# Overlapping functions of APOBEC enzymes in antiviral immunity and cancer

A DISSERTATION SUBMITTED TO THE FACULTY

OF THE UNIVERSITY OF MINNESOTA BY

Gabriel J. Starrett

IN PARTIAL FUFULLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

Adviser: Reuben S. Harris

July 2017

## ACKNOWLEDGMENTS

I would like to thank my adviser Reuben Harris for his enthusiasm, guidance, and expertise during the course of my thesis research. He provided the resources and environment that allowed me to achieve more than I ever anticipated during my graduate career. Additionally, I would like to sincerely thank all of the members of the Harris lab who made my time behind the keyboard and the bench all the better and contributed to an amazing work environment. In particular, I would like to thank my fellow students that joined in 2013, Amber St. Martin, Amy Molan, and Chris Richards. It was great to have company through all of the challenges of graduate school. I would also like to thank my thesis committee, David Largaespada, Doug Yee, and Tim Starr, for their guidance. Lastly, I would like to thank all the mentors that I've had throughout my life, whether family, friends, or teachers. I would not be where I am without any of you and I am forever in your debt.

## DEDICATION

I dedicate this thesis to my girlfriend, Marylyn Presutti. I met you at the beginning of this wild ride and am incredibly grateful for your unwavering love, patience, and support through all of the challenges we have overcome. I could not have done this without you.

## ABSTRACT

APOBEC enzymes are a family of innate antiviral enzymes that form an important barrier against DNA-based pathogens. Encoding and expressing these DNA mutating enzymes, however, is an inherently risky endeavor for the stability of the host genome if not regulated appropriately. These risks have been demonstrated in numerous cancers where APOBEC3B is overexpressed and the APOBEC-associated mutation signature is enriched. Emphasizing the importance of this observation, elevated expression of APOBEC3B and presence of APOBEC-associated mutations has now been consistently linked to aggressive phenotypes and worse outcome in cancer patients. Here I present data demonstrating overlapping functions of APOBEC3 enzymes in antiviral immunity and cancer. In both of these models, APOBEC3 enzymes contribute potentially deleterious and beneficial mutations potentially impacting the survival of tumors and viruses. Additionally, the functions of these enzymes can be modulated by heritable germline mutations in the APOBEC3 locus and viral infections. DNA viruses can also act as valuable molecular probes into the regulation of APOBEC3 enzymes in tumors leading to the development of better therapies.

## TABLE OF CONTENTS

## LIST OF FIGURES

## LIST OF TABLES

# CHAPTER 1:

# Introduction

**Author:** Gabriel Starrett

## Mutation and genome evolution

Mutations are a byproduct of DNA damage minus correct repair. Even in ideal cellular environments, DNA is subject to many sources of damage both endogenous and exogenous. Without appropriate recognition of the DNA lesions resulting from damage, deleterious mutations can accumulate. Many of these mutational processes in cancer have come to be directly tied to mutation signatures observed in cancer, which are mainly defined by single nucleotide substitutions and the trinucleotide contexts in which they occur (1–4). Exogenous sources of DNA damage include UV-induced lesions and bulky DNA adducts from chemical carcinogens, such as from tobacco, aflatoxin and aristolochic acid (**Fig 1.1**) (4–7). These carcinogens typically affect DNA guanines and during DNA replication these adducts are incorrectly base-paired typically resulting in G-to-T mutations. UV irradiation, however, typically seeds mutations at cytosine bases by crosslinking adjacent pyrimidines (4, 8–10). During DNA replication these pyrimidine dimers are bypassed by trans-lesion polymerases such as POLH, which inserts two adjacent alanines regardless of the cross-linked bases (4, 8–12). After another round of replication these are base-paired with thymines and in tumors this is most frequently observed as CC-to-TT dinucleotide substitutions. Reactive oxygen species (ROS) are abundantly generated by metabolism and immune responses and cause abundant damage to both protein and DNA within cells. ROS can cause mutation through oxidation of DNA guanines and subsequent repair of these oxidized bases by nucleotide excision repair (NER) into nearly any other base (13, 14).

An additional factor used in defining signatures is whether these mutations occur with either a transcriptional or replicative strand bias. Transcriptional strand biases can occur through biased repair of DNA lesions, as is observed in the repair of UV-induced damage and previously mentioned guanine modifications, where highly transcribed genes are more frequently repaired due to the recruitment of XP-family ATPases/helicases by RNA polymerase II (RNAPII) stalled at the lesion (4, 15–17). XP-family enzymes preferentially excise from the transcribed stand or template strand resulting in the mutation bias (17). Alternatively, transcriptionally biased mutations can preferentially occur on one strand, which has come to be known as transcription-coupled damage (TCD) (15). TCD is observed with the generally distributed T-to-C mutations associated with a mutation signature found in most types of cancer, but the exact mechanism is still unknown (4, 15).

The other type of mutation bias is based on the replication direction of genomic DNA from various replication origins. Replication timing has been identified as remarkably consistent across cell types in humans. This information has been paramount in identifying any biases when mutations occur during genome replication. Inherently, the polymerases of DNA-based organisms have very low mutation rates because of the presence of proof-reading domains (18, 19). Not surprisingly, error-prone DNA polymerases are also a key source of replication biased mutations in many tumor types (4, 18, 19). These polymerases also provide a mechanism for viral diversity in DNA viruses, yet most have

exceptionally low mutation rates compared to their RNA-based counterparts due to hijacking host high-fidelity DNA polymerases or encode their own (20).

The most abundant source of DNA damage that occurs within virtually every tumor type is the spontaneous deamination of methylcytosines into thymines catalyzed by water molecules. These types of mutations continuously accumulate over the lifespan of an individual. Lastly, a major endogenous source of mutations in both cancer and viral genomes is seeded by an evolutionarily conserved family of innate immune defense proteins known as apolipoprotein B mRNA-editing enzyme catalytic polypeptide-like enzymes (APOBECs). This family of enzymes will be expanded upon in the following sections.

**The APOBEC family of DNA cytosine deaminases**

The apolipoprotein B mRNA-editing enzyme catalytic polypeptide-like (APOBEC) family of enzymes is highly evolutionarily conserved immune-related single-stranded DNA cytosine deaminases that has gained a substantial amount of diversity through gene duplications and rearrangements over time (21, 22). In higher primates and humans, there are 11 APOBEC enzymes that are involved at various stages of the immune response (22–26). These enzymes are activation induced deaminase (AID), APOBEC1, APOBEC2, APOBEC3A-H, and APOBEC4 (26–28). AID and APOBEC2 are the most evolutionarily conserved APOBEC family members, with homologs existing in all vertebrates from present day bony fish to higher primates (22, 28). In mammals, AID is primarily expressed in B-cells and through precise regulation is responsible for somatic

hypermutation and class switch recombination of the immunoglobulin locus that results in antibody maturation. The next most evolutionarily conserved member and the namesake of the family is APOBEC1, which primarily exhibits RNA-editing function, initially discovered at the 6666 position of the *APOB* mRNA in the liver (29–31). The exact functions of APOBEC2 and APOBEC4 are still currently unknown and they have yet to elicit enzymatic activity (27, 28, 32).

The APOBEC3 subfamily of enzymes have gone through a series of duplications and rearrangements during the evolution of primates and in humans there are 7 family members encoded in tandem on chromosome 22 (21, 23, 24, 26). These enzymes are composed of one or two catalytic domains that fall into one of three evolutionarily conserved categories, Z1, Z2 or Z3 (23) (**Fig 1.2A**). These enzymes are unified in their function to target and deaminate cytosines to uracils in single-stranded (ss)DNA, found commonly during many DNA transactions including transcription and replication of host and viral genomes (**Fig 1.2B**). These enzymes are widely expressed in tissues, however, APOBEC3D, F, G and H are most highly expressed in T cells and have all been widely studied in their ability to restrict retroviruses in primates and humans through hypermutation of their ssDNA genome intermediates (33, 34). Overexpression of APOBEC3B, C, F and G have all demonstrated some capacity to restrict retrotransposons via enzymatic and non-enzymatic means (35–43). The effects of these enzymes on retrotransposons has seemingly had strong effects on shaping the evolution of primate genomes with approximately 20% of inherited variants at TCW sites likely originating from APOBEC-mediated deamination (44, 45).

Although many APOBEC3 proteins have overlapping functions, these enzymes are differentially regulated and localized in various stages of the cell cycle (1, 33, 34, 46). The predominantly cytoplasmic enzymes are APOBEC3D, F, G and H and endogenously expressed APOBEC3A in myeloid lineage cells. APOBEC3A and C are expressed cell wide when overexpressed, while APOBEC3B is the only family member that is predominantly nuclear even though it is too large to passively pass through nuclear pores (47–49). APOBEC3B transits into the nucleus through a similar mechanism to how AID is imported and is dependent on a valine at position 54 (49).

The common ability of APOBECs to target cytosine-containing ssDNA makes any pathogen with a DNA stage in its lifecycle a potential target and research supports a working model in which many viruses and endogenous elements are inhibited by APOBEC-mediated deamination. When these enzymes are dysregulated, they can begin to erroneously target the human genome and cause cancer. Primarily, inherited germline mutations, somatic alterations in cancer, and viral proteins form the three primary axes that promote APOBEC3 dysregulation and aberrant function (**Fig 1.3**).

**APOBEC mechanism of mutagenesis and dysregulation in cancer**

Through specific gene knockdown and over-expression studies, as well as somatic mutation signature analysis, APOBEC3B was discovered to be a dominant mutator in breast cancer (1). Through bioinformatic and biochemical studies, the preferred targets of APOBEC3 deamination were determined to be

enriched at cytosines preceded by a 5’ thymine and followed by a 3’ adenosine, guanine, or thymine (50–53). The deamination events at these sites result in a genomic DNA uracil that most commonly will result in a cytosine to thymine transition mutation if allowed to persist through DNA replication. Next most common, excision of this uracil by uracil DNA glycosylase (UDG) and repair of the resulting abasic site by REV1, frequently results in a cytosine to guanine transition mutations because of the propensity for REV1 to insert a cytosine opposite of an abasic site (54–58). Subsequently, APOBEC3B-mediated mutations and elevated expression have been found in many types of cancer, but most prominently in human papillomavirus (HPV)-associated cancers (34, 59, 60) and bladder cancer, which may be also be influenced by DNA viruses (34, 61–65).

APOBEC-signature mutations are enriched on the lagging strand of replicating DNA in bacterial and eukaryotic cell based assays (44, 66–68). Additionally, APOBEC-mediated mutations are enriched in the early replicating portions of the human genome occurring in strand-coordinated clusters (69). These APOBEC-mediated mutations are in part initiated by stalled replication forks, which can be caused by various types of DNA damage (67, 68). Elevated expression of APOBEC3B is associated with more aggressive phenotypes and recurrence in breast cancer, lung cancer and bladder cancers (1, 70–72). Additionally, elevated expression of APOBEC3B associates with greater resistance to the estrogen receptor agonist, tamoxifen, in mouse models and is

supported by human clinical data likely through increased tumor diversity through APOBEC-mediated mutagenesis (73).

Because of the ability of APOBEC3B to augment tumor survival and evolution, APOBEC3B regulation has now become an important in therapeutic target in regards to limiting the effects of the enzyme on tumor survival. APOBEC3B is uniquely upregulated compared to other APOBEC family members in cancer as has been determined via numerous studies using RTqPCR, microarray, and RNAseq transcriptome analyses. The non-canonical PKC/NFκB pathway has been identified as a significant axis to manipulate APOBEC3B expression in normal-like breast epithelial cells and cancer cell lines (74). This pathway can be activated by phorbol-myristate acetate or by agonists of the lymphotoxin beta receptor, which both feed into the non-canonical NF-κB signal transduction pathway (75).

## Germline variants in the *APOBEC3* locus and effects on susceptibility to cancer and viral infections

### *APOBEC3A and APOBEC3B*

APOBEC3A and the carboxyl half of APOBEC3B share some of the highest homology of any APOBEC3 enzymes (92% identical) (22–24). The primary differences between these enzymes have mapped to loops 1 and 3 (with most other loops being 100% identical) (22–24). Specifically, loop 1 has been demonstrated to impact the structure of the active site of these enzymes with APOBEC3B having a longer, flexible loop 1 leading to a more closed

conformation compared to APOBEC3A (76–79). In fact, swapping loop 1 from APOBEC3A into APOBEC3B increases the enzymatic activity of APOBEC3B 5-50 fold (77, 79). It is now known that both APOBEC3A and APOBEC3B interact with ssDNA in a U-shaped hairpin formation (**Fig 1.2C**) (78, 80). While it is well understood that APOBEC3A and APOBEC3B prefer to deaminate TCA context cytosines, it has yet to be fully realized if there is a true difference in the preferred context between these enzymes (78, 80, 81).

There is a geographically distributed *APOBEC3B*-deletion that fuses the coding region of *APOBEC3A* with the 3'UTR of *APOBEC3B* (82). The junctions of this rearrangement occur within a 350 bp region of 100% homology between these genes (82). This variant has been widely studied for potential effects on the replication of pathogens and the development of cancer. This germline variant has paradoxically associated with increased ovarian and breast cancer susceptibility in several populations (83–88). This has been partially attributed to increased stability of the fusion transcript due to the absence of an Alu element in the *APOBEC3B* 3' UTR compared to the *APOBEC3A* 3' UTR (89). This Alu element may be targeted by miRNAs leading to its degradation. More recently the observed APOBEC deaminase activity in the absence of APOBEC3B has been attributed to APOBEC3H-I (90), which is expanded more in a later section. As previously mentioned, the APOBEC3B deletion allele has been studied widely in its potential effect on host-pathogen interactions. From these studies it has been revealed most commonly an increased susceptibility to chronic HBV infection (91–94), which supports the role of APOBEC3B as an innate-immune

restriction factor against HBV (75, 95, 96). Additionally, the deletion allele has been found to be associated with increased incidence and co-infection with HBV of a diverse subset of DNA and RNA viruses including hepatitis C virus (HCV), Torque Teno virus (TTV) and GB virus C (GBV-C). Lastly this allele has even been associated with increased incidence of parasitic infections from *Falciparum malaria* and *Toxoplasma gondii* (92, 97).

Independently of the APOBEC3B deletion allele, a cluster of genetically linked risk variants has been identified that correspond to increased expression of *APOBEC3B* in bladder cancer (98). These variants occur within a 2kb region that is 20kb upstream of *APOBEC3A*. This region is likely a long-distance enhancer for *APOBEC3B* as cell lines with the risk allele demonstrated an ~8-fold increased responsiveness to exogenous stressors such as viral infection and treatment with bleomycin (98). Additionally, patients with the high-expression allele showed increased APOBEC-signature mutations in their tumors which in turn proved to be the best metric for survival in this cohort (98).

