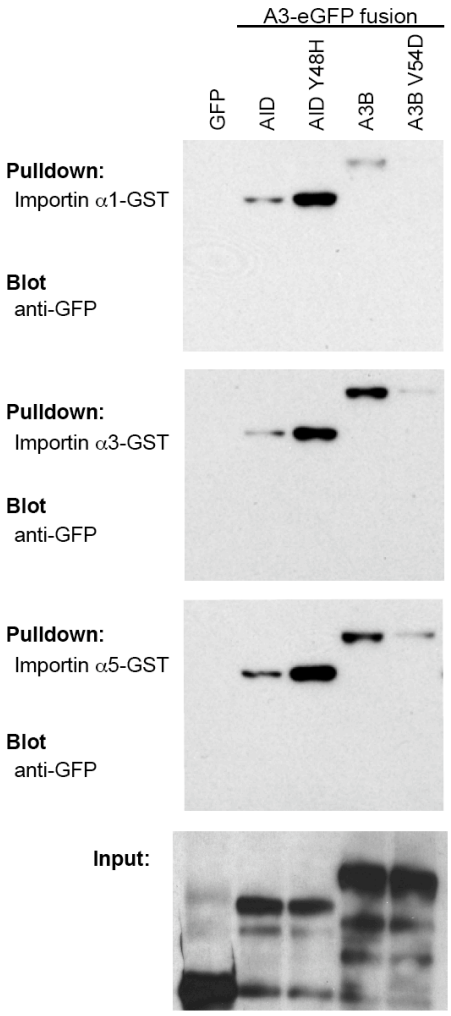

Figure 2-5. Both AID and A3B interact with members of the adaptor importin family. Immunoblots of input HEK293T protein lysates (bottom panel) and pulldown results for eGFP, AID-eGFP, AID Y48H-eGFP, A3B-eGFP and A3B V54D-eGFP with the indicated GST-importin substrates (top panels).

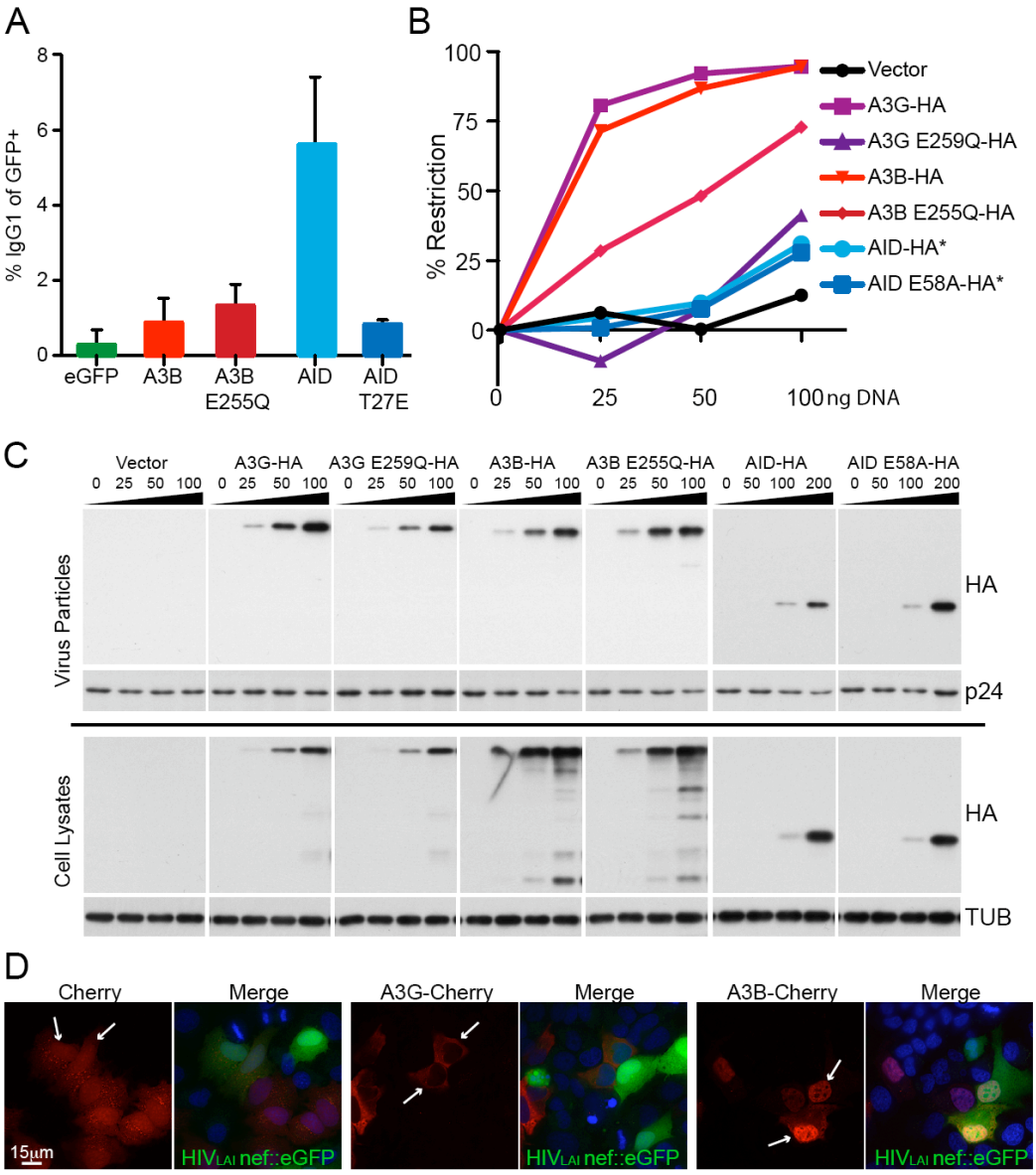

Figure 2-6. A3B does not perform class switch recombination and AID does not restrict HIV-1. (A) Flow cytometry quantification of the isotype switch to IgG in AID-deficient murine B-cells expressing AID, A3B or non-functional mutants (mean and SD are shown for duplicate samples). **(B)** Infectivity of HIV-1 produced in HeLa cells expressing HA tagged APOBEC3G (A3G), A3B, AID and their catalytic mutants. High

levels of restriction correspond to lower levels of fluorescence in a reporter CEM-GFP cell line and a decrease in infectious virus. Asterisks indicate that twice as much DNA was required for adequate expression of AID (0, 50, 100 and 200 ng). **(C)** Immunoblots of viral particle proteins (top) or cell lysate proteins (bottom) probed for anti-HA (APOBEC/AID expression), anti-p24 for a virus loading control, or anti-tubulin for a cellular loading control. **(D)** Representative images of mCherry, A3B-mCherry or AID-mCherry in HeLa cells 24 hrs after infection with replication competent HIV-1_{LAI} nef::eGFP. Cells expressing the protein of interest infected with HIV are indicated by white arrows.

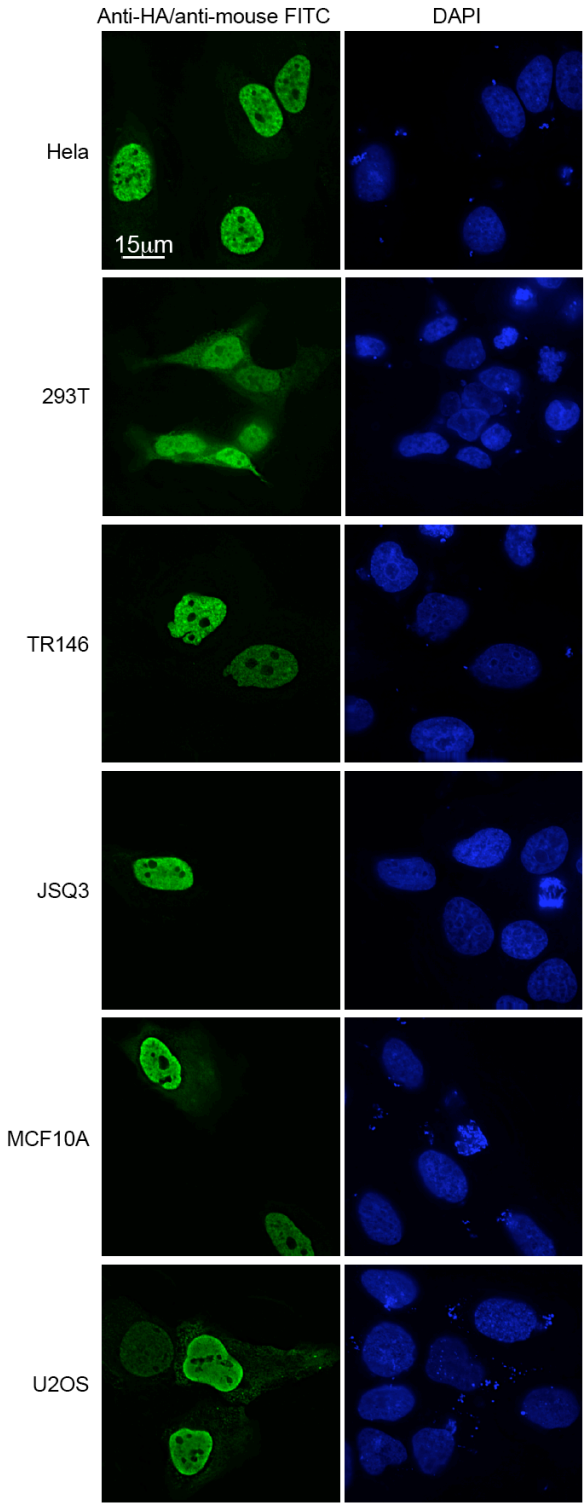

Figure 2-S1. A3B-HA is localized to the nucleus. Representative images showing A3B-

HA localization in HeLa, HEK293T, the head and neck cancer cell lines, TR146 and JSQ3, the breast epithelial cell line, MCF10A, and the osteosarcoma cell line, U2OS.

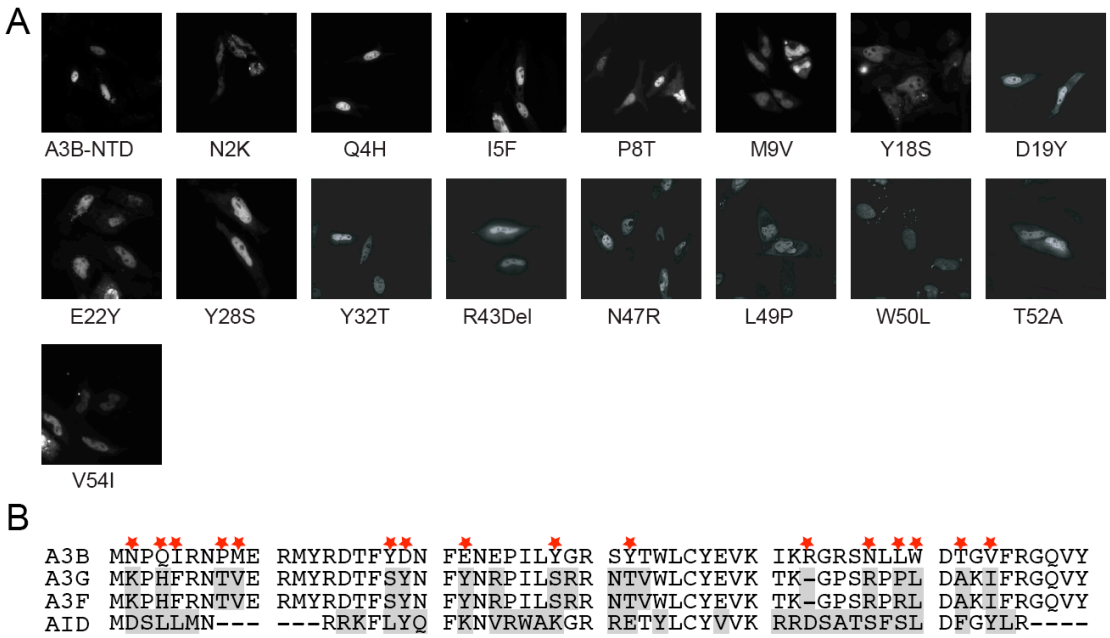

Figure 2-S2. Single point mutants in the N-terminal domain of A3B do not affect its localization. (A) Representative images of A3B NTD-eGFP and point mutants as indicated below the image. **(B)** Alignment of A3B with A3G, APOBEC3F (A3F) and AID. Both A3G and A3F are cytoplasmic proteins. Differences between A3B, A3G/A3F and AID are indicated with gray shading. The red stars indicate mutants shown in (A).

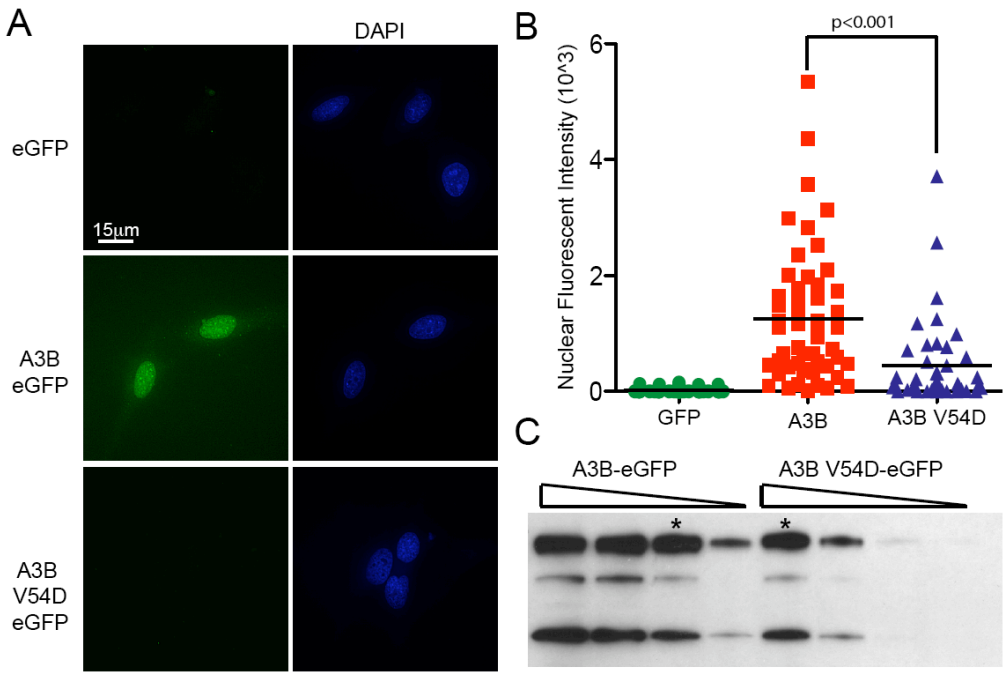

Figure 2-S3. A3B V54D-eGFP does not enter the nucleus after digitonin treatment. **(A)** Representative images of digitonin treated HeLa cells incubated with lysates containing eGFP, A3B-eGFP, or A3B V54D-eGFP. **(B)** Quantification of the results from (A) using the same exposure for all conditions (FITC=1 second; n≥25 cells were examined and the mean nuclear fluorescent intensity is indicated for each condition). **(C)** Anti-GFP immunoblot of cell lysates with asterisks indicating protein expression in the lysate concentration used for quantification.

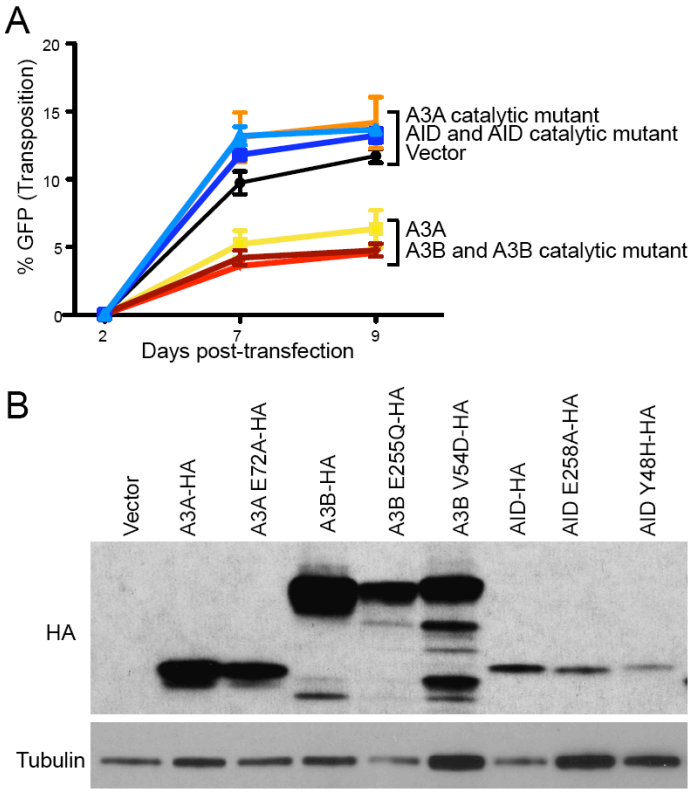

Figure 2-S4. A3B restricts L1 transposition in a deaminase independent manner while AID is a poor restrictor. (A) Flow cytometry quantification of eGFP expressing cells as a marker for retrotransposition in APOBEC/AID expressing populations (mean and SD shown for triplicate experiments). Measurements are shown for vector (black, circles), A3A-HA (yellow, triangles), A3A E72A-HA (orange, triangles), A3B-HA (red, downward triangles), A3B E255Q-HA (maroon, diamonds), AID-HA (blue, squares), and AID E58A-HA (light blue, triangles). **(B)** Anti-HA immunoblot of cell lysates indicating expression of the APOBEC3/AID proteins with anti-tubulin as a loading control.

Movies 1-12 show mCherry, A3B and AID time-lapse microscopy with or without H2B-eGFP. They are available online as supplementary material for Lackey, *et al.* 2012[205] or by request.

Movie 1. mCherry localization during mitosis.

Movie 2. Same as Movie 1, but with mCherry overlaid with H2B-eGFP fluorescence.

Movie 3. Secondary set of mCherry localization during mitosis.

Movie 4. Same as Movie 3, but with mCherry overlaid with H2B-eGFP fluorescence.

Movie 5. A3B-mCherry is excluded during mitosis.

Movie 6. Same as Movie 5, but with A3B-mCherry overlaid with H2B-eGFP.

Movie 7. Secondary set of A3B-mCherry during mitosis.

Movie 8. Same as Movie 7, but with A3B-mCherry overlaid with H2B-eGFP.

Movie 9. AID-mCherry during mitosis.

Movie 10. Same as Movie 9, but with AID-mCherry overlaid with H2B-eGFP.

Movie 11. Secondary set of AID-mCherry during mitosis.

Movie 12. Same as Movie 11, but with AID-mCherry overlaid with H2B-eGFP.

Chapter 3

Subcellular localization of the APOBEC3 proteins during mitosis and implications for genomic DNA deamination

This chapter is part of the manuscript *in review*: Lackey, L., Law, E. K., Brown, W. L. & Harris, R. S. (2012). *Cell Cycle.*

Foreword

Previously, in order to test for active nuclear import by A3B, we documented the subcellular localization of A3B during mitosis (Chapter 2). Surprisingly, we saw that A3B was excluded from condensed chromosomes and only re-entered the nucleus after telophase[205]. This led us to question the behavior of other APOBEC3 proteins during mitosis. For instance, we knew that A3G bound and shifted DNA in electrophoretic mobility shift assays and we wondered whether A3G would enter the nucleus and bind genomic DNA during mitosis[206,207]. At the same time we were becoming more interested in the ability of A3B to deaminate genomic DNA in breast cancers (Chapter 4). We reasoned that if the cytoplasmic APOBEC3 proteins could enter the nucleus during mitosis, they might have a chance to deaminate genomic DNA, like our preliminary evidence suggested for the nuclear A3B (Chapter 4). Here I describe our observations on the subcellular localization of the APOBEC3s and their corresponding ability to affect cell progression.

Contributions

I imaged and analyzed the fixed microscopy of all the APOBEC3 proteins, their derivatives and AID and created time-lapse movies of a subset with the assistance of EKL. I performed the activity assays and cell cycle progression analyses. EKL, WLB and I analyzed the A3B and A3G reactive rabbit antibodies. RSH contributed to experimental design and editing of the manuscript.

Summary

Humans have seven APOBEC3 DNA cytosine deaminases. The activity of these enzymes allows them to restrict a variety of retroviruses and retrotransposons, but may also cause pro-mutagenic genomic uracil lesions. During interphase the APOBEC3 proteins have different subcellular localizations: cell-wide, cytoplasmic and nuclear. This implies that only a few APOBEC3s have contact with nuclear DNA. However, during mitosis, the nuclear envelope breaks down and cytoplasmic proteins may enter what was formerly a privileged zone. To address the hypothesis that all APOBEC3 proteins have access to genomic DNA, we observed the localization of the APOBEC3 proteins during mitosis. We show that APOBEC3A, APOBEC3C and APOBEC3H are excluded from condensed chromosomes, but become cell-wide during telophase. However, APOBEC3B, APOBEC3D, APOBEC3F and APOBEC3G are excluded from chromatin throughout mitosis. After mitosis APOBEC3B becomes nuclear and APOBEC3D, APOBEC3F and APOBEC3G become cytoplasmic. Both structural motifs as well as size may be factors in regulating chromatin exclusion. We also analyzed APOBEC3-induced cell cycle perturbations as a measure of each enzyme's capacity to inflict genomic DNA damage. AID, APOBEC3A and APOBEC3B altered the cell-cycle profile and, unexpectedly, APOBEC3D also caused changes. We conclude that several APOBEC3 family members have access to the nuclear compartment and can impede the cell-cycle, possibly through DNA deamination and the subsequent DNA damage response. Such genomic damage may contribute to carcinogenesis, as demonstrated by AID in B-cell cancers and, recently, APOBEC3B in breast cancers.

Introduction

The seven APOBEC3 proteins are part of a larger polynucleotide cytosine deaminase family that in humans also includes activation induced cytosine deaminase (AID), APOBEC1, APOBEC2 and APOBEC4[5]. All of the APOBEC3 proteins have the ability to convert single-stranded DNA cytosines to uracils and this is the main mechanism by which several of these enzymes restrict the replication of HIV-1 and other retroviruses[5,83]. Relevant APOBEC3s are packaged into viral particles along with the viral RNA genome. During reverse transcription the single-stranded viral cDNA becomes a target for cytosine deamination; replication across this cDNA fixes uracils as point mutations in the genome resulting in hypermutation and either inactivation or degradation of the retrovirus[83]. This mechanism is clear for APOBEC3G (A3G) and APOBEC3F (A3F) on HIV-1, as well as for other APOBEC3 enzymes on a broad number of other viral substrates[83,98]. In addition, APOBEC3 proteins have a role in restricting retrotransposons (including LINE-1, Alu, IAP and MusD), primarily, although not solely, through a deamination process similar to retrovirus restriction[96,208,209].

However, APOBEC3 enzymatic activity may also be problematic for genomic DNA. As an important precedent, AID-dependent uracil lesions in the antibody locus can be processed into chromosomal translocations and these uracil lesions in other genomic areas may also be pro-carcinogenic[25,210,211]. For example, transgenic expression of AID causes T-cell cancer in mice[210]. Likewise, transgenic expression of APOBEC1 can cause hepatocellular carcinoma[65,132]. Although a role for APOBEC3A (A3A) in cancer development is unclear, ectopic expression causes S-phase arrest, γ-H2Ax focus formation and mutational events[133]. In contrast, we recently demonstrated the up-

regulation of A3B and its contribution to the overall mutation load in breast cancer[71]. The role of other APOBEC3 proteins in genomic mutation is less clear.

Subcellular regulation allows cells to compartmentalize proteins that could be genotoxic or cytotoxic. For example, caspase-activated deoxyribonuclease (CAD) is a DNase containing a nuclear localization signal that is complexed with an inhibitory protein in the cytoplasm. Cleavage of this inhibitor by caspase-3 allows CAD to enter the nucleus and degrade the genome as part of the natural apoptotic pathway[212]. Likewise, the transcription factors STAT1 and NFκB are kept in the cytoplasm until activated, when they transport to the nucleus and bind promoters to enhance or suppress transcription[213,214]. Proteins can be actively targeted to a region of the cell through localization sequences within the protein or in an interacting partner. They can also be excluded from the nuclear compartment passively based on size and shape, with a limit of approximately 50 kDa and 6 nm diameter[165].

To complete mitosis, a cell must split its replicated genome between two daughter cells. In mammals, mitosis is facilitated by breaking down the nuclear envelope to allow for spindle formation and physical segregation of the chromosomes. The nuclear envelope re-forms shortly after cytokinesis[215]. Some proteins change localization during mitosis. For example, activated NFAT, which is nuclear during interphase, is excluded from DNA during mitosis and remains cytoplasmic until it is activated again[216]. RUNX proteins, which form nuclear foci, dissolve and reform after mitosis, unlike histones, which remain bound to the DNA[217].

During interphase, the APOBEC3 proteins vary in subcellular localization. A3A, A3C and A3H are the smallest deaminases, each having a single deaminase domain (~25

kDa). A3A and A3C have shown cell-wide distributions, whereas A3H is more variable, but A3H haplotype II is both cytoplasmic and nucleolar[96,111,119,120,151]. As a comparison, AID, which is a similarly sized, single domain deaminase, appears cytoplasmic at steady state, but clearly shuttles between the nuclear and cytoplasmic compartment[41,42,169,205]. APOBEC1 is also a shuttling protein with a predominantly cytoplasmic steady state distribution in the absence of its interacting partner ACF[78,172,218]. The double domain APOBEC3s (~50 kDa) are composed of two conserved deaminase domains and are either nuclear (A3B) or cytoplasmic (A3D, A3F, and A3G) during interphase[87,117,143,144,150,153,155,205,219]. Interestingly, DNA damage has been shown to cause cytoplasmic APOBEC3G to enter the nucleus[157], as well as affecting the shuttling of AID, shifting it from primarily cytoplasmic to more nuclear[44]. It is not clear what the importance of subcellular localization is for the function of the APOBEC3 proteins, although a cytoplasmic distribution may affect HIV-1 inhibition[150,156].

Recently, we analyzed A3B and AID during mitosis[205]. A3B is excluded from nuclear DNA during the entire process of mitosis while AID associates with nuclear DNA during telophase. We thus became interested in the potential for other APOBEC3 enzymes to bind nuclear DNA during mitosis. The present study therefore provides comparative and mechanistic information on the mitotic regulation of the other members of the APOBEC3 family.

Results

A3A, A3C, and A3H have access to genomic DNA during interphase and telophase

We first considered the single domain APOBEC3s, which are similar in size to AID and APOBEC1. The predicted molecular weight for A3A, A3C, and A3H is 25 kDa (50 kDa with eGFP). The interphase localization of these single domain proteins has been described[96,111,119,120,151]. For all our experiments we used the stable A3H haplotype II, which is more cytoplasmic than A3A and A3C, but can be seen in the nucleoli of HeLa cells[122,139,154] (**Fig. 3-S1**). The E72A catalytic mutant was used for A3A, because the wild-type enzyme killed almost all cells after 48 hours of transient expression (concordant with prior reports[131]). As described[205], AID was excluded from chromatin during prophase, metaphase, and anaphase, but co-localized specifically with DNA during telophase (**Fig 3-1A, 3-1B and 3-S2**). As expected, eGFP alone was excluded from the condensed DNA in prophase, metaphase, and anaphase[220] (**Fig. 3-1A and Fig. 3-S2**), but resumed its cell-wide distribution once the chromosomes began to relax during telophase (**Fig. 3-1B**). Likewise, A3A-E72A and A3C were excluded from condensed chromosomes during prophase, metaphase, and anaphase (**Fig. 3-1A and 3-S2**). Once the cells reached telophase A3A-E72A and A3C became fully cell-wide (**Fig. 3-1B**). Thus, A3A-E72A and A3C are excluded from DNA during early mitosis in a manner similar to eGFP, implying that these proteins do not bind to DNA and may simply be physically excluded from highly condensed chromosomes (unlike AID or eGFP-tagged histone 2B (H2B), **Fig. 3-1A and Movie 14**). These data are supported by two-color live cell microscopy using mCherry-tagged A3A-E72A and eGFP-tagged H2B (**Movies 13 and 14**). A3H-eGFP was also excluded from DNA during prophase, metaphase, and anaphase

(**Fig. 3-1A and Fig. 3-S2**). However, during telophase, some A3H-eGFP remained excluded from telophase DNA, but could also been seen in punctate nuclear bodies, possibly part of re-assembling nucleoli (**Fig. 3-1B**). This continuing exclusion of A3H-Hap II during telophase is similar to what we observe for the double domain APOBEC3 proteins (below).

A3B, A3D, A3F and A3G are excluded from the DNA during mitosis

We next considered the double domain APOBEC3s. These APOBEC3 proteins are twice the size of the single domain proteins (~50 kDa native or ~75 kDa with eGFP), and are generally composed of a C-terminal active deaminase domain and a less active or inactive N-terminal deaminase domain. A3B is nuclear during interphase, while A3D, A3F, and A3G are cytoplasmic[87,117,143,150,155,156,219] (**Fig. 3-S3**). Despite A3B's nuclear interphase localization, we had previously seen that it was excluded from mitotic chromosomes[205] (**Fig. 3-2B**). Likewise, A3D, A3F, and A3G were all excluded from condensed chromosomes during prophase, metaphase, anaphase, and telophase (**Fig. 3-2A, 3-2B and 3-S4**). This contrasts with AID, which associates with chromatin during telophase, and with A3A-E72A, A3C, and eGFP alone, which begin to resume their cell-wide distributions during telophase (**Fig. 3-1B**). These results are supported by time-lapse microscopy using A3F-mCherry and H2B-eGFP (**Movies 15 and 16**). We conclude that A3D, A3F, and A3G have little opportunity for contact with genomic DNA during interphase or throughout mitosis.

Size dependent and independent effects on mitotic localization

All the double-domain proteins were excluded from genomic DNA. Since these tagged APOBEC3 proteins are approximately 75 kDa, they are much larger than the tagged 50 kDa single-domain enzymes. This size difference could be an important biological means of control or an effect of the tag. To test this possibility, we analyzed all four double-domain proteins with a smaller HA-tag. Similar to our previous data, we saw that A3B, A3D, A3F, and A3G-HA were all excluded from mitotic DNA during telophase (**Fig. 3-S5**). This triple HA-tag is less than 5 kDa. Since A3A-eGFP is cell-wide (51.3 kDa), A3B-HA (50.1 kDa) should not be impeded by its tag from its normal cellular distribution. To further rule out the effect of a tag on the localization of the double-domain APOBEC3s, we used recently developed rabbit polyclonal sera to analyze the subcellular distribution of the untagged wild-type enzymes. Interphase localization of A3B was nuclear, as expected (**Fig. 3-3A, left**), and A3B was excluded from DNA during telophase and then began to re-enter the nucleus in cells progressing through telophase (**Fig. 3-3A, right**). Untagged A3G was cytoplasmic during interphase and excluded during telophase (**Fig. 3-3B**). The localization of untagged A3B and A3G mimics the localization of HA and eGFP tagged proteins. This supports our conclusion that under normal conditions A3B only co-localizes with DNA during interphase and that A3D, A3F, and A3G are excluded in both interphase and throughout mitosis.

Determinants of double domain APOBEC3 localization

We hypothesized that APOBEC3 exclusion from mitotic DNA is governed by an internal regulatory element. To test the hypothesis that at least one of the two domains in the double-domain proteins harbor subcellular localization determinants, we analyzed the

localization of the N- and C-terminal domains of A3B, A3D, A3F, and A3G separately. During prophase, metaphase, and anaphase all of these proteins were excluded from condensed chromosomes, similar to AID and all the other APOBEC3s (**Fig. 3-1 and 3-2**, data not shown). During telophase, the N-terminal halves of A3B and A3D were both nuclear, while the N-terminal domains of A3G and A3F were cytoplasmic (**Fig. 3-4A**). The C-terminal domains of A3B, A3D, A3F and A3G were cell-wide (**Fig. 3-4B**). As expected[150], the localization of the N-terminal domains during interphase was similar to the localization of the full-length proteins during interphase (i.e. A3B N-terminal and full length were nuclear, while A3F and A3G N-terminal and full-length were cytoplasmic), the only exception being A3D (**Fig. 3-4C**). Interestingly, full-length A3D was cytoplasmic, but its N-terminal domain was nuclear and its C-terminal domain was cell-wide. We tested whether A3D shuttles from the nucleus into the cytoplasm through the CRM1 pathway using the inhibitor lepB[174]. AID appeared cytoplasmic, but lepB caused it to become more nuclear[42,169] (**Fig. 3-S6**). In contrast, A3D remained cytoplasmic in the presence of lepB, implying that it does not use the CRM1 pathway (**Fig. 3-S6**). In addition, the ability of N-terminal A3B to co-localize with DNA during telophase is different from the delayed re-entry of full-length A3B, which remains excluded from DNA during telophase (**Fig. 3-4A and 3-2B**). We conclude that A3F and A3G are likely to be actively excluded from DNA, since their N-terminal domains alone are excluded (**Fig. 3-4A**). A3B and A3D may rely on the size of the double domain or a motif created from the combination of both domains since in both cases the N-terminal half of the protein has access to DNA during telophase while the full-length protein is excluded (**Fig. 3-4A and 3-2B**). Interestingly, forcing two A3C proteins together in an A3C-A3C-

eGFP chimera created a protein that was cytoplasmic in interphase and excluded from genomic DNA in the same way as A3B, A3D, A3F, and A3G (**Fig. 3-S7**). Thus, both *cis* determinants and size seem to affect the telophase DNA exclusion seen for A3B, A3D, A3F, and A3G.

Cell cycle regulation of DNA deaminase activity

We tested for changes in the deaminase activity of APOBECs that co-localize with DNA (A3A, A3B, A3C, and A3H). We hypothesized that in addition to physical exclusion of the APOBECs during mitosis, the deaminase activity of these enzymes might be regulated. Untreated or thymidine-blocked HEK293T cells were transfected with eGFP-tagged A3A, A3B, A3C, and A3H. After 24 hours of expression untreated (asynchronous) cells expressing each APOBEC3 construct were harvested. Subsequently, thymidine synchronized cells followed by a four-hour release were harvested with the cells positioned in S phase, as expected[221] (**Fig. 3-5A**). Nocodazole was added after the second thymidine block to arrest the cells in G2, however, this caused a significant induction of apoptosis (**Fig. 3-5A**). Lysates from these stages were used to test for deaminase activity on fluorophore-conjugated oligonucleotides containing a cytosine base in the appropriate context. Both A3A and A3B are strong deaminases and produced deaminase activity from all samples (**Fig. 3-5B**). Comparatively, A3C and A3H are much weaker deaminases. While A3H could deaminate oligonucleotides, A3C did not demonstrate activity in this assay, even at higher concentrations (data not shown). While deaminase activity has been observed for both A3C and A3H, they are not strong *in vitro* deaminases[222]. None of the lysates from different phases of the cell cycle varied

dramatically in activity. These experiments suggested that A3A, A3B, and A3H are similarly active in both G1 and S-phase.

APOBEC3-induced cell cycle perturbations

The cell cycle is a highly regulated progression with delicate checks and balances that prevent cells from dividing in the presence of DNA damage. We used this innate DNA damage sensing property to test for DNA damage caused by the APOBEC3 proteins. Tetracycline inducible HEK293 and HeLa cells show cell cycle disruptions after expression of A3A and A3B[71,131] (Chapter 4 and Fig. 4-3) and repair deficient B-cells show toxicity after AID expression[223]. We therefore tested all APOBEC3s and AID for cell cycle effects by transient transfection in HEK293T and HeLa cells. Based on the mitotic images described above, we predicted that A3A, A3B, and A3H were mostly likely to alter the cell cycle profile over time, and, based on our activity data, A3C would have little effect even though it is cell-wide. Representative profiles for each APOBEC3 in these two cell lines are shown for 48 hours expression in HEK293T cells and 96 hours expression in HeLa cells (**Fig. 3-6A and 3-S8A**). A3A caused a consistent shift of the G1 peak toward S in both cell lines, and a broadening of the G2 population in HEK293T cells (arrows in **Fig. 3-6A and 3-S8A**). In HEK293T cells, only A3B, A3D and AID caused decreases in mitotic cells (**Fig. 3-6A**), while in HeLa cells only A3B and AID caused a dramatic decreases in DNA content, an indication of apoptotic cells (**Fig. 3-S8A**). Surprisingly, A3H did not cause a reproducible effect on the cell cycle profile, although mild effects were seen in some previous experiments (data not shown). A3C, A3F, and A3G did not cause dramatic changes to the shape or proportions of the cell

cycle profile, indicating that transient over-expression is not the cause of these cell cycle perturbations. From these data it is clear the A3A, A3B, and AID can affect cell cycle progression, as has been shown in different systems[71,131,223], and that A3D and possibly A3H are also able to activate cellular checkpoints. These cell cycle defects have been reported to be dependent on the catalytic activity of both A3A and A3B, supporting our use of cell cycle perturbations as a measure of genomic DNA deamination[71,131].