*APOBEC3C*

APOBEC3C is the most highly and widely expressed APOBEC family member in humans (33, 34, 46). However, despite this near ubiquity, the exact function of APOBEC3C is not known. APOBEC3C demonstrates some of the lowest deaminase activity in the APOBEC3 family in cell-based overexpression assays. However, when overexpressed, APOBEC3C is able to prevent retrotransposition of endogenous elements in a deamination-independent mechanism that is at least partially explained through the direct interaction with

the ORF1 protein of LINE1 elements (36). Investigation into a rare SNP that converts a serine to an isoleucine at position 188 in APOBEC3C has shown a 2-fold increase in the ability of this enzyme to restrict HIV (99).

*APOBEC3G*

Being the most potent restriction factor against HIV-1 in the APOBEC3 family, variants in *APOBEC3G* have been importantly studied in ~~regards to~~ elite control and resistance to HIV-1. The vast majority of *APOBEC3G* variants occur within the 5' UTR and introns and either have no or very weak effects on disease susceptibility and outcome (100). The primary variant in *APOBEC3G* identified to influence HIV-1 disease progression is H186R. R186 results in more rapid CD4+ T cell loss and disease progression to AIDS (101, 102). This variant is more common in African populations (37%) compared to Caucasian populations (5%), yet no significant difference has been observed for APOBEC3G-signature mutations in HIV-1 proviruses in these populations or tied to any other *APOBEC3G* variant (103, 104). These observations indicate that this APOBEC3G variant is still fully functional in hypermutating the HIV-1 genome, but may have other deficiencies, such as packaging or expression later in disease. Although this variant has had no observable effect on HIV-1 susceptibility, a recent study showed that it may have significant effects in individuals with specific genetic backgrounds indicating that in conjunction with other genetic characteristics it may be a risk factor for infection (105).

*APOBEC3H*

In humans, APOBEC3H is the most polymorphic family member (48, 106–109). APOBEC3H amino acid changes have been used to create groups of variants or “haplotypes” for functional characterization. To date, seven APOBEC3H haplotypes have been defined with varying degrees of stability and enzymatic activity. Stable, well-expressed, and enzymatically active haplotypes are II, V and VII. In comparison, haplotype I is expressed at intermediate levels but this variant still elicits full activity in biochemical comparisons (90). In contrast, haplotypes III, IV, and VI are not expressed at the protein level nor active due to a deletion of asparagine 15, which is situated in the middle of an essential alpha helix.

An attenuation of A3H function against retroviruses and retroelements most likely occurred twice during human evolution (108). As alluded to above, a deletion of an asparagine at position 15 results in complete destabilization of the enzyme through the interruption of a critical alpha helix in the structure (48, 106–108). This variant is evenly distributed throughout the globe with an approximate allele frequency of 20-30% (90, 103). A glycine at position 105 results in a protein with intermediate stability as judged by immunoblots, approximately 5-fold lower than variants with N15 intact and an arginine at position 105 (48, 106–108). G015 greatly diminishes the ability of this enzyme to restrict HIV, however, when expressed at equal protein amounts to stable, active forms like haplotype II, it is an equally effective deaminase (90). Haplotypes I and II are inversely distributed around the globe with haplotype I being abundant in Caucasians and East Asians and haplotype II being most abundant in individuals of African descent (90, 103).

In the absence of APOBEC3B, we have identified APOBEC3H-I as a genomic DNA mutating enzyme in breast and lung cancers (90). Supporting of these findings, a study in Han Chinese cohort revealed that there is a variant in APOBEC3H genetically linked with haplotype I that results in increased susceptibility to the development of lung cancer (110).

## APOBEC enzymes and tumor-associated viruses

### *Retroviruses*

APOBEC enzymes are best understood in their ability to restrict retroviruses such as HIV-1 (25, 111, 112). Although there is little direct evidence that HIV-1 mediates the development of cancer, it is well documented that the associated immunosuppression leads to the development of cancers such as Kaposi's sarcoma, which were a hallmark of AIDS early in the HIV epidemic (113–115). APOBEC enzymes package into newly forming virions in infected host cells. After these virions are released from the productively infected cell and infect a new cell, the viral genome and packaged APOBEC3 enzymes reside together in the viral core. During reverse transcription of the HIV genome, the nascent ssDNA strand is subject to enzymatic targeting by APOBEC3 enzymes. This frequently results in hypermutation, when the ssDNA strand accumulates multiple C-to-U lesions, which can be deleterious to the virus and prevent the production of future infectious progeny. When the results are not lethal, these APOBEC-mediated mutations may also contribute to HIV-1 evolution, although this point has been contested. Most studies over the past decade have indicated

that APOBEC3 enzymes contribute to diversity in the HIV-1 genome (112, 116–118). However, two papers have recently shown that APOBEC3 preferred motifs are not significantly depleted in the viral genome, and when replicating the virus in the presence or absence of APOBEC3s that reverse transcriptase is responsible for most of the mutations (119, 120).

Another human retrovirus, human T-lymphotropic virus type 1 (HTLV-1), directly associates with the development of adult T-cell leukemia/lymphoma (ATLL) (121). HTLV-1 is a deltaretrovirus with a much longer evolutionary history in the order of thousands of years rather than the decades that HIV-1 has been in the human population. HTLV-1 is transmitted sexually, blood-to-blood (e.g. intravenous drug use), and mother-to-child through breast milk and is endemic in various regions around the world, including Southern Japan, sub-Saharan Africa, South America and the Caribbean (121). Similar to HIV, HTLV-1 primarily infects CD4+ T cells. APOBEC3A, APOBEC3B, and APOBEC3H-II have all been shown to be able to restrict HTLV in over-expression cell culture studies through editing of the HTLV-1 genome (122). While APOBEC3G appears incapable of packaging into HTLV-1 particles and restricting virus infectivity, it does show evidence of being able to actively edit the viral genome (123–125).

Endogenous retroviruses (ERVs) comprise a large proportion of mammalian genomes including the human genome. Retrotransposition of these DNA pathogens has the potential to interrupt genes and seed deleterious events that can lead to cancer (126, 127). In many ways, host cells respond to these endogenous elements like exogenous retroviruses and most appear to be

susceptible to restriction by several APOBEC family members (35, 36, 38, 39, 41, 43, 128–130).

*Papillomaviruses*

Virtually all cervical cancers and a growing proportion of head and neck cancers have integrated high risk HPV types 16 or 18 (131–133). Human papillomaviruses are a highly diverse family of small dsDNA non-enveloped viruses now containing over 250 subtypes. These viruses have small circular genomes that can be divided into three general regions, early ORFs, late ORFs, and the long control region (**Fig 1.4A**). High risk HPVs differ from their low risk counterparts by the enhanced function of their oncoproteins, E6 and E7, and ability to promote the integration of foreign DNA into the host genome (133, 134). The primary function of HPV E6 is to bind and inhibit p53 and that of HPV E7 is to bind and degrade Rb (135, 136). Both of these functions enhance the ability of the infected cells to survive and produce progeny while having the unintended consequence of increasing chances of surviving cells accumulating APOBEC-mediated chromosomal damage. High-risk HPV E6 has been shown to specifically upregulate APOBEC3B expression in normal immortalized keratinocytes (NIK), hTERT-immortalized keratinocytes, and primary foreskin keratinocytes (59). This mechanism has been explained by both the direct transcriptional activation of A3B by HPV E6 and by the induction of the TEAD family of transcription factors (137, 138).

Although expressed at low levels, increased fold expression of APOBEC3A has been observed in cervical invasive neoplasms (139). This may

be part of the normal response to papillomavirus infection as it has been shown that overexpression of APOBEC3A can mutate and restrict HPV (140, 141). Additionally, overexpression of APOBEC3C has shown the ability to restrict HPV in a deaminase-independent mechanism (141). In support of APOBEC3-mediated mutagenesis of HPV, only alpha-papillomavirus, to which high-risk HPV 16 and 16 belong, genomes generally show evidence of being shaped by APOBEC3 enzymatic activity through the depletion of TC target motifs and the enrichment of TT product motifs throughout the viral genome (142, 143). One of the consequences of using a DNA damaging enzyme like APOBEC3s is increasing the likelihood that DSBs will occur in both the host and viral genomes greatly increasing the chance that these breaks will be repaired to each other using non-homologous end-joining resulting in viral genome integration and partially explaining why high-risk HPVs cause cancer.

*Polyomaviruses*

Papillomaviruses and polyomaviruses were previously classified together in the papovaviridae family due to the similarity that these small dsDNA viruses utilize their compact genome to hijack cellular processes for the production of progeny. It is understandable that these viruses may also have similarities in regards to their relationship with the APOBEC family of enzymes. These viruses are ubiquitous in humans with some populations having nearly 100% seropositivity. The highly related, BK and JC polyomaviruses infect early on in human lifespans and establish latent, subclinical infections in urothelial tissues. Both viruses can reactivate in immunosuppressed individuals leading to fatal

diseases. JCPyV has emerged as the causative agent of progressive multifocal leukoencephalopathy (PML) (144–146); whereas, BKPyV reactivation is the leading cause of kidney rejection in transplant recipients and can cause hemorrhagic cystitis and polyomavirus-associated nephropathy (147, 148).

The polyomavirus genome is approximately 5kb and can be divided into early and late transcription regions (**Fig 1.4B**). The late region contains the structural genes for major capsid protein, VP1, and minor capsid proteins, VP2 and VP3. The early region contains the tumor antigen gene and its splice variants, large T antigen, truncated T antigen, and small T antigen. Small t antigen contains a J-domain that binds to HSC70 and activates its ATPase activity. Together, they as a molecular chaperone that remodels other cellular complexes and regulates their function (149, 150). Large T antigen has several domains that allow the protein to interact with numerous host factors. Similar to E6 and E7, large T antigen inhibits both p53 and pRb through the binding of its DNA-binding domain and LXCXE motif, respectively (135, 151, 152). Truncated T antigen shares much of the sequence of LTAg but lacks the DNA-binding domain (153).

Non-human polyomaviruses have been well understood for their ability to transform normal cells in culture and cause cancer in animal models such as SV40 and murine polyomavirus. While human polyomaviruses can transform cell in culture, they have generally been considered to be non-pathogenic. However, the advent of high-throughput sequencing has dramatically expanded the number of known human polyomaviruses and our understanding of their role in disease

(154–158). Most notably, in 2008, Merkel cell polyomavirus was discovered to be the primary causative agent in about 80% of Merkel cell carcinomas (MCC), a rare, aggressive form of skin cancer (155, 157). Risk factors for MCC not only include viral infection, but also old age and sun exposure.

The large T antigen of BKPyV, JCPyV and MCPyV has been demonstrated to upregulate the cancer-associated enzyme, APOBEC3B. The genome of BK polyomavirus, like alpha-papillomaviruses, shows strong evidence for APOBEC-mediated TC depletion. The growing association of BKPyV seropositivity with bladder cancer, which is an APOBEC-signature high cancer, may be due to chronic infection and reactivation of BKPyV (61, 62). Surprisingly, virus-associated Merkel cell carcinoma shows no evidence of elevated APOBEC3B expression or APOBEC-mediated mutation signatures (159–161). Additional studies into these polyomavirus-associated tumors will help to clarify the complexities of APOBEC regulation and activity in tumor development.

*HBV*

Hepatitis B virus is a small enveloped DNA virus that replicates through an RNA intermediate belonging to the hepadnaviridae family. The viral genome is encapsidated as relaxed circular DNA (rsDNA). After infection of host hepatocytes, the viral genome is converted to covalently closed circular DNA (cccDNA) in the nucleus where transcription of viral genes occurs and persistence is established. Over 350 million people are chronically infected with HBV worldwide and approximately 600,000 die every year from HBV-associated diseases such as liver cirrhosis and hepatocellular carcinoma (162).

APOBEC involvement in the restriction of HBV has become convoluted over the years with several conflicting reports. APOBEC3B, APOBEC3C, APOBEC3F, APOBEC3G, and APOBEC3H are all significantly upregulated by HBV-infected liver cirrhosis (95). Additionally, APOBEC3G preferred motifs are the most frequently mutated in virus isolated from liver cirrhosis and sera from chronically infected patients (95, 163–165). In overexpression studies, the HBV genome is susceptible to targeted deamination by APOBEC3B, APOBEC3C, APOBEC3F and APOBEC3G (95, 96, 163, 164). Together, this evidence points toward APOBEC3G being the primary APOBEC involved in HBV infection.

However, more recent studies have shown that APOBEC3B can be induced by lymphotoxin beta receptor crosslinking and this leads to degradation of the cccDNA genome in HBV-infected hepatocellular carcinoma cell lines (75). Further supporting the involvement of APOBEC3B, the primary HBV oncoprotein HBx can upregulate the ligase MSL2 and target APOBEC3B for polyubiquitination and degradation protecting HBV cccDNA (166). Lending to even more complexity, initially the APOBEC3B deletion has was not associated with HBV infection, but several recent studies have shown that the APOBEC3B deletion does associate with chronic HBV infection in several cohorts (91–93). Lastly, there has been a report that APOBEC3D may act as a molecular sponge, dimerizing with APOBEC3F and APOBEC3G during HBV infection and preventing their ability to restrict the virus (167)

*Herpesviruses*

Herpesviruses are large dsDNA viruses that are ubiquitous in vertebrates. Much like their small DNA virus cousins, herpesvirus infections are predominantly established early in life with herpes simplex virus (HSV) 2 being the primary exception being transmitted through sexual contact. Also like small dsDNA viruses, herpesviruses should be susceptible to deamination by APOBEC3 enzymes. Specifically, Epstein-Barr virus (EBV), HSV1, and murine gammaherpesvirus 68 genomes are able to be targeted by human APOBEC3-mediated DNA-editing (168, 169). Additionally, it has been shown that AID can restrict Kaposi's sarcoma-associated herpesvirus (KSHV) in the absence of UNG2 and KSHV encoded miRNAs are able to target AID transcripts for degradation (170).

EBV is understood to cause diverse malignancies, such as Hodgkin's lymphoma, Burkitt's lymphoma, gastric cancer, and nasopharyngeal carcinoma. Compared to primary effusion lymphoma, Burkitt's lymphoma has generally decreased proportion of APOBEC-signature mutations and expression of most APOBECs indicating a potential inverse correlation with EBV and APOBEC mutagenesis in B cell lymphomas (171). Generally, about 9% of gastric cancers are EBV positive and 14.4% of gastric cancers have significant proportion of APOBEC signature mutations and these groups do not seem to have significant overlap (4, 172, 173). Contrastingly, many nasopharyngeal carcinomas appear to be dominated by APOBEC-signature mutations, but in depth analysis of EBV involvement has not been conducted (174). Together, this may indicate that the

cell type of origin is important for APOBEC involvement in HBV-associated tumors.

**Concluding Introductory Remarks and Chapter Prelude**

The APOBEC3 family of DNA cytosine deaminases is important for mutagenesis in cancer and viral evolution (with restriction being an outcome of viruses that have yet to adapt to the host APOBECs). For all of these enzymes, their expression and function can be impacted by inherited germline mutations, novel somatic mutations in tumors, and the exogenous effects of viral proteins. To address the aforementioned paradox caused by the *APOBEC3B* germline deletion, I investigated germline mutations in the polymorphic antiviral gene, *APOBEC3H*, and identify their association with APOBEC mutagenesis in cancer (**Chapter 2**).

Building off of previous work demonstrating that the E6 oncoprotein from high-risk HPV upregulates APOBEC3B, I tested whether human polyomaviruses are able to modulate the expression of APOBEC3B (**Chapter 3**). This study revealed that the large T antigen of human polyomaviruses share the ability to significantly and specifically upregulate APOBEC3B. To then test if APOBEC3B upregulation by polymaviruses impacts the mutation profiles of MCPyV-associated tumors, we performed RNA and whole genome sequencing (**Chapter 4**). This did not reveal any APOBEC-mediated mutations in the tumors, rather that viral integration events result in focal amplifications of the host genome.

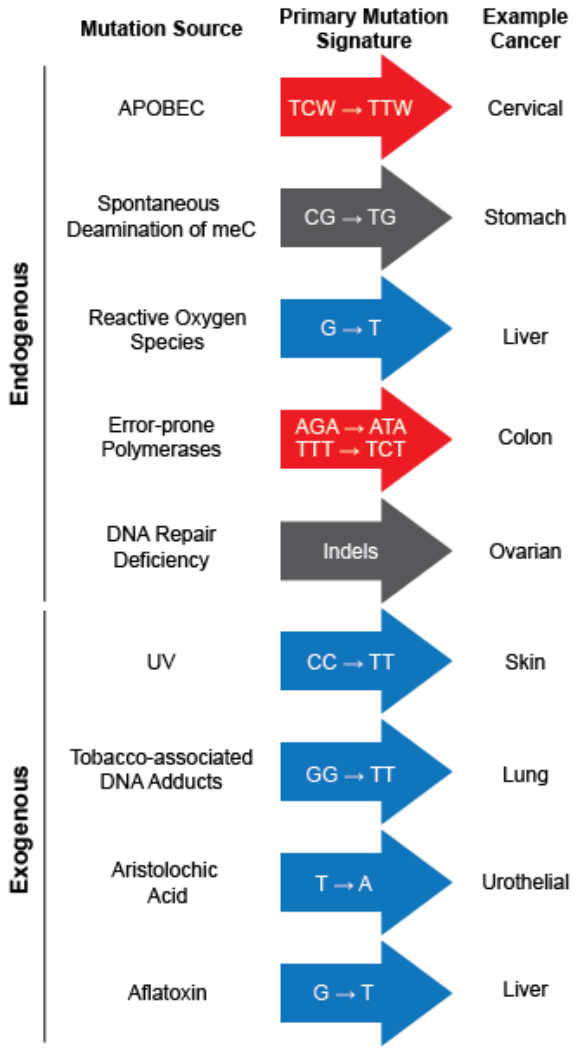



**Figure 1.1. Previously established exogenous and endogenous sources of mutation and their outcomes.** Examples of established exogenous and endogenous sources are listed on the left. The primary mutation context and base substitutions of each source are contained within the colored arrows. Red arrows are for signatures with a replication-based strand bias, blue arrows are for a transcription-based strand bias and grey arrows do not have a strand bias. Examples of cancers that are molded by these processes are on the right.

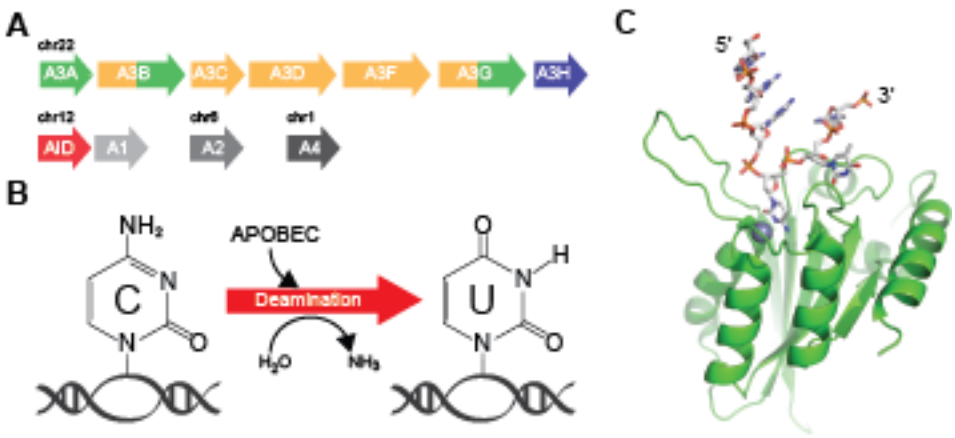



**Figure 1.2. The APOBEC family of polynucleotide cytosine deaminases.** (A) Diagram of the 11 APOBEC family members. Colored arrows represent the evolutionary conservation of the deaminase domain; green is Z1, orange is Z2, blue is Z3. All other colors are unique to that enzyme (B) Schematic of APOBEC-catalyzed deamination in ssDNA. (C) U-shaped ssDNA is shown as a ball and stick model with the target cytosine embedded in the active site of APOBEC3B (green ribbon structure) (PDB: 5TD5).

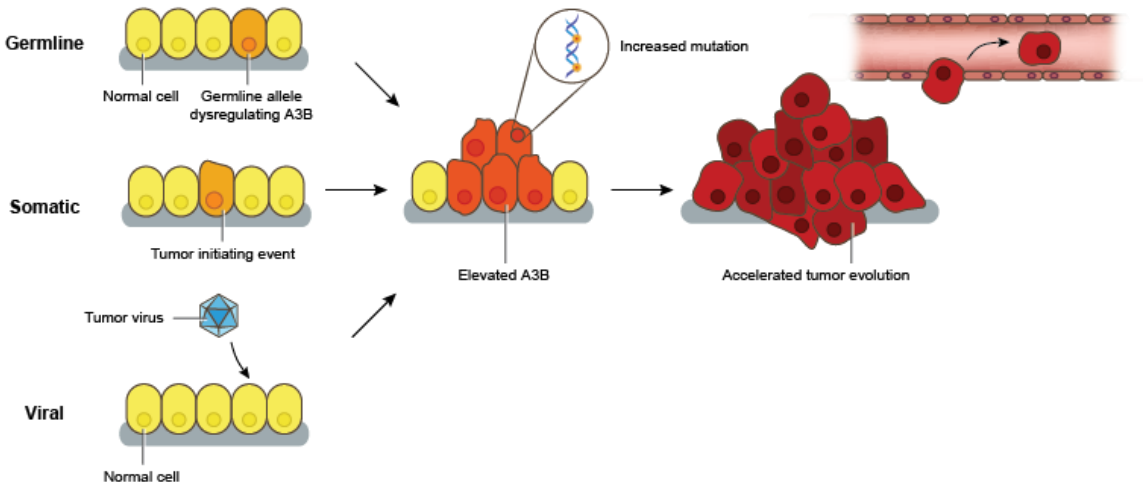



**Figure 1.3. Model for three axes of APOBEC3 regulation.** Normal, healthy cells are in yellow. Cells in the beginning stages of APOBEC3B dysregulation are in light orange. Cells undergoing high APOBEC3B expression and mutagenesis are in dark orange and red.

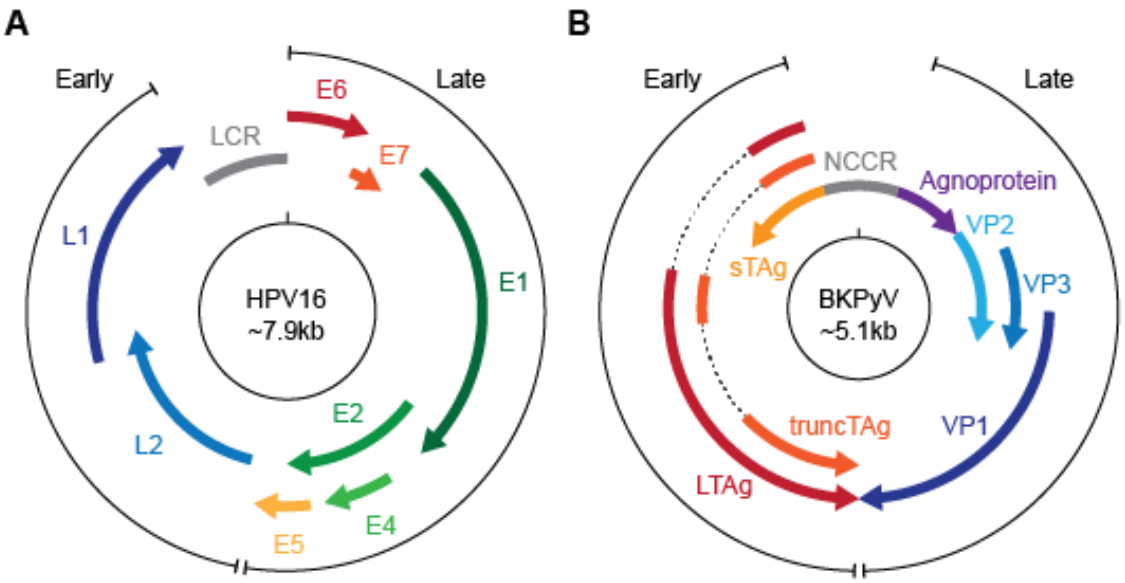



**Figure 1.4. Schematic of the genomes of papillomavirus and polyomavirus**. (A) Genome organization of human papillomavirus 16 (HPV16). (B) Genome organization of BK polyomavirus (BKPyV). Abbreviations: Long control region (LCR), non-coding control region (NCCR), small T antigen (sTAg), Large T antigen (LTAg), truncated T antigen (truncTAg). Approximate genome size is shown underneath the virus name in the inner circle. Starting position of the reference genome (i.e. position 1) is marked with a tick on the inner circle. Early and late expressing regions of the genome are marked appropriately with black lines.

## CHAPTER 2:

# APOBEC3H haplotype I is an active DNA cytosine deaminase that contributes broadly to cancer mutagenesis

*This chapter was adapted with permission from: Starrett et al. (2016) Nat Comm, 7: 12918.*

**Authors:** Gabriel J. Starrett[1,2,3,4], Elizabeth M. Luengas[1,2,3,4], Jennifer L. McCann[1,2,3,4], Diako Ebrahimi[1,2,3], Nuri A. Temiz[1,2], Robin P. Love[5], Yuqing Feng[5], Madison B. Adolph[5], Linda Chelico[5], Emily K. Law[1,2,4,6], Michael A. Carpenter[1,2,3,4,6] & Reuben S Harris[1,2,3,4,6]

**Affiliations:** [1]Department of Biochemistry, Molecular Biology and Biophysics, University of Minnesota, Minneapolis, Minnesota 55455, USA. [2]Masonic Cancer Center, University of Minnesota, Minneapolis, Minnesota 55455, USA. [3]Institute for Molecular Virology, University of Minnesota, Minneapolis, Minnesota 55455, USA. [4]Center for Genome Engineering, University of Minnesota, Minneapolis, Minnesota 55455, USA. [5]Department of Microbiology and Immunology, University of Saskatchewan, Saskatoon, Saskatchewan, Canada S7N 5E5. [6]Howard Hughes Medical Institute, University of Minnesota, Minneapolis, Minnesota 55455, USA

## SUMMARY

Cytosine mutations within TCA/T motifs are common in cancer. A likely cause is the DNA cytosine deaminase APOBEC3B (A3B). However, *A3B*-null breast tumors still have this mutational bias. Here we show that APOBEC3H haplotype-I (A3H-I) provides a likely solution to this paradox. *A3B*-null tumors with this mutational bias have at least one copy of *A3H-I* despite little genetic linkage between these genes. Although deemed inactive previously, A3H-I has robust activity in biochemical and cellular assays, similar to A3H-II after compensation for lower protein expression levels. Gly105 in A3H-I (vs. Arg105 in A3H-II) results in lower protein expression levels and increased nuclear localization, providing a mechanism for accessing genomic DNA. A3H-I also associates with clonal TCA/T-biased mutations in lung adenocarcinoma suggesting this enzyme makes broader contributions to cancer mutagenesis. These studies combine to suggest that A3B and A3H-I, together, explain the bulk of “APOBEC signature” mutations in cancer.


## INTRODUCTION

Tumor genome sequencing studies have identified a number of mutation patterns, or signatures, in cancer that directly reflect the source of the original DNA damage [reviewed by refs. (175–178)]. Prominent established sources include UV-induced C-to-T transition mutations in dipyrimidine motifs in melanoma (UV signature), and spontaneous, water-mediated methyl-C-to-T transition mutations in CG motifs in many tumor types (ageing signature), as well

as a number of other known signatures. However, the most abundant previously unknown mutation signature to emerge from sequencing tumors is “APOBEC”. The signature of APOBEC mutagenesis is C-to-T transition and C-to-G transversion mutations within TCA and TCT trinucleotide motifs (hereafter TCW; TCG motifs may also be targeted by this process but are difficult to distinguish from the ageing signature and therefore discounted in most analyses). APOBEC signature mutations are mostly dispersed throughout the genome, but a fraction occurs in clusters called kataegis (179). APOBEC signature mutations are prevalent in over half of human cancers and often account for the majority of mutations within a single tumor (1, 2, 4, 34, 179–182). For instance, the APOBEC signature often dominates the overall mutation landscapes of breast, lung, head/neck, bladder, and cervical cancers.

Although commonly referred to as a single entity, APOBEC is in fact a family of 11 distinct proteins in humans, and 9 have demonstrated DNA cytosine deaminase activity in a variety of assays with 7 eliciting intrinsic preferences for TC motifs in single-stranded DNA [reviewed by refs.(24, 183, 184)]. Leading candidates to account for the overall APOBEC mutation signature in cancer are APOBEC3A (A3A) and APOBEC3B (A3B). A3A has been attractive because it is arguably the most potent human DNA deaminase that, upon overexpression in heterologous systems, causes genomic DNA damage that results in cell death (1, 81, 89, 185–188). A comparison of the A3A and A3B mutation signatures derived from yeast overexpression experiments and the actual APOBEC signature from tumor genomes suggested that A3A provides the most

parsimonious explanation (81). However, endogenous A3A expression is largely specific to myeloid lineage cell types and, upon natural induction by interferon-α, it localizes to the cytoplasmic compartment and is not cytotoxic (118, 189–191). Consistent with strict developmental regulation and dedicated innate immune function, A3A has yet to be detected at the protein level in cancer cell lines or tumors. Moreover, much of the reported RNA level expression may be due to *A3B* sequence mis-mapping due to >90% nucleotide identity with *A3A*, a likelihood strongly supported by positive linear correlations between *A3A* and *A3B* expression in non-myeloid cancer data sets (**Supplementary Fig. S2.1**) and clear differences in the transcriptional regulatory programs for each of these genes (59, 74, 118, 189–192).