Discussion

We hypothesized that the mitotic breakdown of the nuclear envelope would allow APOBEC3s access to genomic cytosines for deamination. Since several APOBEC3s are positively charged and known to bind DNA[206,207], we expected to see the APOBEC3 proteins interacting with DNA during prophase upon dissolution of the nuclear envelope. Instead, we show that cytoplasmic APOBEC3 proteins do not have full access to genomic DNA, even during mitosis. The mechanism preventing the APOBEC3s from interacting with genomic DNA during prophase, metaphase, and anaphase is unclear, but may be as simple as exclusion from condensed chromatin, since eGFP is excluded in a similar manner. However, this simplistic model does not explain all observations, as DNA exclusion during telophase may be dependent on a combination of N-terminal determinants and protein size. A3F and A3G, for example, have *cis* determinants, since their N-terminal domains alone are cytoplasmic whereas A3B and A3D N-terminal proteins co-localize with DNA during telophase, while their full-length forms are excluded.

We were surprised that separately both domains of A3D had access to the nucleus, as well as by the fact that full-length A3D affected HEK293T and HeLa cell cycle profiles. We theorized that full-length A3D might shuttle between the cytoplasmic and nuclear compartment, but we did not detect CRM1-dependent shuttling. Although we have no evidence that the N- and C-terminal domains of A3D are expressed separately in our transient transfections, it is possible that separate N and C-terminal domains may have a role to play in genomic mutations. Double-domain *APOBEC3* genes in other species (*e.g.*, sheep and pig) can be expressed as either N- or C-terminal domains alone, as well as the full-length double domain[79], while in humans A3F has alternative isoforms[224]. It is therefore possible that other double domain *APOBEC3* genes may express single N or C-terminal domain variants that could have subcellular distributions differing from the full-length enzyme.

We observed a cell cycle progression defect in HeLa and HEK293T cells expressing A3A, A3B, A3D, and AID (and mildly for A3H), implying that deamination and the ensuing DNA damage response affects these cells. Both of A3A and A3B have been reported to induce cell death and the DNA damage response in HEK293 and HeLa cells[71,131]. While A3A is tightly regulated and is confined to cells of the myeloid lineage[128], other APOBEC3 proteins are expressed more broadly[134,135]. Interestingly, A3B, A3D, and A3H all have been inactivated or deleted to a certain extent in the human population. More than 90% of people of Oceanic heritage have an A3B deletion polymorphism that leaves the surrounding A3A and A3C genes intact, but completely removes A3B[193]. Chimpanzee A3D is much more active than human A3D, and this is largely dependent on a single amino acid difference[225,226]. Likewise, there are several

haplotypes of A3H, and the most prevalent among them are unstable[122,139,154]. We used the human A3H haplotype II in these experiments because it is stable and active against HIV-1[96,98]. Thus, A3B, A3D, and A3H may be sufficiently detrimental to cause selective inactivation and loss of these enzymes. Although the experiments we report are transient tests of the capability of the APOBEC3s to affect cell cycle progression, the data suggest that A3A, A3B, A3D, and A3H may be interesting for further studies on genomic mutation. In fact, transient expression of A3A causes genomic mutations in the nuclear and mitochondrial genomes[133] and high levels of endogenous A3B induces genomic mutations in breast cancer cells[71] (Chapter 4).

The long-term effect of genomic deamination may be predisposition to cancer, as suggested by the link between AID expression and B-cell cancer[25,210,227]. Breast cancer genome sequencing has shown that not only do breast tumors have a large number of somatic mutations, but also a high percentage of cytosine-to-thymine (C-to-T) transitions and regions of clustered mutations termed kataegis[228]. These mutation signatures from breast cancers can be linked to A3B, especially in regions of kataegis, where single-stranded DNA may be available for deamination[71]. Many other cancers also have a high proportion of C-to-T mutations, including brain, gastric, head and neck, ovarian, pancreatic, and prostate cancers[229-233]. Over the next few years, with the increasing availability of cancer genome sequences, and additional molecular studies of the APOBEC3 proteins, we predict that minimally A3B and potentially other APOBEC3s will become important for our understanding of cancer mutagenesis.

Methods

Fixed cell microscopy experiments

Microscopy procedures have been described[205]. Briefly, HeLa cells plated on glass coverslips (12-545-85; Fisher Scientific) in 6 well plates were transiently transfected with 400 ng each of eGFP, HA or untagged constructs (Transit-LT1; Mirus) and incubated for 48 hours. The cells were fixed with 4% paraformaldehdye in phosphate buffered saline (PFA in PBS) for 20 minutes at room temperature. APOBEC3-HA and untagged expressing cells were incubated with primary antibody in blocking buffer. Rabbit polyclonal sera were identified as described[136]. Hybridoma media was used for immunofluorescence. The anti-HA antibody (MMS-101P; Covance) was used at 1:200 and visualized with anti-mouse FITC (115095146; Jackson), while untagged A3B was identified with Rb10 14, untagged A3B was identified with Rb10 93 and these samples were visualized with anti-rabbit FITC (111095144; Jackson). All slides were treated with 0.1% Hoechst dye to stain the nuclei. The slides were mounted with 50% glycerol and imaged (Deltavision; Applied Precision). All images were deconvolved using SoftWorks (Applied Precision).

Live cell experiments

Movies were taken as described[205]. Briefly, HeLa cells plated at 40,000 cells/well in 4 well chambers (Nunc) were transiently transfected with 200 ng of eGFP-tagged histone 2B and 400 ng of APOBEC3-mCherry. The cells expressed these constructs for 48 hours before transfer to a heated chamber (37 C) on the microscope (Deltavision; Applied Precision). Images of dividing cells were taken every three minutes for 1-3 hours. These

images were deconvolved and used to create quicktime movies. Treatment with leptomycin B to inhibit CRM1 dependent export has been reported[200,205]. Cells were treated with 40 ng/mL of leptomycin B dissolved in ethanol or ethanol diluted in media alone. After three hours these cells were fixed for imaging.

DNA deaminase oligonucleotide cleavage assays

Synchronization protocols were modified for this experiment[221]. HEK293T cells were treated with 2 mM thymidine (Sigma-Aldrich) to cause an S-phase arrest. After seventeen hours the thymidine was removed and the cells were released in DMEM with fetal bovine serum (media) and transfected with the indicated APOBEC3-eGFP constructs (Transit-LT1, Mirus). After eight hours the media was removed and a second thymidine block was added. After another seventeen hour incubation the thymidine was removed and the cells were released into media. After four hours cells were harvested for S-phase or treated with 50 ng/mL of nocodazole (Sigma-Aldrich). After sixteen hours the nocodazole treated cells were released into media for 1 hour and then harvested for metaphase. A fraction of the cells were analyzed for cell cycle profile (see below). The rest (~ 8 million cells from each condition) were pelleted, washed and resuspended in lysis buffer (25 mM Hepes, pH 7.4, 250 mM NaCl, 10% glycerol, 0.5% Triton X-100, 1 mM EDTA, 1 mM $MgCl_2$, 1 mM $ZnCl_2$) then sonicated three times for 3 seconds. Samples were run on western blots and probed with anti-GFP (1:5000, 632 381; BD Clontech) and anti-tubulin (1:20,000, MMS-407R; Covance). The DNA deamination assay has been described[128]. Lysates were mixed with 6-FAM labeled 43 nucleotide containing oligo containing a TTCC deamination site for two hours at 37 C before addition of uracil DNA glycosylase

and NaOH to create and break an abasic site. The samples were run on 15% acrylimide urea gels for separation and analyzed with a Fuji-FLA 5000 scanner.

Cell cycle profiling experiments

HeLa or HEK293T cells were plated into 6 well plates at 200,000 cells/well and transfected the next day with 300 (A3A, A3B, A3C, A3F, A3G) or 400 ng (A3D, A3H and AID) of eGFP constructs. The cells were harvested and fixed with 4% PFA for cell cycle analysis or lysed in loading buffer for western blots. The cell cycle samples were treated with 0.1% Triton X 100, 20 μg/mL propidium iodide and 40 μg/mL RNase A (Qiagen) in PBS for 30 minutes at room temperature before flow cytometry (BD Biosciences FACS Canto II). GFP positive and GFP negative live cells were analyzed for their PI staining profiles using FloJo and GraphPad Prism. The lysates were run on 4-15% polyacrylamide gels, transferred to pvdf and then blotted with anti-GFP (1:5000, 632 381; BD Clontech) and anti-tubulin (1:20,000, MMS-407R; Covance).

Specific Contributions

We thank J. Mueller for H2B-eGFP construct, M.B. Burns and A. Green for contributions to antibody development, E. Refsland and A.M. Land for assistance with cell cycle profiles, and M. Carpenter for advice with activity assays. This work was supported by grants from the National Institutes of Health R01 AI064046 and P01 GM091743. L. Lackey was supported in part by an NSF pre-doctoral fellowship and subsequently by a position on the Institute for Molecular Virology Training Grant NIH T32 AI083196.

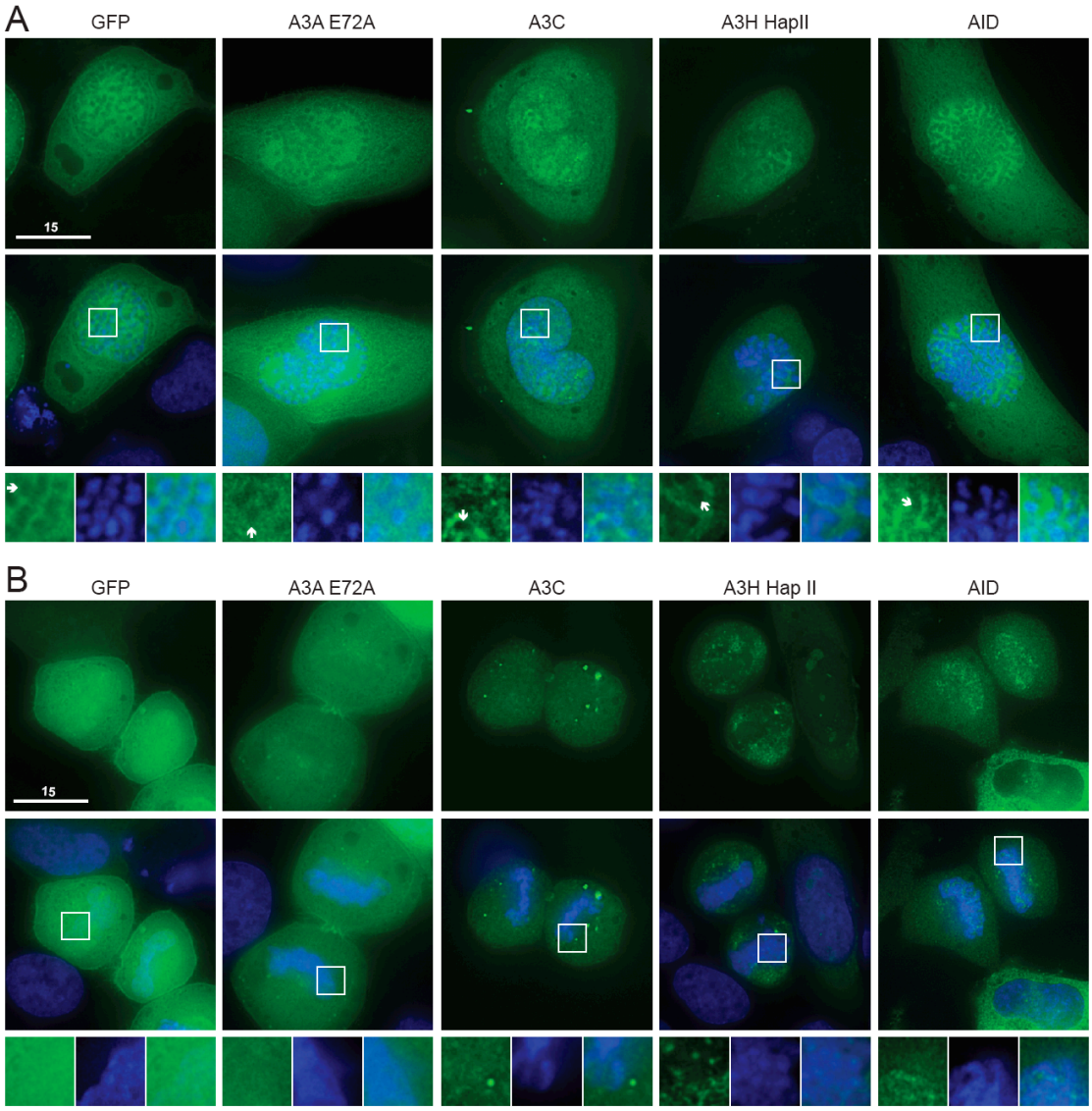

Figure 3-1. A3A, A3C, and A3H are excluded from DNA as the chromosomes condense, but become cell-wide during telophase. (A) Images of HeLa cells in prophase expressing the indicated APOBEC3-eGFP constructs (top) merged with Hoechst stain to visualize the nuclei (merge, middle). Boxed regions (bottom) are magnified below each image with APOBEC3-eGFP exclusion indicated (white arrows). **(B)** Images of HeLa cells in telophase expressing indicated APOBEC3-eGFP constructs (top), merged with nuclear stain (middle) and magnified (bottom). All images are

representative of at least three mitotic cells. See Fig. 3-S2 for APOBEC3-eGFP localization during metaphase and anaphase and Movies 13 and 14 for time-lapse images of A3A-E72A-mCherry localization during mitosis.

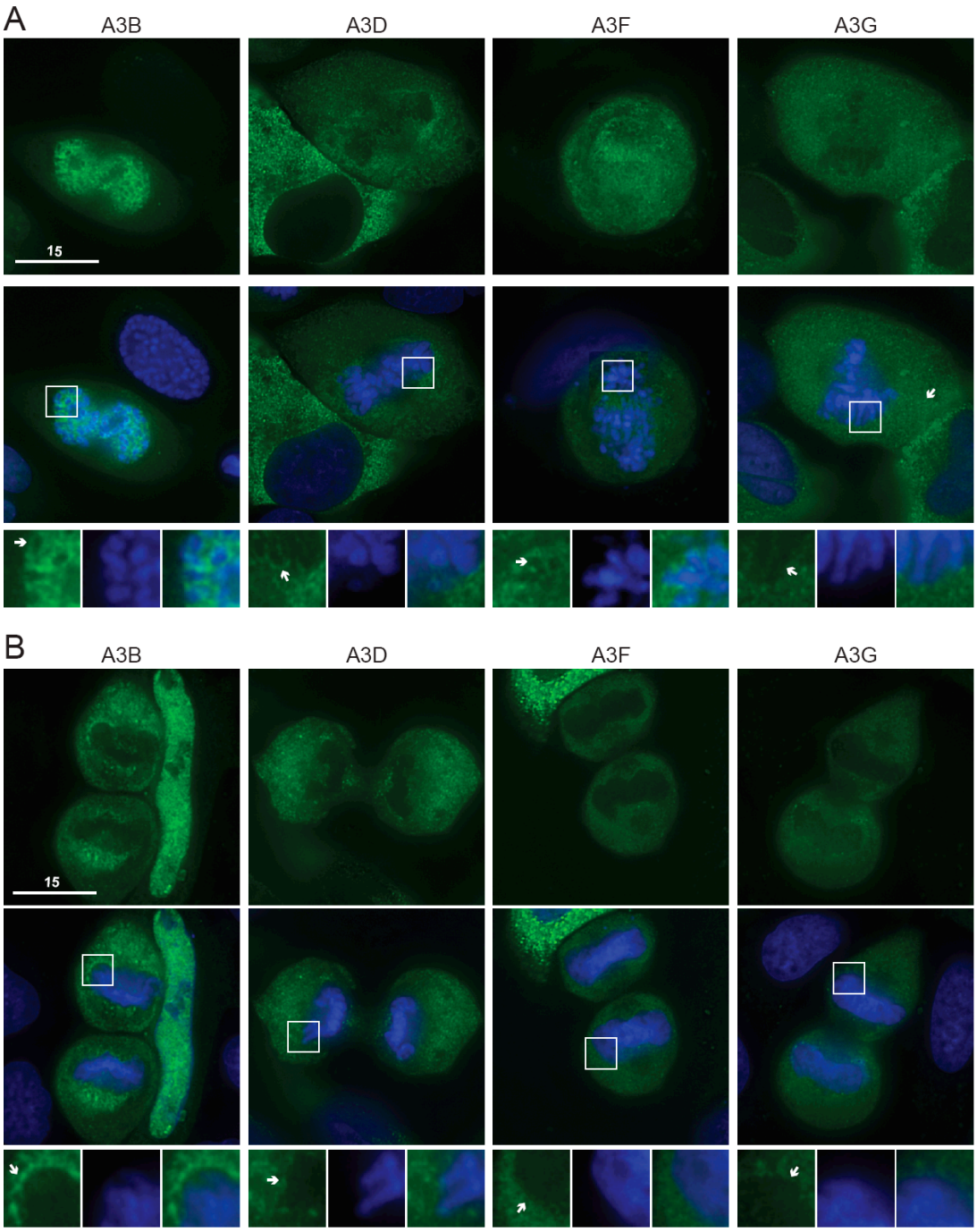

Figure 3-2. A3B, A3D, A3F, and A3G are excluded from DNA during cell division. **(A)** Images of HeLa cells in prophase expressing the indicated APOBEC3-eGFP constructs (top). Cells were stained with Hoechst dye to identify the nuclei (merge, middle). Boxed regions (bottom) are blown up below each image with APOBEC3

exclusion indicated (white arrows). **(B)** Images of HeLa cells in telophase expressing indicated APOBEC3-eGFP constructs (top), merged with nuclear stain (middle) and magnified (bottom). All images are representative of at least three mitotic cells. See Fig. 3-S4 for APOBEC3-eGFP localization during metaphase and anaphase and Movies 15 and 16 for time-lapse images of A3F-mCherry localization during mitosis.

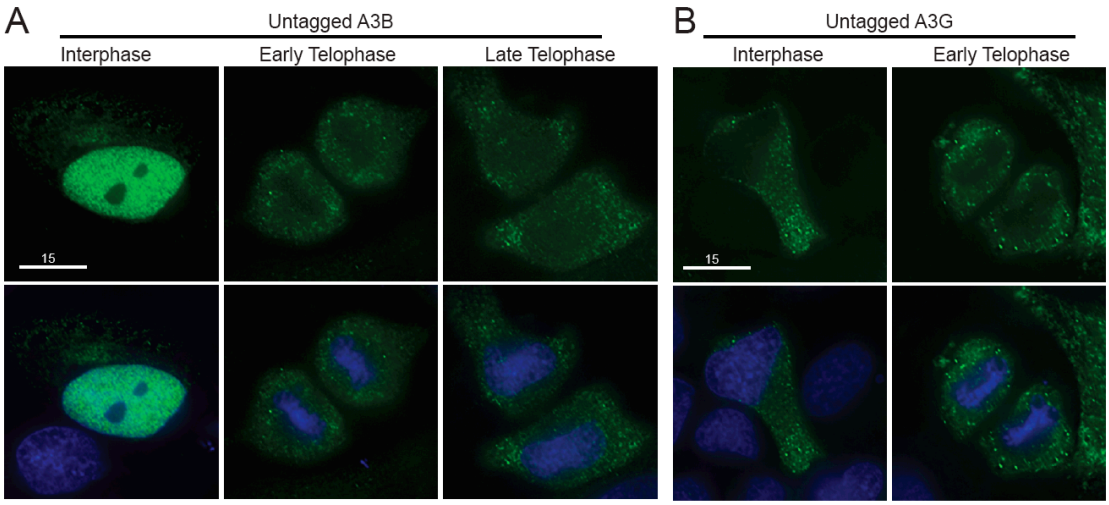

Figure 3-3. Untagged A3B and A3G are excluded from genomic DNA in the same manner as HA and eGFP tagged derivatives. (A) Images of HeLa cells expressing untagged A3B labeled with anti-A3B and anti-rabbit FITC (top) then stained with Hoechst dye to illuminate the DNA (bottom). The indicated progression through telophase (early and late) was based on chromatin condensation. **(B)** Images of HeLa cells expressing untagged A3G labeled with anti-A3G and anti-rabbit FITC (top) and DNA stain as before (bottom). Representative images are based on at least three telophase cells.

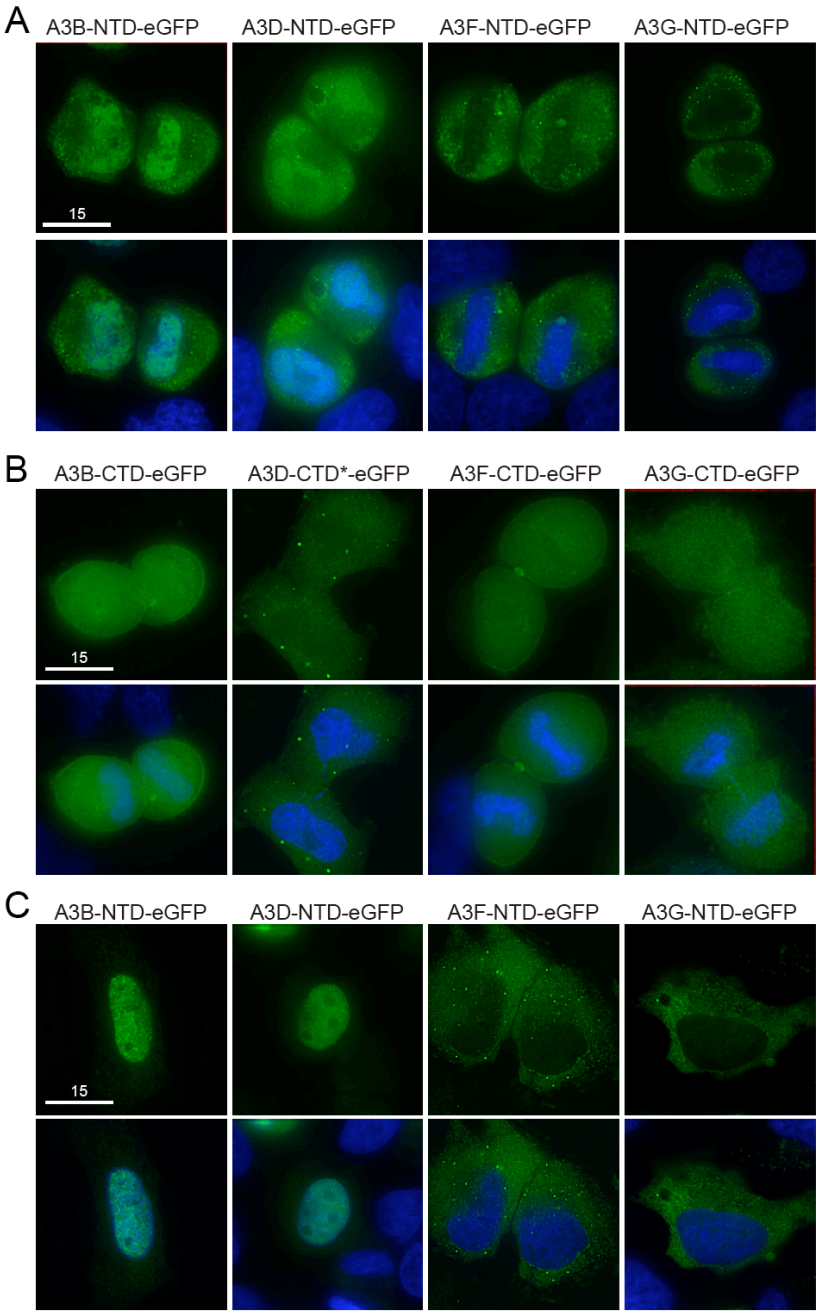

Figure 3-4. Localization of single domain variants of A3B, A3D, A3F, and A3G. (A) Images of HeLa cells in telophase expressing N-terminal or **(B)** C-terminal halves of the indicated APOBEC3-eGFP constructs are shown. Merged images with nuclear stain

below. Wild-type A3D was toxic, so a catalytic mutant was used (A3D-CTD-E264A; indicated by *). All images are representative of at least three telophase cells. **(C)** Representative images of HeLa cells in interphase expressing the N-terminal halves of the indicated APOBEC3-eGFP constructs.

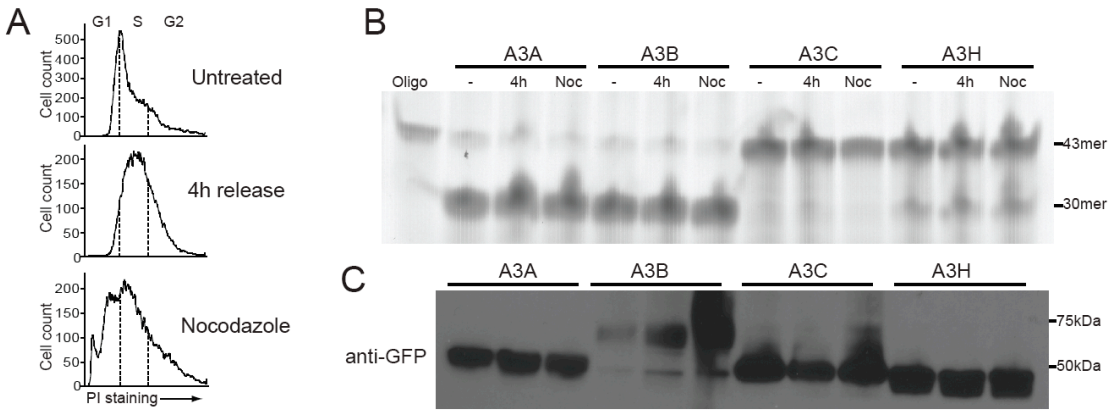

Figure 3-5. A3A, A3B, and A3H are active deaminases in G1 and S phase, while A3C is not active. (A) Representative cell cycle profile of HEK293T cells expressing A3A, A3B, A3C, or A3H-eGFP constructs used to make lysate for deaminase activity assays. **(B)** Lysates were incubated with a fluorescently labeled oligonucleotide containing a cytosine deamination motif. Oligonucleotides were run on a denaturing gel to separate uncleaved from cleaved products. **(C)** Immunoblot of protein lysates probed with anti-GFP or anti-tubulin.

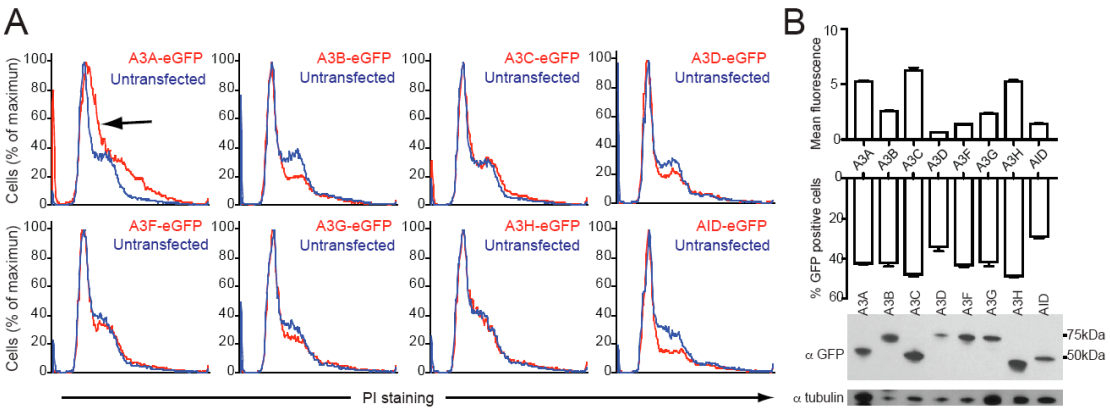

Figure 3-6. APOBEC3 effects on cell cycle progression in HEK293T cells. (A) HEK293T cells transiently transfected with eGFP tagged APOBEC3 proteins were monitored over several days by flow cytometry for PI staining to determine DNA content. Representative profiles from triplicate independent samples are shown at 48 hours. APOBEC3-eGFP expressing cell profiles (red) are overlaid on untransfected cells in the same population (blue). The shift toward S-phase in A3A-eGFP expressing cells is indicated (black arrow). **(B)** Expression of the APOBEC3-eGFP proteins by western blot (left) and mean fluorescence and transfection efficiency by flow cytometry (right) in HEK293T cells.

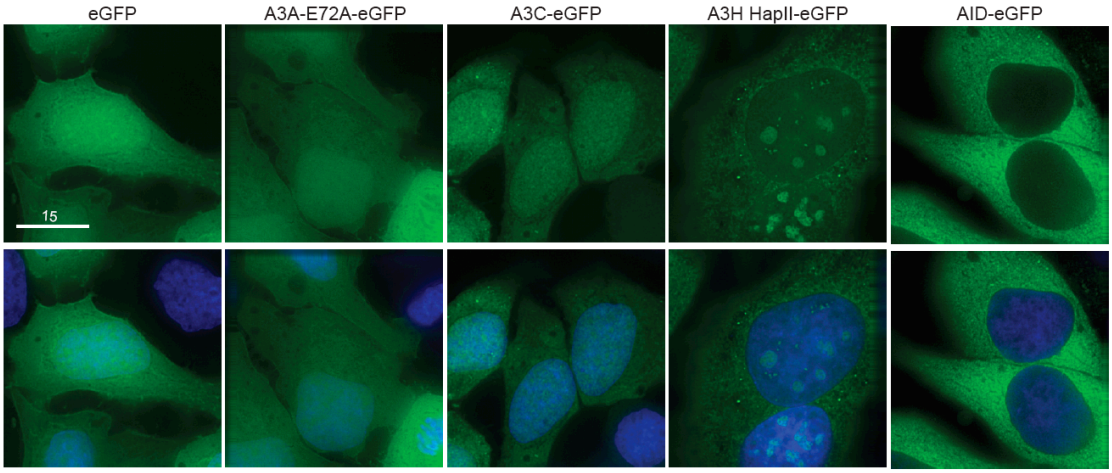

Figure 3-S1. Interphase localization of the single-domain APOBEC3 proteins and AID. HeLa cells were transfected with the indicated eGFP-tagged constructs and imaged by fluorescent microscopy (top). The DNA is indicated by Hoechst staining (merge, bottom).

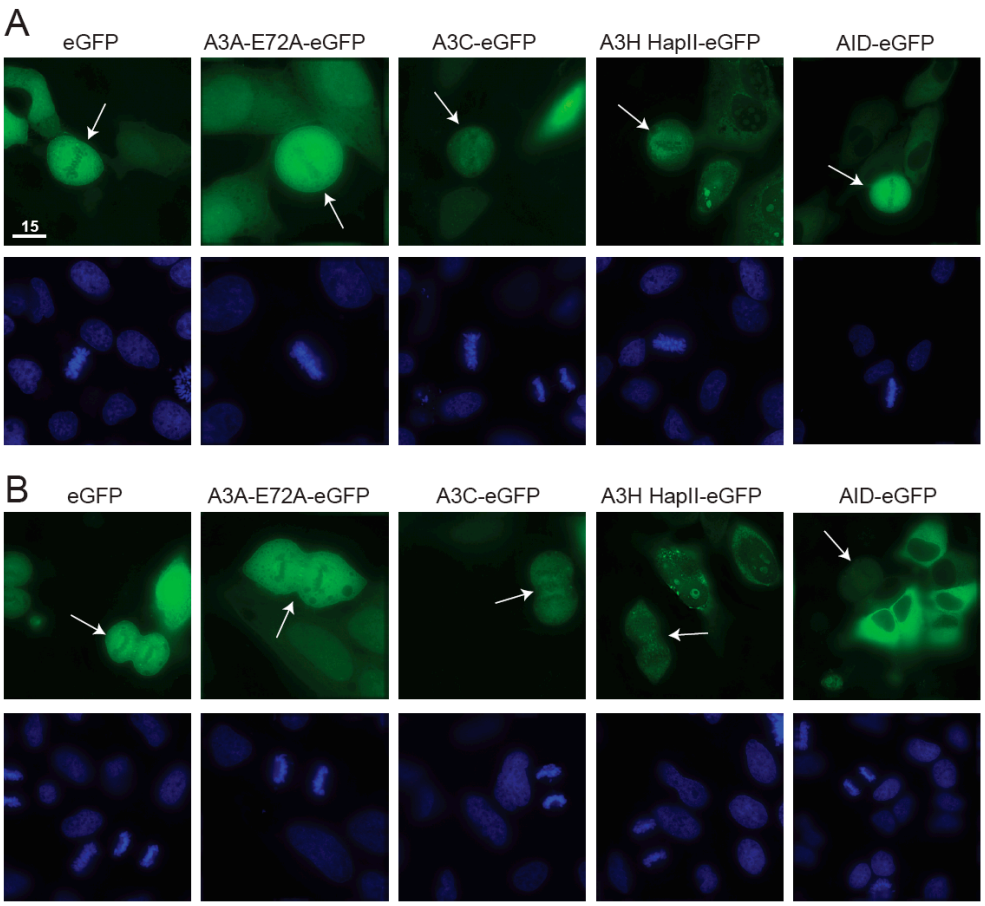

Figure 3-S2. A3A-E72A, A3C, and A3H are excluded during anaphase and metaphase. (A) HeLa cells in metaphase expressing the indicated eGFP-tagged constructs (top). Cells were stained with Hoechst dye to identify the nuclei (bottom). **(B)** HeLa cells in anaphase expressing eGFP tagged APOBEC3s (top) and nuclear stain (bottom). Arrows indicate mitotic cells. All images are representative of at least three mitotic cells.

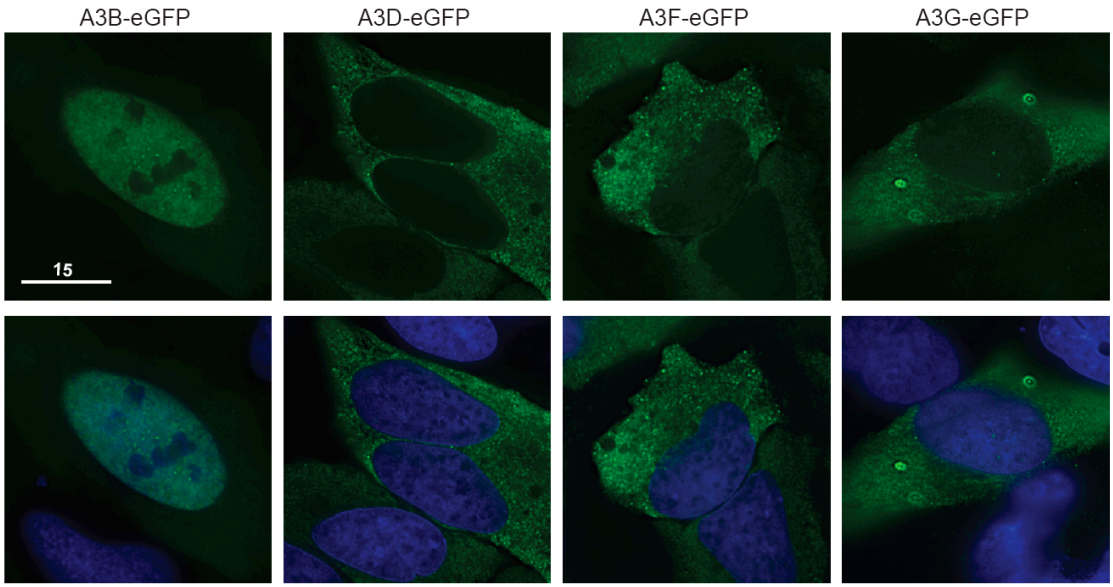

Figure 3-S3. Interphase localization of the double-domain APOBEC3-eGFP proteins. HeLa cells expressing the indicated APOBEC3-eGFP tagged constructs (top) were imaged by fluorescent microscopy. DNA is indicated by Hoechst staining (bottom).

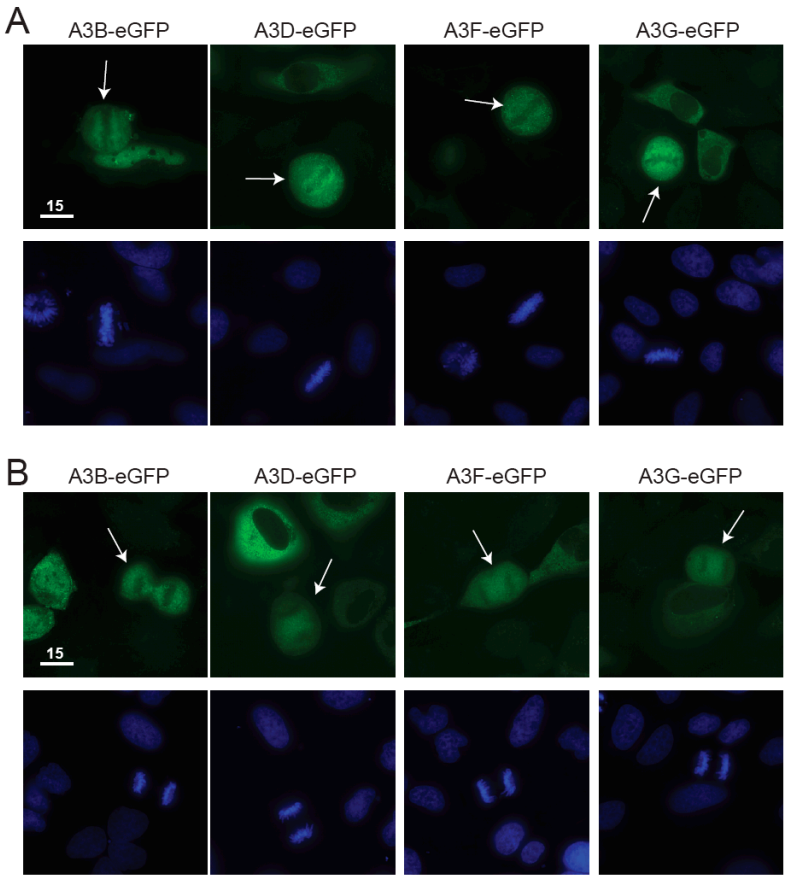

Figure 3-S4. A3B, A3D, A3F, and A3G are excluded from DNA during metaphase and anaphase mitosis. (A) Images of HeLa cells in metaphase expressing the indicated eGFP-tagged constructs (top). Cells were stained with Hoechst dye to identify the nuclei (bottom). **(B)** Images of HeLa cells in anaphase expressing eGFP tagged APOBEC3s (top) and nuclear stain (bottom). Arrows indicate mitotic cells. All images are representative of at least three mitotic cells.

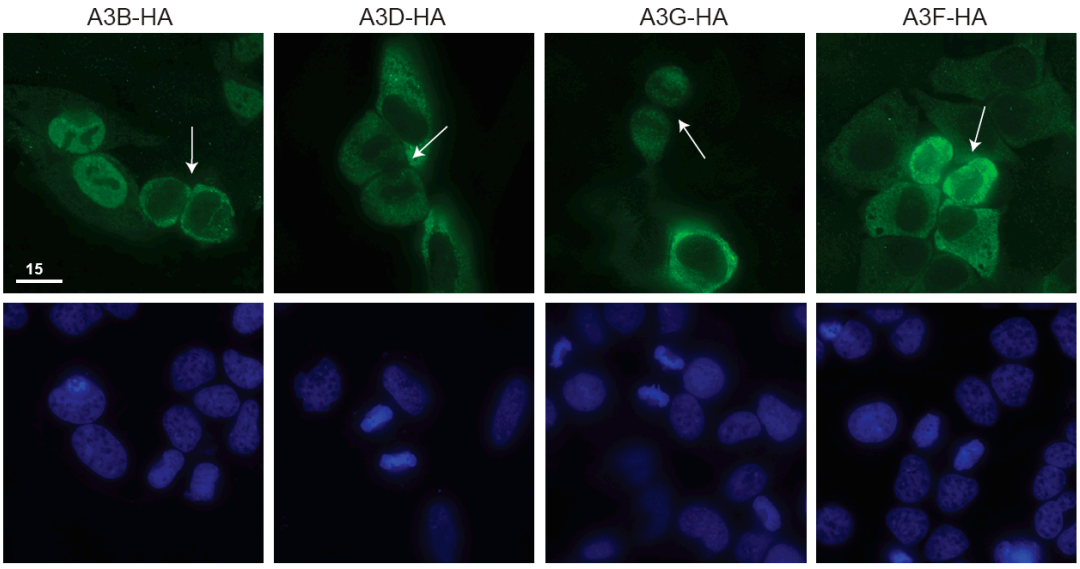

Figure 3-S5. Double domain HA tagged APOBEC3 proteins are excluded from telophase DNA. Telophase HeLa cells were imaged from anti-HA/anti-mouse FITC APOBEC3-HA expressing cells (top). The nuclei were stained with Hoechst dye (bottom). Representative images are shown from at least three telophase cells.

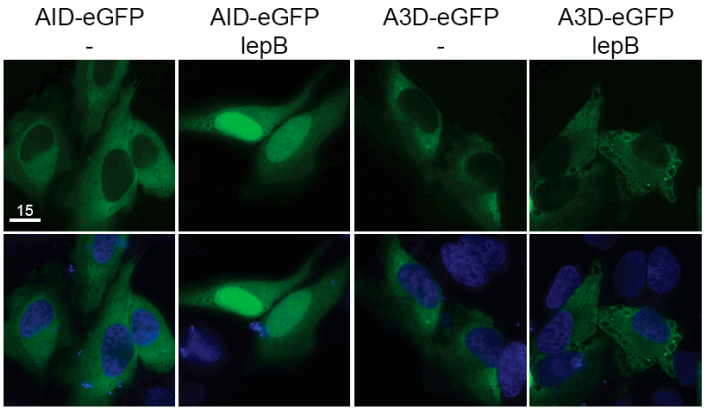

Figure 3-S6. APOBEC3D does not shuttle through the CRM1 export pathway. HeLa cells expressing the indicated constructs were treated with either ethanol or leptomycin B in ethanol for three hours before imaging (top). Nuclei are indicated by Hoechst staining (merge, bottom). Images are representative of at least ten cells.

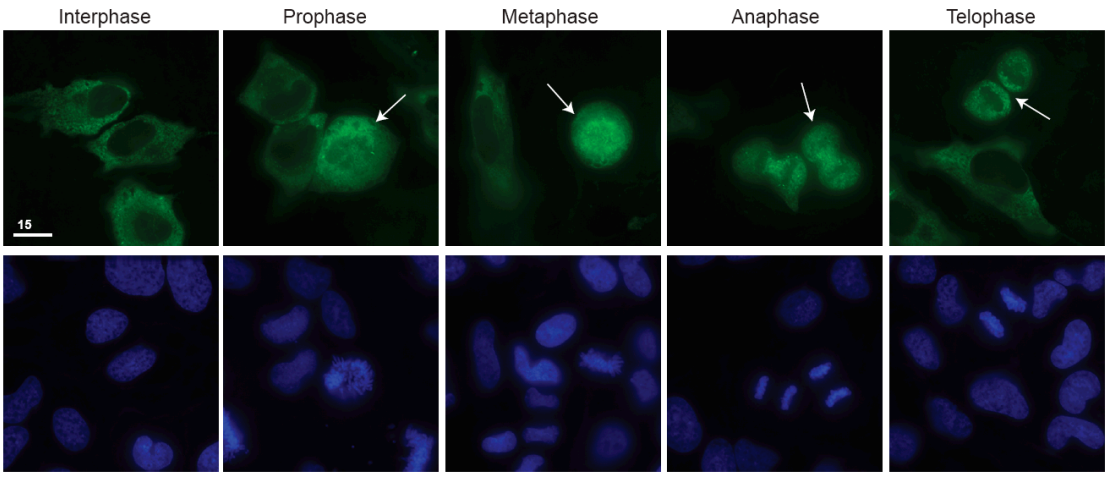

Figure 3-S7. Localization of A3C-A3C-eGFP. Interphase and mitotic stages were imaged for A3C-A3C-eGFP expressed in HeLa cells (top) with DNA stain indicated (bottom). Arrows specify mitotic cells. All images are representative of at least three cells.

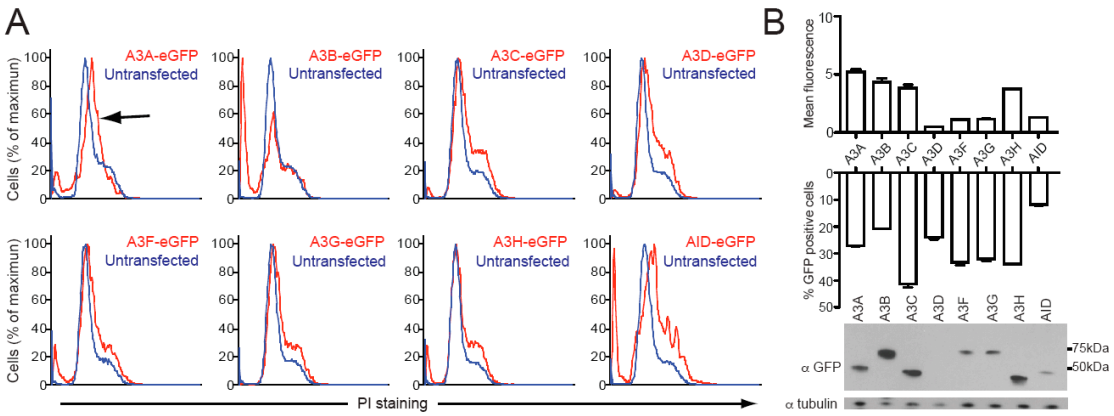

Figure 3-S8. APOBEC3 effects on cell cycle progression in HeLa cells. (A) HeLa cells transiently expressing eGFP tagged APOBEC3 proteins were monitored over several days by flow cytometry. Representative profiles from triplicate independent samples are shown at 96 hours. (B) Expression of the APOBEC3-eGFP proteins by western blot (left) and mean fluorescence and transfection efficiency by flow cytometry (right) in HeLa cells.

Movies 13-16 show HeLa cells expressing A3A E72A-mCherry or A3F-mCherry with or without H2B-eGFP and are available upon request.

Movie 13. A3A-E72A-mCherry localization in a HeLa cell starting in metaphase over a 105 minute period with images taken every 3 minutes. Exclusion is indicated by the shadow of the condensed chromosomes and is visible in metaphase and anaphase. Autofocus changes can obscure this effect in some frames. Immediately upon the separation of the two cells A3A-E72A-mCherry is fully cell-wide (57 minute time-point).

Movie 14. A3A-E72A-mCherry (as in Movie 13) co-expressed with H2B-eGFP in a dividing HeLa cell. The chromosomes can be seen as the cell begins in metaphase followed by their separation in anaphase (39-42 minute time-points). Once the cells separate (57 minute time-point) their chromosomes begin to relax and they re-enter G1.

Movie 15. A3F-mCherry localization in a dividing HeLa cell over a 3 hour period with images taken every 3 minutes. This cell begins in prophase and A3F changes from cytoplasmic to more cell-wide (18 minute time-point). The exclusion of A3F from the condensed chromosomes in metaphase and anaphase is visible as a shadow. Once the cells separate (1h 48 minute time-point), A3F remains excluded from the DNA, even when the cell is re-entering G1. Over time the fluorescence fades as the cell bleaches. This movie is representative of other A3F-mCherry and other double domain APOBEC3-mCherry movies.

Movie 16. A3F-mCherry (as in Movie 15) co-expressed with H2B-eGFP in a dividing HeLa cell. The chromosomes can be seen condensing and forming the metaphase plate (such as at the 45 minute time-point) followed by their separation in anaphase (1h 36-1h 39 minute time-points). Once the cells separate (1h 48 minute time-point) their chromosomes begin to relax and they re-enter G1.

Chapter 4

APOBEC3B is an enzymatic source of mutation in breast cancer

This chapter is part of the manuscript *in review:* Burns, M. B.*, Lackey, L.*, Carpenter, M. A., Rathore, A., Land, A. M., Leonard, B., Refsland, E. W., Kotandeniya, D., Tretyakova, N., Nikas, J. B., Yee, D., Temiz, N. A., Donohue, D. E., McDougle, R. M., Brown, W. L., Law, E. K. & Harris, R. S. (2012). *Nature.* *equal contributions

Foreword

Prior to this work we were interested in A3B because of its unique nuclear localization and because we had seen ectopic expression of A3A (which is highly similar to A3B) and A3B cause cell cycle progression defects[131,205,234]. However, normal tissues do not express significant levels of A3B, and it wasn't clear where to study endogenous A3B activity until we identified high expression of A3B in breast cancers and breast cancer cell lines[71]. From there we studied the role of A3B in deamination and mutation of breast cancer cell lines and correlated breast cancer mutational load and signatures to A3B. A3B is the first APOBEC3 protein linked human cancer.

Contributions

MBB, EWR and BL generated mRNA expression profiles and correlated A3B with tumors. EKL and I performed microscopy on A3B, A3F and A3G in breast cancer cell lines. I also developed and performed fractionation on all cell lines[235] and AR performed activity assays from these lysates. MBB, MAC, DK, and NT assisted with UPLC-MS set-up and uracil quantifications to measure genomic deamination. AML performed TK fluctuations and AR generated 3D-PCR sequences for genomic mutations. I performed the cell cycle, cell death and DNA damage response analysis on the tetracycline inducible cell lines. AR and MAC performed 3D-PCR and comet assay on these lines. MAC performed deaminase assays with recombinant protein and NAT and DED contributed bioinformatic analyses linking A3B with patient breast cancer mutations. JBN conducted the bioinformatic analysis of microarray data and developed a normalization algorithm for this analysis. RSH assisted with experiment design and wrote the manuscript.

Summary

Multiple mutations are required for cancer development, and genome sequencing has revealed that several cancers, including breast, have somatic mutation spectra dominated by C-to-T transitions[228,229,231,233,236-240]. The majority of these mutations are unlikely to be spontaneous because they occur at hydrolytically disfavored[241] non-methylated cytosines throughout the genome, and they are sometimes clustered (kataegis)[228]. Here, we show that the DNA cytosine deaminase A3B is the most likely source of these mutational events. *A3B* mRNA is up-regulated in the majority of primary breast tumors and breast cancer cell lines, but only detected at low levels in normal breast tissues and mammary epithelial cell lines. Endogenous A3B protein is predominantly nuclear and the only detectable source of DNA C-to-U editing activity in breast cancer cell line extracts. Knockdown experiments show that endogenous A3B is responsible for elevated levels of genomic uracil, increased mutation frequencies, and C-to-T transitions. Furthermore, induced A3B over-expression causes cell cycle deviations, cell death, DNA fragmentation, γ-H2AX accumulation, and C-to-T mutations. The preferred deamination signature of recombinant A3B explains a major proportion (~20%) of the entire breast cancer base substitution mutation load. Our data suggest a model in which A3B-catalyzed deamination provides a chronic source of DNA damage in breast cancer that helps explain how some cancers evolve rapidly and manifest gross molecular and clinical heterogeneity. As a non-essential protein[193] with possible overlapping function in innate immunity, A3B may be an attractive marker and therapeutic target for breast cancer and, possibly, non-mammary neoplasms where C-to-T mutations and kataegis also manifest[229,231,233,237,238,240,242].

Results and discussion

Spontaneous hydrolytic deamination of DNA cytosine to uracil (C-to-U) or methyl-cytosine to thymine is a well-established pro-mutagenic process, which, by DNA replication or misrepair, can result in C-to-T transition mutations[243]. Interestingly, methyl-cytosine is 4.5-fold more labile than normal cytosine[241], yet in many cancers, such as breast, only a minority of mutations occur within methyl-CpG contexts[229,236,239]. This incongruence is compounded by reports of localized TC-to-TT biased mutation clusters in breast[228] and other neoplasms[242]. These observations strongly suggest the existence of an underlying non-spontaneous mechanism.

Most humans encode a total of eleven polynucleotide cytosine deaminase family members that might contribute to mutation in cancer – APOBEC1, activation-induced deaminase (AID), APOBEC2, APOBEC3s (A, B, C, D, F, G, and H), and APOBEC4. APOBEC2 and APOBEC4 have not shown activity. APOBEC1 and AID are confined developmentally and implicated in cancers of those tissues, liver and B lymphocytes, respectively[39,132]. We therefore hypothesized that one or more of the seven APOBEC3s may be responsible for C-to-T mutation in other human cancers. This possibility is supported by hybridization experiments suggesting *APOBEC3* up-regulation in some cancers, such as breast[10], and qPCR data demonstrating broad *APOBEC3* expression profiles[134] (**Fig. 4-S1**).

To identify the contributing APOBEC3, we quantified mRNA levels for each of the 11 family members in a panel of breast cancer cell lines (**Fig. 4-S2**). Surprisingly, only *A3B*

mRNA trended toward up-regulation. This analysis was expanded to include a total of 38 independent breast cancer cell lines. *A3B* was up-regulated by ≥3 s.d. relative to controls in 28/38 lines, with levels exceeding 10-fold in 12/38 lines (**Fig. 4-1A and Table 4-S1**). MDA-MB-453, MDA-MB-468, and HCC1569, representative lines used below, showed 20-, 21-, and 61-fold up-regulation, respectively. These results correlate with cell line microarray data (p=0.027; **Fig. 4-1B**). No positive correlation was evident for any other deaminase family member (**Fig. 4-S2**). *A3B* up-regulation is most likely due to an upstream signal transduction event because it is not a frequent site of rearrangement or copy number variation (http://dbCRID.biolead.org) and sequencing failed to reveal promoter activating mutations or CpG islands indicative of epigenetic regulation (**Fig. 4-S3**).

Epitope-tagged A3B localizes to the nucleus of several transfected cell types[205]. To ask whether this is also a property of breast cancer cell lines, an A3B-eGFP construct was transfected into MDA-MB-453, MDA-MB-468, and HCC1569. Live cell images of A3B-eGFP showed nuclear localization, in contrast to the cytoplasmic A3F-eGFP or A3G-eGFP (**Fig. 4-1C and 4-S4**). Corroborating data were obtained for HA-tagged proteins in fixed breast cancer cell lines (**Fig. 4-S4**). To study *endogenous* A3B activity and subcellular compartmentalization, we used a fluorescence-based DNA C-to-U assay. We first found that nuclear fractions of several breast cancer cell lines contain DNA editing activity, which could be ablated by shRNA knockdown of *A3B* (**Fig. 4-1D, E and 4-S5**). Identical results were obtained with an independent shA3B construct (not shown). Protein extracts were then used to assess endogenous A3B's local dinucleotide deamination

preference. Similar to retroviral hypermutation signatures caused by A3B over-expression[83], endogenous A3B showed a strong preference for editing cytosines in the TC dinucleotide context (**Fig. 4-1F**). No deaminase activity was observed for extracts from MCF10A (A3B-low epithelial line) or SK-BR-3 (A3B-null cancer line), although it could be conferred by transfecting an A3B expression construct (**Fig. 4-1F and 4-S6**). Only A3B-eGFP and A3A-eGFP elicited measurable TC-to-TU activity in lysates from transfected HEK293T cells (**Fig. 4-S7**). Since *A3A* is myeloid lineage-specific[128] and non-detectable in breast cancer cell lines (**Figs. 4-S1 and 4-S2**), our studies demonstrate that, of the entire APOBEC/AID family, A3B is the only enzyme appropriately positioned to deaminate breast cancer genomic DNA.

To address whether endogenous A3B damages genomic DNA, we employed a combination of biophysical and genetic assays. We first used a mass spectrometry-based approach to quantify levels of genomic uracil in MDA-MB-453 and HCC1569 with high levels of endogenous A3B (shControl) versus knock-down levels of A3B (shA3B) (**Fig. 4-2A**). Genomic uracil loads decreased by 40% in HCC1569 expressing shA3B, corresponding with weaker knockdown, and by 70% in MDA-MB-453 expressing shA3B, which had stronger knockdown (**Fig. 4-2B and C**). Although these relative differences may seem modest, 20 and 10 uracils per Mbp, respectively, this equates to approximately 60,000 and 30,000 A3B-dependent uracils per haploid genome. These values may underestimate the actual number of A3B-catalyzed pro-mutagenic lesions because mismatch repair and several base excision repair pathways undoubtedly work to counteract this damage.

Second, we used a thymidine kinase-positive (TK^{plus}) to TK^{minus} fluctuation analysis[128] to determine whether up-regulated *A3B* and elevated uracil loads lead to higher levels of mutation (**Fig. 4-2D**). MDA-MB-453 and HCC1569 cells were engineered to express herpes simplex virus *TK*, which confers sensitivity to the drug ganciclovir. TK^{plus} lines were transduced with shA3B or shControl constructs and limiting dilution was used to generate single cell shA3B and shControl sub-clones, respectively (**Fig. 4-2E**). Expanded sub-clones were subjected to ganciclovir selection, and resistant cells were grown to visible colonies. Colony counts enabled the median mutation frequencies to be determined, which revealed that cells with up-regulated A3B accumulate 3-to-5-fold more mutations (**Fig. 4-2F**).

Third, 3D-PCR[128,133] was used to ask whether C/G-to-T/A transition mutations accumulate at three genomic loci from cells transduced with shA3B and shControl viruses resulting in $A3B^{low}$ and $A3B^{high}$ pools of HCC1569 cells. This technique enables qualitative estimates of genomic mutation within a population of cells because DNA sequences with higher T/A content amplify at lower denaturation temperatures than parental sequences. Lower temperature amplicons were detected for *TP53* and *c-MYC*, but not *CDKN2B* (**Fig. 4-2G**). Individual low temperature amplicons were cloned and sequenced, and more C-to-T transition mutations were observed in $A3B^{high}$ in comparison to $A3B^{low}$ samples (**Fig. 4-2H**). Other types of base substitution mutations were rare. Some C-to-T transitions were still evident in the $A3B^{low}$ samples, possibly due to residual deaminase activity and/or amplification of spontaneous events.

Next, we asked whether A3B triggers additional hallmarks of cancer[244]. We first tried and failed to stably express A3B in MCF10A and HEK293 cells. To circumvent toxicity, we constructed a panel of HEK293 clones with doxycycline (Dox)-inducible A3B, A3B-E68A-E255Q, A3A, or A3A-E72A eGFP fusions. As measured by flow cytometry, A3-eGFP levels were barely detectable without Dox and induced in nearly 100% of cells with Dox (**Fig. 4-3A**). A3A over-expression caused rapid S-phase arrest, cytotoxicity, and γ-H2AX focus formation, as reported[131] (**Fig. 4-3B to F**). In comparison, A3B induction caused a delayed cell cycle arrest, a more pronounced formation of abnormal anucleate and multinucleate cells, and eventual cell death (**Fig. 4-3B to E**). A3B induction also caused γ-H2AX focus formation, DNA fragmentation, as evidenced by visible comets, and C-to-T mutations (**Fig. 4-3E to H**).

We next asked whether our cell-based results could be extended to primary tumors. First, we quantified mRNA levels for each of the 11 family members in 21 randomly chosen breast tumor specimens, in parallel with matched normal tissue procured simultaneously from an adjacent area or the contralateral breast. Only *A3B* was expressed preferentially in tumors (p=0.0003) (**Fig. 4-S8**). We confirmed this analysis by measuring *A3B* levels in 31 additional tumor/normal matched tissue sets. In total, *A3B* was up-regulated by ≥3 s.d. in 20/52 tumors in comparison to the patient-matched normal tissue mean, and in 44/52 tumors in comparison to the reduction mammoplasty tissue mean (**Fig. 4-4A**, $p=7.1x10^{-7}$ and $p=2x10^{-5}$; **Table 4-S2** for patient information). These values, though highly significant and broadly reflected by microarray data (**Supplementary Discussion, Fig. 4-**

S9 and Tables 4-S3 to 4-S10), are underestimates because tumor specimens have varying fractions of non-*A3B* expressing normal cells. Some of the matched 'normal' samples may also be contaminated by tumor cells, as judged by comparisons to mean levels in mammoplasty samples (**Fig. 4-4A**; p=0.002). The related deaminase, *A3G*, was not expressed differentially in the same tumor panel, indicating that these observations are not due to immune cells known to express multiple A3s[134] (**Fig. 4-4B**; p=0.591).

Finally, we determined the impact of A3B on the breast tumor genome by correlating recombinant A3B's deamination signature *in vitro* and the somatic mutation spectra accumulated during tumor development *in vivo*. Using a series of single-stranded DNA substrates varying only at the immediate 5' or 3' position relative to the target cytosine (underlined), we found that recombinant A3B prefers T<u>C</u>>C<u>C</u>>G<u>C</u>=A<u>C</u> (**Fig. 4-S10,**; similar to endogenous A3B in **Fig. 4-1F**) and <u>C</u>A=<u>C</u>G=<u>C</u>T>C<u>C</u> (**Fig. 4-4C**). These local sequence contexts were then compared to those for C-to-T transitions reported for breast[228,239], melanoma[245], liver[246], and lung[247] tumors. Consistent with non-spontaneous origins, C-to-T transition loads are much greater in melanoma (~80%) and breast (~40%) than liver (~20%) and lung (~20%) tumors (**Fig. 4-4D**). The local sequence contexts for C-to-T transitions are even more striking, with flanking T and C in melanoma (T<u>C</u>C) and T and A in breast cancer (T<u>C</u>A) (**Fig. 4-4E and 4-S11**). The melanoma pattern is expected due to error-prone DNA synthesis (A insertion) opposite UV-induced pyrimidine dimers. In contrast, the preferred context of C-to-T transitions in two independent breast cancer somatic mutation data sets, one including kataegis, mirrors the *in vitro* preference of recombinant A3B (**Fig. 4-4E and 4-S11**).

Taken together, we conclude that A3B is the only DNA deaminase family member with expression and activity profiles consistent with a role in deaminating the breast cancer genome and causing the reported C-to-T mutation biases and localized kataegis events[228,229,236,239]. Other mutational patterns are evident in breast cancers suggesting additional processes at work[228]. However, some of these other patterns may also be due to A3B activity via further processing by 'repair' enzymes into transitions, transversions, and even DNA breaks that may precipitate larger-scale rearrangements (**Fig. 4-S12**). Future work is needed to understand the regulation of *A3B* and its interplay with other oncogenes and tumor suppressors. For example, a possible mechanistic linkage between *A3B* and *TP53* is suggested by a near-significant correlation (p=0.071) between *A3B* up-regulation and *TP53* inactivation in the ATCC breast cancer cell line panel, whereas other common markers do not correlate (**Tables 4-S1, 4-S2 and Fig. 4-S13**). *TP53* inactivation helps reconcile our observations that many tumor cells and cell lines are able to tolerate up-regulated A3B, whereas other cell lines succumb to toxic genomic DNA damage upon enforced A3B over-expression (*e.g.*, **Fig. 4-3**).

Although the A3B-dependent mechanism described here may impact the majority of breast cancers, this is not always the case as an *A3B* deletion[193] is homozygous in 2/38 breast cancer cell lines and 1/52 tumors (**Figs. 4-1 and 4-4**). This natural variation creates opportunities for clinical studies. For instance, one study reported a *non-significant* negative correlation with *A3B* status and breast cancer in Japan[248], where *A3B* deletions are prevalent[193]. Large cohorts will be needed to address whether *A3B* status and/or

expression levels correlate with molecular and clinical features of breast cancer. It will be interesting to ask whether the 2-to-3-fold higher incidence of breast cancer in the United States versus Japan[249] is partly attributable to the differential prevalence of A3B.

We provide the first direct evidence for active involvement of the DNA deaminase A3B in breast cancer. Conceptually supportive of the original mutator hypothesis[250], A3B-catalyzed genomic DNA deamination could provide a major source of genetic fuel for cancer development, metastasis, and even resistance to therapy. We propose that A3B is a dominant underlying factor that contributes to tumor heterogeneity by broadly affecting multiple pathways and phenotypes. A3B represents a new marker for breast cancer and a strong candidate for targeted intervention, especially given its non-essential nature[193]. A3B inhibition in evolving tumors may decrease mutation rates and thereby stabilize the cellular targets of existing therapeutics and anti-cancer immune responses (*e.g.*, A3B inhibitors used in combination with therapies that are frequently undermined by resistance mutations).

Supplemental Discussion

Why has *A3B* eluded identification as an oncogene prior to this study? The most likely explanation is that the *A3B* gene shares a high level of sequence identity (in some regions nearly 100%) with the 10 other APOBEC family members. Therefore, the short oligonucleotides used as probes on microarrays are not capable of identifying any single *APOBEC*, simply an overall total for different cross-hybridizing mRNA species. This issue is illustrated in tabular format in **Tables 4-S3 to 4-S10**. For instance, the commonly used Affymetrix Genechip Human Genome Array U133A has 11 probes intended for *A3B* detection (**Table 4-S3 and 4-S5**). Of these probes, *nine are not specific*, with 22/25 or 23/25 nucleotides identity to *A3A* and/or *A3G*. Similar non-specificities (and even complete off-target designs) were evident for the other *APOBEC3* probe clusters (**Tables 4-S3 to 4-S10**).

Nevertheless, with knowledge of these limitations, useful information can still be derived from published microarray data sets. In particular, robust comparisons with microarray data become possible for breast cancer cell lines, which are clonal and do not express *A3A* (this gene is only expressed in myeloid lineage cells[106,128,134,251]) (**Fig. 4-S1 and 4-S2**). A strong, positive correlation is evident between our *A3B* qPCR measurements and reported microarray values for *A3B* in the ATCC breast cancer cell line panel (**Fig. 4-1B**; Cancer Cell Line Encyclopedia, http://www.broadinstitute.org/ccle/home).

However, the situation is more complex for microarray studies of human neoplasms, which are invariably a montage of tumor and multiple surrounding/infiltrating normal cell types. Moreover, depending on the stringency of hybridization and the

particular sample being analyzed, *A3A* and *A3G* sequences may easily outcompete potential *A3B* target sequences (*e.g.*, *A3G* is higher than *A3B* in most samples that we analyzed; **Fig. 4-4 and 4-S8**). Regardless, in comparisons of large published microarray data sets, we were still able to detect significant *A3B* up-regulation in tumor versus normal tissues (n=285 and n=22; p-value $<10^{-6}$; **Table 4-S3**). As expected by the non-specificity of several probe sets, highly significant differences were also seen for the "*A3A*" and "*A3F,G*" probe sets, which are both predicted to cross-hybridize with *A3B* mRNA (**Tables 4-S4 and 4-S7**). In comparison, probe sets with low identity to *A3B* showed no significant correlation (*e.g.*, *A3C*; **Table 4-S6**). As shown in **Fig. 4-S9**, near-identical expression values for 62 housekeeping genes between different microarray data sets provides strong confidence that this approach is detecting over-expression of an *APOBEC3* gene in tumor versus normal samples. This situation mirrors our original hybridization results[10]. However, combined with the data sets shown here, we are confident in our conclusion that this up-regulated *APOBEC3* gene is indeed only *A3B*.

A secondary explanation for why *A3B* has proven elusive up to now is that the short read lengths generated during deep-sequencing (RNA-Seq) are difficult to assign unambiguously to members of repetitive gene families such as the *APOBEC3*s, resulting in sequence mis-assignment or exclusion at the grooming stage of bioinformatic analysis. A final explanation is that the *A3B* gene is not a hotspot for gross chromosome abnormalities (database of Chromosomal Rearrangements In Diseases[252], http://dbCRID.biolead.org), which might have been found by classical cytogenetic techniques[253] or, more recently, by deep sequencing[228,239].

Methods

Flash frozen breast tumor and matched normal tissue pairs were obtained from the University of Minnesota Tissue Procurement Facility. Samples were chosen randomly with breast cancer and available matched normal tissue being the only selection criteria. Mammary reduction samples were used as non-cancer controls. These studies were performed in accordance with IRB guidelines (IRB study number 1003E78700). The breast cancer cell line panel 30-4500K was obtained from the ATCC and cultured as recommended. RNA isolation, cDNA synthesis, and qPCR procedures were performed as reported[134] (**Table 4-S11**). Knockdown and control shRNA constructs were obtained from Open Biosystems. Microscopy, cellular fractionation and deaminase activity assays were done as described[128,205]. Genomic uracil was quantified by treating DNA samples with uracil DNA glycosylase, purifying the nucleobase from the remaining DNA and analyzing the samples by mass spectrometry. The TK assay and 3D PCR have been described and were modified for use with breast cancer cell lines[128]. Dox-inducible cells were obtained from Invitrogen and stables were created with the indicated constructs. These lines were analyzed for cell cycle arrest using propidium iodide staining and cell viability with crystal violet staining and the MTS assay. DNA damage was measured by the comet assay and by flow cytometry and microscopy of cells immunostained for γ-H2AX. Recombinant A3B195-382-mycHis was purified and used for deamination kinetics as described[136] using 5'-ATTATTATTATNCNAATGGATTTATTTATTTATTTATTTATTT-6-FAM (NCA and TCN for 5' and 3' preference experiments, respectively). The somatic single nucleotide mutation frequencies with local sequence contexts were determined from published

primary tumor genomes[228,239,245-247]. Potential mechanistic overlap with hydrolytic deamination of 5-methyl-cytosines was avoided by excluding CpG dinucleotides from mutational preference calculations.

Additional Methods

RNA isolation, cDNA synthesis, and qPCR.

Matched tumor/normal breast tumors and mammary reduction samples from the University of Minnesota TPF and breast cancer cell lines 30-4500K from the ATCC were used for RNA isolation, cDNA synthesis and qPCR as described[134]. Tissue RNA was from 100 mg flash-frozen tissue disrupted by a 2 h water bath sonication in 1 mL of Qiazol Lysis Reagent (RNeasy, Qiagen). Cell RNA was made using Qiashredder (RNeasy, Qiagen). qPCR was performed on a Roche Lightcycler 480 instrument. The housekeeping gene *TBP* was used for normalization. Statistical analyses for matched tissues were done using the Wilcoxson signed-rank test, and unmatched sets with the Mann-Whitney U-test (Graphpad Prism). Primer and probe sequences are listed in **Table 4-S11**.

Knockdown constructs.

A3B shRNA and shControl lentiviral constructs were from Open Biosystems through the BMGC RNAi Core (TRCN0000157469, TRCN0000140546, and scramble). Helper plasmids pdelta-NRF, containing HIV-1 *gag*, *pol*, *rev*, and *tat* genes, and pMDG, containing the VSV-G *env* gene, were co-transfected in HEK293T cells. Cell-free

supernatants were harvested and concentrated by centrifugation (14,000 g x 2 h). Stable transductants were selected with puromycin (1 μg/ml).

Cell fractionation and DNA deaminase activity assays.

Cellular fractionation was performed as described by syringe treatment of 10^7 cells in 0.5 mL of hypotonic buffer[235]. Nuclei were lysed by sonication in lysis buffer (25 mM Hepes, pH7.4, 250 mM NaCl, 10% glycerol, 0.5% Triton X-100, 1 mM EDTA, 1 mM $MgCl_2$, 1 mM $ZnCl_2$). Anti-histone H3 (1:2000; Abcam) and anti-tubulin (1:10,000; Covance) followed by anti-mouse 800 or anti-rabbit 680 (1:5000; Licor) immunoblots were used to assess fractionation. Lysates were tested in a fluorescence-based deaminase activity assay[128]. Dilutions were incubated 2 h at 37°C with a DNA oligonucleotide 5'-(6-FAM)-AAA-TTC-TAA-TAG-ATA-ATG-TGA-(TAMRA). Fluorescence was measured on SynergyMx plate reader (BioTek). Local dinucleotide preferences in extracts were analyzed similarly using 5'-AC, CC, GC, or TC at the NN position of 5'-(6-FAM)-ATA-ANN-AAA-TAG-ATA-AT-(TAMRA).

Genomic uracil quantifications.

Genomic DNA was prepared from shA3B of shControl cells transduced and cultured for 21 days. Samples were spiked with heavy (+6)-labeled uracil (C^{13} and N^{15}; Cambridge Isotopes) and treated with UDG (NEB). Uracil was purified using 3,000 MWCO columns (Pall Scientific) and SPE (Carbograph, Grace). Samples were resuspended in water containing 0.1% formic acid. Analyses were performed on a capillary HPLC-ESI+-MS/MS (Thermo-Finnigan Ultra TSQ mass spectrometer, Waters nanoACQUITY

HPLC). The MS was operated in positive ion mode, with 3.0 kV typical spray voltage, 250ºC capillary temperature, 67 V tube lens offset, and nitrogen sheath gas (25 counts). Argon collision gas was used at 1.1 mTorr. MS/MS analyses were performed with a scan width of 0.4 m/z and scan time of 0.1 s. The Hypercarb HPLC column (0.5 mm x100 mm, 5 μm, Thermo Scientific) was maintained at 40ºC and a flowrate of 15 μL/min. Solvents were 0.1% formic acid and acetonitrile. A linear gradient of 0% to 8% acetonitrile in 8 min was used, followed by an increase to 80% acetonitrile over 7 min. Uracils eluted at 11.5 min. Selected reaction monitoring was conducted with collision energy of 20V using the transitions: m/z 113.08 [M+H+]→70.08 [M-CONH]+ and m/z 96.08 [M-NH2]+ for uracil, while the internal standard ([15N-2, 13C-4]-uracil) was monitored by the transitions m/z 119.08 [M+H+]→m/z 74.08 [M-CONH]+ and m/z 101.08 [M-NH2]+ respectively. Internal standards were used for quantification.