A3B was first implicated as an endogenous mutagen in breast cancer (1). A3B is overexpressed in approximately 50% of breast tumors and the majority of breast cancer cell lines (1). A3B is constitutively nuclear (47, 49, 193), and the only detectable DNA deaminase activity in breast cancer cell extracts (1). A3B overexpression also induces DNA damage and cytotoxicity, but timing is delayed in comparison to A3A overexpression and more cells become multinucleated before dying (1). *A3B* expression levels correlate positively with overall cytosine mutation loads in breast cancer (1). *A3B* overexpression has also been documented in many other tumor types, although overexpression alone is not the sole factor determining whether or not a tumor will have a visible APOBEC mutation signature (1, 2, 34, 60, 180, 194). High *A3B* expression levels have been associated with poor clinical outcomes for estrogen receptor-positive breast

cancer, lung cancer, multiple myeloma, and renal cell carcinoma (70–72, 195–198). Altogether, these data are consistent with a model in which A3B causes mutations that fuel the evolution of multiple different tumor types and negatively influences clinical outcomes (199–201).

However, the importance of A3B in cancer has been questioned with the observation that APOBEC signature mutations are still clearly evident in *A3B*-null breast tumors (202). A 29.5 kbp deletion that removes the entire *A3B* coding sequence and fuses the 3' untranslated regions (UTR) of *A3A* and *A3B* occurs at different frequencies in different populations around the world ranging from <5% in Caucasian and African populations to >30% in many Southeast Asian and Polynesian populations (82). Here, we extend this result to the largest available breast tumor data sets, and test the hypothesis that stably expressed variants of the only other functionally dimorphic DNA deaminase family member, APOBEC3H (A3H), are responsible for APOBEC signature mutations in the absence of A3B.

A3H is the most polymorphic APOBEC3 protein in humans with 7 reported haplotypes (107, 108). Haplotypes II, V, and VII are stable, highly active enzymes with potent retrovirus restriction and hypermutation activities (48, 107, 108, 122, 203). Haplotypes III, IV, and VI are unstable proteins with no reported function, most likely due to a structure-corrupting deletion of Asn15 in the conserved α1-helix (48, 107, 108, 203). Haplotype I is poorly expressed at the protein level due to Gly105 (versus Arg105 in stable haplotypes) and, until now, has failed to show reproducible activity (48, 107, 108, 203, 204). This functional

dimorphism is apparent in immunoblots comparing overexpressed A3H constructs in 293 cells (107) or endogenous A3H proteins in primary T lymphocytes (203), in which the stable haplotypes show robust expression and unstable haplotypes have no detectable expression. An exception to these phenotypes is the A3H-I protein, which is expressed at steady state levels approximately 10-fold lower than stable variants and is difficult to detect by immunoblotting.

Surprisingly, during the course of testing the hypothesis that stable A3H haplotypes contribute to cancer mutagenesis, we found that the poorly expressed haplotype, A3H-I, associates with the APOBEC mutation signature in breast tumors lacking A3B. Linkage studies showed that the responsible SNP encoding Gly105, as well as surrounding genetic variations in *A3H*, are genetically unlinked to most of the remaining *APOBEC3* locus including *A3A* and *A3B*, indicating that other *APOBEC3* genes are unlikely to be responsible. Enzyme activity assays and HIV-1 restriction and mutation experiments demonstrated that A3H-I has strong DNA cytosine deaminase activity with a clear preference for TC dinucleotide substrates. Moreover, after compensation for lower protein expression levels, A3H-I and A3H-II showed similar enzymatic activities and local motif preferences. Subcellular localization images demonstrated that A3H-I has a significantly greater tendency to localize to the nuclear compartment in comparison to A3H-II, suggesting that Gly105 may disrupt an interaction with a cytoplasmic retaining factor and thereby provide a plausible mechanism for somatic mutation in cancer. The potential generality of this mechanism was

indicated by an additional, statistically significant, correlation between A3H-I and early-arising, clonal mutations in lung adenocarcinoma. Overall, together with prior studies on A3B, we propose that the combination of A3B and A3H-I explains the full APOBEC signature in breast and lung cancer, with a strong likelihood of extending more broadly to somatic mutation in other tumor types given the ~50% global frequency of A3H-I, the general nature of the underlying molecular mechanism, and the fact that global searches have failed to identify an APOBEC-hypermutated tumor data set without at least one copy of either A3B or A3H-I.

## RESULTS

### *A3H-I associates with APOBEC signature mutations*

To test the hypothesis that stable forms of A3H (haplotypes II, V, and VII) contribute to breast cancer mutagenesis in the absence of A3B, we computationally screened and manually confirmed all available breast tumor data sets from TCGA to identify specimens with two copies of the *A3B* deletion allele, deduce *A3H* haplotypes, and assess association with APOBEC signature mutations (gene schematics in **Fig. 2.1a, b**). A total of 17 *A3B*-null tumors were identified and, as reported previously from analyses of 13 null-tumors (202), the overall APOBEC signature in these tumors is visibly more pronounced than the composite cytosine mutation spectrum from 577 breast tumors with 2 confirmed intact copies of *A3B* (boxed in **Fig. 2.1c**).

The 17 independent A3B-null tumors could be divided into 4 distinct A3H

haplotype groupings for additional analyses. Contrary to our initial hypothesis, *A3B*-null tumors with an A3H-II/III haplotype combination, one active allele and one inactive allele, showed the lowest (not the highest) proportion of TC-biased cytosine mutations (**Fig. 2.1d**). This result indicated that A3H-II is unlikely to be contributing to the APOBEC mutation signature in breast cancer. We were additionally surprised to find that *A3B*-null tumors with one copy of A3H-I (A3H-I/II or I/III) or two copies of A3H-I showed remarkably strong APOBEC signatures (boxed in **Fig. 2.1d** with statistical significance for non-A3H-I versus A3H-I mutation proportion comparisons in **Supplementary Table S2.1**). This relationship was especially strong and statistically significant for C-to-T transitions and C-to-G transversion mutations in TCA trinucleotide contexts, which represent the largest proportion of APOBEC signature mutations in cancer. These results suggested that the previously classified “unstable” and “inactive” A3H-I protein could be the source of APOBEC signature mutations in breast tumors lacking A3B.

*A3H linkage analysis*

We next performed a genetic linkage analysis using all available 1000 genomes SNP data within the 7-gene human *APOBEC3* locus, arranged tandemly *A3A* through *A3H* in ~120 kbp segment of chromosome 22, to assess the simple possibility that the *A3B* deletion and *A3H-I* allele may be linked genetically and, therefore, that one of the other family members in the same linkage group may be responsible for the observed APOBEC signature mutations

in *A3B*-null, *A3H-I* breast tumors. SNP data from the 1000 genomes project (103) were used to determine linkage disequilibrium ($r^2$) between all *A3H* SNPs and all SNPs for the six other *APOBEC3* genes. The $r^2$ values for linkage disequilibrium between two SNPs range from 0 for no correlation to 1 for a perfect correlation. An $r^2$ of 0.9 indicates that knowledge of one allele will correctly predict the second allele 90% of the time.

Interestingly, we found that the *A3H* gene is only linked strongly to itself and to the immediate upstream gene, *A3G*. Linkage is weaker with genes further upstream, *A3D* and A3F, and completely absent with *A3A*, *A3B*, and *A3C* (**Fig. 2.2a**). Although *A3G* is linked to *A3H*, it is very unlikely to be involved in cancer mutagenesis because the encoded enzyme has an intrinsic preference for the 3' cytosine in C<u>C</u> and CC<u>C</u> motifs (205, 206), which are rarely mutated in cancer (*e.g*., **Fig. 2.1** for cytosine mutations in breast cancer). Importantly, this linkage analysis shows that the *A3H* gene is unlinked to *A3A* (or the *A3A-B* chimeric gene resulting from the *A3B* deletion), suggesting that the significant association with *A3H-I* and APOBEC signature mutations in *A3B*-null tumors may be due directly to A3H-I enzymatic activity and not to A3A [as favored by recent studies (1, 89, 185–188, 207)] nor to any other TC-preferring enzyme encoded by the locus.

Our analyses of the 1000 genomes project data also revealed an interesting inverse correlation between the presence of *A3B* and *A3H-I* in different human populations around the world (-0.76, Spearman's rho; p=0.0015; **Fig. 2.2b**). In areas such as Africa where *A3B* dominates (>95%), *A3H-I* occurs

at a very low frequency (<10%). In other parts of the world, such as Southeast Asia where the *A3B* deletion allele is found at higher frequencies (>30%), *A3H-I* is found at higher frequencies (~65%). In other words, if *A3B* is present then *A3H-I* is rare, and if *A3B* is absent then *A3H-I* is more common. This correlation implies that *A3B* and *A3H-I* may have a redundant physiological function, possibly in antiviral immunity, or that the two genes together may incur a fitness penalty. In addition, when *A3H-I* is common then *A3H-II* is rare (Southeast Asia), whereas when *A3H-I* is rare then *A3H-II* is common (Africa). The null *A3H* haplotypes (III/IV/VI) occur at similar frequencies regardless of population implying the existence of an as yet unknown balancing selective pressure.

*A3H-I is an active DNA cytosine deaminase.*

Prior studies comparing the DNA deamination activity of A3H-I and A3H-II have reported that the former protein is inactive (107, 108, 204) or weakly active (203). Obviously, an inactive enzyme cannot contribute to cancer mutagenesis. We therefore interrogated the activity of this enzyme using multiple independent approaches.

A3H-I, A3H-II, and catalytic mutant derivatives (E56A) were purified from human 293T cells, normalized to be equimolar, and assayed for DNA cytosine deaminase activity using a gel-based single-stranded DNA deamination assay (1, 51, 59, 77, 194) (schematic in **Fig. 2.3a**). In agreement with prior studies (48, 107, 108, 203, 204), A3H-I showed lower overall protein expression levels in cellular lysates in comparison to A3H-II (**Fig. 2.3b**). However, similar

concentrations of each protein could be achieved by concentrating A3H-I, as evidenced by near-equivalent A3H band intensities upon direct visualization of SDS-PAGE fractionated proteins (**Fig. 4.3c**). A head-to-head comparison of A3H-I and A3H-II using a ssDNA with a single TCA target motif indicated near-equivalent enzymatic activities (**Fig. 4.3d**). Importantly, the E56A catalytic mutant derivatives purified and analyzed in parallel had no detectable catalytic activity, which demonstrates that all of the observed activity is due to active A3H-I or A3H-II (and not, for instance, to a co-purifying factor from 293T cells). In independent experiments, recombinant A3H-I from *Sf*9 insect cells also elicited single-stranded DNA C-to-U editing activity indicating no other human factors are required (**Fig. 4.3e**).

Next, HIV-1 restriction and mutation assays were used as biological read-outs for A3H-I activity (schematic in **Fig. 4.4a**). Vif-deficient HIV-1 particles were produced in 293T cells with a range of untagged A3H concentrations up to maximally tolerated amounts. As expected (24, 183), A3H-II and A3G-HA caused strong dose-responsive decreases in virus infectivity (**Fig. 4.4b**). In comparison, overexpression of A3H-I caused more modest, but still significant, drops in virus infectivity (200 ng, $p=0.055$; 400 ng, $p=0.0059$, Welch's t-test). These virus restriction phenotypes correlated with the overall amounts of A3H and A3G-HA proteins expressed in cells and packaged into nascent viral particles (**Fig. 4.4b**). It is notable that the highest quantity of A3H-I yielded an HIV-1 restriction phenotype similar to the lowest amount of A3H-II, and that these transfected amounts produced similar steady-state protein levels by immunoblotting (red

boxed data in **Fig. 4.4b**). These HIV-1 restriction phenotypes lend further support to the *in vitro* biochemical results above and to the surprising finding that A3H-I elicits a level of catalytic activity similar to the better-expressed A3H-II enzyme.

To extend these results, DNA was purified from infected target cells (400 ng condition), and high-fidelity PCR was used to amplify proviral sequences for mutation analyses. A3H-I and A3H-II inflicted an average of 3.4 and 18 C-to-T mutations per kilobase, respectively (>32 independent 276 bp sequences per condition). Interestingly, although A3H-I has lower virus restriction activity per unit of transfected expression plasmid, its intrinsic DNA cytosine deamination preferences strongly resemble those of A3H-II with clear biases for TC dinucleotides (**Fig. 4.4c**). As expected, A3G caused high levels of mutation within C<u>C</u> dinucleotide contexts, which are rarely mutated in cancer. It is noteworthy that the APOBEC3-catalyzed viral cDNA uracils in this system are not subject to normal cellular DNA repair processes because DNA deamination and reverse transcription occur within the physical confines of the capsid-encased viral core, and reverse transcription effectively immortalizes these uracil lesions as viral genomic strand G-to-A mutations prior to integration into the genomic DNA of a susceptible host cell [mechanism reviewed by refs. (24, 183, 184)].

*A3H-I has increased nuclear localization*

We next asked if A3H-I is capable of accessing the nuclear compartment, which is another property likely to be essential for cancer mutagenesis. Prior studies have lacked consensus with epitope-tagged A3H-I showing variable

subcellular localization (48, 108, 109, 204). However, one study proposed that A3H-II (Arg105) is retained in the cytoplasm by interacting with a specific host factor, and that the Gly105 amino acid characteristic of A3H-I dislodges this mechanism and enables entry into the nuclear compartment by passive diffusion (48).

To clarify and advance this important point, we used immunofluorescent microscopy to quantify the localization of untagged A3H-I versus A3H-II in cell lines with varying endogenous A3B levels (null, SK-BR-3; low, HeLa; high, U2OS) (1). A3B-HA and A3G-HA were used as controls for predominantly nuclear and cytoplasmic localization, respectively (**Fig. 2.5**). We found the overall subcellular distributions of A3H-I and A3H-II to be consistent in all cell types, regardless of endogenous A3B levels, with the former enzyme invariably appearing more nuclear than the latter (representative images, **Fig. 2.5a**; quantification, **Fig. 2.5b**). These data show that A3H-I is proficient at entering the nuclear compartment, and advance the general model in which Gly105 disrupts a cytoplasmic retention mechanism and enables the A3H-I enzyme to breach the nuclear compartment and mutagenize genomic DNA.

*A3H-I explains many clonal mutations in lung adenocarcinoma*

Given the results detailed above, particularly the potent enzymatic activity of A3H-I with an intrinsic preference for cytosines in a TC context, the capacity of A3H-I to mutate a variety of substrates, and the ability of A3H-I to breach the nuclear compartment, we next performed a comprehensive analysis of all

available TCGA tumor data sets with n-values >400 exomes to begin to assess whether A3H-I is a general source of mutation in cancer or a unique mechanism that compensates for the loss of A3B in a limited subset of breast cancers. To distinguish between these possibilities, we determined whether A3H-I associates with common mutation signatures including APOBEC, smoking, and ageing (respectively, C-to-T in TCW motifs, C-to-A in any motif, and C-to-T in CG motifs). We predicted an association of A3H-I with APOBEC signature in some tumor types, but not with smoking or ageing in any tumor type because these signatures are clearly due to independent mechanisms. Mutations with the highest frequency of occurrence in a given tumor normalized by copy number and tumor purity are considered early-arising and clonal, whereas any mutation occurring at a lower frequency is considered later-arising and subclonal (**Fig. 2.6a**). Such temporal relationships are critical because different mutational processes have been shown to promote different stages of tumor evolution (177).