TK fluctuations.

TK-neo was introduced into MDA-MB-453 and HCC1569 cells as described[128]. TK^{plus} cells were transduced with shA3B or shControl lentiviruses and subcloned by limiting dilution. 10^6 cells from each expanded subclone population were subjected to ganciclovir and incubated until colonies outgrew. Frequencies were determined by applying the method of the median[254].

3D-PCR and sequencing.

DNA was harvested from Ugi-expressing[255] T-REx-293 clones or HCC1569 cells transduced with shA3B or shControl lentiviruses. 3D-PCR was done using Taq (Denville

Scientific) as described[128]. Primers sequences available upon request. PCR products were analyzed by gel electrophoresis with ethidium bromide, PCR purified (Epoch), blunt-end cloned into pJET (Fermentas), sequenced with T7 primer (BMGC), and aligned and analyzed with Sequencher software (Gene Codes Corporation).

Cell cycle experiments.

T-REx-293 cells (Invitrogen) were transfected with pcDNA5/TO A3-GFP using TransIT-LT1 (Mirus) followed by clone selection using hygromycin. Cells were induced with 1 μg/mL Dox (MP Biomedicals 198955) for the indicated times then trypsinized and fixed with 4% paraformaldehyde in PBS. Cell pellets were resuspended in 0.1% Triton X 100, 20 μg/mL propidium iodide and 40 μg/mL RNase A (Qiagen) in PBS for 30 min and the DNA content and GFP induction measured by flow cytometry (BD Biosciences FACS Canto II) and analyzed with FlowJo and GraphPad Prism.

Cell viability assays.

Cells were plated into multiple 96 well plates (2500 cells/well) and measured at the days indicated. The MTS reagent and PMS reagents were used as directed (Promega, Celltiter Aq 96). Absorbance was measured at 490 nm (PerkinElmer 1420 Victor 3V). The results were normalized to untreated cells. For crystal violet staining wells of a six well plate were plated with 2 x 10^5 cells. Half of the wells were induced with 1 ug/mL Dox. A crystal violet (0.5%), methanol (49.5%), water (50%) solution was used to stain cells after seven days.

DNA damage experiments.

Flow cytometric analysis of γ-H2AX foci was adapted[256]. Fixed cells were incubated overnight in 0.2% Triton-X 100, 1 % BSA in PBS (blocking buffer) with 1:100 rabbit anti-γ-H2AX (Bethyl A300-081A). Secondary incubation was with goat anti-rabbit TRITC (Jackson 111025144) for 3 hrs before flow cytometry (BD Biosciences FACS Canto II) and analysis (FloJo and GraphPad). For microscopy, HEK293 cells were induced with 1 ug/mL of Dox before fixation with 4% paraformaldehyde and incubation with 1:50 anti-γ-H2AX conjugated to Alexa 647 (Cell Signaling 20E3) in blocking buffer for 3 hours. The cells were stained with 0.1% Hoechst dye and imaged at 20x or 60x (Deltavision) and deconvolved (SoftWoRx, Applied Precision).

Comet Assays.

As described[257], microscope slides were coated with 1.5% agarose and dried. Low melting agarose (0.5% in PBS) was combined 1:1 with HEK293T cells transfected with A3A-eGFP (1 d) or A3B-eGFP (6 d). 10,000 cells were added to coated slides and the cells were lysed overnight in 10 mM Tris, 100 mM EDTA, 2.5 M NaCl, 1% Triton X-100. Slides were incubated for 10 min in running buffer (300 mM NaOH, 1 mM EDTA pH 13.1) then run at 0.75 V/cm 30 min. Gels were neutralized with 0.4 M Tris-Cl pH 7.5 and treated with RNase A (Qiagen). The microgels were allowed to dry and comets were visualized using propidium iodide.

Microarray comparisons. Affymetrix GeneChip microarray data were reported previously by others. Tripathi *et al*.[258] (GEO ID GSE9574) and Graham *et al*.[259] (GEO ID GSE20437) reported data for 15 and 7 reduction mammoplasty samples, respectively.

Tabchy *et al*.[260] (GEO ID GSE20271) reported data for 178 stage I-III breast cancers (procured at 6 sites worldwide), and Lasham *et al*.[261] (GEO ID GSE36771) reported data for 107 primary breast tumors. NCBI GEO resources were used to obtain raw data sets for additional analyses (CEL files). Next, we used the RMA algorithm (510K FDA approved) of the Expression Console Software (Affymetrix) with the standard settings to re-analyze the data for all 307 subjects. Since data sets from multiple independent studies were used, we normalized all tumor data with respect to the normal data in order to be able to perform comparisons. More specifically, we projected all tumor data into the space of the normal data by performing a non-linear normalization employing the following mathematical function:

$$\mathbf{X_n} = \frac{\mathbf{R_n}}{\mathbf{1}+\mathbf{e}^{\left(\frac{\mathbf{X_o}-\mathbf{m}}{\mathbf{R_o}}\right)}} + \mathbf{N}_{\min}$$

(1)

In Eq. (1), X_n is the new, normalized variable; X_o is the old variable; R_n is the magnitude of the range of the new space; R_o is the magnitude of the range of the old space; m is the median of the old variable; and N_{min} is the minimum of the range of the new space. 62 housekeeping genes were used to assess these normalization methods, and a strong positive correlation was found between each independent data set (*e.g.*, **Fig. 4-S9**). Having performed the same normalization method to all *APOBEC3* genes, we were able to obtain expression data for the tumor versus normal samples (**Table 4-S3**). As previously[262-266], we assessed statistical significance using three different methods: i) t-

Test (Mann-Whitney for non-parametric variables) with the significance level adjusted to $\alpha = 0.007143$ to account for seven comparisons, ii) fold-change defined as the ratio of the mean expression of the cancer group over the mean expression of the normal group (FC=C/N), and iii) ROC AUC. We performed ROC curve analysis on all seven *APOBEC3* probe clusters to assess their discriminating power with respect to the two groups (cancer versus normal). As can be seen in **Table 4-S3**, the probe sets corresponding to *A3A*, *A3B*, and *A3(F,G)* are deemed to have significant differential expression according to all three methods.

Specific Contributions

We thank J. Hultquist for assistance with statistics, M. Li for help with DNA deaminase assays, R. Nzara for promoter sequencing, V. Polunovsky for hTERT-HMEC, V. Simon for an shRNA construct, S. Kaufmann, C. Lange, and D. Largaespada for expert consultation, and the University of Minnesota Masonic Cancer Center for purchasing the ATCC breast cancer panel. Tissues were obtained from the University of Minnesota Tissue Procurement Facility, which is part of BioNet, supported by the Academic Health Center and NIH grants P30 CA77598 (D. Yee), P50 CA101955 (D. Buchsbaum) and KL2 RR033182 (B. Blazar). M.B.B. was supported in part by a Cancer Biology Training Grant (NIH NCI T32 CA009138) and a Department of Defense Breast Cancer Research Program Predoctoral Fellowship (BC101124). L. Lackey was supported in part by an NSF Predoctoral Fellowship and by a position on the Institute for Molecular Virology Training Grant NIH T32 AI083196. A.M. Land was supported by a CIHR Postdoctoral Fellowship. Computational analyses (N.A.T, D.E.D) were supported by federal funds

from the National Cancer Institute, National Institutes of Health, CBIIT/caBIG ISRCE yellow task #09-260. The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government. Harris laboratory operational support was provided in part by the Children's Cancer Research Fund (Minneapolis, MN), the University of Minnesota College of Biological Sciences, and a seed grant from the University of Minnesota Clinical and Translational Science Institute (supported by NIH 1UL1RR033183).

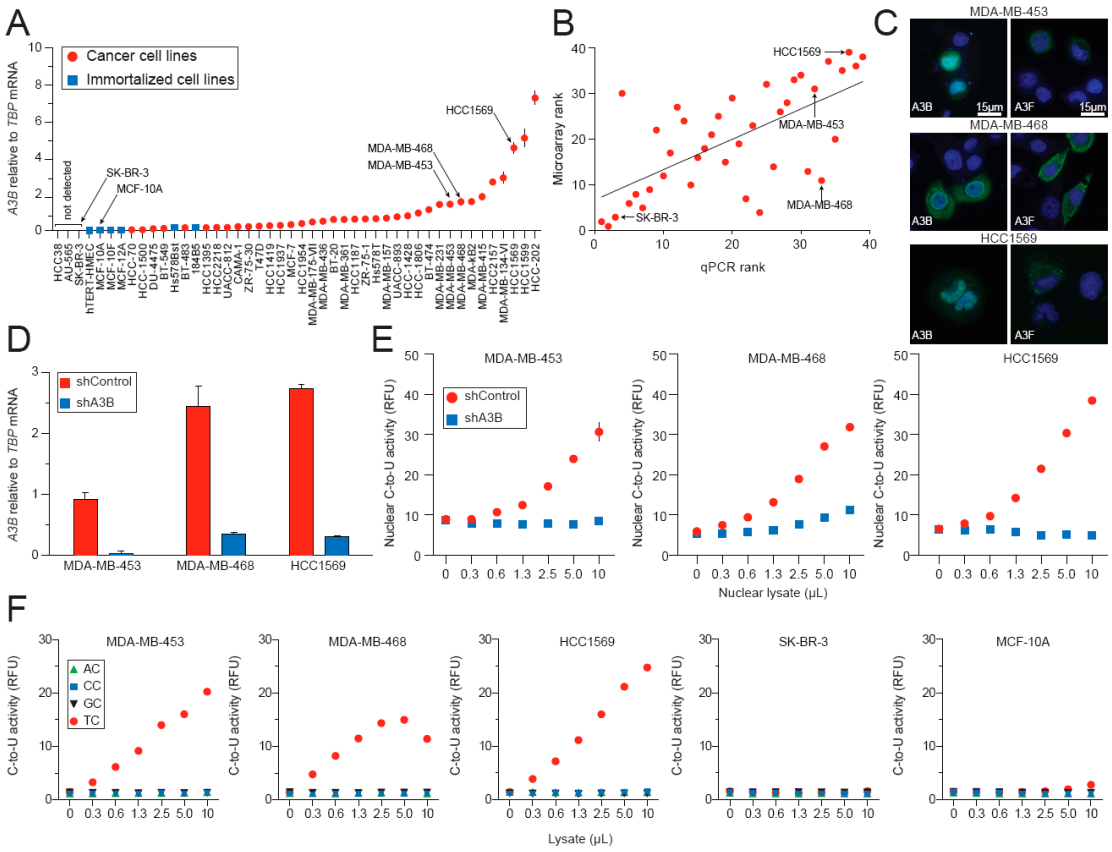

Figure 4-1. *A3B* up-regulation and activity in breast cancer cell lines. (A) *A3B* levels in the indicated breast cancer cell lines (red circles, n=40 including two sister pairs SK-BR-3/AU565 and MDA-MB-453/MDA-kb2) and non-cancerous cell lines (blue squares, n=6 including one sister pair MCF10A/F). Each data point is the mean *A3B* level of three qPCR reactions presented relative to mRNA levels of the constitutive housekeeping gene *TBP* (s.d. shown unless smaller than the data point). Data are arranged from lowest to highest *A3B* expression level. Cell lines used in mechanistic studies are labeled. **(B)** Positive correlation between *A3B* qPCR data and microarray data (n=39; see Supplementary Discussion). Cell lines used in mechanistic studies are labeled. **(C)** A3B-

eGFP (green) co-localizes with nuclear DNA (Hoescht-stained blue), whereas A3F-eGFP is cytoplasmic, in the indicated breast cancer cell lines. MDA-MB-468 shows some cytoplasmic A3B-eGFP localization, but is still predominantly nuclear. **(D)** *A3B* mRNA levels in the indicated breast cancer cell lines stably transduced with shControl or shA3B lentiviruses. **(E)** Nuclear DNA C-to-U activity in extracts from the indicated breast cancer cell lines transduced as in (d) (n=3). s.d. shown unless smaller than data point. **(F)** Intrinsic dinucleotide DNA deamination preference of enzyme(s) in soluble extracts from the indicated cell lines (n=3; s.d. is smaller than each data point).

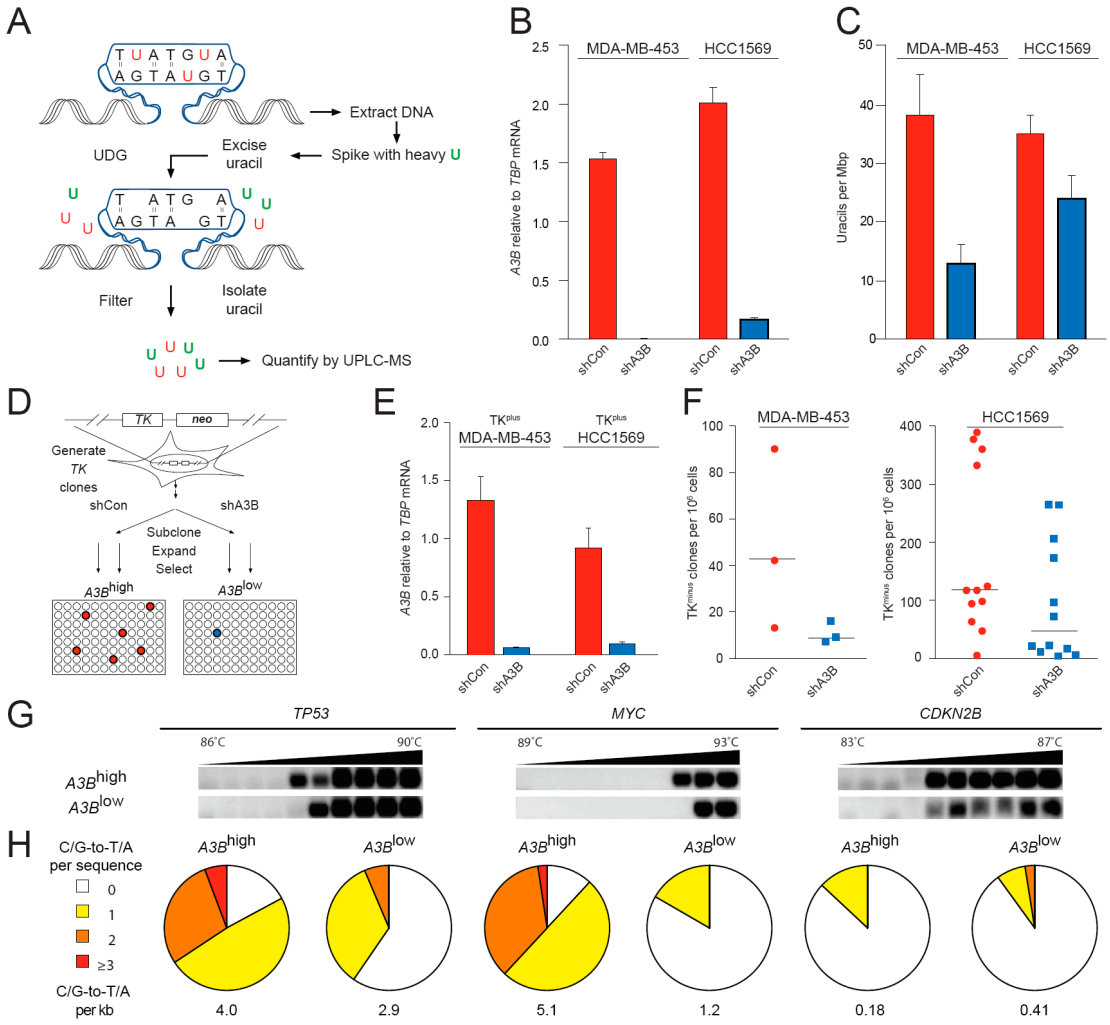

Figure 4-2. A3B-dependent uracil lesions and mutations in breast cancer genomic DNA. (A) Workflow for genomic uracil quantification by UPLC-MS. **(B)** *A3B* mRNA levels in the indicated breast cancer cell lines stably transduced with shControl or shA3B lentiviruses. **(C)** Steady-state genomic uracil loads per mega-basepair (Mbp) in the indicated breast cancer cell lines expressing shControl or shA3B constructs. **(D)** Workflow for TK fluctuation analysis. **(E)** *A3B* mRNA levels in TK^{plus} MDA-MB-453 and HCC1569 lines expressing shControl or shA3B constructs. **(F)** Dot plots depicting

the TK^{minus} mutation frequencies of MDA-MB-453 and HCC1569 subclones expressing shControl or shA3B constructs. Each dot corresponds to one subclone, and *median* values are indicated for each condition. **(G)** Agarose gel analysis of 3D-PCR amplicons obtained using primers specific for the indicated target genes and genomic DNA prepared from HCC1569 cells expressing shControl or shA3B constructs. The denaturation temperature range is indicated above each gel. **(H)** Pie charts depicting the C/G-to-T/A mutation load in 3D-PCR products after cloning and sequencing ($n \geq 35$ per condition). Charts align with target genes labeled in (G).

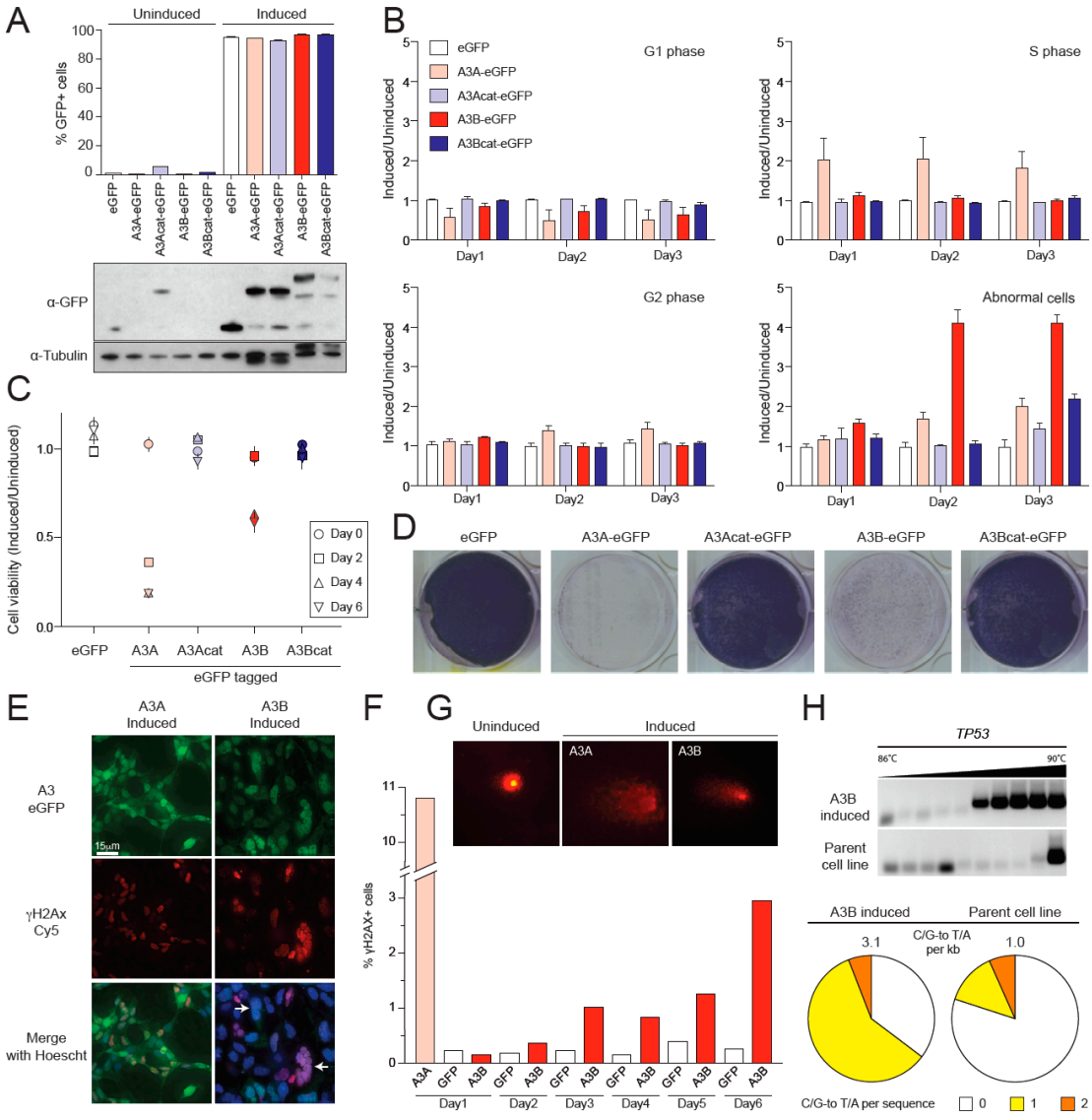

Figure 4-3. Cancer phenotypes triggered by inducing A3B over-expression. (A) The percent fluorescence for the indicated HEK293-derived A3-eGFP cell lines in absence or presence of Dox (corresponding immunoblot below). **(B)** Cell cycle status of the indicated cell lines at the indicated time points (comparisons made relative to each line uninduced). **(C)** Cell viability of the indicated cell populations using the MTS assay at the indicated time points post-Dox addition (comparisons made relative to each line uninduced). **(D)** Toxicity of induced A3A-eGFP and A3B-eGFP. Colonies were stained

with crystal violet 8 days post-Dox treatment. **(E, F)** Representative fields of cells imaged for γ-H2AX and A3A-eGFP (1 day) or A3B-eGFP (3 days) post-Dox treatment and quantitative γ-H2AX flow cytometry over several days. Abnormal, multinuclear cell clusters are typical of induced A3B-eGFP expression (highlighted by white arrows; classified as abnormal cells in 'B'). **(G)** Representative images of comets due to DNA fragmentation induced by A3A-eGFP or A3B-eGFP. **(H)** Pie diagram of C/G-to-T/A mutations in *TP53* detected by sequencing 3D-PCR products from A3A- or A3B-eGFP expressing cells, 2 or 4 days post-Dox treatment, respectively ($n \geq 12$ clones per condition).

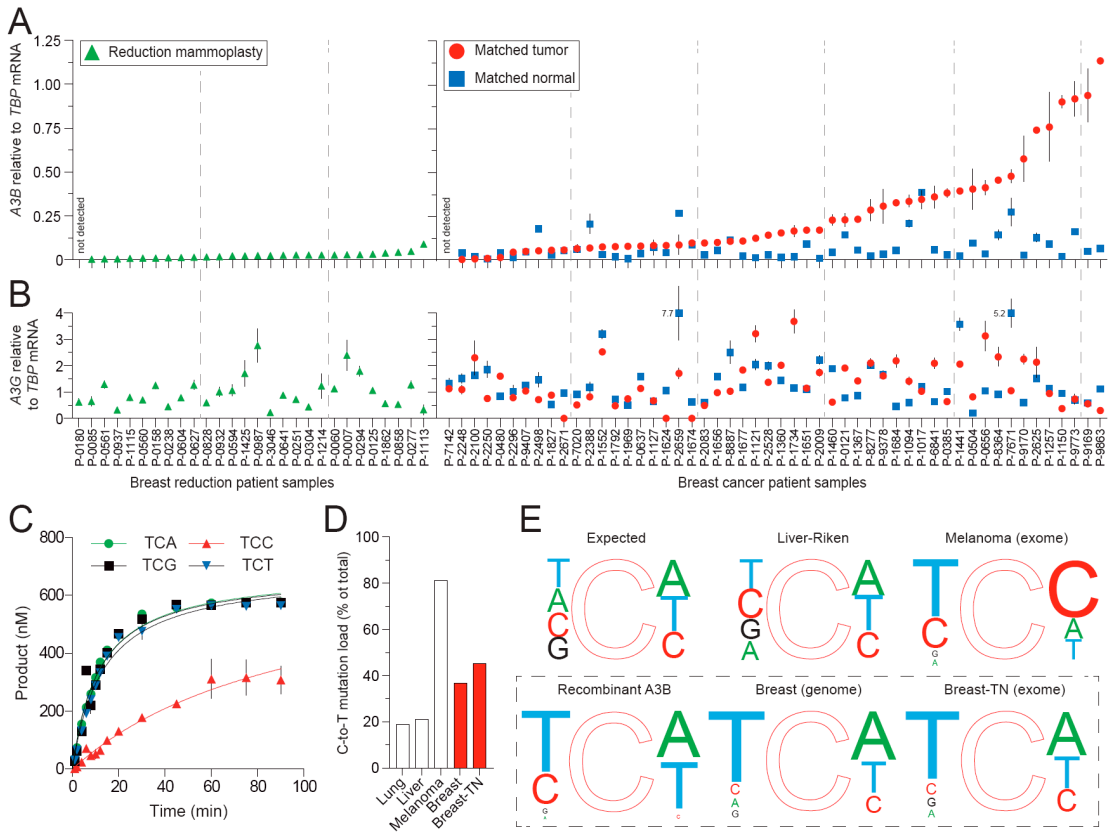

Figure 4-4. *A3B* up-regulation and mutational signatures in breast tumors.

(A) *A3B* mRNA levels in reduction mammoplasty samples (green triangles, n=28) and matched sets of breast tumor in comparison to adjacent or contralateral normal tissue (red circles and blue squares, respectively, n=52). Each data point is the mean *A3B* level of three qPCR reactions presented relative to mRNA levels of the constitutive housekeeping gene *TBP* (s.d. shown unless smaller than data point). Mean *A3B* levels are 0.025 (reduction), 0.26 (tumor), and 0.072 (normal). Data are arranged from lowest to highest *A3B* expression level. **(B)** *A3G* levels in the same samples as in (a). Average *A3G* levels are 1.00 (reduction), 1.45 (tumor), and 1.31 (normal) relative to *TBP*. **(C)** A3B catalytic

domain deamination kinetics using single-stranded DNA substrates that vary only at the 3' position relative to the target cytosine. **(D)** Histogram depicting the percentage C-to-T of the total number of somatic mutations in the indicated tumors (total somatic mutations: lung 50489, liver 654879, melanoma 2798, breast 183916, breast triple negative 6964). **(E)** Nucleotide frequencies proportional to font size immediately 5' and 3' of genomic cytosines (expected) or the C-to-T mutated cytosine in cancers of the liver, skin, breast, or breast (triple negative). The next nucleotide preferences for recombinant A3B are derived experimentally from catalytic efficiencies (panel C & **Fig. 4-S10**; additional comparisons in **Fig. 4-S11**).

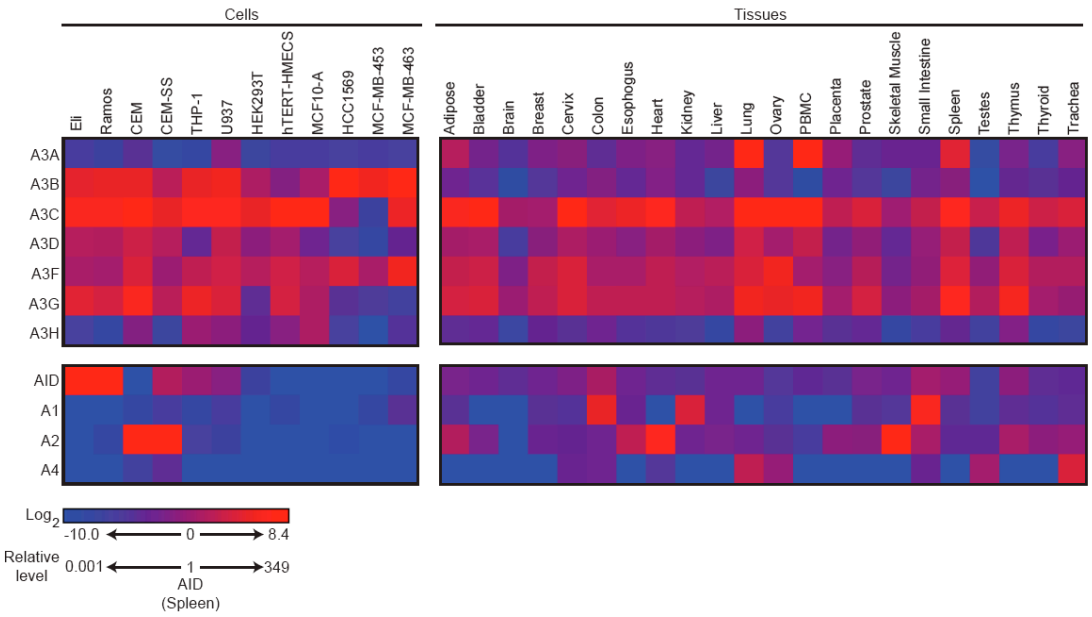

Figure 4-S1. Expression profiles for *APOBEC* family members in human cell lines and tissues. A heat-map summary of qPCR data showing relative *APOBEC3* (*A3)*, *AID*, *APOBEC1* (*A1*), *APOBEC2* (*A2*), and *APOBEC4* (*A4*) mRNA expression levels in the indicated cell lines and tissues. The data are relative to the median *AID* mRNA level in spleen and presented in $\log_2$ format. The average of three independent qPCR reactions was used for each condition. Data for the normal tissues, excluding PBMCs and breast tissue, were reported previously [Refsland *et al*. (Ref. 17)]. They were recalculated and presented here in $\log_2$ format for comparative purposes and to emphasize the general observation that *A3B* is low or almost undetectable in every normal tissue that we have examined to date.

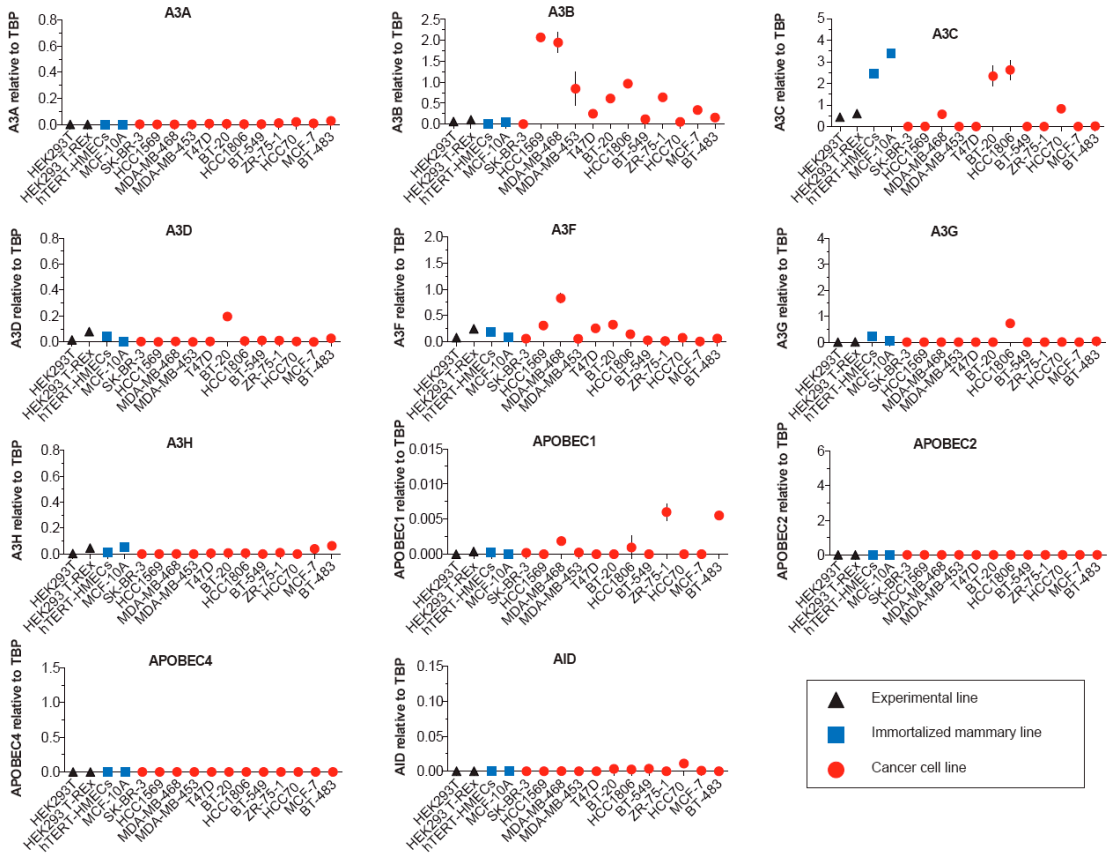

Figure 4-S2. Full expression profiles for *APOBEC* family members in a panel of representative cell lines. The indicated cell lines were used to generate cDNA for qPCR analyses of the full human *APOBEC* repertoire. Each data point is mean mRNA level of three qPCR reactions presented relative to mRNA levels of the constitutive housekeeping gene *TBP* (s.d. shown as a bar unless smaller than the data point). Relevant *A3B* data are also presented in Fig. 4-1A in the context of the full panel of normal and breast cancer cell lines.

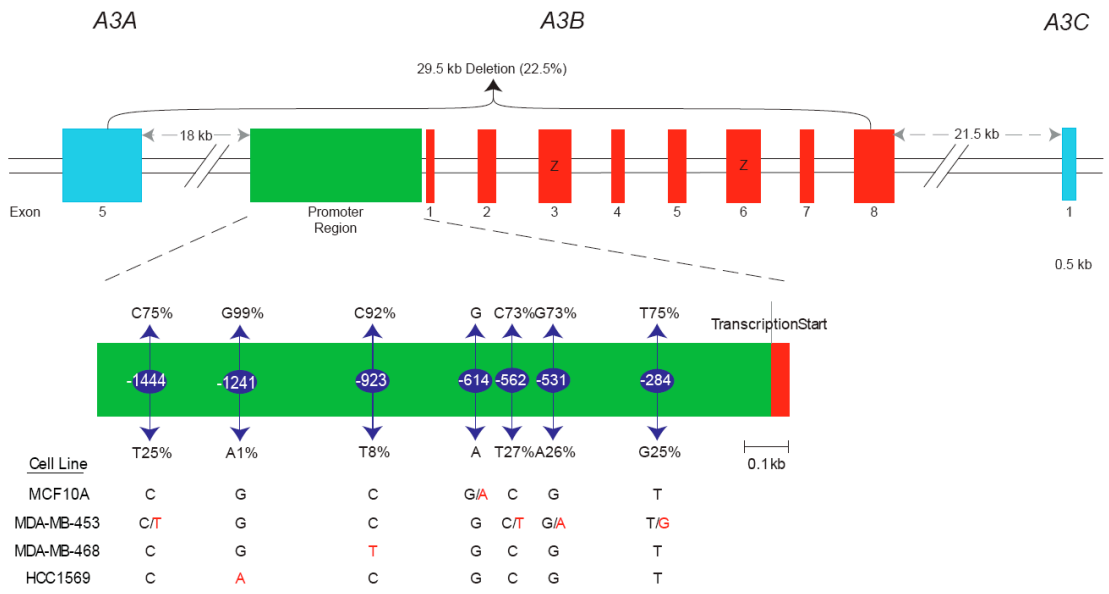

Figure 4-S3. *A3B* promoter region sequence analysis. Schematic of the *A3B* genomic locus depicting flanking genes (blue), exons (red), deaminase domain exons (red with Z label), promoter region (green), and position of the 29.5kb deletion allele. Below, an enlarged schematic of the *A3B* promoter region showing the most common SNPs (above) and minor alleles (below). Allele frequencies are indicated as percentages (www.ncbi.nlm.nih.gov/projects/SNP/). Nucleotide positions are labeled relative to the transcription start site (+1). The promoter regions of the indicated cell lines are identical except at the nucleotides shown.

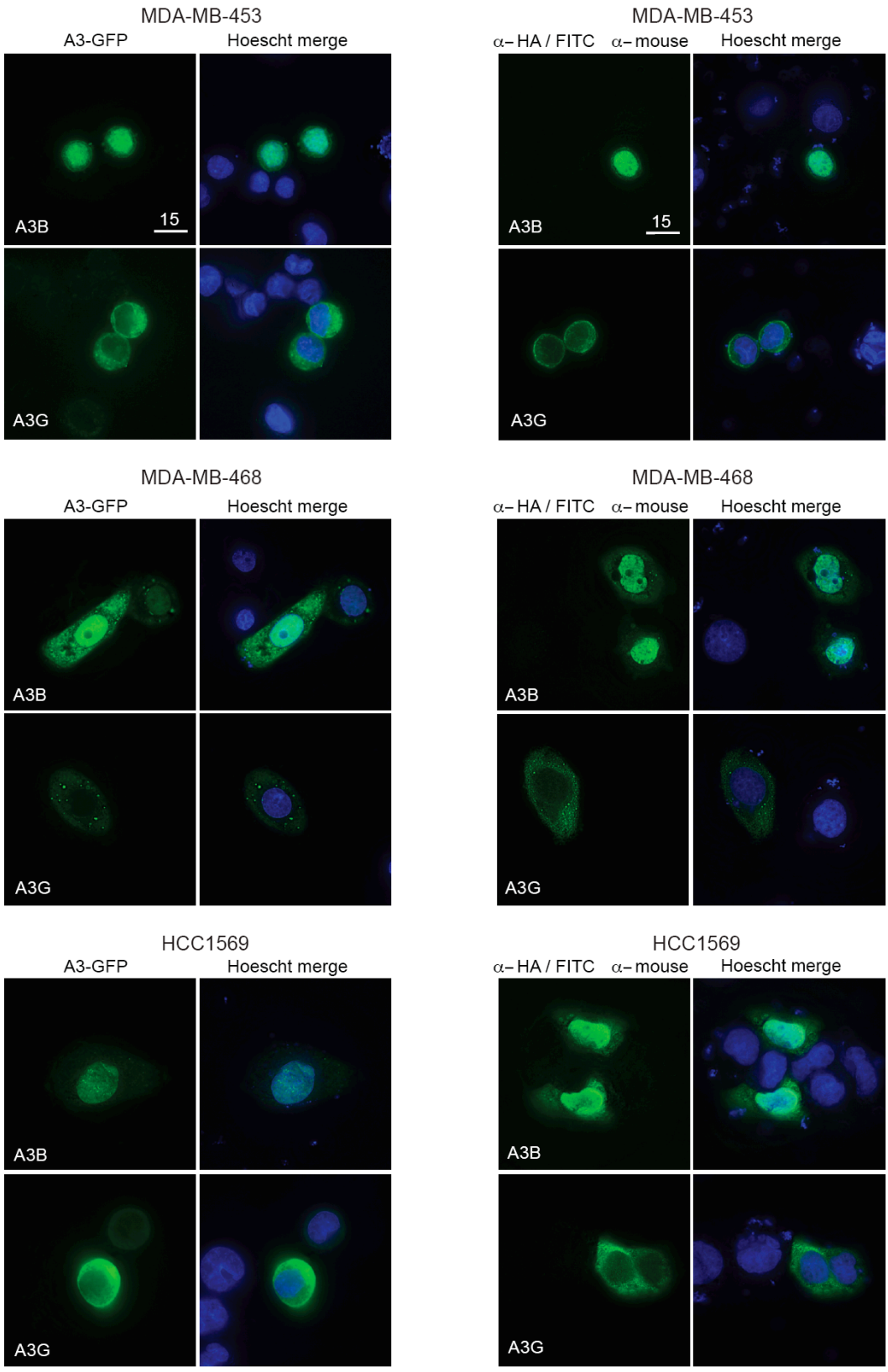

Figure 4-S4. Additional live and fixed breast cancer cell localization

data. A3B-eGFP (green) co-localizes with nuclear DNA (Hoescht-stained blue), whereas A3G-eGFP is cytoplasmic, in the indicated breast cancer cell lines. MDA-MB-468 shows some cytoplasmic A3B-eGFP localization, but is still predominantly nuclear. A3B-HA, A3G-HA, and A3F-HA (not shown) in fixed cells have localization patterns similar to those of live cell eGFP-tagged proteins. In many cases, A3B-HA is more nuclear, perhaps owing to background caused by internal translation initiation and cell-wide expression of the eGFP protein alone.

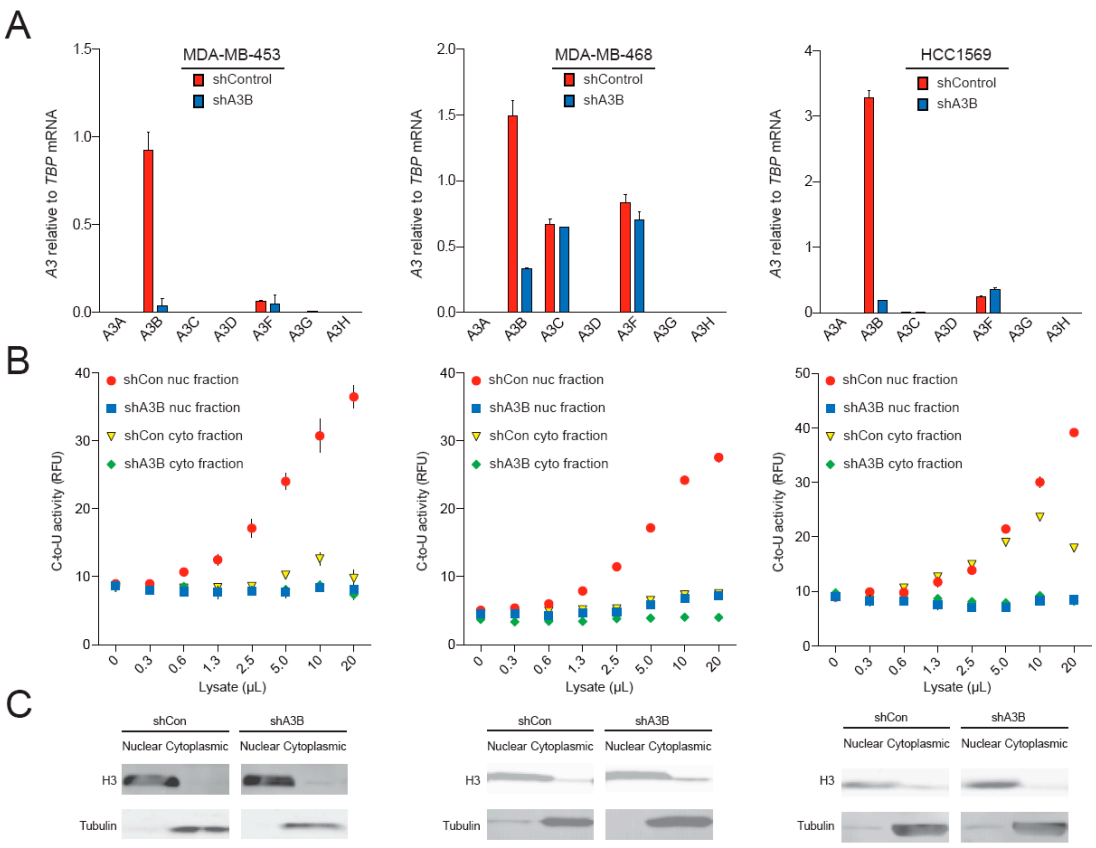

Figure 4-S5. A3B is active in the nuclear protein fraction of multiple breast cancer cell lines. (A) *A3* mRNA levels in the indicated breast cancer cell lines. Each column is mean +/- s.d. of three qPCR reactions presented relative to mRNA levels of the constitutive housekeeping gene *TBP*. Red and blue bars represent expression data from cells stably transduced with shControl or shA3B lentivirus, respectively. **(B)** A3B-dependent DNA deaminase activity in the nuclear (Nuc) and cytoplasmic (Cyt) fractions obtained from the cell lines in (A). The fractionation was cleaner in MDA-MB-453 and MDA-MB-468 lines than HCC1569, but all detectable deaminase activity was still dependent on A3B. **(C)** Immunoblots showing the distribution of histone H3, a nuclear

protein, and tubulin, a cytoplasmic protein, in the protein preparations used in (B) to confirm efficient sub-cellular fractionation.

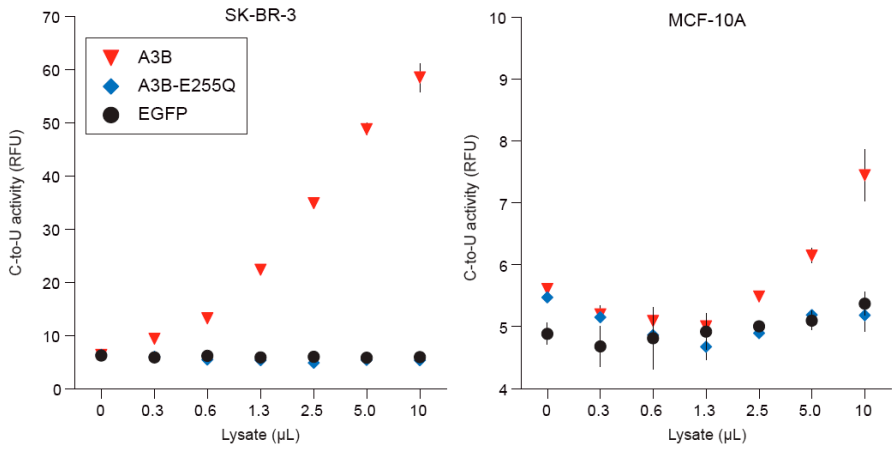

Figure 4-S6. DNA deaminase activity in A3B-low cell types.

Nuclear DNA C-to-U activity in extracts from SK-BR-3 and MCF10A transfected transiently with A3B-eGFP, A3B-E255Q-eGFP, or eGFP expression constructs. The higher activity levels in SK-BR-3 nuclear lysates are due to higher transfection efficiencies (30-40%), in comparison to MCF10A (1-5%). Mean values are shown with s.d. indicated unless smaller than the symbol (n=3).

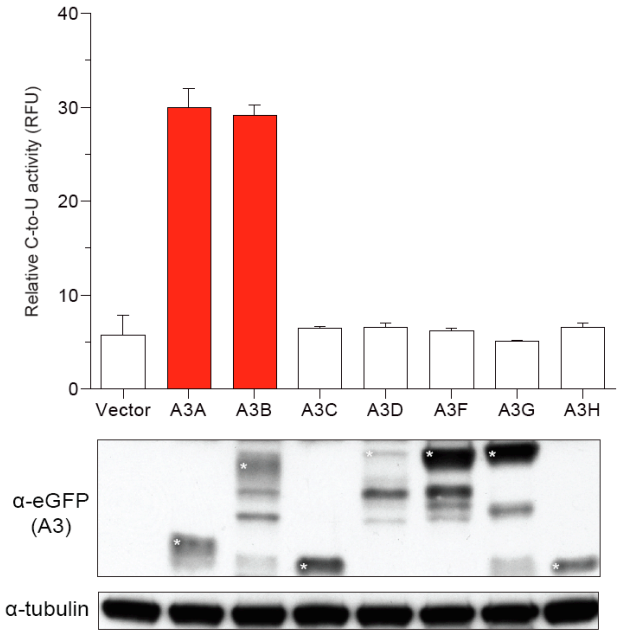

Figure 4-S7. Deaminase activity of HEK293T cell extracts with individual over-expressed A3 proteins. Mean DNA C-to-U activity in whole cell extracts of HEK293T cells transfected with the indicated A3-eGFP expression constructs (n=3 per condition; s.d. shown). Activity was only detected in lysates from cells transfected with A3A- or A3B-eGFP. The corresponding anti-GFP immunoblot shows levels of each A3 (white asterisks), and the anti-tubulin blot indicates similar protein levels in each lysate.

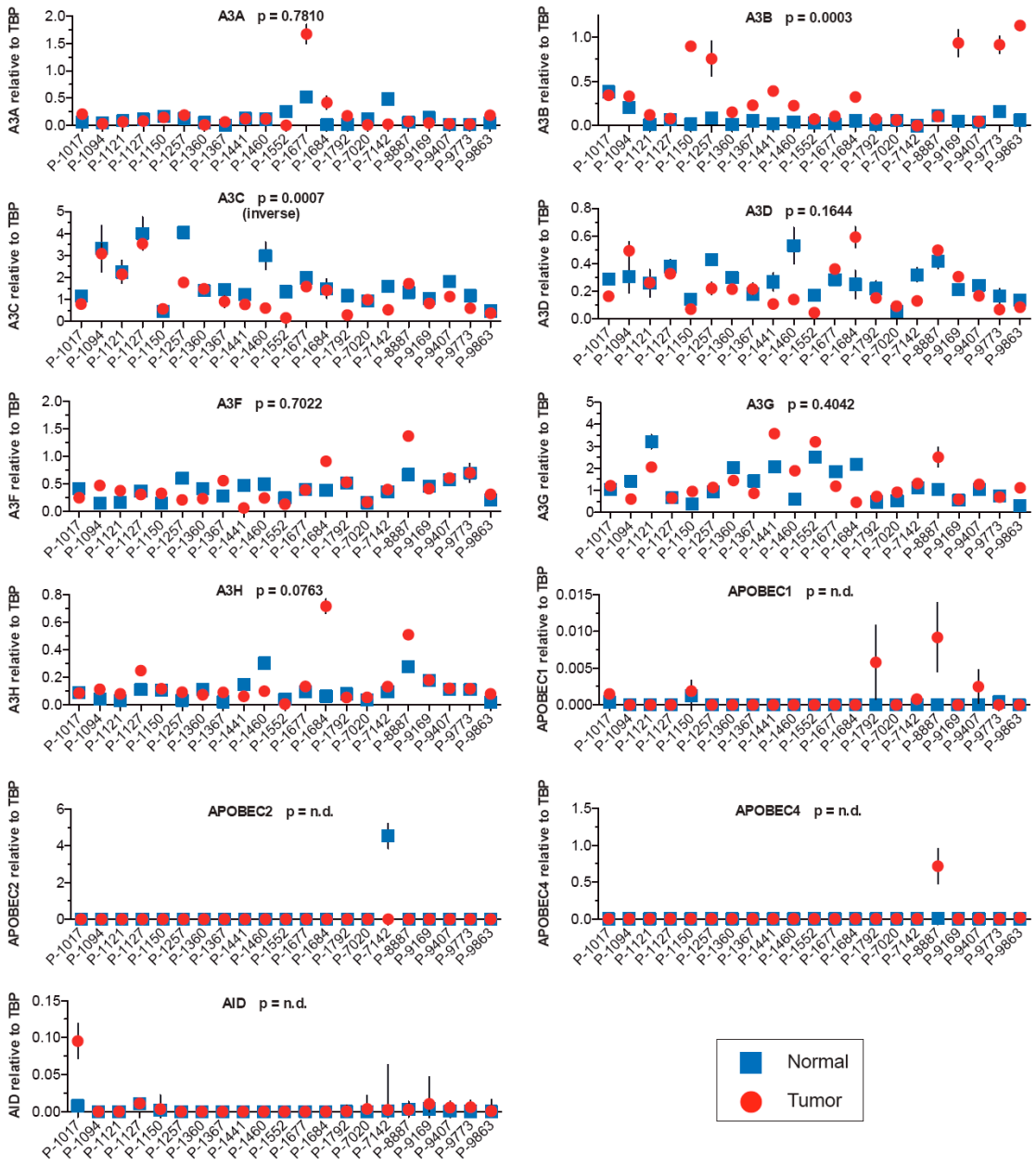

Figure 4-S8. Discovery data set - APOBEC family member expression profiles for 21 randomly selected sets of matched breast tumor and normal tissue. 21 representative breast tumor samples and the matched normal control tissues were used to synthesize cDNA for qPCR analyses of the full human *APOBEC* repertoire. Each data point is the mean mRNA level of three qPCR reactions presented relative to mRNA levels of the constitutive housekeeping gene *TBP* (s.d. shown as a bar unless smaller than

the data point). P-values are indicated except those *AID, A1*, *A2*, and *A4* where the majority of samples had no detectable mRNA for these targets. *A3B* emerges as the only differentially up-regulated family member in tumor versus matched normal tissues. *A3C* shows an inverse correlation. Samples are presented in order of an arbitrarily assigned patient number. The *A3B* and *A3G* data were merged with 31 validation set samples for presentation in **Fig. 4-4**.

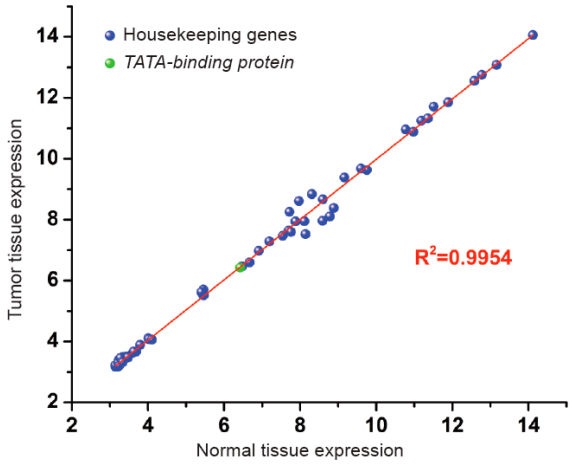

Figure 4-S9. Microarray housekeeping gene comparisons. Expression values for 62 housekeeping genes between different microarray data sets.

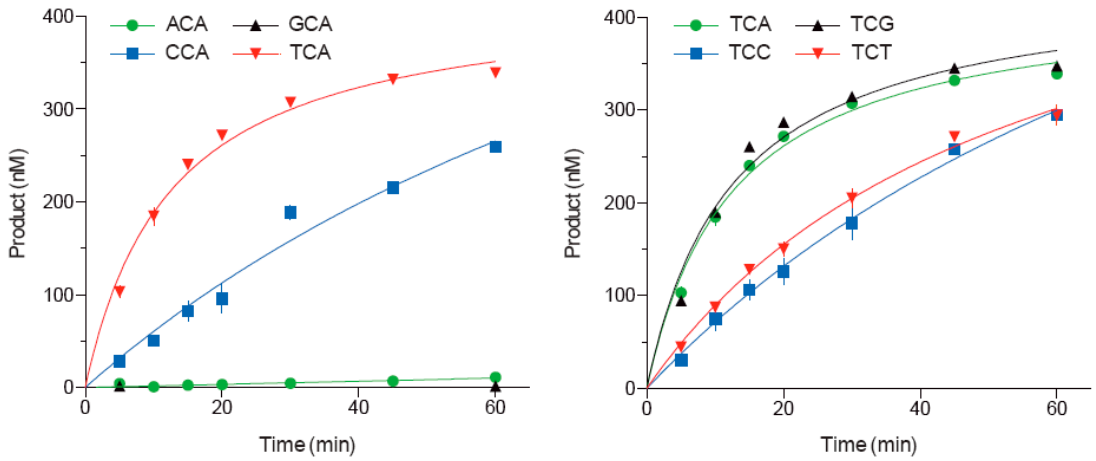

Figure 4-S10. A3B local deamination preferences. Oligo cleavage assays testing the effect of the 5' or 3' nucleotide (TC>GC>AC=CC) on A3B cytosine deaminase activity.

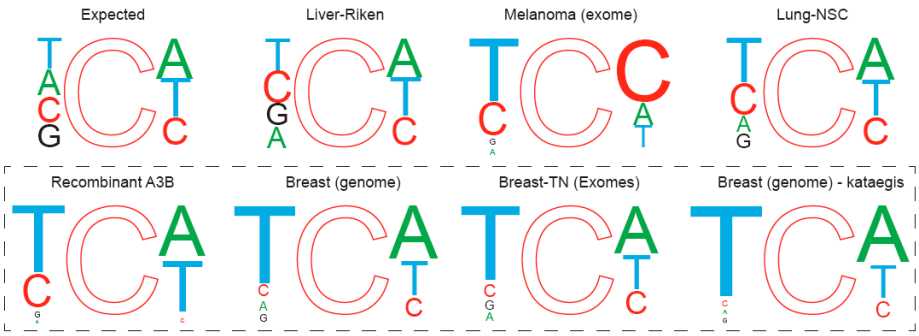

Figure 4-S11. C-to-T transition mutation contexts *in vivo* versus *in vitro*. Data for non-small cell (NSC) lung cancer and breast cancer genomic kataegis events are included here, in addition to data from Fig. 4e, for comparison. The distribution of C-to-T transition mutation contexts in melanoma[245] is consistent with the established mechanism of error-prone DNA polymerase misinsertion of A's opposite UV-induced pyrimidine dimers, which through replication or repair become transition mutations. In contrast, the distribution of C-to-T transition mutation contexts in liver and lung cancers[246,247] strongly resembles the actual distribution of cytosines in the human genome (*i.e.*, the C-to-T transition mutations in these tumors appear context independent consistent with other underlying mutational mechanisms). In further contrast, the distribution of C-to-T transition mutation contexts in 21 full breast tumor genome sequences[228] and 100 triple-negative breast tumor exome sequences[239] strongly resembles the local deamination preference of recombinant A3B with prominent 5'-T and 3'-A or T (not C) flanking nucleotide biases. This bias is even further exaggerated in regions of kataegis, which may represent exposed single-stranded DNA regions *in vivo* that may be subjected to high-frequency A3B-dependent DNA deamination and different repair mechanisms.

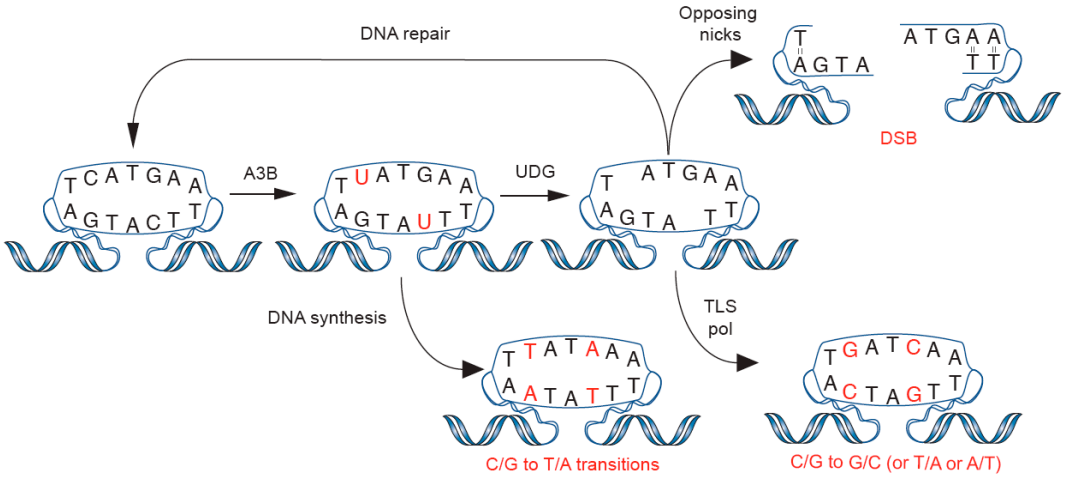

Figure 4-S12. DNA deamination model for A3B in cancer. Deamination of genomic DNA cytosines by up-regulated A3B leads to uracil lesions, which can be repaired faithfully or can lead to at least three possible outcomes: i) C-to-T transitions by direct DNA synthesis (lower outcome), ii) DNA double-stranded breaks by uracil excision and opposing abasic site cleavage (upper right outcome), and iii) transversions or transition mutations by error-prone DNA synthesis or aberrant repair (TLS pol = translesion synthesis DNA polymerase).

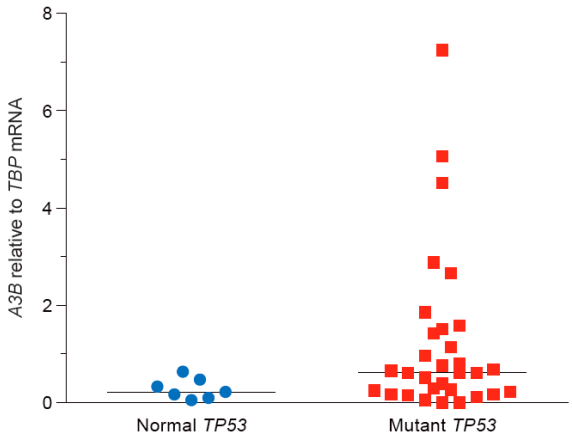

Figure 4-S13. *A3B* up-regulation and *TP53* inactivation in the ATCC breast cancer cell line panel. A simple plot of *A3B* mRNA levels in *TP53* positive versus *TP53* mutant breast cancer cell lines from the ATCC (n=38; full list of cell lines in Table 4-S1).

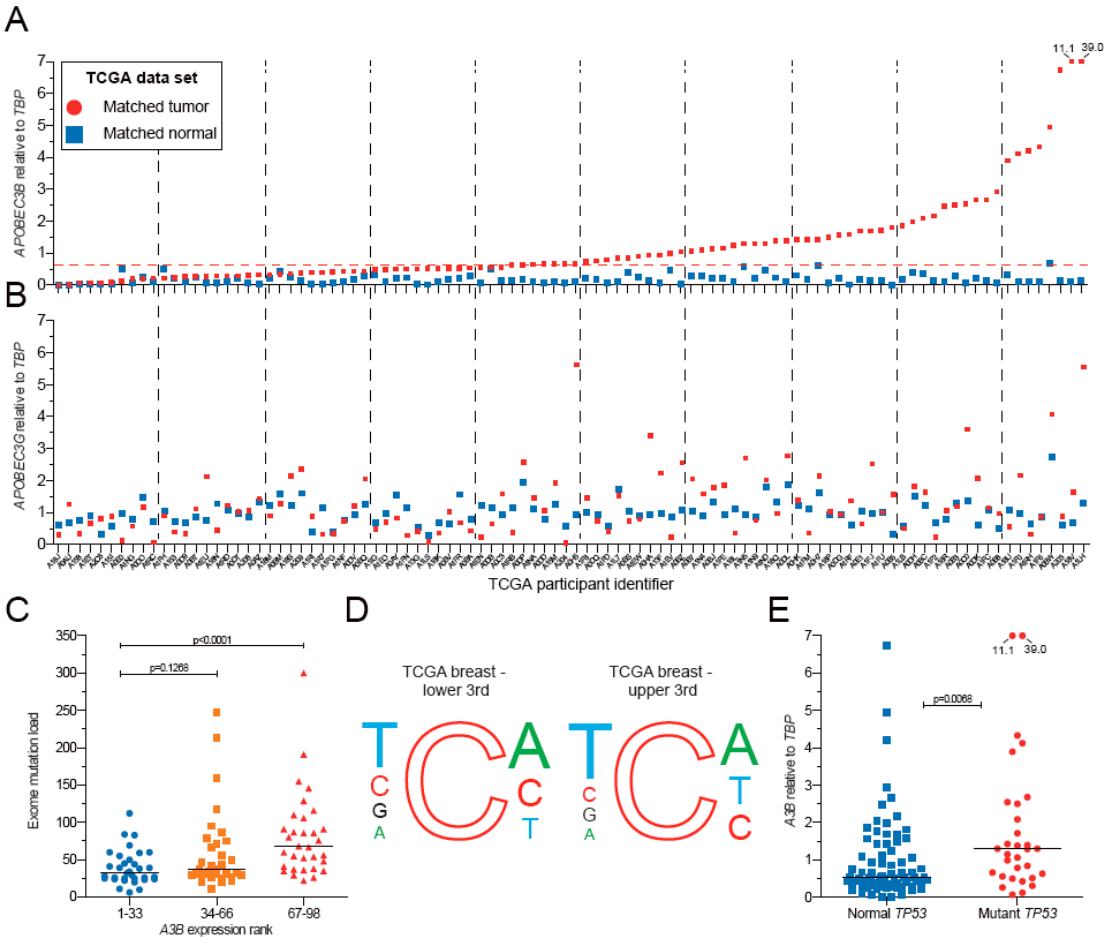

Figure 4-S14. Evidence for A3B up-regulation and genomic mutation in TCGA breast cancer data sets. (A) *A3B* expression relative to *TBP* in 98 breast tumors (red circles) and matched normal tissues (blue squares) as determined by RNAseq data in the TCGA database (https://tcga-data.nci.nih.gov/tcga/). The red dashed line indicates 3 s.d. above the mean *A3B* relative to *TBP* level in the normal samples. Based on this stringent cut-off, *A3B* is up-regulated in 50% of tumors ($p<0.0001$ by Wilcoxon-signed-rank). These data sets were selected based on availability of matched tumor and normal RNAseq data and tumor exome sequencing data. *A3B* levels relative to *TBP* were

calculated using the normalized counts from RNAseq for each gene. **(B)** *A3G* expression in the same tumor and normal samples, determined as above. *A3G* expression is not significantly different between tumor and matched normal samples (p=0.0680 by Wilcoxon signed-rank). **(C)** Dot plots of the total somatic mutation load per exome for the 98 samples in (A) and (B). The data divided into three groups, as shown, from A3B-low to A3B-high. The median mutation loads for the lowest and middle groups are not significantly different, whereas the median between the lowest and the highest A3B-expressing group is significantly different (median values of 32 and 68 somatic mutations per exome, respectively; p values determined by Mann Whitney U test). **(D)** Nucleotide frequencies proportional to font size immediately 5' and 3' of genomic cytosines (expected) or the C-to-T mutated cytosine in the bottom and top one-third of A3B-expressing tumors from above. There are significant increases in both 5'T and 3'T in the A3B-high tumors, which is a signature that strongly resembles the A3B preferences *in vitor* and is apparent in other breast cancer genomic mutation data sets (Fig. 4-4 and 4-S11). **(E)** *A3B* levels are significantly higher in tumors with *TP53* mutations (p values determined by Mann Whitney U test). Exome sequences (MAF files) were used to identify potentially inactivating mutations in the *TP53* gene, which enabled the A3B data points to be separated into two groups. Silent mutations were ignored, and other mechanisms for TP53 inactivation were not considered (*e.g.,* MDM2 over-expression). These data are analogous to similar trends observed with breast cancer cell lines (Fig. 4-S13).

Table 4-S1

Cell Line	Derivation	Site of Origin	ER	PR	Her2/neu	TP53
hTERT-HMEC	Immortalized	Mammary gland	n.a.	n.a.	n.a.	normal
MCF-10A (MCF-10F)*	Immortalized	Mammary Gland	n.a.	n.a.	n.a.	normal
MCF-10F (MCF-10A)*	Immortalized	Mammary Gland	n.a.	n.a.	n.a.	normal
MCF-12A	Immortalized	Mammary Gland	n.a.	n.a.	n.a.	normal
Hs578Bst	Immortalized	Mammary Gland	-	n.a.	n.a.	normal
184B5	Immortalized	Mammary Gland	n.a.	n.a.	n.a.	normal
HCC38	Cancer	Primary Ductal Carcinoma	-	-	-	mutant
AU-565 (SK-BR-3)*	Cancer	Metastatic Adenocarcinoma; Pleural Effusion	n.a.	n.a.	+	mutant
SK-BR-3 (AU-565)*	Cancer	Adenocarcinoma; Pleural effusion	n.a.	n.a.	n.a.	mutant
HCC70	Cancer	Primary Ductal Carcinoma	+	-	-	mutant
HCC1500	Cancer	Primary Ductal Carcinoma	+	+	-	normal
DU4475	Cancer	Mammary Gland	n.a.	n.a.	n.a.	normal
BT-549	Cancer	Papillary, Invasive Ductal Tumor	n.a.	n.a.	n.a.	mutant
BT-483	Cancer	Ductal Carcinoma	n.a.	n.a.	n.a.	mutant
HCC1395	Cancer	Primary Ductal Carcinoma	-	-	-	mutant
HCC2218	Cancer	Primary Ductal Carcinoma	-	n.a.	+	mutant
UACC-812	Cancer	Primary Ductal Carcinoma	-	-	+	normal
CAMA-1	Cancer	Pleural Effusion	n.a.	n.a.	n.a.	mutant
ZR-75-30	Cancer	Ascites	n.a.	n.a.	n.a.	normal
T47D	Cancer	Ductal Carcinoma	+	+	-	mutant
HCC1419	Cancer	Primary Ductal Carcinoma	-	-	+	mutant
HCC1937	Cancer	Primary Ductal Carcinoma	-	-	-	mutant
MCF-7	Cancer	Adenocarcinoma; Pleural Effusion	+	+	-	normal
HCC1954	Cancer	Primary Ductal Carcinoma	-	-	+	mutant
MDA-MB-175-VII	Cancer	Metastic Ductal Carcinoma; Pleural Effusion	n.a.	n.a.	n.a.	normal
MDA-MB-436	Cancer	Metastatic Adenocarcinoma; Pleural Effusion	n.a.	n.a.	n.a.	mutant
BT-20	Cancer	Mammary Gland Carcinoma	-	n.a.	n.a.	mutant
MDA-MB-361	Cancer	Metastatic Adenocarinoma	n.a.	n.a.	n.a.	mutant
HCC1187	Cancer	Primary Ductal Carcinoma	n.a.	-	-	mutant
ZR-75-1	Cancer	Ascites	+	n.a.	n.a.	normal
Hs578T	Cancer	Mammary Gland Carcinoma	-	n.a.	n.a.	mutant
MDA-MB-157	Cancer	Medulallary Carcinoma	n.a.	n.a.	n.a.	mutant
UACC-893	Cancer	Primary Ductal Carcinoma	-	-	+	mutant
HCC1428	Cancer	Adenocarcinoma; Pleural Effusion cells	n.a.	n.a.	-	mutant
HCC1806	Cancer	Primary Squamous Cell Carcinoma	-	-	-	mutant
BT-474	Cancer	Invasive Ductal Carcinoma	n.a.	n.a.	n.a.	mutant
MDA-MB-231	Cancer	Metastatic adenocarcinoma; Pleural Effusion	-	-	-	mutant
MDA-MB-453 (MDA-kb2)*	Cancer	Metastatic Pericardial Effusion	-	-	+	mutant
MDA-MB-468	Cancer	Metastatic Adenocarcinoma; Pleural Effusion	-	-	-	mutant
MDA-kb2 (MDA-MB-453)*	Cancer	Metastatic Pericardial Effusion	n.a.	n.a.	n.a.	mutant
MDA-MB-415	Cancer	Adenocarcinoma; Pleural Effusion	n.a.	n.a.	n.a.	mutant
HCC2157	Cancer	Primary Ductal Carcinoma	-	+	+	mutant
MDA-MB-134-VI	Cancer	Pleural Effusion	n.a.	n.a.	n.a.	mutant
HCC1569	Cancer	Primary Metaplastic Carcinoma	-	-	+	mutant
HCC1599	Cancer	Primary Ductal Carcinoma	-	-	-	mutant
HCC202	Cancer	Primary Ductal Carcinoma	-	-	+	mutant

*Related lines
Unk Unknown
n.a. not available

Table 4-S2

Patient ID	Age	Ethnicity	Age	ER	PR	Her2/neu	Type	Grade
P-7142	40	Caucasian	40	+	-	+	IDC	3
P-2248	51	African American	51	+	-	-	IDC	2
P-2100	75	Caucasian	75	+	+	-	IDC	2
P-2250	76	Caucasian	76	+	+	-	IDC	2
P-0480	51	Caucasian	51	-	-	-	IDC	3
P-2296	49	Caucasian	49	+	+	-	IDC	2
P-9407	38	Caucasian	38	+	+	-	IDC	2
P-2498	40	Caucasian	40	+	+	-	IDC	2
P-1827	37	Caucasian	37	+	-	-	IDC/ILC	2
P-2671	61	Caucasian	61	+	+	-	ILC	2
P-7020	40	Caucasian	40	+	+	-	IDC/ILC	1
P-2388	47	Caucasian	47	+	+	-	IDC	1
P-1552	58	Caucasian	58	+	+	-	IDC	3
P-1792	44	Caucasian	44	+	+	n.a.	DCIS	1
P-1969	77	Caucasian	77	+	-	+	IDC	2
P-0637	70	Caucasian	70	+	-	+	IDC	2
P-1127	68	Caucasian	68	+	+	n.a.	DCIS	1
P-1624	49	Caucasian	49	+	+	-	IDC	2
P-2659	58	Caucasian	58	+	-	+	IDC	3
P-1674	64	Caucasian	64	+	-	-	ILC	2
P-2083	39	Caucasian	39	+	+	-	IDC	2
P-1656	74	Caucasian	74	+	+	-	ILC	2
P-8887	45	Native American	45	+	+	-	LC	2
P-1677	49	Caucasian	49	+	+	-	IDC	2
P-1121	75	Caucasian	75	+	+	-	ILC	1
P-2528	51	Caucasian	51	+	+	-	ILC	2
P-1360	66	Caucasian	66	+	+	-	IDC	2
P-1734	47	Caucasian	47	+	+	-	IDC	2
P-1651	51	Caucasian	51	+	+	-	IMC	2
P-2009	62	Caucasian	62	+	+	-	IDC	1
P-1460	62	Caucasian	62	+	+	+	IDC	3
P-0121	77	Caucasian	77	+	+	-	IDC	2
P-1367	43	Caucasian	43	+	+	-	IDC	2
P-8277	54	Caucasian	54	+	-	-	IDC	1
P-9378	68	Caucasian	68	+	-	-	ILC	2
P-1684	45	Caucasian	45	+	+	-	IDC/ILC	2
P-1094	51	Caucasian	51	+	+	-	IDC	2
P-1017	40	Caucasian	40	+	+	+	IDC	3
P-6841	68	Caucasian	68	+	+	+	IDC	3
P-0385	56	Caucasian	56	+	+	-	ILC	2
P-1441	70	Caucasian	70	-	-	-	IDC	3
P-0504	56	Caucasian	56	+	-	-	IDC	2
P-0656	39	Caucasian	39	-	-	+	IDC	3
P-8364	42	Caucasian	42	+	-	n.a.	DCIS	1
P-7671	48	Caucasian	48	+	-	-	DCIS	1
P-9170	55	Caucasian	55	+	-	-	IDC	2
P-2625	72	Caucasian	72	+	+	+	IDC	3
P-1257	77	Caucasian	77	+	-	+	IDC	2
P-1150*	30	Caucasian	30	+	+	+	IDC	3
P-9773	37	Caucasian	37	-	-	-	IDC	3
P-9169	62	Caucasian	62	+	+	-	IDC/ILC	1
P-9863	46	Caucasian	46	+	+	-	IDC	2

* Male patient

DCIS	Ductal carcinoma *in situ*
IDC	Invasive ductal carcinoma
ILC	Invasive lobular carcinoma
IDC/ILC	Invasive ductal carcinoma with lobular features
IMC	Invasive mucinous carcinoma

Table 4-S3. Statistical results of the differential expression of the genes APOBEC3 (A, B, C, F, & G) between 22 normal subjects (N) and 285 subjects with breast cancer (C). The significance level is set at: α=0.007143. The differential expression of two of the 62 house genes [(1) AFFX-HSAC07/X00351_M_at (ACTB) (HGNC:132) and (2) AFFX-HUMGAPDH/M33197_M_at (GAPDH) (HGNC:4141)] that are used in the Affymetrix Quality Control Metrics Report is also shown.