Three distinct scenarios emerged from these analyses (**Fig. 2.6b** and **Supplementary Fig. S2.2**). In the first, A3H-I associates with the occurrence of APOBEC signature mutations despite contributions from multiple mutational processes. Nearly 3-fold more early-clonal APOBEC-signature mutations were evident in lung adenocarcinomas with at least 1 copy of A3H-I in comparison to tumors with any other A3H haplotype ($p=0.0024$, Welch's two-tailed t-test; top left histogram in **Fig. 2.6b**). In contrast and as expected, clonal mutation signatures attributable to smoking and ageing did not correlate with A3H haplotype ($p=0.48$ and $p=0.37$, respectively). Subclonal APOBEC-signature mutations also occurred

independently of A3H haplotype (p=0.13) and, based on prior studies (2, 4, 34, 180), are most likely due to the temporally late upregulation of A3B. Clonal APOBEC signature mutations also appeared to be enriched in breast tumors with A3H-I versus those with any other haplotype, but this difference is not statistically significant (p=0.17, Welch's two-tailed t-test). This result could easily be due to A3B upregulation occurring at variable times during the progression of individual breast tumors because the trend completely disappears in the analysis of subclonal APOBEC signature mutations where A3B has already been implicated strongly (1, 2, 4, 34, 179, 180, 182).

In the second scenario, both clonal and subclonal APOBEC mutation signatures are stronger, as judged by the higher percentages of total cytosine mutations, and there are no correlations with the presence or absence of A3H-I. The clearest example is cervical cancer, where HPV infection is an established early event (132), and HPV infection has been mechanistically linked to A3B upregulation (59, 60). Here, approximately 10% of overall APOBEC signature mutations are clonal for both A3H-I and non-A3H-I cervical tumors, and this proportion rises to nearly 15% for APOBEC signature subclonal mutations (p=0.69 and p=0.10, respectively, Welch's two tailed t-test; **Fig. 2.6b**). Thus, in this virus-induced scenario, we propose that chronic levels of A3B eclipse any mutational contributions from A3H-I or, alternatively, that the *A3H* gene is not expressed at the mRNA level in these tumor types [supported by cell-based HPV studies (59)]. Again, as expected, both clonal and subclonal smoking and ageing mutation signatures are unrelated to A3H haplotype.

In the final scenario, a proportion of cancer types simply do not manifest an APOBEC mutation signature (2, 4, 34, 181, 182, 208). For instance, prostate cancer and low-grade glioma lack an APOBEC signature at any time point (**Supplementary Fig. S2.2**). There are many possible molecular explanations for why the general mutation mechanism described here for A3H-I may not be operational in these cancer types. The simplest is that both basal and induced levels of *A3H-I* mRNA expression are likely to be constrained by cell type and developmental program, and therefore differences between protein-level haplotypes may be irrelevant and incapable of contributing to the overall mutation spectrum.

## DISCUSSION

The studies presented here are the first to implicate A3H-I as an enzymatic contributor to “APOBEC signature” mutations in breast and lung cancer, with a general mechanism that could extend to other cancer types. This work was initiated with the goal of providing a molecular explanation to the paradox that APOBEC signature mutations still exist in *A3B*-null breast cancers (202). We had independently reached and expanded upon the same conclusion (data summarized in **Fig. 2.1c**), and we set out to test the hypothesis that stably expressed variants of the only other functionally dimorphic APOBEC3 family member, A3H, would be responsible. To our surprise, stable *A3H* haplotypes did not explain the APOBEC signature mutations in *A3B*-null tumors; instead, a variant previously deemed unstable and inactive, *A3H-I*, showed a statistically

significant association (**Fig. 2.1d**). A linkage analysis further strengthened this possibility and, simultaneously, discounted participation by other APOBEC3 family members including wild-type A3A and A3A produced by the chimeric *A3A-B* fusion gene generated by the 29.5 kbp *A3B* deletion (**Fig. 2.2**). Biochemical and cell-based studies contrasted with prior reports and clearly demonstrated that A3H-I is not only catalytically active but, once protein concentrations were equalized, it also proved similarly active to the A3H-II enzyme and yielded a near-identical TC-biased single-stranded DNA deamination preference (**Figs. 2.3-2.4**). Subcellular localization studies revealed the likely molecular mechanism responsible for enzymatically active A3H-I gaining access to the nuclear compartment, with A3H-I Gly105 disrupting a likely cytoplasmic retention mechanism (**Fig. 2.5**). Finally, an informatics approach was used to significantly associate A3H-I with clonal APOBEC signature mutations in lung cancer (**Fig. 2.6**). The high frequency of A3H-I in global populations, 48% (54% in Caucasians) based on 1000 genome project data (103), and the general nature of the mechanism described here (mislocalization of an active DNA cytosine deaminase) combine to suggest relevance to other cancer types.

As described in the **Introduction**, previous work has favored A3A and/or A3B as the predominant sources of the APOBEC signature mutation in cancer. The studies presented here are the first to implicate A3H-I and, simultaneously, further support A3B and cast doubt on a possible role of A3A and other APOBEC family members in cancer mutagenesis. For instance, if A3A or another family member caused the APOBEC signature in *A3B*-null tumors, then statistically

significant associations with *A3H-I* would never have emerged from the comprehensive analyses in **Fig. 2.1** for breast cancer and **Fig. 2.6** for lung cancer. Moreover, because the APOBEC signature is not found in *A3B*-null/non-*A3H-I* breast tumors (**Fig. 2.1d**), it is possible that A3B and A3H-I account for all APOBEC signature mutations in breast cancer. In support of this idea, an exhaustive analysis of all TCGA tumor data sets failed to find a single tumor with an APOBEC mutation signature that did not have either one copy of A3B or A3H-I (n=6863 tumors over 15 cancer types). Thus our studies are also the first to positively link A3B and A3H-I at the genetic level to the APOBEC mutation signature observed broadly in cancer. Future population-focused studies will either confirm these data or unambiguously implicate another APOBEC enzyme [*e.g.*, analyses of A3B-null/A3H-inactive (haplotypes III, IV, VI) tumors could be informative although identifying cohorts with statistically significant numbers will be non-trivial].

Many factors contribute to the composite mutation spectra observed in cancer and comparisons with models systems are, at worst, misleading and, at best, challenging and inaccurate. For instance, the APOBEC signature of cytosine mutations in TCA and TCT motifs (**Fig. 2.1**) is the net result of the intrinsic single-stranded DNA preference of the deaminase (53, 77, 209) *and* the biased uracil excision activity of nuclear UNG2 (likely also influenced by contributions from other uracil DNA glycosylases) (57, 210), *and* the biased insertion of adenine or cytosine bases opposite abasic sites by DNA polymerization enzymes (most DNA polymerases follow the A-rule with the

exception of REV1, which preferentially inserts cytosine and accounts for APOBEC-induced C-to-G transversions) (55, 56, 211). The biochemical DNA deamination preferences with recombinant A3H-I in **Fig. 2.3d** are concordant with the observed APOBEC signature mutations in cancer with TCA/T/G targets preferred over TCC. Although A3H-I (this study) and A3B (1, 77, 194) can efficiently deaminate cytosines in TCG, mutations in tumors within this trinucleotide context are numerically less common than those in TCA/T because CpG motifs are underrepresented in exomic regions of the human genome and methylation of these motifs decreases the efficiency of enzymatic deamination *in vitro* by at least 5-fold (51, 76, 212) [further complicated by the fact that methylation increases the efficiency of spontaneous hydrolytic deamination (213)]. However, if a correction factor for trinucleotide motif frequency is applied to the observed mutation frequencies in **Fig. 2.1d**, then cytosine mutations within TCG motifs become enriched relative to the abundance of this motif the human genome, fully consistent with the biochemical preferences of A3H (**Supplementary Fig. S2.3**). A3H-catalyzed mutation in Vif-deficient HIV occurs most frequently within TCG and TCA motifs and least frequently within TCC and TCT motifs, and this result becomes even clearer upon examination of weighted mutation frequencies (**Supplementary Fig. S2.4**) and more closely resembles the weighted mutation frequencies of the breast tumor data sets (**Supplementary Fig. S2.3**). These results do not exactly mirror the biochemical data, likely because many experimental variables differ (enzyme sources, substrates, reaction conditions, *etc*.). Nevertheless, each approach still

contributes an important piece to the overall picture, and the biochemical approach was invaluable for demonstrating that A3H-I is active and has essentially the same intrinsic DNA deamination preferences as A3H-II.

How can prior work be further reconciled? One recent study compared mutation patterns induced by A3A and A3B in yeast with the actual APOBEC signature in human cancer and, relying heavily on comparing frequencies of the -2 base relative to the mutated cytosine, concluded that A3A was 10-times more important than A3B in cancer mutagenesis (207). Unfortunately, this work did not consider A3H and other human APOBEC3 enzymes preferring TC substrates. As discussed above, it difficult to account for the fact that yeast and human genomic structures and DNA repair processes differ and that this is likely to strongly influence the distribution of initial uracil lesions that ultimately become mutations. Moreover, a re-analysis of breast cancer TCGA WES data sets for C-to-T mutations in YTCA versus RTCA contexts revealed that tumors with this bias have at least one copy of A3H-I (**Supplementary Fig. S2.5**). Another recent paper also favored A3A and advanced the idea that the *A3A-B* chimeric gene created by the 29.5 kbp *A3B* deletion disrupts *A3A* regulation and results in elevated A3A protein levels, genomic DNA damage, and cancer mutagenesis (89). If this were the case then A3H-I would not have shown statistically significant associations with breast and lung cancer APOBEC mutation signatures (**Fig. 2.1** and **Fig. 2.6**). This study also relied heavily upon analysis of transfected reporter constructs (not endogenous *A3A*) and failed to do the critical cause/effect experiment of specifically depleting endogenous *A3A* expression in

*A3B*-null SK-BR-3 breast cancer cells and asking if it alone is responsible for the DNA damage phenotypes caused by PMA treatment. We investigated this possibility using CRISPR to recreate the human *A3B* deletion in an isogenic system, the breast cancer cell line MCF-7, and found no significant change in expression of the *A3A-B* fusion gene after deleting *A3B* (**Supplementary Fig. S2.6**). An additional critical point is that many studies purporting to document *A3A* expression in tumors may be undermined by *A3B* RNAseq reads mismapping to *A3A* and/or by *A3B* cDNA cross-hybridizing to *A3A* probes on microarrays and/or by *A3A* expressing immune cells infiltrating the analyzed tumor tissues (1, 194, 207, 214) (**Supplementary Fig. S2.1**). Accordingly, to our knowledge, no studies to-date have documented A3A protein or enzymatic activity in primary tumor samples or tumor-derived cell lines. However, despite mounting evidence against A3A and other family members, we acknowledge that it is currently difficult to eliminate the possibility that one or more of these proteins may be able to contribute to cancer mutagenesis (*e.g*., in a tumor type not addressed here due genetic and statistical constraints of TCGA data sets).

Our results suggest distinct temporal models for the generation of APOBEC signature mutations in cancer. In the first, in *A3B*-null cancers such as in a subset of breast tumors described here, A3H-I may provide a low mutator activity that over a long period of time results in the observed APOBEC signature mutation spectrum and load (continuous mutator model in **Fig. 2.7a**). This mutation program may be prone to periodic “flares” (not depicted) because at least one virus (HIV-1) has been shown to induce A3H expression in primary

cells (25, 203), and other viruses may have similar stimulatory effects. This model may be particularly relevant to Southeast Asian populations with high frequencies of both *A3H-I* and the *A3B* deletion allele (**Fig. 2.2**). In the second model, in A3B overexpressing tumors such as many breast cancers and HPV-positive cancers, the mutational impact of A3B may be early, strong, constitutive, and additive to that of A3H-I, and the powerful effect of A3B and its similar TC target preferences may rapidly eclipse the A3H-I contribution [activated (early) mutator model in **Fig. 2.7b**]. HPV infection provides a mechanism for A3B upregulation in virus-positive tumor types, but the mechanisms responsible for early A3B induction in virus-negative tumors are less uniform and less clear (*e.g.*, ref. 35). In the third model, the impact of A3H-I is evident among early-arising clonal mutation spectrum in lung adenocarcinomas but it eventually becomes eclipsed by A3B overexpression at a later point in tumor development [continuous mutator plus activated (late) mutator model in **Fig. 2.7c**]. In a variant of this model, the early continuous mutator effect is absent in tumors lacking A3H-I [activated (late) mutator model in **Fig. 2.7d**]. In all of the models, apart from those depicting an early smoking signature, other prominent sources of mutation are excluded for purposes of focusing on the APOBEC signature and the different contributions of A3H-I and A3B observed in this study. Such additional sources of mutation are of course capable of contributing to the overall mutation loads and spectra in various tumor types. These models may extend to APOBEC signature cancers beyond those highlighted here, and future studies should be designed to isolate and quantify the mutagenic contributions of A3H-I,

A3B, and possibly other family members. Future studies should also examine the clinical impact of these different mutational sources together, as well as in isolation, in appropriate populations. Indeed, an analysis of lung cancer incidence in China indicated that unstable/inactive forms of A3H may be protective and, therefore, that A3H-I as the predominant allele in China (~70%) could be a significant risk factor (110).

## METHODS

*APOBEC genotyping and tumor mutation analyses*

TCGA RNAseq and WES data were obtained between April 2014 and January 2016 (https://cghub.ucsc.edu/). Somatic mutation and CNV data were obtained in January 2016 (https://confluence.broadinstitute.org/display/GDAC/Download). Initial candidate *A3B*-null patients were identified using TCGA SNP CNV data with a sample segment mean of -1.5 or less with probe start positions 37693565, 37693563, or 37693530 on chromosome 22. Additional *A3B*-null candidates were identified using WES of the *A3A* and *A3B* genes and flanking exons preserved with the deletion allele. *A3B*-null patients were confirmed by manual inspection of WES alignments (*e.g.*, **Supplementary Fig. S2.7**). A3H genotypes were determined by extracting base calls, quality, and coverage data from WES alignments spanning the *APOBEC* locus for tumor/normal samples acquired from the TCGA. Global *A3H* allele frequencies were estimated using phased variants from all available data as of December 2015 from 1000 Genomes Project (103).

All somatic mutations were isolated with adjacent 5' and 3' nucleotides using the hg19 reference genome. Mutations were binned by trinucleotide context and proportions calculated compared to total somatic mutations. No statistical difference in the total mutation loads were evident between the various genotypic groups of *A3B* and *A3H* breast tumors, which is not particularly surprising given the fact that the durations of tumor growth are unknown (**Supplementary Fig. S2.8**). To take motif abundance into consideration, we first determined the frequency of each trinucleotide motif in the reference sequence by counting each observed motif and dividing by the total number of trinucleotide motifs possible in the reference (length of sequence - 2). Counts of mutations at each trinucleotide motif were then divided by the previously calculated frequency of that trinucleotide in the reference and these adjusted counts were divided by the sum of all adjusted counts to derive weighted frequencies.