Gene	N (mean ± SD) n=22	C (mean ± SD) n=285	t-Test P value	Fold Change (C/N)	ROC AUC
210873_x_at (APOBEC3A)	3.554 ± 0.237	3.698 ±0.042	$< 1x10^{-6}$	1.041	0.836
206632_s_at (APOBEC3B)	4.049 ± 0.386	4.404 ± 0.082	$< 1x10^{-6}$	1.088	0.900
209584_x_at (APOBEC3C)	4.977 ± 0.226	4.901 ± 0.038	0.144	0.985	0.594
214995_s_at (APOBEC3F,G)	3.858 ± 0.190	4.012 ±0.037	$1x10^{-6}$	1.040	0.816
214994_at (APOBEC3F)	3.968 ± 0.228	3.894 ± 0.041	0.008	0.981	0.670
204205_at (APOBEC3G)	5.535 ± 0.491	5.422 ± 0.071	0.011	0.980	0.663
215579_at (APOBEC3G)	5.845 ± 0.187	5.897 ±0.037	0.001	1.010	0.705
House Gene 1	6.107 ± 0.312	6.039 ± 0.050	0.183	0.990	0.585
House Gene 2	3.053 ± 0.643	3.128 ± 0.080	0.438	1.025	0.550

Table 4-S4. Affymetrix microarray HG-U133A A3A probe (cross)hybridization within the APOBEC3 family.

		Probe identity to APOBEC3A-H						
		# identities/probe length (%)						
Intended target gene* (RefSeq)	Probe set 210873_x_at	A	B	C	D*	F	G	H*
APOBEC3A NM_145699	GCTCACAGACGCCAGCAAAGCAGTA	25/25	22/25	≤20/25	≤20/25	≤20/25	≤20/25	≤20/25
	GACGCCAGCAAAGCAGTATGCTCCC	25/25	≤20/25	≤20/25	≤20/25	≤20/25	≤20/25	≤20/25
	GCAGTATGCTCCCGATCAAGTAGAT	25/25	22/25	≤20/25	≤20/25	≤20/25	≤20/25	≤20/25
	AAAAAATCAGAGTGGGCCGGGCGCG	25/25	≤20/25	≤20/25	≤20/25	≤20/25	≤20/25	≤20/25
	GAGGCAGGAGAGTACGTGAACCCGG	24/25	≤20/25	≤20/25	≤20/25	≤20/25	≤20/25	≤20/25
	AACTGAAAATTTCTCTTATGTTCCA	25/25	24/25	≤20/25	≤20/25	≤20/25	≤20/25	≤20/25
	CTCTTATGTTCCAAGGTACACAATA	25/25	≤20/25	≤20/25	≤20/25	≤20/25	≤20/25	≤20/25
	GATTATGCTCAATATTCTCAGAATA	25/25	24/25	≤20/25	≤20/25	≤20/25	≤20/25	≤20/25
	TTTGGCTTCATATCTAGACTAACAC	24/25	≤20/25	≤20/25	≤20/25	≤20/25	≤20/25	≤20/25
	GAATCTTCCATAATTGCTTTTGCTC	25/25	21/25	≤20/25	≤20/25	≤20/25	≤20/25	≤20/25
	TAATTGCTTTTGCTCAGTAACTGTG	25/25	≤20/25	≤20/25	≤20/25	≤20/25	≤20/25	≤20/25

* A3D and A3H are not represented intentionally in the U133 probe set.

Table 4-S5. Affymetrix microarray HG-U133A A3B probe (cross)hybridization within the APOBEC3 family.

		Probe identity to APOBEC3A-H						
		# identities/probe length (%)						
Intended target gene* (RefSeq)	Probe set 206632_s_at	A	B	C	D*	F	G	H*
A3B NM_004900	CTACGATGAGTTTGAGTACTGCTGG	22/25	25/25	≤20/25	≤20/25	≤20/25	≤20/25	≤20/25
	CACCTTTGTGTACCGCCAGGGATGT	23/25	25/25	≤20/25	≤20/25	≤20/25	23/25	≤20/25
	GAAATGCAAACGAGCCGTTCACCAC	22/25	22/25	≤20/25	≤20/25	≤20/25	22/25	≤20/25
	ACCAGCAAAGCAATGTGCTCCTGAT	≤20/25	25/25	≤20/25	≤20/25	≤20/25	22/25	≤20/25
	AGCAATGTGCTCCTGATCAAGTAGA	22/25	25/25	≤20/25	≤20/25	≤20/25	22/25	≤20/25
	ATGTGCTCCTGATCAAGTAGATTTT	22/25	25/25	≤20/25	≤20/25	≤20/25	≤20/25	≤20/25
	TGTTCCAAGTGTACAAGAGTAAGAT	22/25	25/25	≤20/25	≤20/25	≤20/25	≤20/25	≤20/25
	TTATGCTCAATATTCCCAGAATAGT	23/25	25/25	≤20/25	≤20/25	≤20/25	≤20/25	≤20/25
	ATTCCCAGAATAGTTTTCAATGTAT	23/25	25/25	≤20/25	≤20/25	≤20/25	≤20/25	≤20/25
	GAAGTGATTAATTGGCTCCATATTT	≤20/25	25/25	≤20/25	≤20/25	≤20/25	≤20/25	≤20/25
	TAATTGGCTCCATATTTAGACTAAT	≤20/25	25/25	≤20/25	≤20/25	≤20/25	≤20/25	≤20/25

* A3D and A3H are not represented intentionally in the U133 probe set.

Table 4-S6. Affymetrix microarray HG-U133A A3C probe (cross)hybridization within the APOBEC3 family.

		Probe identity to APOBEC3A-H						
		# identities/probe length (%)						
Intended target gene* (RefSeq)	Probe set 209584_x_at	A	B	C	D*	F	G	H*
A3C NM_14508	AAGGGGTCGCTGTGGAGATCATGGA	≤20/25	≤20/25	25/25	≤20/25	≤20/25	≤20/25	≤20/25
	TAATGAGCCATTCAAGCCTTGGGAA	≤20/25	≤20/25	24/25	23/25	23/25	≤20/25	≤20/25
	CCAACTTTCGACTTCTGAAAAGAAG	≤20/25	≤20/25	25/25	25/25	≤20/25	≤20/25	≤20/25
	AAGAAGGCTACGGGAGAGTCTCCAG	≤20/25	≤20/25	25/25	24/25	≤20/25	≤20/25	≤20/25
	GGGAGAGTCTCCAGTGAGGGGTCTC	≤20/25	≤20/25	25/25	24/25	22/25	≤20/25	≤20/25
	CTCCCCAGCATAACCAAATCTTACT	≤20/25	≤20/25	25/25	23/25	≤20/25	≤20/25	≤20/25
	TTACTAAACTCATGCTAGGCTGGGC	≤20/25	≤20/25	24/25	≤20/25	≤20/25	≤20/25	≤20/25
	TAGGCTGGGCATGGTGACTCACGCC	≤20/25	≤20/25	25/25	22/25	≤20/25	≤20/25	≤20/25
	GGTGGGAGAATCGCGTGAGCCCAGG	≤20/25	≤20/25	25/25	23/25	23/25	≤20/25	≤20/25
	AGCCCAGGAGTTCCAGACCAGGCTG	≤20/25	≤20/25	25/25	≤20/25	22/25	≤20/25	≤20/25
	TCCAGACCAGGCTGGGTCACATGAC	≤20/25	≤20/25	25/25	≤20/25	≤20/25	≤20/25	≤20/25

* A3D and A3H are not represented intentionally in the U133 probe set.

Table 4-S7. Affymetrix microarray HG-U133A A3F/A3G (1) probe (cross)hybridization within the APOBEC3 family.

		Probe identity to APOBEC3A-H						
		# identities/probe length (%)						
Intended target gene* (RefSeq)	Probe set 214995_s_at	A	B	C	D*	F	G	H*
A3F / A3G_1 NM_145298 / NM_021822	GAAAGTGAAACCCTGGTGCTCCAGA	≤20/25	≤20/25	≤20/25	≤20/25	25/25	25/25	≤20/25
	GGTGCTCCAGACAAAGATCTTAGTC	≤20/25	≤20/25	≤20/25	≤20/25	25/25	25/25	≤20/25
	AGATCTTAGTCGGGACTAGCCGGCC	≤20/25	≤20/25	≤20/25	≤20/25	25/25	25/25	≤20/25
	GGGACTAGCCGGCCAAGGATGAAGC	≤20/25	≤20/25	≤20/25	≤20/25	25/25	25/25	≤20/25
	GAAGCCTCACTTCAGAAACACAGTG	≤20/25	≤20/25	≤20/25	≤20/25	25/25	25/25	≤20/25
	AGTGGAGCGAATGTATCGAGACACA	≤20/25	23/25	≤20/25	23/25	25/25	25/25	≤20/25
	ACACATTCTCCTACAACTTTTATAA	≤20/25	≤20/25	≤20/25	≤20/25	25/25	25/25	≤20/25
	TATAATAGACCCATCCTTTCTCGTC	≤20/25	≤20/25	≤20/25	≤20/25	25/25	25/25	≤20/25
	CTTTCTCGTCGGAATACCGTCTGGC	≤20/25	≤20/25	≤20/25	≤20/25	25/25	25/25	≤20/25
	TACCGTCTGGCTGTGCTACGAAGTG	≤20/25	≤20/25	≤20/25	≤20/25	25/25	25/25	≤20/25
	GGACGCAAAGATCTTTCGAGGCCAG	≤20/25	≤20/25	≤20/25	≤20/25	25/25	25/25	≤20/25

* A3D and A3H are not represented intentionally in the U133 probe set.

Table 4-S8. Affymetrix microarray HG-U133A A3F/A3G (2) probe (cross)hybridization within the APOBEC3 family.

		Probe identity to APOBEC3A-H						
		# identities/probe length (%)						
Intended target gene* (RefSeq)	Probe set 214994_at	A	B	C	D*	F	G	H*
A3F / A3G_2 NM_145298 / NM_021822	CACCACATGGGACAGCGCAGGTCCA	≤20/25	≤20/25	≤20/25	≤20/25	≤20/25	≤20/25	≤20/25
	CACATGGGACAGCGCAGGTCCAGTG	≤20/25	≤20/25	≤20/25	≤20/25	≤20/25	≤20/25	≤20/25
	CCAGCTGACCGCAGGCAGGGAACAA	≤20/25	≤20/25	≤20/25	≤20/25	≤20/25	≤20/25	≤20/25
	GGCAGGGAACAAGGCAGACCCTAGA	≤20/25	≤20/25	≤20/25	≤20/25	≤20/25	≤20/25	≤20/25
	AAGGCAGACCCTAGAGGGCCAGGCC	≤20/25	≤20/25	≤20/25	≤20/25	≤20/25	≤20/25	≤20/25
	TGCCAGAATTCACGCATGAGGCTCT	≤20/25	≤20/25	≤20/25	≤20/25	≤20/25	≤20/25	≤20/25
	GCATGAGGCTCTGAACAGGGCTGGG	≤20/25	≤20/25	≤20/25	≤20/25	≤20/25	≤20/25	≤20/25
	TGAACAGGGCTGGGAAAACTTCCAA	≤20/25	≤20/25	≤20/25	≤20/25	≤20/25	≤20/25	≤20/25
	AAGCTCATGTCTTGGTGCACTTTGT	≤20/25	≤20/25	≤20/25	≤20/25	≤20/25	≤20/25	≤20/25
	CACTTTGTGATGATGCTTCAACAGC	≤20/25	≤20/25	≤20/25	≤20/25	≤20/25	≤20/25	≤20/25
	GCTTCAACAGCAGGACTGAGATGGG	≤20/25	≤20/25	≤20/25	≤20/25	≤20/25	≤20/25	≤20/25

* A3D and A3H are not represented intentionally in the U133 probe set.

Table 4-S9. Affymetrix microarray HG-U133A A3G (1) probe (cross)hybridization within the APOBEC3 family.

		Probe identity to APOBEC3A-H						
		# identities/probe length (%)						
Intended target gene* (RefSeq)	Probe set 204205_at	A	B	C	D*	F	G	H*
A3G_1 NM_021822	GCCCGCATCTATGATGATCAAGGAA	≤20/25	≤20/25	≤20/25	≤20/25	≤20/25	25/25	≤20/25
	AAGATGTCAGGAGGGGCTGCGCACC	≤20/25	≤20/25	≤20/25	≤20/25	≤20/25	25/25	≤20/25
	ACCAGCAAAGCAATGCACTCCTGAC	≤20/25	22/25	≤20/25	≤20/25	≤20/25	25/25	≤20/25
	GCAATGCACTCCTGACCAAGTAGAT	≤20/25	22/25	≤20/25	≤20/25	≤20/25	25/25	≤20/25
	GCACTCCTGACCAAGTAGATTCTTT	≤20/25	≤20/25	≤20/25	≤20/25	≤20/25	25/25	≤20/25
	ATTAGAGTGCATTACTTTGAATCAA	≤20/25	≤20/25	≤20/25	≤20/25	≤20/25	25/25	≤20/25
	TAAAGTACTAAGATTGTGCTCAATA	≤20/25	≤20/25	≤20/25	≤20/25	≤20/25	25/25	≤20/25
	GTTTCAAACCTACTAATCCAGCGAC	≤20/25	≤20/25	≤20/25	≤20/25	≤20/25	25/25	≤20/25
	AAACCTACTAATCCAGCGACAATTT	≤20/25	≤20/25	≤20/25	≤20/25	≤20/25	25/25	≤20/25
	ATCCAGCGACAATTTGAATCGGTTT	≤20/25	≤20/25	≤20/25	≤20/25	≤20/25	25/25	≤20/25
	GAATCGGTTTTGTAGGTAGAGGAAT	≤20/25	≤20/25	≤20/25	≤20/25	≤20/25	25/25	≤20/25

* A3D and A3H are not represented intentionally in the U133 probe set.

Table 4-S10. Affymetrix microarray HG-U133A A3G (2) probe (cross)hybridization within the APOBEC3 family.

		Probe identity to APOBEC3A-H						
		# identities/probe length (%)						
Intended target gene* (RefSeq)	Probe set 215579_at	A	B	C	D*	F	G	H*
A3G_2 NM_021822	TTTCCAAATACAGCCACCCTTTGAG	≤20/25	≤20/25	≤20/25	≤20/25	≤20/25	≤20/25	≤20/25
	ACAGCCACCCTTTGAGGGAGCGGGG	≤20/25	≤20/25	≤20/25	≤20/25	≤20/25	≤20/25	≤20/25
	TGAGGGAGCGGGGGTTAAGGCTTCA	≤20/25	≤20/25	≤20/25	≤20/25	≤20/25	≤20/25	≤20/25
	GGGGGTTAAGGCTTCAATACATTGA	≤20/25	≤20/25	≤20/25	≤20/25	≤20/25	≤20/25	≤20/25
	AGAAACAGTGAAGGCCACGGCAAGA	≤20/25	≤20/25	≤20/25	≤20/25	≤20/25	≤20/25	≤20/25
	AGAAGCTGCAGTCATTGTGGGCGGG	≤20/25	≤20/25	≤20/25	≤20/25	≤20/25	≤20/25	≤20/25
	TTCCCAGGGGAGTCCTGACCTGACT	≤20/25	≤20/25	≤20/25	≤20/25	≤20/25	≤20/25	≤20/25
	TCTGGGGTCCGGACATGACCCCTCA	≤20/25	≤20/25	≤20/25	≤20/25	≤20/25	≤20/25	≤20/25
	GTCCTATCAAAGGTGGCATCCTCCC	≤20/25	≤20/25	≤20/25	≤20/25	≤20/25	≤20/25	≤20/25
	GCCTCTGCACTGGGTGCTAATAATT	≤20/25	≤20/25	≤20/25	≤20/25	≤20/25	≤20/25	≤20/25
	GGGTGCTAATAATTCACTTTTACCT	≤20/25	≤20/25	≤20/25	≤20/25	≤20/25	≤20/25	≤20/25

* A3D and A3H are not represented intentionally in the U133 probe set.

Chapter 5

Discussion

Summary

The APOBEC3 proteins are DNA cytosine deaminases that have beneficial roles in innate immunity, but may also cause detrimental genomic deamination. The data I present in this thesis provide evidence for a model in which nuclear APOBEC3 proteins can deaminate genomic DNA, resulting in mutations and affecting carcinogenesis. In Chapter 2, I describe data indicating that A3B is actively imported into the nuclear compartment, similar to AID. In Chapter 3, I explain our data showing the exclusion of the APOBEC3 proteins from condensed chromosomes during mitosis, indicating that even without the barrier of the nuclear envelope many of the APOBEC3 proteins do not interact with DNA. In Chapter 4, I show our data proving that A3B is deaminating breast cancer genomic DNA, illustrating that not only can the APOBEC3 proteins deaminate genomic DNA, but that this is an important and previously unknown aspect of cancer biology. In this section, I summarize these results and place them in a broader context.

Active nuclear import of A3B

When our studies on A3B localization and import began, we were not aware of its role in breast cancer genomic deamination and mutation. We were intrigued by reports on the active nuclear import of AID and the fact that A3B is a relatively small 45.9 kDa (close to the limit for passive diffusion of ~50 kDa)[41,165]. We wondered if we could use a comparative approach to determine whether A3B and AID shared the same nuclear import pathway. Additionally, we hoped to uncover any shared functional roles between A3B and AID.

As described in Chapter 2, I tested for active nuclear import of A3B using two separate assays – *in vitro* transport of A3B through the nuclear pore and time-lapse microscopy following A3B entry after cell division. A3B was capable of active transport in both assays[205]. In addition, we further linked A3B to active import by demonstrating that it interacts with adaptor importin proteins, similar to AID[205]. In addition, both AID and A3B required the same β2 region for efficient nuclear import[205]. Thus, A3B and AID are both actively imported through a characteristic adaptor importin network and guided by a similar region in both of these proteins.

In the course of our research we also noticed several variations between AID and A3B indicating that these two proteins are regulated differently. For example, AID co-localizes with telophase chromosomes during mitosis, while A3B is excluded during this stage of cell division[205]. In addition, over-expression of A3B in AID deficient B-cells cannot restore class switching and AID cannot function to deaminate and restrict HIV-1 like A3B[205]. While we could not find evidence for shared functions between A3B and AID, we conclusively showed that A3B is consistently nuclear in a wide range of cell lines and is actively imported into the nuclear compartment. Interestingly, we know that A3B is actively transported into the nucleus, and we now know that A3B deaminates genomic DNA. Such active localization of a potentially mutagenic protein seems risky without an important functional role for A3B in the nucleus.

While testing the ability of AID and A3B to inhibit HIV-1 infection, we also explored the possibility that viral infection causes A3B to change localization, since its restriction of HIV-1 is thought to occur in the cytoplasm[151,156]. Although access to substrate seems a simple requirement (*i.e.,* nuclear A3B should not have cytoplasmic

functions and cytoplasmic A3B should not have nuclear functions), this link remains surprisingly blurry. We did not see a change in localization in A3B localization in HIV-1 infected cells and found that A3B restricts HIV-1 equally well in HEK293T and HeLa cells, despite a difference in localization between these lines[96,205]. Although other stimuli may affect the nuclear localization of A3B, our results show that it is a consistently nuclear enzyme.

Dynamic subcellular localization of the APOBEC family

One initial hypothesis for A3B's nuclear localization was that it interacted and bound to DNA during mitosis after the nuclear envelope disintegrated. This could make A3B appear nuclear without any need for active nuclear import. As shown in Chapters 2 and 3, this was definitely not the case for A3B, with the normally nuclear protein actually being excluded from mitotic DNA and only re-entering after the cells completed cytokinesis[205,234]. However, it is well known that A3G is capable of binding to single-stranded DNA during electrophoretic mobility shift assays[206,207]. Although this A3G is normally cytoplasmic, the mitotic breakdown of the nucleus envelope could theoretically allow cytoplasmic proteins, especially DNA-binding proteins like A3G, access to genomic DNA[215]. With this in mind we decided to test all the APOBEC3 proteins for mitotic localization, using AID for comparison.

As described in Chapter 3, we saw that, surprisingly, the APOBEC3 enzymes had very little interaction with DNA during mitosis and appeared to be excluded from condensed chromosomes[220,234]. Only subtle differences distinguished the single domain enzymes, which were excluded from DNA for the majority of mitosis and re-entered the

nucleus during telophase, from the double-domain enzymes, which remained excluded throughout mitosis[234]. With these results, it seems likely that some APOBEC3s, like A3G and A3F, rarely (if ever) have full access to genomic DNA. Other APOBEC3s that can deaminate genomic DNA are not likely to act during mitosis. Several publications suggest that AID deamination and error-prone repair occur during G1 phase; however, the telophase localization of AID on chromatin is intriguing and may indicate a window for extensive interaction between AID and genomic DNA[267-270].

Our localization results are consistent with APOBEC3 effects on cell cycle profiles. AID, A3A, A3B and A3H all have access to genomic DNA and all affect cell cycle progression, while A3C, which has access but appears to be a less active deaminase, does not affect cell cycle progression[234]. Interestingly, we noticed a mild effect with A3D on cell cycle profiles. While full-length A3D is cytoplasmic, the N-terminal domain of A3D is nuclear[234]. This suggests that A3D may have some previously undetected access to the nuclear compartment. Further work is needed to understand more specifically when and where the APOBEC3 proteins deaminate genomic DNA and how the exclusion phenotype seen during mitosis is mediated.

The vast majority of APOBEC localization data report static images in unperturbed cells. While these are useful for initial analyses, we suggest that some of the most interesting localization data occur in response to certain stimuli or over a period of time. Indeed, AID and A3G have been reported to change localization after DNA damage, both becoming more nuclear in response to irradiation[44,157]. In addition, our time-lapse imaging demonstrates the unexpected movement of the APOBEC3s during cell division[234]. Further research will demonstrate whether A3B or the other APOBEC3s

are affected by signals, like interferon induction, DNA damage and apoptosis. Although straightforward imaging has its uses, more sophisticated experiments may be necessary in the future for understanding the normal and abnormal activities of the APOBECs.

APOBEC oncogenes

The best-studied oncogenic deaminase is *AID*. This body of work provides a framework for studies on the oncogenic potential of other *APOBECs*. We know that while AID-dependent deamination causes beneficial recombination at the antibody locus[17], it can cause oncogenic translocations such as *c-myc*/IgH characteristic of B-cell cancers[211,227,271-277]. In addition, off-target AID-dependent point mutations can be identified in transcriptionally active genes like *MYC, PIM1* and *PAX5,* outside of the antibody variable regions[24,166,277,278]. These deamination events in mice cause a variety of cancers and in humans have been strongly correlated with B-cell cancers[25,210,279,280]. AID is the gold standard in the APOBEC field for an APOBEC oncogene as both ectopic and endogenous expression of human AID causes **1)** deamination of mammalian genomic DNA[281], **2)** translocations and mutations[26,277] and **3)** cancer development[282].

A3A have displayed some of the criteria for an APOBEC oncogene[132,163,283]. For instance, ectopic expression of A3A has been shown to induce genomic mutations in several human genes as well as in the mitochondria[131,133]. A3A also causes cell cycle defects and activation of the DNA damage response dependent on uracil DNA glycosylase activity, similar to the toxic effects of AID in repair deficient B-cells, implying that A3A is deaminating genomic DNA[131,223]. However, we have no evidence for genomic mutation or cell toxicity in primary macrophages or monocyte cell lines

expressing endogenous A3A[128,136]. In addition, A3A has not been linked with any human cancers.

Since we knew that A3B is actively imported into the nuclear compartment and spends the majority of the cell cycle in close proximity to genomic DNA, it was always a strong candidate for a role in on human mutation and cancer[205,234]. As illustrated in Chapter 4, we demonstrated that ectopic A3B expression causes cell cycle defects[234]. Using a tetracycline inducible system where we could express A3B in nearly 100% of cells, we detected A3B-induced toxicity and induction of the DNA damage response as well as genomic mutations[71]. However, the biggest difference between the prior work implicating A3A in genomic mutation was that we uniquely showed that *endogenous* A3B can deaminate genomic DNA[71]. Furthermore, we moved out of cell culture and correlated high levels of *endogenous* A3B expression with highly mutated cancerous breast tumors, but not normal breast tissues[71]. This is not adequate to conclusively define A3B as an oncogene, especially since it remains a possibility that A3B may function in a form of deamination-dependent apoptosis like AID and A3G[283,284]. However, we provide the first evidence, apart from AID, linking a human APOBEC enzyme to cancer and approach the gold standard for defining A3B as an APOBEC oncogene.

Future studies on APOBEC3-induced genomic mutations

Similar to breast cancer genome sequences, the mutations in many other human cancer sequences, including ovarian and brain cancers, have a high proportion of C-to-T transitions[228-233,239]. This preponderance of C-to-T mutations implies that A3B may have a role in breast cancer and other human cancers. If correct, this changes our perception of

mutations in cancer. In addition to spontaneous and environmental damage, we also have a common endogenous source of mutations that may be accelerating cancer progression and drug resistance. While additional studies on A3B in breast cancer will be informative, cytosine deamination by A3B may be a broader mechanism in carcinogenesis than we presently understand.

Interestingly, *A3B* is a non-essential gene and an *A3B* deletion allele is common world-wide[193]. It should be possible to study breast cancers, as well as other cancers, in humans with and without A3B. Such clinical studies will help define the effect of genomic cytosine deamination on cancer phenotypes. In combination with further molecular studies on mechanistic aspects of A3B, such as transcriptional regulation, these future experiments will help us understand the extent of A3B's contribution to genomic mutations as well as what effect A3B has on cancer prognosis.

While A3B is clearly involved in breast cancer, additional APOBEC3 proteins may also be involved in cancers. In support of this, we know that many APOBEC3 proteins are broadly expressed across a wide range of tissues[134,135]. Although we believe that our data show that A3B is the most important APOBEC in breast cancers, other APOBEC proteins are frequently expressed in breast tumors, often at higher levels than A3B, albeit not in a way that correlates them with tumor tissue[71]. These APOBECs could contribute to genomic mutations in a sporadic manner or in specific cancer sub-types. From transient expression experiments, we have identified AID, A3A, A3B, A3D and A3H as potential genomic deaminases[234]. Additional sequence information, which is becoming much more accessible and common, combined with expression data for the

APOBEC genes, will help correlate deamination mutations with cancer mutations, potentially in breast cancers, but also in different types of human cancers.

In conclusion, my thesis research has resulted in the description of a new role for A3B in genomic DNA deamination and breast cancers. I have implicated other APOBEC3 proteins in genomic deamination and correlated this ability to target genomic DNA with nuclear access through subcellular localization. My contributions in this area provide the groundwork for future investigation on the role of APOBEC3 proteins in genomic mutations and cancer.

References

1. Cohen, R. M. & Wolfenden, R. (1971). Cytidine deaminase from Escherichia coli. Purification, properties and inhibition by the potential transition state analog 3,4,5,6-tetrahydrouridine. *J Biol Chem* **246**, 7561-5.
2. Conticello, S. G., Thomas, C. J., Petersen-Mahrt, S. K. & Neuberger, M. S. (2005). Evolution of the AID/APOBEC family of polynucleotide (deoxy)cytidine deaminases. *Mol Biol Evol* **22**, 367-77.
3. Löffler, M., Fairbanks, L. D., Zameitat, E., Marinaki, A. M. & Simmonds, H. A. (2005). Pyrimidine pathways in health and disease. *Trends Mol Med* **11**, 430-7.
4. Conticello, S. G. (2012). Creative deaminases, self-inflicted damage, and genome evolution. *Ann N Y Acad Sci* **1267**, 79-85.
5. Conticello, S. G. (2008). The AID/APOBEC family of nucleic acid mutators. *Genome Biol* **9**, 229.
6. Gerber, A. P. & Keller, W. (1999). An adenosine deaminase that generates inosine at the wobble position of tRNAs. *Science* **286**, 1146-9.
7. Conticello, S. G., Ganesh, K., Xue, K., Lu, M., Rada, C. & Neuberger, M. S. (2008). Interaction between antibody-diversification enzyme AID and spliceosome-associated factor CTNNBL1. *Mol Cell* **31**, 474-84.
8. Teng, B., Burant, C. F. & Davidson, N. O. (1993). Molecular cloning of an apolipoprotein B messenger RNA editing protein. *Science* **260**, 1816-9.
9. Navaratnam, N., Morrison, J. R., Bhattacharya, S., Patel, D., Funahashi, T., Giannoni, F., Teng, B. B., Davidson, N. O. & Scott, J. (1993). The p27 catalytic subunit of the apolipoprotein B mRNA editing enzyme is a cytidine deaminase. *J Biol Chem* **268**, 20709-12.
10. Harris, R. S., Petersen-Mahrt, S. K. & Neuberger, M. S. (2002). RNA editing enzyme APOBEC1 and some of its homologs can act as DNA mutators. *Mol Cell* **10**, 1247-53.
11. Pancer, Z., Amemiya, C. T., Ehrhardt, G. R., Ceitlin, J., Gartland, G. L. & Cooper, M. D. (2004). Somatic diversification of variable lymphocyte receptors in the agnathan sea lamprey. *Nature* **430**, 174-80.
12. Boehm, T. (2012). Evolution of vertebrate immunity. *Curr Biol* **22**, R722-32.
13. Rogozin, I. B., Iyer, L. M., Liang, L., Glazko, G. V., Liston, V. G., Pavlov, Y. I., Aravind, L. & Pancer, Z. (2007). Evolution and diversification of lamprey antigen receptors: evidence for involvement of an AID-APOBEC family cytosine deaminase. *Nat Immunol* **8**, 647-56.
14. Wakae, K., Magor, B. G., Saunders, H., Nagaoka, H., Kawamura, A., Kinoshita, K., Honjo, T. & Muramatsu, M. (2006). Evolution of class switch recombination function in fish activation-induced cytidine deaminase, AID. *Int Immunol* **18**, 41-7.
15. Kato, L., Stanlie, A., Begum, N. A., Kobayashi, M., Aida, M. & Honjo, T. (2012). An evolutionary view of the mechanism for immune and genome diversity. *J Immunol* **188**, 3559-66.

16. Xian-guang, H., Aldridge, R. J., Siveter, D. J. & Xiang-hong, F. (2002). New evidence on the anatomy and phylogeny of the earliest vertebrates. *Proc Biol Sci* **269**, 1865-9.
17. Di Noia, J. M. & Neuberger, M. S. (2007). Molecular mechanisms of antibody somatic hypermutation. *Annu Rev Biochem* **76**, 1-22.
18. Muramatsu, M., Kinoshita, K., Fagarasan, S., Yamada, S., Shinkai, Y. & Honjo, T. (2000). Class switch recombination and hypermutation require activation-induced cytidine deaminase (AID), a potential RNA editing enzyme. *Cell* **102**, 553-63.
19. Muramatsu, M., Sankaranand, V. S., Anant, S., Sugai, M., Kinoshita, K., Davidson, N. O. & Honjo, T. (1999). Specific expression of activation-induced cytidine deaminase (AID), a novel member of the RNA-editing deaminase family in germinal center B cells. *J Biol Chem* **274**, 18470-6.
20. Revy, P., Muto, T., Levy, Y., Geissmann, F., Plebani, A., Sanal, O., Catalan, N., Forveille, M., Dufourcq-Labelouse, R., Gennery, A., Tezcan, I., Ersoy, F., Kayserili, H., Ugazio, A. G., Brousse, N., Muramatsu, M., Notarangelo, L. D., Kinoshita, K., Honjo, T., Fischer, A. & Durandy, A. (2000). Activation-induced cytidine deaminase (AID) deficiency causes the autosomal recessive form of the Hyper-IgM syndrome (HIGM2). *Cell* **102**, 565-75.
21. Stavnezer, J., Guikema, J. E. & Schrader, C. E. (2008). Mechanism and regulation of class switch recombination. *Annu Rev Immunol* **26**, 261-92.
22. Quartier, P., Bustamante, J., Sanal, O., Plebani, A., Debré, M., Deville, A., Litzman, J., Levy, J., Fermand, J. P., Lane, P., Horneff, G., Aksu, G., Yalcin, I., Davies, G., Tezcan, I., Ersoy, F., Catalan, N., Imai, K., Fischer, A. & Durandy, A. (2004). Clinical, immunologic and genetic analysis of 29 patients with autosomal recessive hyper-IgM syndrome due to Activation-Induced Cytidine Deaminase deficiency. *Clin Immunol* **110**, 22-9.
23. Imai, K., Zhu, Y., Revy, P., Morio, T., Mizutani, S., Fischer, A., Nonoyama, S. & Durandy, A. (2005). Analysis of class switch recombination and somatic hypermutation in patients affected with autosomal dominant hyper-IgM syndrome type 2. *Clin Immunol* **115**, 277-85.
24. Liu, M., Duke, J. L., Richter, D. J., Vinuesa, C. G., Goodnow, C. C., Kleinstein, S. H. & Schatz, D. G. (2008). Two levels of protection for the B cell genome during somatic hypermutation. *Nature* **451**, 841-5.
25. Pasqualucci, L., Bhagat, G., Jankovic, M., Compagno, M., Smith, P., Muramatsu, M., Honjo, T., Morse, H. C., 3rd, Nussenzweig, M. C. & Dalla-Favera, R. (2008). AID is required for germinal center-derived lymphomagenesis. *Nat Genet* **40**, 108-12.
26. Unniraman, S. & Schatz, D. G. (2006). AID and Igh switch region-Myc chromosomal translocations. *DNA Repair (Amst)* **5**, 1259-64.
27. Litinskiy, M. B., Nardelli, B., Hilbert, D. M., He, B., Schaffer, A., Casali, P. & Cerutti, A. (2002). DCs induce CD40-independent immunoglobulin class switching through BLyS and APRIL. *Nat Immunol* **3**, 822-9.

28. Kinoshita, K., Harigai, M., Fagarasan, S., Muramatsu, M. & Honjo, T. (2001). A hallmark of active class switch recombination: transcripts directed by I promoters on looped-out circular DNAs. *Proc Natl Acad Sci U S A* **98**, 12620-3.
29. Dedeoglu, F., Horwitz, B., Chaudhuri, J., Alt, F. W. & Geha, R. S. (2004). Induction of activation-induced cytidine deaminase gene expression by IL-4 and CD40 ligation is dependent on STAT6 and NFkappaB. *Int Immunol* **16**, 395-404.
30. Yadav, A., Olaru, A., Saltis, M., Setren, A., Cerny, J. & Livák, F. (2006). Identification of a ubiquitously active promoter of the murine activation-induced cytidine deaminase (AICDA) gene. *Mol Immunol* **43**, 529-41.
31. Delker, R. K., Fugmann, S. D. & Papavasiliou, F. N. (2009). A coming-of-age story: activation-induced cytidine deaminase turns 10. *Nat Immunol* **10**, 1147-53.
32. Dorsett, Y., McBride, K. M., Jankovic, M., Gazumyan, A., Thai, T. H., Robbiani, D. F., Di Virgilio, M., Reina San-Martin, B., Heidkamp, G., Schwickert, T. A., Eisenreich, T., Rajewsky, K. & Nussenzweig, M. C. (2008). MicroRNA-155 suppresses activation-induced cytidine deaminase-mediated Myc-Igh translocation. *Immunity* **28**, 630-8.
33. Teng, G., Hakimpour, P., Landgraf, P., Rice, A., Tuschl, T., Casellas, R. & Papavasiliou, F. N. (2008). MicroRNA-155 is a negative regulator of activation-induced cytidine deaminase. *Immunity* **28**, 621-9.
34. de Yébenes, V. G., Belver, L., Pisano, D. G., González, S., Villasante, A., Croce, C., He, L. & Ramiro, A. R. (2008). miR-181b negatively regulates activation-induced cytidine deaminase in B cells. *J Exp Med* **205**, 2199-206.
35. Storck, S., Aoufouchi, S., Weill, J. C. & Reynaud, C. A. (2011). AID and partners: for better and (not) for worse. *Curr Opin Immunol* **23**, 337-44.
36. Pavri, R., Gazumyan, A., Jankovic, M., Di Virgilio, M., Klein, I., Ansarah-Sobrinho, C., Resch, W., Yamane, A., Reina San-Martin, B., Barreto, V., Nieland, T. J., Root, D. E., Casellas, R. & Nussenzweig, M. C. (2010). Activation-induced cytidine deaminase targets DNA at sites of RNA polymerase II stalling by interaction with Spt5. *Cell* **143**, 122-33.
37. Stavnezer, J. (2011). Complex regulation and function of activation-induced cytidine deaminase. *Trends Immunol* **32**, 194-201.
38. Yamane, A., Resch, W., Kuo, N., Kuchen, S., Li, Z., Sun, H. W., Robbiani, D. F., McBride, K., Nussenzweig, M. C. & Casellas, R. (2011). Deep-sequencing identification of the genomic targets of the cytidine deaminase AID and its cofactor RPA in B lymphocytes. *Nat Immunol* **12**, 62-9.
39. Pavri, R. & Nussenzweig, M. C. (2011). AID targeting in antibody diversity. *Adv Immunol* **110**, 1-26.
40. Patenaude, A. M. & Di Noia, J. M. (2010). The mechanisms regulating the subcellular localization of AID. *Nucleus* **1**, 325-31.
41. Patenaude, A. M., Orthwein, A., Hu, Y., Campo, V. A., Kavli, B., Buschiazzo, A. & Di Noia, J. M. (2009). Active nuclear import and cytoplasmic retention of activation-induced deaminase. *Nat Struct Mol Biol* **16**, 517-27.
42. Ito, S., Nagaoka, H., Shinkura, R., Begum, N., Muramatsu, M., Nakata, M. & Honjo, T. (2004). Activation-induced cytidine deaminase shuttles between

nucleus and cytoplasm like apolipoprotein B mRNA editing catalytic polypeptide 1. *Proc Natl Acad Sci U S A* **101**, 1975-80.
43. McBride, K. M., Gazumyan, A., Woo, E. M., Barreto, V. M., Robbiani, D. F., Chait, B. T. & Nussenzweig, M. C. (2006). Regulation of hypermutation by activation-induced cytidine deaminase phosphorylation. *Proc Natl Acad Sci U S A* **103**, 8798-803.
44. Brar, S. S., Watson, M. & Diaz, M. (2004). Activation-induced cytosine deaminase (AID) is actively exported out of the nucleus but retained by the induction of DNA breaks. *J Biol Chem* **279**, 26395-401.
45. Rada, C., Jarvis, J. M. & Milstein, C. (2002). AID-GFP chimeric protein increases hypermutation of Ig genes with no evidence of nuclear localization. *Proc Natl Acad Sci U S A* **99**, 7003-8.
46. Mikl, M. C., Watt, I. N., Lu, M., Reik, W., Davies, S. L., Neuberger, M. S. & Rada, C. (2005). Mice deficient in APOBEC2 and APOBEC3. *Mol Cell Biol* **25**, 7270-7.
47. Sawyer, S. L., Emerman, M. & Malik, H. S. (2004). Ancient adaptive evolution of the primate antiviral DNA-editing enzyme APOBEC3G. *PLoS Biol* **2**, E275.
48. Liao, W., Hong, S. H., Chan, B. H., Rudolph, F. B., Clark, S. C. & Chan, L. (1999). APOBEC-2, a cardiac- and skeletal muscle-specific member of the cytidine deaminase supergene family. *Biochem Biophys Res Commun* **260**, 398-404.
49. Etard, C., Roostalu, U. & Strähle, U. (2010). Lack of Apobec2-related proteins causes a dystrophic muscle phenotype in zebrafish embryos. *J Cell Biol* **189**, 527-39.
50. Sato, Y., Probst, H. C., Tatsumi, R., Ikeuchi, Y., Neuberger, M. S. & Rada, C. (2010). Deficiency in APOBEC2 leads to a shift in muscle fiber type, diminished body mass, and myopathy. *J Biol Chem* **285**, 7111-8.
51. Vonica, A., Rosa, A., Arduini, B. L. & Brivanlou, A. H. (2011). APOBEC2, a selective inhibitor of TGFβ signaling, regulates left-right axis specification during early embryogenesis. *Dev Biol* **350**, 13-23.
52. Anant, S., Mukhopadhyay, D., Sankaranand, V., Kennedy, S., Henderson, J. O. & Davidson, N. O. (2001). ARCD-1, an apobec-1-related cytidine deaminase, exerts a dominant negative effect on C to U RNA editing. *Am J Physiol Cell Physiol* **281**, C1904-16.
53. Prochnow, C., Bransteitter, R., Klein, M. G., Goodman, M. F. & Chen, X. S. (2007). The APOBEC-2 crystal structure and functional implications for the deaminase AID. *Nature* **445**, 447-51.
54. Bransteitter, R., Prochnow, C. & Chen, X. S. (2009). The current structural and functional understanding of APOBEC deaminases. *Cell Mol Life Sci* **66**, 3137-47.
55. Severi, F., Chicca, A. & Conticello, S. G. (2010). Analysis of reptilian APOBEC1 suggests that RNA editing may not be its ancestral function. *Mol Biol Evol* **28**, 1125-9.
56. Benton, M. J. & Donoghue, P. C. (2007). Paleontological evidence to date the tree of life. *Mol Biol Evol* **24**, 26-53.

57. Powell, L. M., Wallis, S. C., Pease, R. J., Edwards, Y. H., Knott, T. J. & Scott, J. (1987). A novel form of tissue-specific RNA processing produces apolipoprotein-B48 in intestine. *Cell* **50**, 831-40.
58. Chen, S. H., Habib, G., Yang, C. Y., Gu, Z. W., Lee, B. R., Weng, S. A., Silberman, S. R., Cai, S. J., Deslypere, J. P., Rosseneu, M. & et al. (1987). Apolipoprotein B-48 is the product of a messenger RNA with an organ-specific in-frame stop codon. *Science* **238**, 363-6.
59. Lellek, H., Kirsten, R., Diehl, I., Apostel, F., Buck, F. & Greeve, J. (2000). Purification and molecular cloning of a novel essential component of the apolipoprotein B mRNA editing enzyme-complex. *J Biol Chem* **275**, 19848-56.
60. Mehta, A., Kinter, M. T., Sherman, N. E. & Driscoll, D. M. (2000). Molecular cloning of apobec-1 complementation factor, a novel RNA-binding protein involved in the editing of apolipoprotein B mRNA. *Mol Cell Biol* **20**, 1846-54.
61. Chester, A., Somasekaram, A., Tzimina, M., Jarmuz, A., Gisbourne, J., O'Keefe, R., Scott, J. & Navaratnam, N. (2003). The apolipoprotein B mRNA editing complex performs a multifunctional cycle and suppresses nonsense-mediated decay. *EMBO J* **22**, 3971-82.
62. Rosenberg, B. R., Hamilton, C. E., Mwangi, M. M., Dewell, S. & Papavasiliou, F. N. (2011). Transcriptome-wide sequencing reveals numerous APOBEC1 mRNA-editing targets in transcript 3' UTRs. *Nat Struct Mol Biol* **18**, 230-6.
63. Blanc, V., Xie, Y., Luo, J., Kennedy, S. & Davidson, N. O. (2012). Intestine-specific expression of Apobec-1 rescues apolipoprotein B RNA editing and alters chylomicron production in Apobec1-/- mice. *J Lipid Res*.
64. Hirano, K., Young, S. G., Farese, R. V., Jr., Ng, J., Sande, E., Warburton, C., Powell-Braxton, L. M. & Davidson, N. O. (1996). Targeted disruption of the mouse apobec-1 gene abolishes apolipoprotein B mRNA editing and eliminates apolipoprotein B48. *J Biol Chem* **271**, 9887-90.
65. Nakamuta, M., Chang, B. H., Zsigmond, E., Kobayashi, K., Lei, H., Ishida, B. Y., Oka, K., Li, E. & Chan, L. (1996). Complete phenotypic characterization of APOBEC-1 knockout mice with a wild-type genetic background and a human apolipoprotein B transgenic background, and restoration of apolipoprotein B mRNA editing by somatic gene transfer of APOBEC-1. *J Biol Chem* **271**, 25981-8.
66. Petersen-Mahrt, S. K. & Neuberger, M. S. (2003). In vitro deamination of cytosine to uracil in single-stranded DNA by apolipoprotein B editing complex catalytic subunit 1 (APOBEC1). *J Biol Chem* **278**, 19583-6.
67. Ikeda, T., Abd El Galil, K. H., Tokunaga, K., Maeda, K., Sata, T., Sakaguchi, N., Heidmann, T. & Koito, A. (2011). Intrinsic restriction activity by apolipoprotein B mRNA editing enzyme APOBEC1 against the mobility of autonomous retrotransposons. *Nucleic Acids Res* **39**, 5538-54.
68. Ikeda, T., Ohsugi, T., Kimura, T., Matsushita, S., Maeda, Y., Harada, S. & Koito, A. (2008). The antiretroviral potency of APOBEC1 deaminase from small animal species. *Nucleic Acids Res* **36**, 6859-71.

69. Petit, V., Guetard, D., Renard, M., Keriel, A., Sitbon, M., Wain-Hobson, S. & Vartanian, J. P. (2009). Murine APOBEC1 is a powerful mutator of retroviral and cellular RNA in vitro and in vivo. *J Mol Biol* **385**, 65-78.
70. Greeve, J., Altkemper, I., Dieterich, J. H., Greten, H. & Windler, E. (1993). Apolipoprotein B mRNA editing in 12 different mammalian species: hepatic expression is reflected in low concentrations of apoB-containing plasma lipoproteins. *J Lipid Res* **34**, 1367-83.
71. Burns, M. B., Lackey, L., Carpenter, M. A., Rathore, A., Land, A. M., Leonard, B., Refsland, E. W., Kotandeniya, D., Tretyakova, N., Nikas, J. B., Yee, D., Temiz, N. A., Donohue, D. E., McDougle, R. M., Brown, W. L., Law, E. K. & Harris, R. S. (2012). APOBEC3B is an enzymatic source of mutation in breast cancer. *Nature* **In revision**.
72. Funahashi, T., Giannoni, F., DePaoli, A. M., Skarosi, S. F. & Davidson, N. O. (1995). Tissue-specific, developmental and nutritional regulation of the gene encoding the catalytic subunit of the rat apolipoprotein B mRNA editing enzyme: functional role in the modulation of apoB mRNA editing. *J Lipid Res* **36**, 414-28.
73. von Wronski, M. A., Hirano, K. I., Cagen, L. M., Wilcox, H. G., Raghow, R., Thorngate, F. E., Heimberg, M., Davidson, N. O. & Elam, M. B. (1998). Insulin increases expression of apobec-1, the catalytic subunit of the apolipoprotein B mRNA editing complex in rat hepatocytes. *Metabolism* **47**, 869-73.
74. Chester, A., Scott, J., Anant, S. & Navaratnam, N. (2000). RNA editing: cytidine to uridine conversion in apolipoprotein B mRNA. *Biochim Biophys Acta* **1494**, 1-13.
75. Lau, P. P., Xiong, W. J., Zhu, H. J., Chen, S. H. & Chan, L. (1991). Apolipoprotein B mRNA editing is an intranuclear event that occurs posttranscriptionally coincident with splicing and polyadenylation. *J Biol Chem* **266**, 20550-4.
76. Sowden, M. P., Ballatori, N., Jensen, K. L., Reed, L. H. & Smith, H. C. (2002). The editosome for cytidine to uridine mRNA editing has a native complexity of 27S: identification of intracellular domains containing active and inactive editing factors. *J Cell Sci* **115**, 1027-39.
77. Yang, Y. & Smith, H. C. (1997). Multiple protein domains determine the cell type-specific nuclear distribution of the catalytic subunit required for apolipoprotein B mRNA editing. *Proc Natl Acad Sci U S A* **94**, 13075-80.
78. Blanc, V., Kennedy, S. & Davidson, N. O. (2003). A novel nuclear localization signal in the auxiliary domain of apobec-1 complementation factor regulates nucleocytoplasmic import and shuttling. *J Biol Chem* **278**, 41198-204.
79. LaRue, R. S., Jónsson, S. R., Silverstein, K. A., Lajoie, M., Bertrand, D., El-Mabrouk, N., Hötzel, I., Andrésdóttir, V., Smith, T. P. & Harris, R. S. (2008). 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* **9**, 104.
80. Münk, C., Willemsen, A. & Bravo, I. G. (2012). An ancient history of gene duplications, fusions and losses in the evolution of APOBEC3 mutators in mammals. *BMC Evol Biol* **12**, 71.

81. MacDuff, D. A., Demorest, Z. L. & Harris, R. S. (2009). AID can restrict L1 retrotransposition suggesting a dual role in innate and adaptive immunity. *Nucleic Acids Res* **37**, 1854-67.
82. Gourzi, P., Leonova, T. & Papavasiliou, F. N. (2006). A role for activation-induced cytidine deaminase in the host response against a transforming retrovirus. *Immunity* **24**, 779-86.
83. Albin, J. S. & Harris, R. S. (2010). Interactions of host APOBEC3 restriction factors with HIV-1 in vivo: implications for therapeutics. *Expert Rev Mol Med* **12**, e4.
84. Arias, J. F., Koyama, T., Kinomoto, M. & Tokunaga, K. (2012). Retroelements versus APOBEC3 family members: No great escape from the magnificent seven. *Front Microbiol* **3**, 275.
85. Sheehy, A. M., Gaddis, N. C., Choi, J. D. & Malim, M. H. (2002). Isolation of a human gene that inhibits HIV-1 infection and is suppressed by the viral Vif protein. *Nature* **418**, 646-50.
86. Harris, R. S., Bishop, K. N., Sheehy, A. M., Craig, H. M., Petersen-Mahrt, S. K., Watt, I. N., Neuberger, M. S. & Malim, M. H. (2003). DNA deamination mediates innate immunity to retroviral infection. *Cell* **113**, 803-9.
87. Mangeat, B., Turelli, P., Caron, G., Friedli, M., Perrin, L. & Trono, D. (2003). Broad antiretroviral defence by human APOBEC3G through lethal editing of nascent reverse transcripts. *Nature* **424**, 99-103.
88. Zhang, H., Yang, B., Pomerantz, R. J., Zhang, C., Arunachalam, S. C. & Gao, L. (2003). The cytidine deaminase CEM15 induces hypermutation in newly synthesized HIV-1 DNA. *Nature* **424**, 94-8.
89. Wiegand, H. L., Doehle, B. P., Bogerd, H. P. & Cullen, B. R. (2004). A second human antiretroviral factor, APOBEC3F, is suppressed by the HIV-1 and HIV-2 Vif proteins. *EMBO J* **23**, 2451-8.
90. Bishop, K. N., Holmes, R. K. & Malim, M. H. (2006). Antiviral potency of APOBEC proteins does not correlate with cytidine deamination. *J Virol* **80**, 8450-8.
91. Newman, E. N., Holmes, R. K., Craig, H. M., Klein, K. C., Lingappa, J. R., Malim, M. H. & Sheehy, A. M. (2005). Antiviral function of APOBEC3G can be dissociated from cytidine deaminase activity. *Curr Biol* **15**, 166-70.
92. Mariani, R., Chen, D., Schröfelbauer, B., Navarro, F., König, R., Bollman, B., Münk, C., Nymark-McMahon, H. & Landau, N. R. (2003). Species-specific exclusion of APOBEC3G from HIV-1 virions by Vif. *Cell* **114**, 21-31.
93. Marin, M., Rose, K. M., Kozak, S. L. & Kabat, D. (2003). HIV-1 Vif protein binds the editing enzyme APOBEC3G and induces its degradation. *Nat Med* **9**, 1398-403.
94. Bogerd, H. P., Doehle, B. P., Wiegand, H. L. & Cullen, B. R. (2004). A single amino acid difference in the host APOBEC3G protein controls the primate species specificity of HIV type 1 virion infectivity factor. *Proc Natl Acad Sci U S A* **101**, 3770-4.

95. Stopak, K., de Noronha, C., Yonemoto, W. & Greene, W. C. (2003). HIV-1 Vif blocks the antiviral activity of APOBEC3G by impairing both its translation and intracellular stability. *Mol Cell* **12**, 591-601.
96. Hultquist, J. F., Lengyel, J. A., Refsland, E. W., LaRue, R. S., Lackey, L., Brown, W. L. & Harris, R. S. (2011). Human and rhesus APOBEC3D, APOBEC3F, APOBEC3G, and APOBEC3H demonstrate a conserved capacity to restrict Vif-deficient HIV-1. *J Virol* **85**, 11220-34.
97. Refsland, E. W., Hultquist, J. F. & Harris, R. S. (2012). Endogenous origins of HIV-1 G-to-A hypermutation and restriction in the nonpermissive T cell line CEM2n. *PLoS Pathog* **8**, e1002800.
98. LaRue, R. S., Lengyel, J., Jónsson, S. R., Andrésdóttir, V. & Harris, R. S. (2010). Lentiviral Vif degrades the APOBEC3Z3/APOBEC3H protein of its mammalian host and is capable of cross-species activity. *J Virol* **84**, 8193-201.
99. Münk, C., Beck, T., Zielonka, J., Hotz-Wagenblatt, A., Chareza, S., Battenberg, M., Thielebein, J., Cichutek, K., Bravo, I. G., O'Brien, S. J., Löchelt, M. & Yuhki, N. (2008). Functions, structure, and read-through alternative splicing of feline APOBEC3 genes. *Genome Biol* **9**, R48.
100. Zielonka, J., Marino, D., Hofmann, H., Yuhki, N., Löchelt, M. & Münk, C. (2010). Vif of feline immunodeficiency virus from domestic cats protects against APOBEC3 restriction factors from many felids. *J Virol* **84**, 7312-24.
101. Ooms, M., Krikoni, A., Kress, A. K., Simon, V. & Münk, C. (2012). APOBEC3A, APOBEC3B, and APOBEC3H haplotype 2 restrict human T-lymphotropic virus type 1. *J Virol* **86**, 6097-108.
102. Mahieux, R., Suspène, R., Delebecque, F., Henry, M., Schwartz, O., Wain-Hobson, S. & Vartanian, J. P. (2005). Extensive editing of a small fraction of human T-cell leukemia virus type 1 genomes by four APOBEC3 cytidine deaminases. *J Gen Virol* **86**, 2489-94.
103. Sasada, A., Takaori-Kondo, A., Shirakawa, K., Kobayashi, M., Abudu, A., Hishizawa, M., Imada, K., Tanaka, Y. & Uchiyama, T. (2005). APOBEC3G targets human T-cell leukemia virus type 1. *Retrovirology* **2**, 32.
104. Günther, S., Sommer, G., Plikat, U., Iwanska, A., Wain-Hobson, S., Will, H. & Meyerhans, A. (1997). Naturally occurring hepatitis B virus genomes bearing the hallmarks of retroviral G-->A hypermutation. *Virology* **235**, 104-8.
105. Bonvin, M. & Greeve, J. (2008). Hepatitis B: modern concepts in pathogenesis--APOBEC3 cytidine deaminases as effectors in innate immunity against the hepatitis B virus. *Curr Opin Infect Dis* **21**, 298-303.
106. Chen, H., Lilley, C. E., Yu, Q., Lee, D. V., Chou, J., Narvaiza, I., Landau, N. R. & Weitzman, M. D. (2006). APOBEC3A is a potent inhibitor of adeno-associated virus and retrotransposons. *Curr Biol* **16**, 480-5.
107. Narvaiza, I., Linfesty, D. C., Greener, B. N., Hakata, Y., Pintel, D. J., Logue, E., Landau, N. R. & Weitzman, M. D. (2009). Deaminase-independent inhibition of parvoviruses by the APOBEC3A cytidine deaminase. *PLoS Pathog* **5**, e1000439.
108. Petit, V., Vartanian, J. P. & Wain-Hobson, S. (2009). Powerful mutators lurking in the genome. *Philos Trans R Soc Lond B Biol Sci* **364**, 705-15.

109. Burns, K. H. & Boeke, J. D. (2012). Human transposon tectonics. *Cell* **149**, 740-52.
110. Turelli, P., Vianin, S. & Trono, D. (2004). The innate antiretroviral factor APOBEC3G does not affect human LINE-1 retrotransposition in a cell culture assay. *J Biol Chem* **279**, 43371-3.
111. Muckenfuss, H., Hamdorf, M., Held, U., Perkovic, M., Löwer, J., Cichutek, K., Flory, E., Schumann, G. G. & Münk, C. (2006). APOBEC3 proteins inhibit human LINE-1 retrotransposition. *J Biol Chem* **281**, 22161-72.
112. Esnault, C., Heidmann, O., Delebecque, F., Dewannieux, M., Ribet, D., Hance, A. J., Heidmann, T. & Schwartz, O. (2005). APOBEC3G cytidine deaminase inhibits retrotransposition of endogenous retroviruses. *Nature* **433**, 430-3.
113. Esnault, C., Millet, J., Schwartz, O. & Heidmann, T. (2006). Dual inhibitory effects of APOBEC family proteins on retrotransposition of mammalian endogenous retroviruses. *Nucleic Acids Res* **34**, 1522-31.
114. Bogerd, H. P., Wiegand, H. L., Doehle, B. P., Lueders, K. K. & Cullen, B. R. (2006). APOBEC3A and APOBEC3B are potent inhibitors of LTR-retrotransposon function in human cells. *Nucleic Acids Res* **34**, 89-95.
115. Bogerd, H. P., Wiegand, H. L., Hulme, A. E., Garcia-Perez, J. L., O'Shea, K. S., Moran, J. V. & Cullen, B. R. (2006). Cellular inhibitors of long interspersed element 1 and Alu retrotransposition. *Proc Natl Acad Sci U S A* **103**, 8780-5.
116. Schumacher, A. J., Nissley, D. V. & Harris, R. S. (2005). APOBEC3G hypermutates genomic DNA and inhibits Ty1 retrotransposition in yeast. *Proc Natl Acad Sci U S A* **102**, 9854-9.
117. Stenglein, M. D. & Harris, R. S. (2006). APOBEC3B and APOBEC3F inhibit L1 retrotransposition by a DNA deamination-independent mechanism. *J Biol Chem* **281**, 16837-41.
118. Lee, Y. N. & Bieniasz, P. D. (2007). Reconstitution of an infectious human endogenous retrovirus. *PLoS Pathog* **3**, e10.
119. Kinomoto, M., Kanno, T., Shimura, M., Ishizaka, Y., Kojima, A., Kurata, T., Sata, T. & Tokunaga, K. (2007). All APOBEC3 family proteins differentially inhibit LINE-1 retrotransposition. *Nucleic Acids Res* **35**, 2955-64.
120. Niewiadomska, A. M., Tian, C., Tan, L., Wang, T., Sarkis, P. T. & Yu, X. F. (2007). Differential inhibition of long interspersed element 1 by APOBEC3 does not correlate with high-molecular-mass-complex formation or P-body association. *J Virol* **81**, 9577-83.
121. Hulme, A. E., Bogerd, H. P., Cullen, B. R. & Moran, J. V. (2007). Selective inhibition of Alu retrotransposition by APOBEC3G. *Gene* **390**, 199-205.
122. OhAinle, M., Kerns, J. A., Li, M. M., Malik, H. S. & Emerman, M. (2008). Antiretroelement activity of APOBEC3H was lost twice in recent human evolution. *Cell Host Microbe* **4**, 249-59.
123. Tan, L., Sarkis, P. T., Wang, T., Tian, C. & Yu, X. F. (2009). Sole copy of Z2-type human cytidine deaminase APOBEC3H has inhibitory activity against retrotransposons and HIV-1. *FASEB J* **23**, 279-87.

124. Khatua, A. K., Taylor, H. E., Hildreth, J. E. & Popik, W. (2010). Inhibition of LINE-1 and Alu retrotransposition by exosomes encapsidating APOBEC3G and APOBEC3F. *Virology* **400**, 68-75.
125. Wissing, S., Montano, M., Garcia-Perez, J. L., Moran, J. V. & Greene, W. C. (2011). Endogenous APOBEC3B restricts LINE-1 retrotransposition in transformed cells and human embryonic stem cells. *J Biol Chem*.
126. Ostertag, E. M., Prak, E. T., DeBerardinis, R. J., Moran, J. V. & Kazazian, H. H., Jr. (2000). Determination of L1 retrotransposition kinetics in cultured cells. *Nucleic Acids Res* **28**, 1418-23.
127. Babushok, D. V. & Kazazian, H. H., Jr. (2007). Progress in understanding the biology of the human mutagen LINE-1. *Hum Mutat* **28**, 527-39.
128. Stenglein, M. D., Burns, M. B., Li, M., Lengyel, J. & Harris, R. S. (2010). APOBEC3 proteins mediate the clearance of foreign DNA from human cells. *Nat Struct Mol Biol* **17**, 222-9.
129. Schumann, G. G. (2007). APOBEC3 proteins: major players in intracellular defence against LINE-1-mediated retrotransposition. *Biochem Soc Trans* **35**, 637-42.
130. Zhang, J. & Webb, D. M. (2004). Rapid evolution of primate antiviral enzyme APOBEC3G. *Hum Mol Genet* **13**, 1785-91.
131. Landry, S., Narvaiza, I., Linfesty, D. C. & Weitzman, M. D. (2011). APOBEC3A can activate the DNA damage response and cause cell-cycle arrest. *EMBO Rep* **12**, 444-50.
132. Yamanaka, S., Balestra, M. E., Ferrell, L. D., Fan, J., Arnold, K. S., Taylor, S., Taylor, J. M. & Innerarity, T. L. (1995). Apolipoprotein B mRNA-editing protein induces hepatocellular carcinoma and dysplasia in transgenic animals. *Proc Natl Acad Sci U S A* **92**, 8483-7.
133. Suspène, R., Aynaud, M. M., Guétard, D., Henry, M., Eckhoff, G., Marchio, A., Pineau, P., Dejean, A., Vartanian, J. P. & Wain-Hobson, S. (2011). Somatic hypermutation of human mitochondrial and nuclear DNA by APOBEC3 cytidine deaminases, a pathway for DNA catabolism. *Proc Natl Acad Sci U S A* **108**, 4858-63.
134. Refsland, E. W., Stenglein, M. D., Shindo, K., Albin, J. S., Brown, W. L. & Harris, R. S. (2010). Quantitative profiling of the full APOBEC3 mRNA repertoire in lymphocytes and tissues: implications for HIV-1 restriction. *Nucleic Acids Res* **38**, 4274-84.
135. Koning, F. A., Newman, E. N., Kim, E. Y., Kunstman, K. J., Wolinsky, S. M. & Malim, M. H. (2009). Defining APOBEC3 expression patterns in human tissues and hematopoietic cell subsets. *J Virol* **83**, 9474-85.
136. Carpenter, M. A., Li, M., Rathore, A., Lackey, L., Law, E. K., Land, A. M., Leonard, B., Shandilya, S. M., Bohn, M. F., Schiffer, C. A., Brown, W. L. & Harris, R. S. (2012). Methyl- and normal-cytosine deamination by the foreign DNA restriction enzyme APOBEC3A. *J Biol Chem* **Aug. 15 Online.**
137. Jarmuz, A., Chester, A., Bayliss, J., Gisbourne, J., Dunham, I., Scott, J. & Navaratnam, N. (2002). An anthropoid-specific locus of orphan C to U RNA-editing enzymes on chromosome 22. *Genomics* **79**, 285-96.

138. Morgan, H. D., Dean, W., Coker, H. A., Reik, W. & Petersen-Mahrt, S. K. (2004). Activation-induced cytidine deaminase deaminates 5-methylcytosine in DNA and is expressed in pluripotent tissues: implications for epigenetic reprogramming. *J Biol Chem* **279**, 52353-60.
139. OhAinle, M., Kerns, J. A., Malik, H. S. & Emerman, M. (2006). Adaptive evolution and antiviral activity of the conserved mammalian cytidine deaminase APOBEC3H. *J Virol* **80**, 3853-62.
140. Zhang, W., Zhang, X., Tian, C., Wang, T., Sarkis, P. T., Fang, Y., Zheng, S., Yu, X. F. & Xu, R. (2008). Cytidine deaminase APOBEC3B interacts with heterogeneous nuclear ribonucleoprotein K and suppresses hepatitis B virus expression. *Cell Microbiol* **10**, 112-21.
141. Chiu, Y. L., Witkowska, H. E., Hall, S. C., Santiago, M., Soros, V. B., Esnault, C., Heidmann, T. & Greene, W. C. (2006). High-molecular-mass APOBEC3G complexes restrict Alu retrotransposition. *Proc Natl Acad Sci U S A* **103**, 15588-93.
142. Kozak, S. L., Marin, M., Rose, K. M., Bystrom, C. & Kabat, D. (2006). The anti-HIV-1 editing enzyme APOBEC3G binds HIV-1 RNA and messenger RNAs that shuttle between polysomes and stress granules. *J Biol Chem* **281**, 29105-19.
143. Gallois-Montbrun, S., Holmes, R. K., Swanson, C. M., Fernández-Ocaña, M., Byers, H. L., Ward, M. A. & Malim, M. H. (2008). Comparison of cellular ribonucleoprotein complexes associated with the APOBEC3F and APOBEC3G antiviral proteins. *J Virol* **82**, 5636-42.
144. Gallois-Montbrun, S., Kramer, B., Swanson, C. M., Byers, H., Lynham, S., Ward, M. & Malim, M. H. (2007). Antiviral protein APOBEC3G localizes to ribonucleoprotein complexes found in P bodies and stress granules. *J Virol* **81**, 2165-78.
145. Huang, J., Liang, Z., Yang, B., Tian, H., Ma, J. & Zhang, H. (2007). Derepression of microRNA-mediated protein translation inhibition by apolipoprotein B mRNA-editing enzyme catalytic polypeptide-like 3G (APOBEC3G) and its family members. *J Biol Chem* **282**, 33632-40.
146. Liu, C., Zhang, X., Huang, F., Yang, B., Li, J., Liu, B., Luo, H., Zhang, P. & Zhang, H. (2012). APOBEC3G Inhibits microRNA-mediated repression of translation by interfering with the interaction between Argonaute-2 and MOV10. *J Biol Chem* **287**, 29373-83.
147. Phalora, P. K., Sherer, N. M., Wolinsky, S. M., Swanson, C. M. & Malim, M. H. (2012). HIV-1 replication and APOBEC3 antiviral activity are not regulated by P-bodies. *J Virol*.
148. Aynaud, M. M., Suspène, R., Vidalain, P. O., Mussil, B., Guétard, D., Tangy, F., Wain-Hobson, S. & Vartanian, J. P. (2012). Human Tribbles 3 protects nuclear DNA from cytidine deamination by APOBEC3A. *J Biol Chem*.
149. Bennett, R. P., Diner, E., Sowden, M. P., Lees, J. A., Wedekind, J. E. & Smith, H. C. (2006). APOBEC-1 and AID are nucleo-cytoplasmic trafficking proteins but APOBEC3G cannot traffic. *Biochem Biophys Res Commun* **350**, 214-9.

150. Stenglein, M. D., Matsuo, H. & Harris, R. S. (2008). Two regions within the amino-terminal half of APOBEC3G cooperate to determine cytoplasmic localization. *J Virol* **82**, 9591-9.
151. Marin, M., Golem, S., Rose, K. M., Kozak, S. L. & Kabat, D. (2008). Human immunodeficiency virus type 1 Vif functionally interacts with diverse APOBEC3 cytidine deaminases and moves with them between cytoplasmic sites of mRNA metabolism. *J Virol* **82**, 987-98.
152. Goila-Gaur, R., Khan, M. A., Miyagi, E., Kao, S. & Strebel, K. (2007). Targeting APOBEC3A to the viral nucleoprotein complex confers antiviral activity. *Retrovirology* **4**, 61.
153. Bonvin, M., Achermann, F., Greeve, I., Stroka, D., Keogh, A., Inderbitzin, D., Candinas, D., Sommer, P., Wain-Hobson, S., Vartanian, J. P. & Greeve, J. (2006). Interferon-inducible expression of APOBEC3 editing enzymes in human hepatocytes and inhibition of hepatitis B virus replication. *Hepatology* **43**, 1364-74.
154. Wang, X., Abudu, A., Son, S., Dang, Y., Venta, P. J. & Zheng, Y. H. (2011). Analysis of human APOBEC3H haplotypes and anti-human immunodeficiency virus type 1 activity. *J Virol* **85**, 3142-52.
155. Wichroski, M. J., Robb, G. B. & Rana, T. M. (2006). Human retroviral host restriction factors APOBEC3G and APOBEC3F localize to mRNA processing bodies. *PLoS Pathog* **2**, e41.
156. Pak, V., Heidecker, G., Pathak, V. K. & Derse, D. (2011). The role of amino-terminal sequences in cellular localization and antiviral activity of APOBEC3B. *J Virol* **85**, 8538-47.
157. Nowarski, R., Wilner, O. I., Cheshin, O., Shahar, O. D., Kenig, E., Baraz, L., Britan-Rosich, E., Nagler, A., Harris, R. S., Goldberg, M., Willner, I. & Kotler, M. (2012). APOBEC3G enhances lymphoma cell radioresistance by promoting cytidine deaminase-dependent DNA repair. *Blood* **120**, 366-75.
158. Collin, S. P. (2009). Early evolution of vertebrate photoreception: lessons from lampreys and lungfishes. *Integr Zool* **4**, 87-98.
159. Rogozin, I. B., Basu, M. K., Jordan, I. K., Pavlov, Y. I. & Koonin, E. V. (2005). APOBEC4, a new member of the AID/APOBEC family of polynucleotide (deoxy)cytidine deaminases predicted by computational analysis. *Cell Cycle* **4**, 1281-5.
160. Zielonka, J., Bravo, I. G., Marino, D., Conrad, E., Perković, M., Battenberg, M., Cichutek, K. & Münk, C. (2009). Restriction of equine infectious anemia virus by equine APOBEC3 cytidine deaminases. *J Virol* **83**, 7547-59.
161. LaRue, R. S., Andrésdóttir, V., Blanchard, Y., Conticello, S. G., Derse, D., Emerman, M., Greene, W. C., Jónsson, S. R., Landau, N. R., Löchelt, M., Malik, H. S., Malim, M. H., Münk, C., O'Brien, S. J., Pathak, V. K., Strebel, K., Wain-Hobson, S., Yu, X. F., Yuhki, N. & Harris, R. S. (2009). Guidelines for naming nonprimate APOBEC3 genes and proteins. *J Virol* **83**, 494-7.
162. Xu, Z., Zan, H., Pone, E. J., Mai, T. & Casali, P. (2012). Immunoglobulin class-switch DNA recombination: induction, targeting and beyond. *Nat Rev Immunol* **12**, 517-31.