The chronological timing of somatic mutations was determined using published methods and definitions (180, 215) with conservative modifications. Specifically, tumor purity as calculated by ESTIMATE from TCGA RNAseqV2 and GISTIC2 copy-number calls from Broad were used instead of ASCAT data to adjust variant allele frequencies (216). 95% confidence intervals (CI) of adjusted allele frequencies were determined by resampling the number of reads supporting the reference and alternate alleles at half of the observed coverage (10,000 bootstraps). Early-clonal mutations were any mutation occurring clonally before a copy number altering event. This was defined as a mutation with an adjusted allele frequency lower 95% confidence interval that is greater than 1

copy (*i.e*., AF > 0.50) in diploid regions of the genome and any mutation that corresponds to 2 or more copies in regions of copy number amplifications (*e.g*., CN = 3, AF = 0.667). Late-clonal mutations (after copy number altering event) are identified as those occurring in diploid regions or regions of copy number amplification, but the 95% CI of the allele frequencies overlap 1 copy of the allele. Non-clonal mutations were defined as any mutation with an upper 95% CI allele frequency that falls below one copy indicating that it is present in a subpopulation of the cells. Proportion of each base substitution in each trinucleotide was calculated on a per patient per temporal category basis. Any patient with fewer than 4 somatic mutations in a particular category was excluded from the analysis to avoid potential biases.

*DNA constructs*

Epitope (C-terminal 2xStrep-3xFlag)-tagged A3H-I and A3H-II constructs for purification from human cells were cloned as PCR fragments into pcDNA4/TO restriction sites HindIII-EcoRV using 5'-NNN-AAG-CTT-ATG-GCT-CTG-TTA-ACA-GCC and 5'-NNN-GAT-ATC-GGC-GGG-ACT-GCT-TTA-TCC and vectors described below as amplification templates. Catalytic mutant derivatives were constructed by standard site-directed mutation. The GST-A3H-I construct was made by PCR subcloning *A3H-I* cDNA from Open Biosystems (GenBank BC069023) using 5' CCC GGG AAT TGG A AT GGC TCT GTT AAC and 5' GCG GCC GC T CAG GAC TTT ATC CTC TC into SmaI/NotI digested pFastBac1-GST (Life Technologies). The A3B-HA, A3G-HA, A3H-I (untagged), and A3H-II (untagged) expression constructs are based on pcDNA3.1+ (Invitrogen). The

A3B-HA, A3G-HA, and untagged A3H-II were constructed by PCR subcloning coding sequencing matching Genbank accessions NM004900, NM021822, and FJ376614.1, respectively (25, 217). The untagged A3H-I construct was made by PCR subcloning *A3H-I* cDNA (GenBank FJ376611) using 5'-NNN-NGA-GCT-CGG-TAC-CAC-CAT-GGC-TCT-GTT-AAC-AGC-CGA-AAC and 5'-NNN-NGT-CGA-CTC-AGG-ACT-GCT-TTA-TCC-TCT-CAA-GCC-GTC into KpnI/XhoI-digested pcDNA3.1+ (Invitrogen). Sanger sequencing was used to confirm the integrity of all constructs.

*Protein purification and DNA deaminase activity assays*

HEK293T cells were transfected with pcDNA4/TO-A3H-I:2xStrep3xFlag or pcDNA4/TO-A3H-II:2xStrep3xFlag (or catalytic mutants). Cells were harvested 48 hrs post-transfection and lysed in 50 mM Tris-HCl pH 8, 1% (v/v) NP-40, 150 mM NaCl, 0.5% (w/v) deoxycholate, 0.1% (w/v) SDS, 5 mM EDTA, 1x EDTA-free Protease Inhibitor Cocktail (Roche), and then further disrupted by sonication. A3H proteins were purified using Strep-tactin resin (IBA). Samples were washed in high salt buffer (20 mM Tris-HCl pH 7.5, 1.5 mM $MgCl_2$, 1 M NaCl, 0.5 mM DTT and 5% glycerol) followed by low salt buffer (high salt with 150 mM NaCl) and elution using 2.5 mM desthiobiotin. Normalized amounts of each purified protein were incubated with a ssDNA substrate containing a target TCA motif and uracil DNA glycosylase for 1 hour, and then treated with mild hydroxide to cleave the deaminated/de-uracilated substrate at the position of the abasic site as described (1, 51, 59, 77, 194). The products of the assay were run on a 15%

urea gel and were imaged using the Typhoon FLA 7000 fluorescent imager (GE Healthcare Life Sciences).

GST-A3H-I was produced using the Bac-to-Bac expression system (Life Technologies). *Sf9* cells were infected with recombinant GST-A3H-I baculovirus at an MOI of 20. Cells were harvested 40 hrs post-infection and lysed as described (218). Cleared lysates were incubated with glutathione sepharose resin (GE Healthcare) and subjected to washes of PBS with 250 mM NaCl in the presence or absence of 1% Triton-X 100. The resin was then resuspended to a 50% slurry in 100 mM Tris pH 7.5, 100 mM NaCl, 10% glycerol, 1 μM $ZnCl_2$, and 5 mM DTT. Aliquots of the 50% slurry with bound GST-A3H were incubated with 100 nM ssDNA substrate for 2 hrs. Deamination events were detected by uracil DNA glycosylase and mild alkaline treatment as described (219). Gel images were obtained using a FX fluorescence scanner (BioRad) and background subtraction and integrated gel band intensities were quantified using ImageQuant (GE Healthcare). A3H-I showed no activity after elution of the enzyme from the affinity resin, and activity of the resin-bound enzyme diminished with time at 4 °C.

Virus infectivity and hypermutation experiments were done using 50% confluent 293T cells in DMEM (HyClone) with 10% FBS, and 0.5% Pen/Strep. Cells were co-transfected with 300 ng pΔNRF (gag-pol-rev-tat), 100 ng pMDG (vesicular stomatitis virus G protein), 100 ng pCS-CG (lentiviral transfer vector encoding GFP), and a titration (50, 100, 200, and 400 ng) of APOBEC3 expression plasmids (vector, A3G-HA, A3H I, or A3H II) using TransIT-LT1 (Mirus Bio). After 24 hrs, the media was removed and replaced with fresh media

containing 50 U/ml DNase to digest free plasmid DNA. After another 24 hrs, virus containing supernatants were purified with 0.45 μm PVDF filters, treated with DNase at 50 U/ml, and used to infect 293T target cells. Virus particles were isolated through a 20% sucrose cushion from the remaining supernatant. Producer cells were used for flow cytometry (FACSCanto II Ruo; BD Biosciences) and analyzed (FlowJo) to monitor transfection efficiency. The remaining virus particles and producer cells were lysed in 2.5x Laemmli sample buffer and used for immunoblotting. A3H was detected using a rabbit anti-A3H polyclonal antibody (Novus Biologicals, NBP1-91682), and A3A-HA and A3G-HA were detected using mouse anti-HA monoclonal antibody (Covance). Tubulin and p24 served as loading controls and were detected by a mouse anti-tubulin mononclonal antibody (Covance) and a mouse anti-p24 monoclonal antibody (NIH AIDS Reagent Program). Primary antibodies were detected with anti-rabbit and anti-mouse HRP-conjugated secondary antibodies (Jackson ImmunoResearch). After 48 hrs infection, target cells were harvested and used to analyze infectivity by GFP flow cytometry and to obtain genomic DNA for viral hypermutation analyses.

Genomic DNA from infected 293T target cells was used for viral hypermutation studies. Genomic DNA was first treated with DpnI (NEB) to remove plasmids that may have carried over from the original transfection. Phusion high-fidelity DNA polymerase (NEB) was used in nested PCR reactions to amplify a region in *GFP* with outer primers 5′-CCTRAARTTCATCTRCACCA and 5′-CACRCTRCCRTCCTC, followed by the inner primers 5′-

CCRCTACCCCRACCAC and 5′-TCACCTTRATRCCRTTCTTC. Amplicons were cloned into pJET1.2 (Fermentas) and subjected to Sanger sequencing. Each sequence was aligned to the *GFP* reference sequence using BWA (220) and mutations were quantified using a custom perl script available from https://github.com/gjstarrett/countSangerMuts (**Supplementary Software S2.1**).

*Immunoblotting*

Transfected 293T cells were lysed with Laemmli sample buffer and quantified for total protein by Lowry Assay (Sigma). 35 µg whole cell lysate was separated on a 12% SDS-PAGE. A3H was detected using a mouse anti-A3H mAb diluted 1:1000 (203) (NIH AIDS Reagent, Cat#12155). Alpha-tubulin was detected with a rabbit polyclonal serum (Thermo, PA1-20988). Fluorescent anti-mouse (680nm, Licor, 926-68070) and anti-rabbit (800nm, Licor, 32211) secondary antibodies were used for detection. Scans were performed on a Licor Odyssey infrared scanner.

*Immunofluorescent microscopy*

Cell lines were obtained from the ATCC. HeLa cells were cultured in DMEM (HyClone) and SKBR3 and U2OS cells in McCoy's 5A (Corning), each supplemented with 10% fetal bovine serum and 0.5% Pen/Strep. 50% confluent cells were transfected with 200 ng of A3B-HA, A3G-HA, A3H-I, or A3H-II with TransIT-LT1 according to the manufacturer's instructions (Mirus Bio). After 24 hrs incubation, cells were fixed in 4% paraformaldehyde and incubated overnight with either rabbit anti-HA (Cell Signaling, #3724; 1:250) or rabbit anti-A3H (Novus Biologicals, NBP1-91682; 1:200) followed by a 2 hr incubation with anti-rabbit

TRITC conjugated antibody (Jackson ImmunoResearch). The nuclei were stained using a 0.1% Hoechst solution. Cells were imaged using 1000X magnification on a Nikon Inverted TiE Deconvolution Microscope. Nuclear versus cytoplasmic localization of APOBEC3 proteins was calculated from these images using MATLAB (MathWorks) with the assistance of the University of Minnesota Imaging Center.

*Statistics.*

Statistical significance between categorical data was calculated by Welch's two-tailed t-test using both Graphpad Prism and the R statistical computing environment. Trinucleotide enrichments were calculated by Fisher's exact test.

**ACKNOWLEDGEMENTS**

We thank TCGA for access to RNAseq and mutation data, and C. Richards, N. Shaban, and J. Wang for comments. Salary support for G.J.S. was provided by a NSF Graduate Research Fellowship (Grant No. 00039202). The University of Minnesota University Imaging Centers assisted with fluorescent imaging and quantification. Cancer studies in the Harris laboratory have been supported by grants from the Department of Defense Breast Cancer Research Program (BC121347), the Jimmy V Foundation for Cancer Research, and the National Cancer Institute (R21 CA206309). A3H work in the Chelico laboratory is supported by a NSERC Discovery Grant. R.S.H. is an Investigator of the Howard Hughes Medical Institute.

## AUTHOR CONTRIBUTIONS

R.S.H and G.J.S conceived the original hypothesis and experimental plans. G.J.S, D.E., and N.A.T analyzed TCGA data. R.P.L. performed genomic mutation experiments and G.J.S. analyzed the data. J.L.M., Y.F., M.B.A., L.C. produced recombinant enzymes and together with M.A.C., J.L.M., and G.J.S. performed deaminase assays. E.M.L performed HIV infectivity and hypermutation assays as well as imaging experiments. E.K.L. generated and characterized the A3B-null MCF-7L clones. R.S.H and G.J.S drafted the manuscript and all authors contributed revisions.

## ADDITIONAL INFORMATION

**Competing financial interests:** R.S.H. is a cofounder of ApoGen Biotechnologies Inc. The other authors declare no competing financial interests.

**Data Availability:** Data that support this study were downloaded from https://confluence.broadinstitute.org/display/GDAC/Download, https://cghub.ucsc.edu, or https://gdc.nci.nih.gov. All of the remaining data are included within the Article and its supplementary files or available from the author upon request.

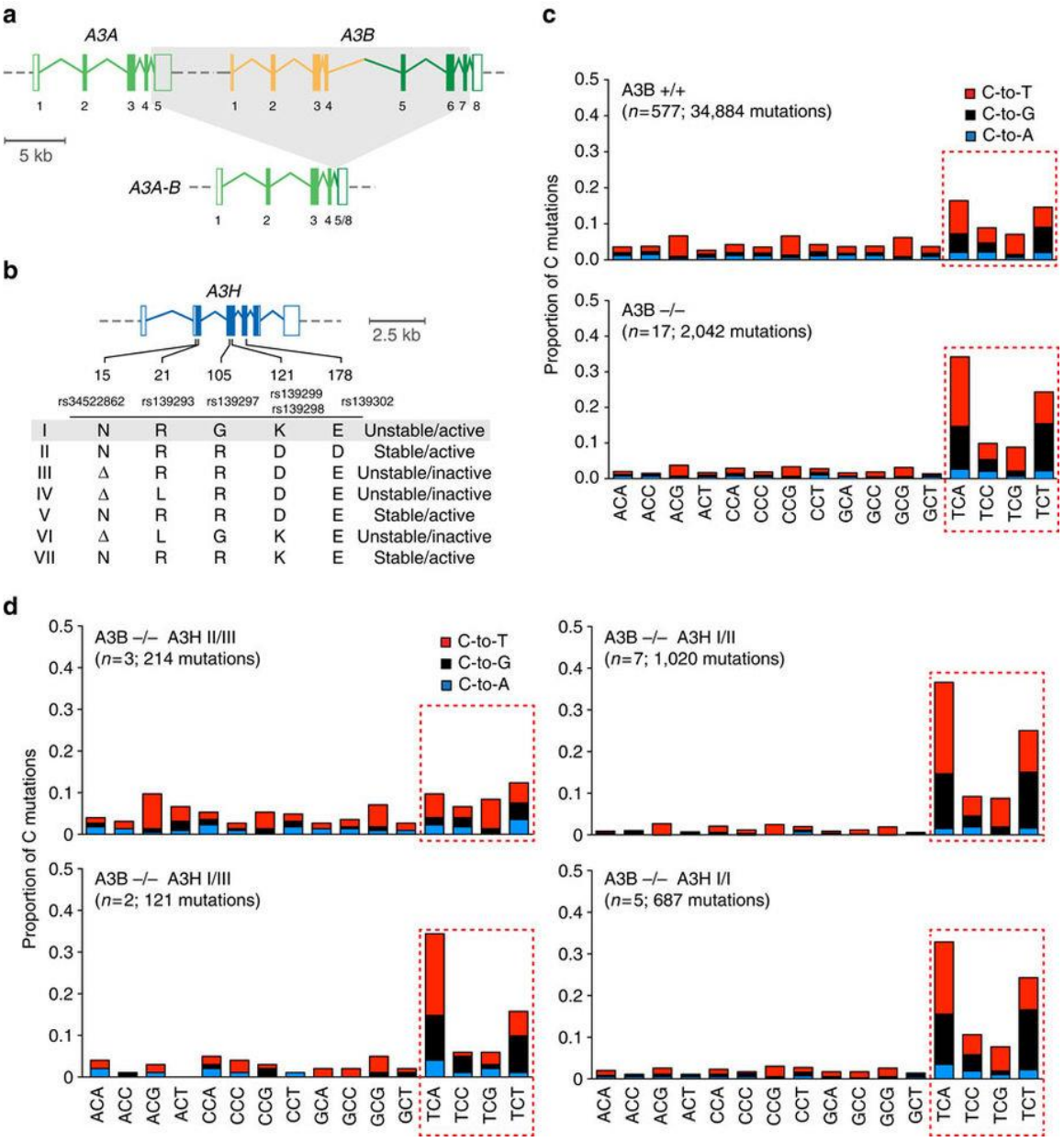



**Figure 2.1. A3H haplotype I accounts for APOBEC signature mutations in *A3B*-null breast tumors.**

**a**, Schematic of *A3A*, *A3B*, and the *A3A-B* fusion gene. Exons are numbered and indicated by boxes, and coding regions are shaded. A 5 kbp scale is indicated.