163. Blanc, V. & Davidson, N. O. (2010). APOBEC-1-mediated RNA editing. *Wiley Interdiscip Rev Syst Biol Med* **2**, 594-602.
164. MacDuff, D. A. & Harris, R. S. (2006). Directed DNA deamination by AID/APOBEC3 in immunity. *Curr Biol* **16**, R186-9.
165. Yoneda, Y. (2000). Nucleocytoplasmic protein traffic and its significance to cell function. *Genes Cells* **5**, 777-87.
166. Liu, M. & Schatz, D. G. (2009). Balancing AID and DNA repair during somatic hypermutation. *Trends Immunol* **30**, 173-81.
167. Hirano, M., Das, S., Guo, P. & Cooper, M. D. (2011). The evolution of adaptive immunity in vertebrates. *Adv Immunol* **109**, 125-57.
168. Harris, R. S. & Liddament, M. T. (2004). Retroviral restriction by APOBEC proteins. *Nat Rev Immunol* **4**, 868-77.
169. McBride, K. M., Barreto, V., Ramiro, A. R., Stavropoulos, P. & Nussenzweig, M. C. (2004). Somatic hypermutation is limited by CRM1-dependent nuclear export of activation-induced deaminase. *J Exp Med* **199**, 1235-44.
170. Ichikawa, H. T., Sowden, M. P., Torelli, A. T., Bachl, J., Huang, P., Dance, G. S., Marr, S. H., Robert, J., Wedekind, J. E., Smith, H. C. & Bottaro, A. (2006). Structural phylogenetic analysis of activation-induced deaminase function. *J Immunol* **177**, 355-61.
171. Barreto, V. M. & Magor, B. G. (2011). Activation-induced cytidine deaminase structure and functions: a species comparative view. *Dev Comp Immunol* **35**, 991-1007.
172. Yang, Y., Sowden, M. P. & Smith, H. C. (2001). Intracellular trafficking determinants in APOBEC-1, the catalytic subunit for cytidine to uridine editing of apolipoprotein B mRNA. *Exp Cell Res* **267**, 153-64.
173. Mehta, A., Banerjee, S. & Driscoll, D. M. (1996). Apobec-1 interacts with a 65-kDa complementing protein to edit apolipoprotein-B mRNA in vitro. *J Biol Chem* **271**, 28294-9.
174. Wolff, B., Sanglier, J. J. & Wang, Y. (1997). Leptomycin B is an inhibitor of nuclear export: inhibition of nucleo-cytoplasmic translocation of the human immunodeficiency virus type 1 (HIV-1) Rev protein and Rev-dependent mRNA. *Chem Biol* **4**, 139-47.
175. Gasiorowski, J. Z. & Dean, D. A. (2003). Mechanisms of nuclear transport and interventions. *Adv Drug Deliv Rev* **55**, 703-16.
176. Mohr, D., Frey, S., Fischer, T., Güttler, T. & Görlich, D. (2009). Characterisation of the passive permeability barrier of nuclear pore complexes. *EMBO J* **28**, 2541-53.
177. Ribbeck, K. & Görlich, D. (2001). Kinetic analysis of translocation through nuclear pore complexes. *EMBO J* **20**, 1320-30.
178. Adam, S. A., Marr, R. S. & Gerace, L. (1990). Nuclear protein import in permeabilized mammalian cells requires soluble cytoplasmic factors. *J Cell Biol* **111**, 807-16.
179. Kutay, U., Izaurralde, E., Bischoff, F. R., Mattaj, I. W. & Görlich, D. (1997). Dominant-negative mutants of importinβ block multiple pathways of import and export through the nuclear pore complex. *EMBO J* **16**, 1153-63.

180. Aoufouchi, S., Faili, A., Zober, C., D'Orlando, O., Weller, S., Weill, J. C. & Reynaud, C. A. (2008). Proteasomal degradation restricts the nuclear lifespan of AID. *J Exp Med* **205**, 1357-68.
181. Arnold, K., Bordoli, L., Kopp, J. & Schwede, T. (2006). The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics* **22**, 195-201.
182. Bordoli, L., Kiefer, F., Arnold, K., Benkert, P., Battey, J. & Schwede, T. (2009). Protein structure homology modeling using SWISS-MODEL workspace. *Nat Protoc* **4**, 1-13.
183. Shandilya, S. M., Nalam, M. N., Nalivaika, E. A., Gross, P. J., Valesano, J. C., Shindo, K., Li, M., Munson, M., Royer, W. E., Harjes, E., Kono, T., Matsuo, H., Harris, R. S., Somasundaran, M. & Schiffer, C. A. (2010). Crystal structure of the APOBEC3G catalytic domain reveals potential oligomerization interfaces. *Structure* **18**, 28-38.
184. Pemberton, L. F. & Paschal, B. M. (2005). Mechanisms of receptor-mediated nuclear import and nuclear export. *Traffic* **6**, 187-98.
185. Goldfarb, D. S., Corbett, A. H., Mason, D. A., Harreman, M. T. & Adam, S. A. (2004). Importin α: a multipurpose nuclear-transport receptor. *Trends Cell Biol* **14**, 505-14.
186. Görlich, D. & Kutay, U. (1999). Transport between the cell nucleus and the cytoplasm. *Annu Rev Cell Dev Biol* **15**, 607-60.
187. Zheng, Y. H., Irwin, D., Kurosu, T., Tokunaga, K., Sata, T. & Peterlin, B. M. (2004). Human APOBEC3F is another host factor that blocks human immunodeficiency virus type 1 replication. *J Virol* **78**, 6073-6.
188. Bishop, K. N., Holmes, R. K., Sheehy, A. M., Davidson, N. O., Cho, S. J. & Malim, M. H. (2004). Cytidine deamination of retroviral DNA by diverse APOBEC proteins. *Curr Biol* **14**, 1392-6.
189. Schumacher, A. J., Haché, G., MacDuff, D. A., Brown, W. L. & Harris, R. S. (2008). The DNA deaminase activity of human APOBEC3G is required for Ty1, MusD, and human immunodeficiency virus type 1 restriction. *J Virol* **82**, 2652-60.
190. Wissing, S., Galloway, N. L. & Greene, W. C. (2010). HIV-1 Vif versus the APOBEC3 cytidine deaminases: an intracellular duel between pathogen and host restriction factors. *Mol Aspects Med* **31**, 383-97.
191. Bogerd, H. P., Wiegand, H. L., Doehle, B. P. & Cullen, B. R. (2007). The intrinsic antiretroviral factor APOBEC3B contains two enzymatically active cytidine deaminase domains. *Virology* **364**, 486-93.
192. Rosenberg, B. R. & Papavasiliou, F. N. (2007). Beyond SHM and CSR: AID and related cytidine deaminases in the host response to viral infection. *Adv Immunol* **94**, 215-44.
193. Kidd, J. M., Newman, T. L., Tuzun, E., Kaul, R. & Eichler, E. E. (2007). Population stratification of a common APOBEC gene deletion polymorphism. *PLoS Genet* **3**, e63.

194. Eto, T., Kinoshita, K., Yoshikawa, K., Muramatsu, M. & Honjo, T. (2003). RNA-editing cytidine deaminase Apobec-1 is unable to induce somatic hypermutation in mammalian cells. *Proc Natl Acad Sci U S A* **100**, 12895-8.
195. Krause, K., Marcu, K. B. & Greeve, J. (2006). The cytidine deaminases AID and APOBEC-1 exhibit distinct functional properties in a novel yeast selectable system. *Mol Immunol* **43**, 295-307.
196. Han, L., Masani, S. & Yu, K. (2010). Cutting edge: CTNNBL1 is dispensable for Ig class switch recombination. *J Immunol* **185**, 1379-81.
197. Ganesh, K., Adam, S., Taylor, B., Simpson, P., Rada, C. & Neuberger, M. (2011). CTNNBL1 is a novel nuclear localization sequence-binding protein that recognizes RNA-splicing factors CDC5L and Prp31. *J Biol Chem* **286**, 17091-102.
198. Maeda, K., Singh, S. K., Eda, K., Kitabatake, M., Pham, P., Goodman, M. F. & Sakaguchi, N. (2010). GANP-mediated recruitment of activation-induced cytidine deaminase to cell nuclei and to immunoglobulin variable region DNA. *J Biol Chem* **285**, 23945-53.
199. Chen, Z., Eggerman, T. L., Bocharov, A. V., Baranova, I. N., Vishnyakova, T. G., Kurlander, R. J., Csako, G. & Patterson, A. P. (2012). Hypermutation of ApoB mRNA by Rat APOBEC-1 Overexpression Mimics APOBEC-3 Hypermutation. *J Mol Biol*.
200. Demorest, Z. L., Li, M. & Harris, R. S. (2011). Phosphorylation directly regulates the intrinsic DNA cytidine deaminase activity of activation-induced deaminase and APOBEC3G protein. *J Biol Chem* **286**, 26568-75.
201. Kamentsky, L., Jones, T. R., Fraser, A., Bray, M. A., Logan, D. J., Madden, K. L., Ljosa, V., Rueden, C., Eliceiri, K. W. & Carpenter, A. E. (2011). Improved structure, function and compatibility for CellProfiler: modular high-throughput image analysis software. *Bioinformatics* **27**, 1179-80.
202. Carpenter, A. E., Jones, T. R., Lamprecht, M. R., Clarke, C., Kang, I. H., Friman, O., Guertin, D. A., Chang, J. H., Lindquist, R. A., Moffat, J., Golland, P. & Sabatini, D. M. (2006). CellProfiler: image analysis software for identifying and quantifying cell phenotypes. *Genome Biol* **7**, R100.
203. Benkert, P., Biasini, M. & Schwede, T. (2011). Toward the estimation of the absolute quality of individual protein structure models. *Bioinformatics* **27**, 343-50.
204. Humphrey, W., Dalke, A. & Schulten, K. (1996). VMD: visual molecular dynamics. *J Mol Graph* **14**, 33-8, 27-8.
205. Lackey, L., Demorest, Z. L., Land, A. M., Hultquist, J. F., Brown, W. L. & Harris, R. S. (2012). APOBEC3B and AID have similar nuclear import mechanisms. *J Mol Biol* **419**, 301-14.
206. Chelico, L., Pham, P., Calabrese, P. & Goodman, M. F. (2006). APOBEC3G DNA deaminase acts processively 3' --> 5' on single-stranded DNA. *Nat Struct Mol Biol* **13**, 392-9.
207. Chen, K. M., Harjes, E., Gross, P. J., Fahmy, A., Lu, Y., Shindo, K., Harris, R. S. & Matsuo, H. (2008). Structure of the DNA deaminase domain of the HIV-1 restriction factor APOBEC3G. *Nature* **452**, 116-9.

208. Aguiar, R. S. & Peterlin, B. M. (2008). APOBEC3 proteins and reverse transcription. *Virus Res* **134**, 74-85.
209. Malim, M. H. & Emerman, M. (2008). HIV-1 accessory proteins--ensuring viral survival in a hostile environment. *Cell Host Microbe* **3**, 388-98.
210. Okazaki, I. M., Hiai, H., Kakazu, N., Yamada, S., Muramatsu, M., Kinoshita, K. & Honjo, T. (2003). Constitutive expression of AID leads to tumorigenesis. *J Exp Med* **197**, 1173-81.
211. Robbiani, D. F., Bothmer, A., Callen, E., Reina-San-Martin, B., Dorsett, Y., Difilippantonio, S., Bolland, D. J., Chen, H. T., Corcoran, A. E., Nussenzweig, A. & Nussenzweig, M. C. (2008). AID is required for the chromosomal breaks in c-myc that lead to c-myc/IgH translocations. *Cell* **135**, 1028-38.
212. Enari, M., Sakahira, H., Yokoyama, H., Okawa, K., Iwamatsu, A. & Nagata, S. (1998). A caspase-activated DNase that degrades DNA during apoptosis, and its inhibitor ICAD. *Nature* **391**, 43-50.
213. Reich, N. C. & Liu, L. (2006). Tracking STAT nuclear traffic. *Nat Rev Immunol* **6**, 602-12.
214. Pelzer, C. & Thome, M. (2011). IKKα takes control of canonical NF-κB activation. *Nat Immunol* **12**, 815-6.
215. Güttinger, S., Laurell, E. & Kutay, U. (2009). Orchestrating nuclear envelope disassembly and reassembly during mitosis. *Nat Rev Mol Cell Biol* **10**, 178-91.
216. Estrada-Gelonch, A., Aramburu, J. & López-Rodriguez, C. (2009). Exclusion of NFAT5 from mitotic chromatin resets its nucleo-cytoplasmic distribution in interphase. *PLoS One* **4**, e7036.
217. Zaidi, S. K., Young, D. W., Pockwinse, S. M., Javed, A., Lian, J. B., Stein, J. L., van Wijnen, A. J. & Stein, G. S. (2003). Mitotic partitioning and selective reorganization of tissue-specific transcription factors in progeny cells. *Proc Natl Acad Sci U S A* **100**, 14852-7.
218. Blanc, V., Henderson, J. O., Kennedy, S. & Davidson, N. O. (2001). Mutagenesis of apobec-1 complementation factor reveals distinct domains that modulate RNA binding, protein-protein interaction with apobec-1, and complementation of C to U RNA-editing activity. *J Biol Chem* **276**, 46386-93.
219. Bennett, R. P., Presnyak, V., Wedekind, J. E. & Smith, H. C. (2008). Nuclear Exclusion of the HIV-1 host defense factor APOBEC3G requires a novel cytoplasmic retention signal and is not dependent on RNA binding. *J Biol Chem* **283**, 7320-7.
220. Martin, R. M. & Cardoso, M. C. (2010). Chromatin condensation modulates access and binding of nuclear proteins. *FASEB J* **24**, 1066-72.
221. Harper, J. V. (2005). Synchronization of cell populations in G1/S and G2/M phases of the cell cycle. *Methods Mol Biol* **296**, 157-66.
222. Yu, Q., Chen, D., König, R., Mariani, R., Unutmaz, D. & Landau, N. R. (2004). APOBEC3B and APOBEC3C are potent inhibitors of simian immunodeficiency virus replication. *J Biol Chem* **279**, 53379-86.
223. Hasham, M. G., Donghia, N. M., Coffey, E., Maynard, J., Snow, K. J., Ames, J., Wilpan, R. Y., He, Y., King, B. L. & Mills, K. D. (2010). Widespread genomic

breaks generated by activation-induced cytidine deaminase are prevented by homologous recombination. *Nat Immunol* **11**, 820-6.
224. Lassen, K. G., Wissing, S., Lobritz, M. A., Santiago, M. & Greene, W. C. (2010). Identification of two APOBEC3F splice variants displaying HIV-1 antiviral activity and contrasting sensitivity to Vif. *J Biol Chem* **285**, 29326-35.
225. Duggal, N. K., Malik, H. S. & Emerman, M. (2011). The breadth of antiviral activity of Apobec3DE in chimpanzees has been driven by positive selection. *J Virol* **85**, 11361-71.
226. Dang, Y., Abudu, A., Son, S., Harjes, E., Spearman, P., Matsuo, H. & Zheng, Y. H. (2011). Identification of a single amino acid required for APOBEC3 antiretroviral cytidine deaminase activity. *J Virol* **85**, 5691-5.
227. Ramiro, A. R., Jankovic, M., Eisenreich, T., Difilippantonio, S., Chen-Kiang, S., Muramatsu, M., Honjo, T., Nussenzweig, A. & Nussenzweig, M. C. (2004). AID is required for c-myc/IgH chromosome translocations in vivo. *Cell* **118**, 431-8.
228. Nik-Zainal, S., Alexandrov, L. B., Wedge, D. C., Van Loo, P., Greenman, C. D., Raine, K., Jones, D., Hinton, J., Marshall, J., Stebbings, L. A., Menzies, A., Martin, S., Leung, K., Chen, L., Leroy, C., Ramakrishna, M., Rance, R., Lau, K. W., Mudie, L. J., Varela, I., McBride, D. J., Bignell, G. R., Cooke, S. L., Shlien, A., Gamble, J., Whitmore, I., Maddison, M., Tarpey, P. S., Davies, H. R., Papaemmanuil, E., Stephens, P. J., McLaren, S., Butler, A. P., Teague, J. W., Jönsson, G., Garber, J. E., Silver, D., Miron, P., Fatima, A., Boyault, S., Langerød, A., Tutt, A., Martens, J. W., Aparicio, S. A., Borg, Å., Salomon, A. V., Thomas, G., Børresen-Dale, A. L., Richardson, A. L., Neuberger, M. S., Futreal, P. A., Campbell, P. J. & Stratton, M. R. (2012). Mutational processes molding the genomes of 21 breast cancers. *Cell* **149**, 979-93.
229. Greenman, C., Stephens, P., Smith, R., Dalgliesh, G. L., Hunter, C., Bignell, G., Davies, H., Teague, J., Butler, A., Stevens, C., Edkins, S., O'Meara, S., Vastrik, I., Schmidt, E. E., Avis, T., Barthorpe, S., Bhamra, G., Buck, G., Choudhury, B., Clements, J., Cole, J., Dicks, E., Forbes, S., Gray, K., Halliday, K., Harrison, R., Hills, K., Hinton, J., Jenkinson, A., Jones, D., Menzies, A., Mironenko, T., Perry, J., Raine, K., Richardson, D., Shepherd, R., Small, A., Tofts, C., Varian, J., Webb, T., West, S., Widaa, S., Yates, A., Cahill, D. P., Louis, D. N., Goldstraw, P., Nicholson, A. G., Brasseur, F., Looijenga, L., Weber, B. L., Chiew, Y. E., DeFazio, A., Greaves, M. F., Green, A. R., Campbell, P., Birney, E., Easton, D. F., Chenevix-Trench, G., Tan, M. H., Khoo, S. K., Teh, B. T., Yuen, S. T., Leung, S. Y., Wooster, R., Futreal, P. A. & Stratton, M. R. (2007). Patterns of somatic mutation in human cancer genomes. *Nature* **446**, 153-8.
230. Agrawal, N., Frederick, M. J., Pickering, C. R., Bettegowda, C., Chang, K., Li, R. J., Fakhry, C., Xie, T. X., Zhang, J., Wang, J., Zhang, N., El-Naggar, A. K., Jasser, S. A., Weinstein, J. N., Treviño, L., Drummond, J. A., Muzny, D. M., Wu, Y., Wood, L. D., Hruban, R. H., Westra, W. H., Koch, W. M., Califano, J. A., Gibbs, R. A., Sidransky, D., Vogelstein, B., Velculescu, V. E., Papadopoulos, N., Wheeler, D. A., Kinzler, K. W. & Myers, J. N. (2011). Exome sequencing of head and neck squamous cell carcinoma reveals inactivating mutations in NOTCH1. *Science* **333**, 1154-7.

231. Jones, S., Wang, T. L., Shih Ie, M., Mao, T. L., Nakayama, K., Roden, R., Glas, R., Slamon, D., Diaz, L. A., Jr., Vogelstein, B., Kinzler, K. W., Velculescu, V. E. & Papadopoulos, N. (2010). Frequent mutations of chromatin remodeling gene ARID1A in ovarian clear cell carcinoma. *Science* **330**, 228-31.
232. Yachida, S., Jones, S., Bozic, I., Antal, T., Leary, R., Fu, B., Kamiyama, M., Hruban, R. H., Eshleman, J. R., Nowak, M. A., Velculescu, V. E., Kinzler, K. W., Vogelstein, B. & Iacobuzio-Donahue, C. A. (2010). Distant metastasis occurs late during the genetic evolution of pancreatic cancer. *Nature* **467**, 1114-7.
233. Berger, M. F., Lawrence, M. S., Demichelis, F., Drier, Y., Cibulskis, K., Sivachenko, A. Y., Sboner, A., Esgueva, R., Pflueger, D., Sougnez, C., Onofrio, R., Carter, S. L., Park, K., Habegger, L., Ambrogio, L., Fennell, T., Parkin, M., Saksena, G., Voet, D., Ramos, A. H., Pugh, T. J., Wilkinson, J., Fisher, S., Winckler, W., Mahan, S., Ardlie, K., Baldwin, J., Simons, J. W., Kitabayashi, N., MacDonald, T. Y., Kantoff, P. W., Chin, L., Gabriel, S. B., Gerstein, M. B., Golub, T. R., Meyerson, M., Tewari, A., Lander, E. S., Getz, G., Rubin, M. A. & Garraway, L. A. (2011). The genomic complexity of primary human prostate cancer. *Nature* **470**, 214-20.
234. Lackey, L., Law, E. K., Brown, W. L. & Harris, R. S. (2012). Subcellular localization of the APOBEC3 proteins during mitosis and implications for genomic DNA deamination. *Cell Cycle* **in review**.
235. Shlyakhtenko, L. S., Lushnikov, A. Y., Li, M., Lackey, L., Harris, R. S. & Lyubchenko, Y. L. (2011). Atomic force microscopy studies provide direct evidence for dimerization of the HIV restriction factor APOBEC3G. *J Biol Chem* **286**, 3387-95.
236. Sjoblom, T., Jones, S., Wood, L. D., Parsons, D. W., Lin, J., Barber, T. D., Mandelker, D., Leary, R. J., Ptak, J., Silliman, N., Szabo, S., Buckhaults, P., Farrell, C., Meeh, P., Markowitz, S. D., Willis, J., Dawson, D., Willson, J. K., Gazdar, A. F., Hartigan, J., Wu, L., Liu, C., Parmigiani, G., Park, B. H., Bachman, K. E., Papadopoulos, N., Vogelstein, B., Kinzler, K. W. & Velculescu, V. E. (2006). The consensus coding sequences of human breast and colorectal cancers. *Science* **314**, 268-74.
237. Kumar, A., White, T. A., MacKenzie, A. P., Clegg, N., Lee, C., Dumpit, R. F., Coleman, I., Ng, S. B., Salipante, S. J., Rieder, M. J., Nickerson, D. A., Corey, E., Lange, P. H., Morrissey, C., Vessella, R. L., Nelson, P. S. & Shendure, J. (2011). Exome sequencing identifies a spectrum of mutation frequencies in advanced and lethal prostate cancers. *Proc Natl Acad Sci U S A* **108**, 17087-92.
238. Stransky, N., Egloff, A. M., Tward, A. D., Kostic, A. D., Cibulskis, K., Sivachenko, A., Kryukov, G. V., Lawrence, M. S., Sougnez, C., McKenna, A., Shefler, E., Ramos, A. H., Stojanov, P., Carter, S. L., Voet, D., Cortes, M. L., Auclair, D., Berger, M. F., Saksena, G., Guiducci, C., Onofrio, R. C., Parkin, M., Romkes, M., Weissfeld, J. L., Seethala, R. R., Wang, L., Rangel-Escareno, C., Fernandez-Lopez, J. C., Hidalgo-Miranda, A., Melendez-Zajgla, J., Winckler, W., Ardlie, K., Gabriel, S. B., Meyerson, M., Lander, E. S., Getz, G., Golub, T. R., Garraway, L. A. & Grandis, J. R. (2011). The mutational landscape of head and neck squamous cell carcinoma. *Science* **333**, 1157-60.

239. Stephens, P. J., Tarpey, P. S., Davies, H., Van Loo, P., Greenman, C., Wedge, D. C., Nik-Zainal, S., Martin, S., Varela, I., Bignell, G. R., Yates, L. R., Papaemmanuil, E., Beare, D., Butler, A., Cheverton, A., Gamble, J., Hinton, J., Jia, M., Jayakumar, A., Jones, D., Latimer, C., Lau, K. W., McLaren, S., McBride, D. J., Menzies, A., Mudie, L., Raine, K., Rad, R., Chapman, M. S., Teague, J., Easton, D., Langerød, A., Lee, M. T., Shen, C. Y., Tee, B. T., Huimin, B. W., Broeks, A., Vargas, A. C., Turashvili, G., Martens, J., Fatima, A., Miron, P., Chin, S. F., Thomas, G., Boyault, S., Mariani, O., Lakhani, S. R., van de Vijver, M., van 't Veer, L., Foekens, J., Desmedt, C., Sotiriou, C., Tutt, A., Caldas, C., Reis-Filho, J. S., Aparicio, S. A., Salomon, A. V., Børresen-Dale, A. L., Richardson, A. L., Campbell, P. J., Futreal, P. A. & Stratton, M. R. (2012). The landscape of cancer genes and mutational processes in breast cancer. *Nature* **486**, 400-4.

240. Parsons, D. W., Jones, S., Zhang, X., Lin, J. C., Leary, R. J., Angenendt, P., Mankoo, P., Carter, H., Siu, I. M., Gallia, G. L., Olivi, A., McLendon, R., Rasheed, B. A., Keir, S., Nikolskaya, T., Nikolsky, Y., Busam, D. A., Tekleab, H., Diaz, L. A., Jr., Hartigan, J., Smith, D. R., Strausberg, R. L., Marie, S. K., Shinjo, S. M., Yan, H., Riggins, G. J., Bigner, D. D., Karchin, R., Papadopoulos, N., Parmigiani, G., Vogelstein, B., Velculescu, V. E. & Kinzler, K. W. (2008). An integrated genomic analysis of human glioblastoma multiforme. *Science* **321**, 1807-12.

241. Ehrlich, M., Norris, K. F., Wang, R. Y., Kuo, K. C. & Gehrke, C. W. (1986). DNA cytosine methylation and heat-induced deamination. *Biosci Rep* **6**, 387-93.

242. Roberts, S. A., Sterling, J., Thompson, C., Harris, S., Mav, D., Shah, R., Klimczak, L. J., Kryukov, G. V., Malc, E., Mieczkowski, P. A., Resnick, M. A. & Gordenin, D. A. (2012). Clustered mutations in yeast and in human cancers can arise from damaged long single-strand DNA regions. *Mol Cell* **46**, 424-35.

243. Lindahl, T. (1993). Instability and decay of the primary structure of DNA. *Nature* **362**, 709-15.

244. Hanahan, D. & Weinberg, R. A. (2011). Hallmarks of cancer: the next generation. *Cell* **144**, 646-74.

245. Wei, X., Walia, V., Lin, J. C., Teer, J. K., Prickett, T. D., Gartner, J., Davis, S., Stemke-Hale, K., Davies, M. A., Gershenwald, J. E., Robinson, W., Robinson, S., Rosenberg, S. A. & Samuels, Y. (2011). Exome sequencing identifies GRIN2A as frequently mutated in melanoma. *Nat Genet* **43**, 442-6.

246. Zhang, J., Baran, J., Cros, A., Guberman, J. M., Haider, S., Hsu, J., Liang, Y., Rivkin, E., Wang, J., Whitty, B., Wong-Erasmus, M., Yao, L. & Kasprzyk, A. (2011). International Cancer Genome Consortium Data Portal--a one-stop shop for cancer genomics data. *Database (Oxford)* **2011**, bar026.

247. Lee, W., Jiang, Z., Liu, J., Haverty, P. M., Guan, Y., Stinson, J., Yue, P., Zhang, Y., Pant, K. P., Bhatt, D., Ha, C., Johnson, S., Kennemer, M. I., Mohan, S., Nazarenko, I., Watanabe, C., Sparks, A. B., Shames, D. S., Gentleman, R., de Sauvage, F. J., Stern, H., Pandita, A., Ballinger, D. G., Drmanac, R., Modrusan, Z., Seshagiri, S. & Zhang, Z. (2010). The mutation spectrum revealed by paired genome sequences from a lung cancer patient. *Nature* **465**, 473-7.

248. Komatsu, A., Nagasaki, K., Fujimori, M., Amano, J. & Miki, Y. (2008). Identification of novel deletion polymorphisms in breast cancer. *Int J Oncol* **33**, 261-70.
249. Forouzanfar, M. H., Foreman, K. J., Delossantos, A. M., Lozano, R., Lopez, A. D., Murray, C. J. & Naghavi, M. (2011). Breast and cervical cancer in 187 countries between 1980 and 2010: a systematic analysis. *Lancet* **378**, 1461-84.
250. Loeb, L. A., Springgate, C. F. & Battula, N. (1974). Errors in DNA replication as a basis of malignant changes. *Cancer Res* **34**, 2311-21.
251. Peng, G., Greenwell-Wild, T., Nares, S., Jin, W., Lei, K. J., Rangel, Z. G., Munson, P. J. & Wahl, S. M. (2007). Myeloid differentiation and susceptibility to HIV-1 are linked to APOBEC3 expression. *Blood* **110**, 393-400.
252. Kong, F., Zhu, J., Wu, J., Peng, J., Wang, Y., Wang, Q., Fu, S., Yuan, L. L. & Li, T. (2011). dbCRID: a database of chromosomal rearrangements in human diseases. *Nucleic Acids Res* **39**, D895-900.
253. Edwards, P. A. (2010). Fusion genes and chromosome translocations in the common epithelial cancers. *J Pathol* **220**, 244-54.
254. Lea, D. E. & Coulson, C. A. (1949). The distribution of the numbers of mutants in bacterial populations. *Journal of Genetics* **49**, 264-285.
255. Di Noia, J. & Neuberger, M. S. (2002). Altering the pathway of immunoglobulin hypermutation by inhibiting uracil-DNA glycosylase. *Nature* **419**, 43-8.
256. Huang, X. & Darzynkiewicz, Z. (2006). Cytometric assessment of histone H2AX phosphorylation: a reporter of DNA damage. *Methods Mol Biol* **314**, 73-80.
257. Fairbairn, D. W., Olive, P. L. & O'Neill, K. L. (1995). The comet assay: a comprehensive review. *Mutat Res* **339**, 37-59.
258. Tripathi, A., King, C., de la Morenas, A., Perry, V. K., Burke, B., Antoine, G. A., Hirsch, E. F., Kavanah, M., Mendez, J., Stone, M., Gerry, N. P., Lenburg, M. E. & Rosenberg, C. L. (2008). Gene expression abnormalities in histologically normal breast epithelium of breast cancer patients. *Int J Cancer* **122**, 1557-66.
259. Graham, K., de las Morenas, A., Tripathi, A., King, C., Kavanah, M., Mendez, J., Stone, M., Slama, J., Miller, M., Antoine, G., Willers, H., Sebastiani, P. & Rosenberg, C. L. (2010). Gene expression in histologically normal epithelium from breast cancer patients and from cancer-free prophylactic mastectomy patients shares a similar profile. *Br J Cancer* **102**, 1284-93.
260. Tabchy, A., Valero, V., Vidaurre, T., Lluch, A., Gomez, H., Martin, M., Qi, Y., Barajas-Figueroa, L. J., Souchon, E., Coutant, C., Doimi, F. D., Ibrahim, N. K., Gong, Y., Hortobagyi, G. N., Hess, K. R., Symmans, W. F. & Pusztai, L. (2010). Evaluation of a 30-gene paclitaxel, fluorouracil, doxorubicin, and cyclophosphamide chemotherapy response predictor in a multicenter randomized trial in breast cancer. *Clin Cancer Res* **16**, 5351-61.
261. Lasham, A., Samuel, W., Cao, H., Patel, R., Mehta, R., Stern, J. L., Reid, G., Woolley, A. G., Miller, L. D., Black, M. A., Shelling, A. N., Print, C. G. & Braithwaite, A. W. (2012). YB-1, the E2F pathway, and regulation of tumor cell growth. *J Natl Cancer Inst* **104**, 133-46.

262. Nikas, J. B., Boylan, K. L., Skubitz, A. P. & Low, W. C. (2011). Mathematical prognostic biomarker models for treatment response and survival in epithelial ovarian cancer. *Cancer Inform* **10**, 233-47.
263. Nikas, J. B. & Low, W. C. (2011). Application of clustering analyses to the diagnosis of Huntington disease in mice and other diseases with well-defined group boundaries. *Comput Methods Programs Biomed* **104**, e133-47.
264. Nikas, J. B. & Low, W. C. (2011). ROC-supervised principal component analysis in connection with the diagnosis of diseases. *Am J Transl Res* **3**, 180-96.
265. Nikas, J. B. & Low, W. C. (2012). Linear Discriminant Functions in Connection with the micro-RNA Diagnosis of Colon Cancer. *Cancer Inform* **11**, 1-14.
266. Nikas, J. B., Low, W. C. & Burgio, P. A. (2012). Prognosis of treatment response (pathological complete response) in breast cancer. *Biomark Insights* **7**, 59-70.
267. Faili, A., Aoufouchi, S., Guéranger, Q., Zober, C., Léon, A., Bertocci, B., Weill, J. C. & Reynaud, C. A. (2002). AID-dependent somatic hypermutation occurs as a DNA single-strand event in the BL2 cell line. *Nat Immunol* **3**, 815-21.
268. Schrader, C. E., Guikema, J. E., Linehan, E. K., Selsing, E. & Stavnezer, J. (2007). Activation-induced cytidine deaminase-dependent DNA breaks in class switch recombination occur during G1 phase of the cell cycle and depend upon mismatch repair. *J Immunol* **179**, 6064-71.
269. Sharbeen, G., Yee, C. W., Smith, A. L. & Jolly, C. J. (2012). Ectopic restriction of DNA repair reveals that UNG2 excises AID-induced uracils predominantly or exclusively during G1 phase. *J Exp Med* **209**, 965-74.
270. Ordinario, E. C., Yabuki, M., Larson, R. P. & Maizels, N. (2009). Temporal regulation of Ig gene diversification revealed by single-cell imaging. *J Immunol* **183**, 4545-53.
271. Franco, S., Gostissa, M., Zha, S., Lombard, D. B., Murphy, M. M., Zarrin, A. A., Yan, C., Tepsuporn, S., Morales, J. C., Adams, M. M., Lou, Z., Bassing, C. H., Manis, J. P., Chen, J., Carpenter, P. B. & Alt, F. W. (2006). H2AX prevents DNA breaks from progressing to chromosome breaks and translocations. *Mol Cell* **21**, 201-14.
272. Ramiro, A. R., Jankovic, M., Callen, E., Difilippantonio, S., Chen, H. T., McBride, K. M., Eisenreich, T. R., Chen, J., Dickins, R. A., Lowe, S. W., Nussenzweig, A. & Nussenzweig, M. C. (2006). Role of genomic instability and p53 in AID-induced c-myc-Igh translocations. *Nature* **440**, 105-9.
273. Chiarle, R., Zhang, Y., Frock, R. L., Lewis, S. M., Molinie, B., Ho, Y. J., Myers, D. R., Choi, V. W., Compagno, M., Malkin, D. J., Neuberg, D., Monti, S., Giallourakis, C. C., Gostissa, M. & Alt, F. W. (2011). Genome-wide translocation sequencing reveals mechanisms of chromosome breaks and rearrangements in B cells. *Cell* **147**, 107-19.
274. Klein, I. A., Resch, W., Jankovic, M., Oliveira, T., Yamane, A., Nakahashi, H., Di Virgilio, M., Bothmer, A., Nussenzweig, A., Robbiani, D. F., Casellas, R. & Nussenzweig, M. C. (2011). Translocation-capture sequencing reveals the extent and nature of chromosomal rearrangements in B lymphocytes. *Cell* **147**, 95-106.

275. Staszewski, O., Baker, R. E., Ucher, A. J., Martier, R., Stavnezer, J. & Guikema, J. E. (2011). Activation-induced cytidine deaminase induces reproducible DNA breaks at many non-Ig Loci in activated B cells. *Mol Cell* **41**, 232-42.
276. Tsai, A. G., Lu, H., Raghavan, S. C., Muschen, M., Hsieh, C. L. & Lieber, M. R. (2008). Human chromosomal translocations at CpG sites and a theoretical basis for their lineage and stage specificity. *Cell* **135**, 1130-42.
277. Nussenzweig, A. & Nussenzweig, M. C. (2010). Origin of chromosomal translocations in lymphoid cancer. *Cell* **141**, 27-38.
278. Pasqualucci, L., Neumeister, P., Goossens, T., Nanjangud, G., Chaganti, R. S., Kuppers, R. & Dalla-Favera, R. (2001). Hypermutation of multiple proto-oncogenes in B-cell diffuse large-cell lymphomas. *Nature* **412**, 341-6.
279. Robbiani, D. F., Bunting, S., Feldhahn, N., Bothmer, A., Camps, J., Deroubaix, S., McBride, K. M., Klein, I. A., Stone, G., Eisenreich, T. R., Ried, T., Nussenzweig, A. & Nussenzweig, M. C. (2009). AID produces DNA double-strand breaks in non-Ig genes and mature B cell lymphomas with reciprocal chromosome translocations. *Mol Cell* **36**, 631-41.
280. Küppers, R. (2005). Mechanisms of B-cell lymphoma pathogenesis. *Nat Rev Cancer* **5**, 251-62.
281. Maul, R. W., Saribasak, H., Martomo, S. A., McClure, R. L., Yang, W., Vaisman, A., Gramlich, H. S., Schatz, D. G., Woodgate, R., Wilson, D. M., 3rd & Gearhart, P. J. (2011). Uracil residues dependent on the deaminase AID in immunoglobulin gene variable and switch regions. *Nat Immunol* **12**, 70-6.
282. Robbiani, D. F. & Nussenzweig, M. C. (2012). Chromosome Translocation, B Cell Lymphoma, and Activation-induced Cytidine Deaminase. *Annu Rev Pathol*.
283. Narvaiza, I., Landry, S. & Weitzman, M. D. (2012). APOBEC3 proteins and genomic stability: the high cost of a good defense. *Cell Cycle* **11**, 33-8.
284. Zaheen, A., Boulianne, B., Parsa, J. Y., Ramachandran, S., Gommerman, J. L. & Martin, A. (2009). AID constrains germinal center size by rendering B cells susceptible to apoptosis. *Blood* **114**, 547-54.