**b**, Schematic of the *A3H* gene with haplotype-defining amino acid variants and SNP numbers listed below. Labeled as in panel a, except the scale indicates 2.5

kbp.

**c**, Bar plots depicting the proportions of cytosine mutations occurring in the indicated trinucleotide motifs in A3B$^{+/+}$ and A3B$^{-/-}$ breast tumors with n-values and total mutation numbers in parentheses. C-to-T, C-to-G, and C-to-A are represented by red, black, and blue shading, respectively.

**d**, Bar plots depicting the proportions of cytosine mutations occurring in the indicated trinucleotide motifs in A3B$^{-/-}$ breast cancers with the indicated *A3H* haplotype combinations (n-values and total mutation numbers in parentheses). C-to-T, C-to-G, and C-to-A are represented by red, black, and blue shading, respectively.

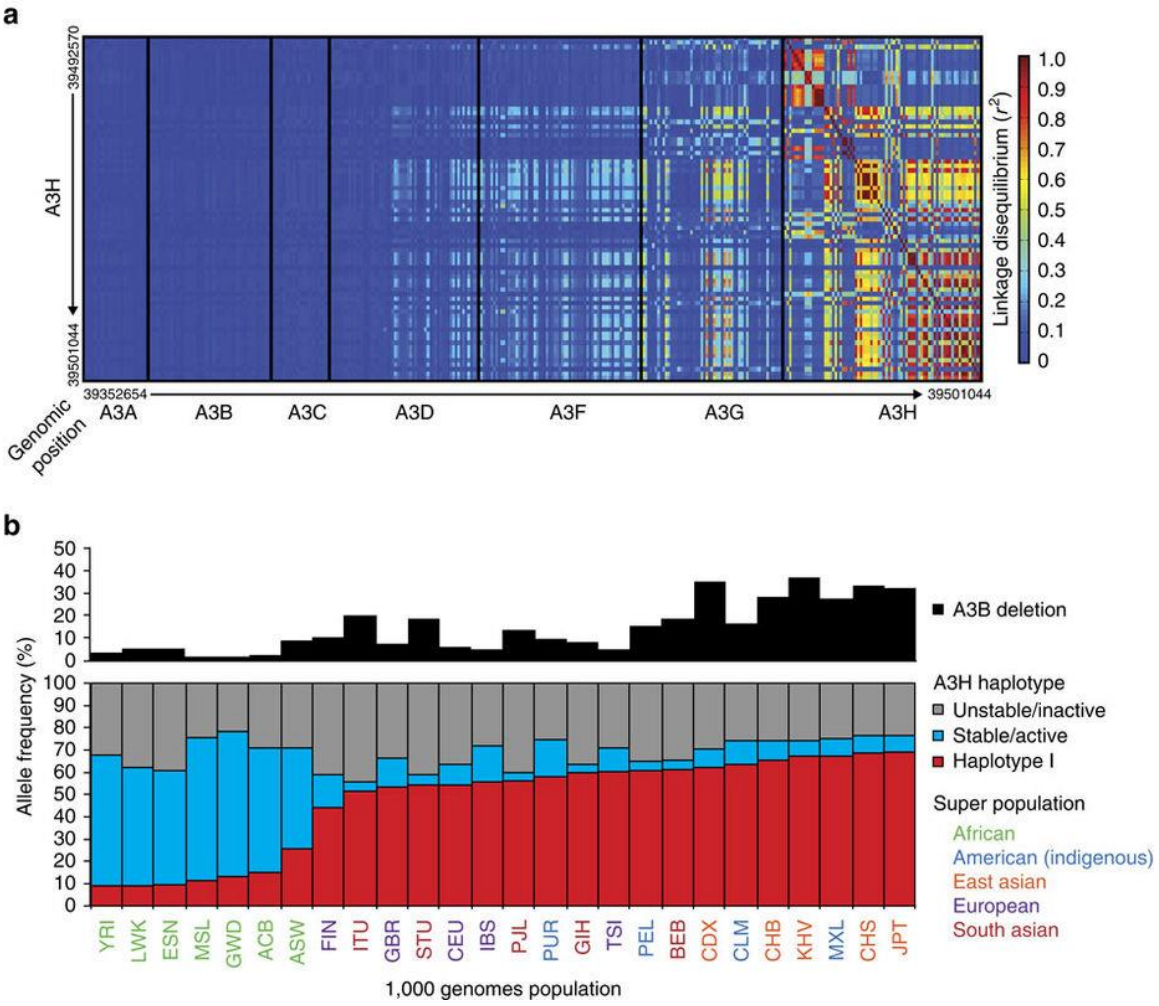



**Figure 2.2. Polymorphisms in A3H are not in linkage disequilibrium with A3A or A3B.**

**a**, Heatmap showing the strength of linkage ($r^2$) of SNPs located within the *A3H* gene versus the rest of the *APOBEC3* locus.

**b**, Bar plots of the *A3B* deletion and *A3H* haplotype frequencies for the indicated populations (A3H-I in red; stable A3H-II/V/VII in blue, and unstable A3H-III/IV/VI in gray). Superpopulations are color coded for visualization of larger geographic areas, and individual 3-letter population identifiers are from the 1000 genomes project

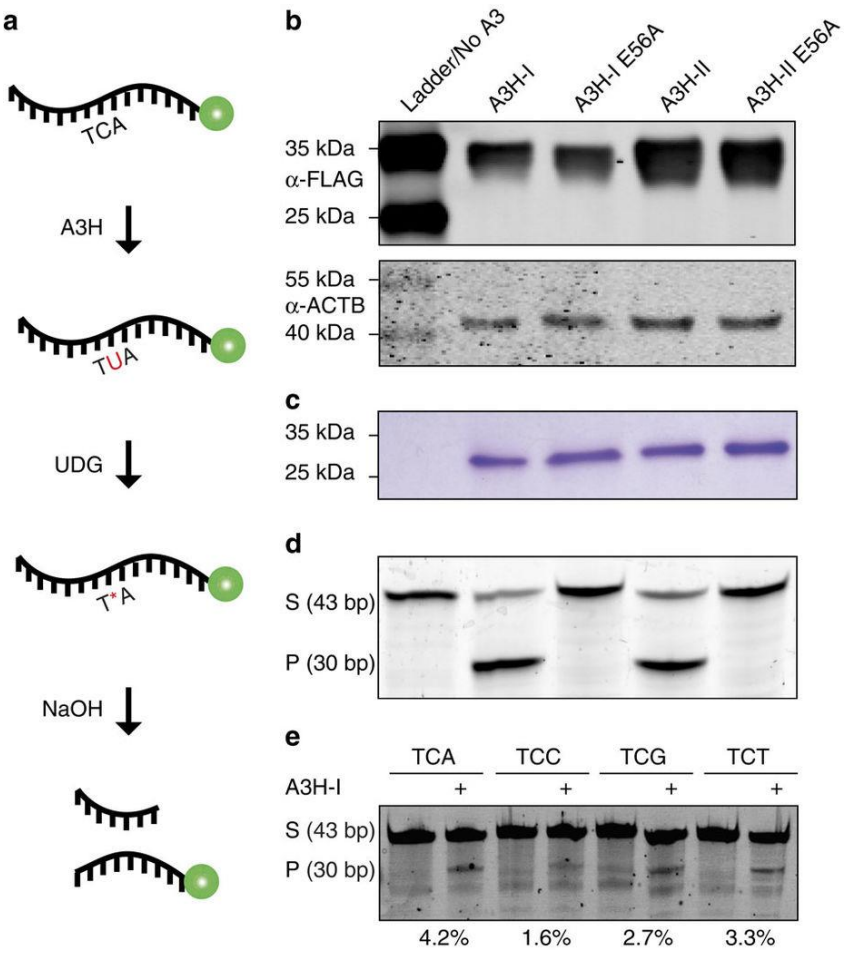



**Figure 2.3. A3H haplotype I is an active DNA cytosine deaminase.**

**a**, Schematic of the single-stranded DNA deamination assay. A3H-mediated deamination yields a uracil that, upon excision by excess uracil DNA glycosylase, is converted into a hydroxide-labile abasic site.

**b**, Image of a Coomassie-stained gel with approximately equal amounts of A3H-I, A3H-II, and catalytic mutant derivatives purified from 293T cells.

**c**, Activity data for the recombinant A3H proteins shown in panel b (S, substrate; P, product).

**d**, Activity of GST-A3H-I purified from insect cells using the indicated trinucleotide containing single-stranded DNA substrates (S, substrate; P, product).

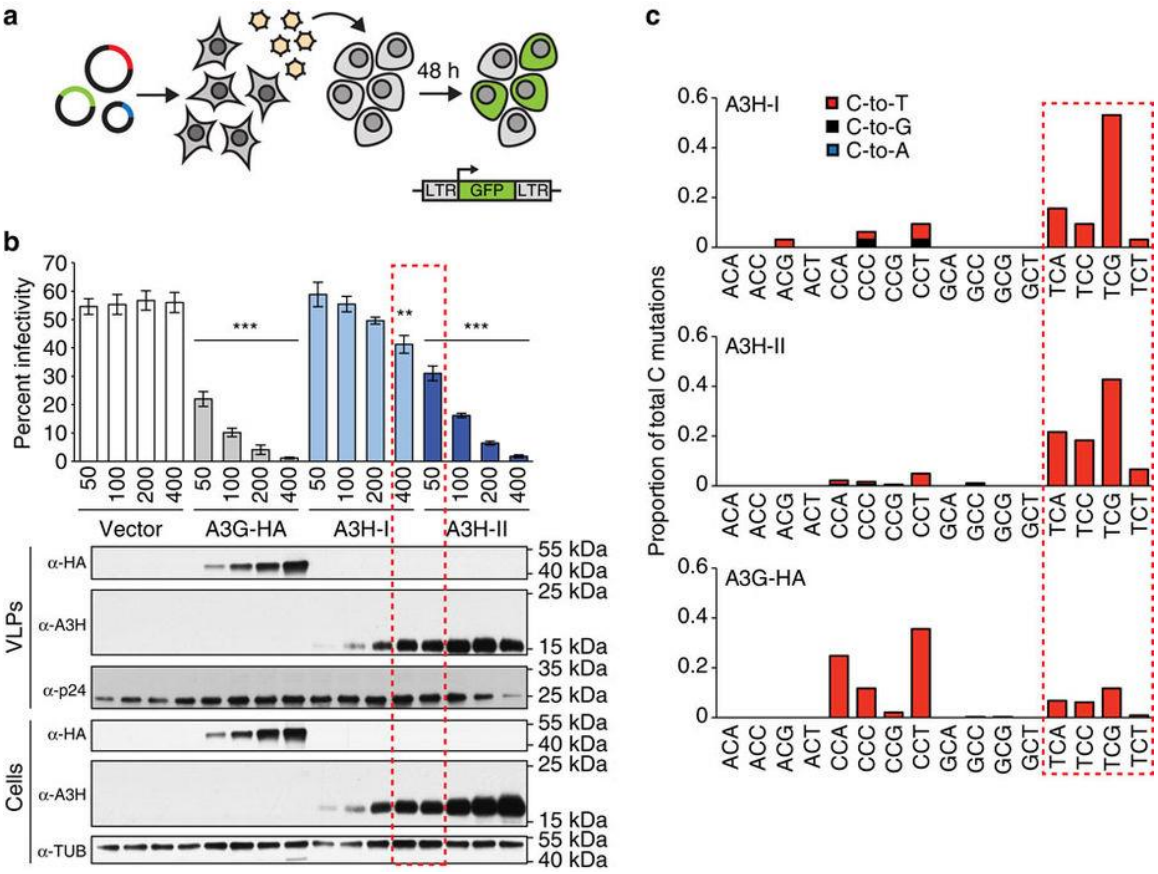



**Figure 2.4. A3H haplotype I has enzymatic activity against viral DNA.**

**a**, HIV-1 infectivity assay. Viral and A3 expression vectors are transfected into 293 producer cells and, after 48 hrs, virus-containing supernatants are titered by infecting CEM-GFP reporter cells in which an integrated LTR-GFP cassette is activated by the Tat protein expressed from newly integrated viruses. Infectivity is quantified by flow cytometry and calculating the percentage of GFP-positive reporter cells.

**b**, Mean and SEM plotted for three biological replicates of HIV-1 infectivity data for Vif-deficient viruses produced in 293 cells expressing a vector control, A3G-HA, A3H-I, or A3H-II (*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, Welch's two sided t-test). Immunoblots for the indicated proteins in cell lysates and virus containing

supernatants are shown below.

**c**, C-to-T mutation distribution in viral DNA sequences recovered from CEM-GFP reporter cells. C-to-T, C-to-G, and C-to-A are represented by red, black, and blue shading, respectively. The mutations are reported for the viral cDNA strand, rather than the conventional genomic strand to facilitate comparisons with tumors

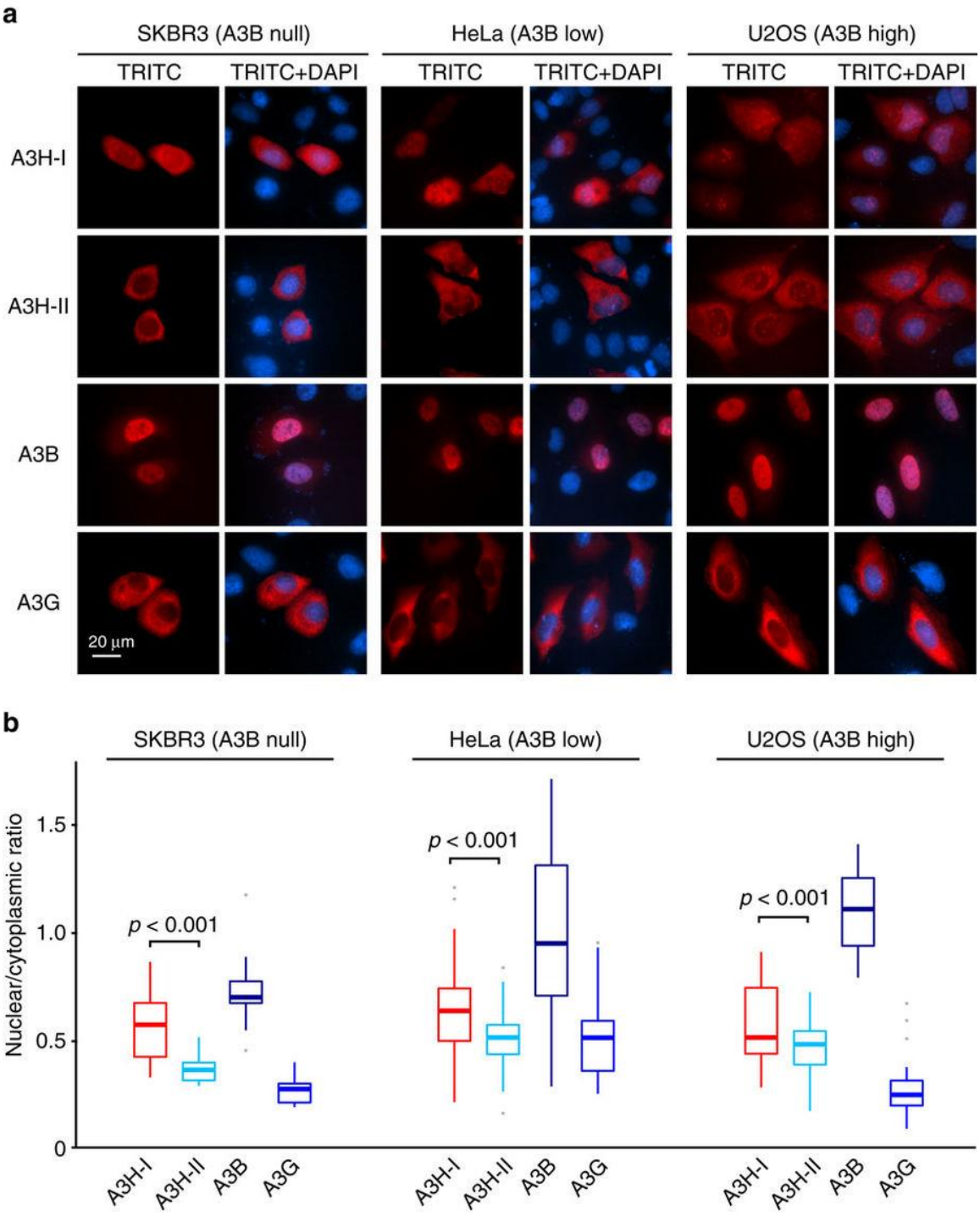



**Figure 2.5. A3H haplotype I has greater nuclear localization than haplotype II.**

**a**, Representative images of A3H-I (untagged), A3H-II (untagged), A3B-HA, and

A3G-HA in SK-BR-3, HeLa, and U2OS cells. The 20 μm scale applies to all images.

**b**, Whisker plots quantifying the subcellular localization data as nuclear-to-cytoplasmic ratios for n > 50 cells per condition. The average is shown, the error box represents the first and third quartiles, and the whiskers extend to the highest value within 1.5x the interquartile range (p-values determined by two-tailed Welch's t-test).

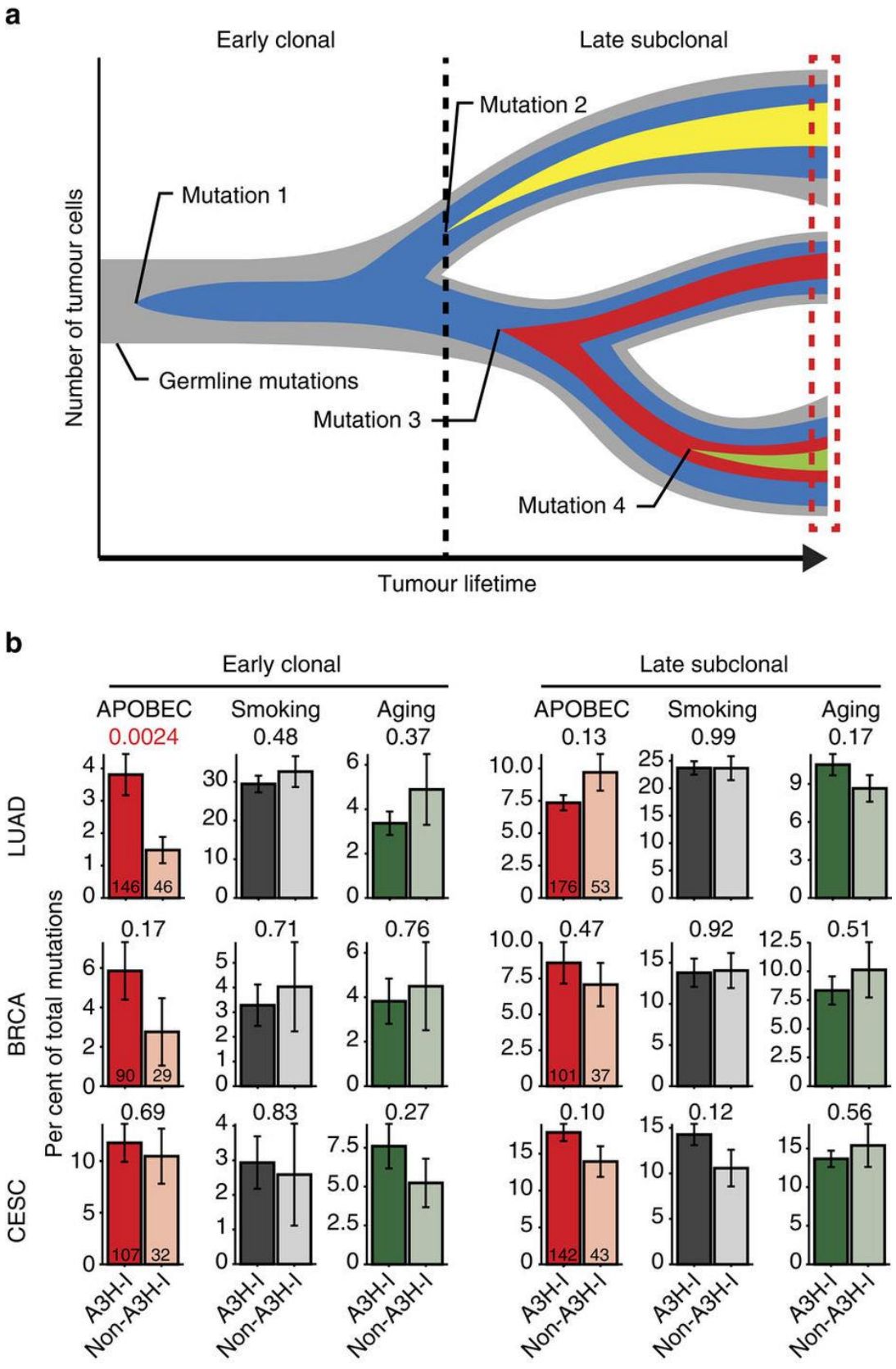

a
Early clonal
Late subclonal
Mutation 2
Mutation 1
Number of tumour cells
Germline mutations
Mutation 3
Mutation 4
Tumour lifetime
b
Early clonal
Late subclonal
APOBEC
Smoking
Aging
APOBEC
Smoking
Aging
0.0024
0.48
0.37
0.13
0.99
0.17
LUAD
146
46
176
53
0.17
0.71
0.76
0.47
0.92
0.51
BRCA
90
29
101
37
0.69
0.83
0.27
0.10
0.12
0.56
CESC
107
32
142
43
Per cent of total mutations
A3H-I
Non-A3H-I

**Figure 2.6. A3H haplotype I contributes APOBEC signature mutations to lung cancer.**

**a**, Tree-based cartoon of tumor evolution. To represent the heterogeneity of each tumor deep sequencing data set (boxed area), 1 trunk mutation (blue) and 3 branch mutations (other colors) are depicted on a background of normal germline DNA (grey). Trunk mutations occur early and are found in every tumor branch, whereas branch mutations occur later in one or more branches (*i.e*., clonal versus subclonal).

**b**, Bar plots showing the frequency of clonal and subclonal mutations in the indicated cancer types attributable to APOBEC, smoking, and ageing (LUAD, lung adenocarcinoma; BRCA, breast cancer; CESC, cervical cancer). Each bar represents the average proportion +/- SEM of signature mutations occurring within each *A3H* haplotype group (*i.e*., *A3H-I* versus non-*A3H-I*). The total number of tumors with 1 or 2 copies of G105 (A3H-I) or 2 copies of R105 (A3H-II and other haplotypes) is indicated within the first set of histogram bars. Welch's two-tailed t-test for each category was used to calculate the p-value above each graph.

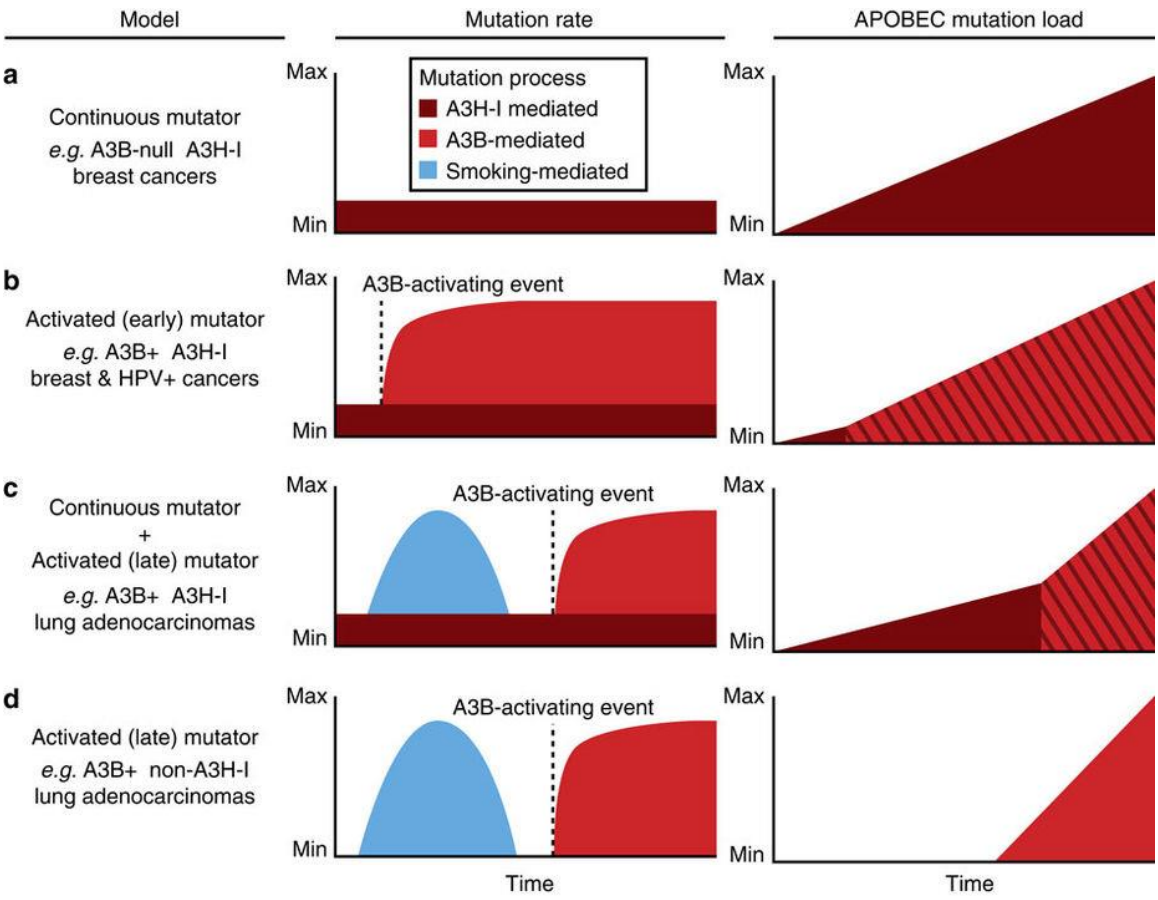



**Figure 2.7. Models for differential APOBEC mutation accumulation in cancer.**

The far left column describes the *A3B* and *A3H-I* genotypes of each model as well as examples of relevant tumor types. The middle columns show the average mutation rate over time for each model with sources of mutations highlighted in different colors, smoking (blue), A3B (red), and A3H-I (maroon). The far right column depicts the accumulation of somatic APOBEC signature mutations over time, with mutations mediated by A3B and A3H-I represented in red and maroon, respectively. Somatic mutations from both APOBEC3 enzymes are shown as

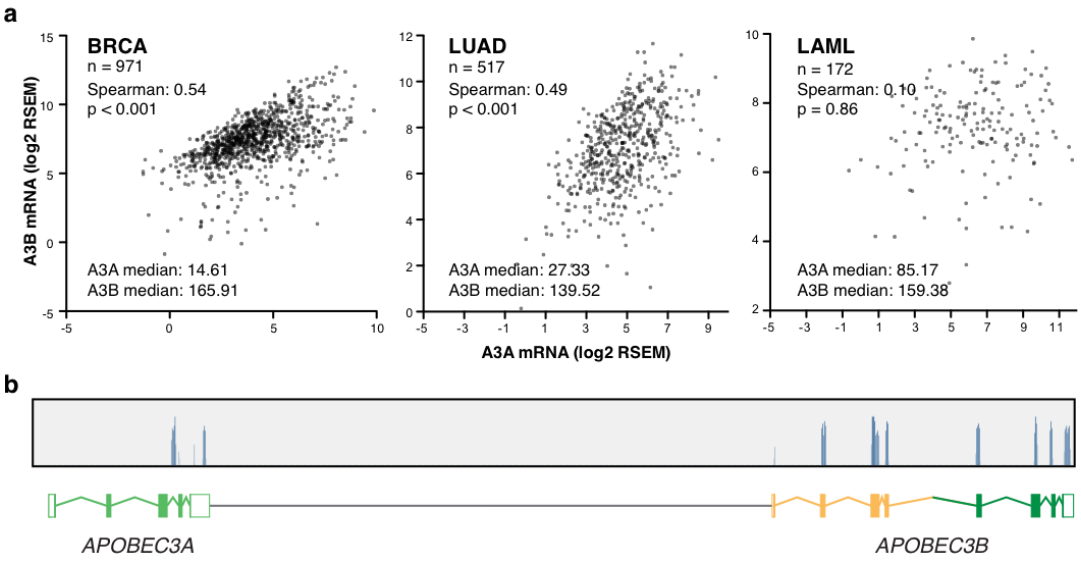
red/maroon diagonal stripes to highlight that these mutations from these sources are not clearly distinguishable.

**a**, The continuous mutator model depicts constant A3H-I mediated mutagenesis and subsequent accumulation of APOBEC-signature mutations over time in the absence of A3B as may be occurring in some breast cancers.

**b**, The activated (early) mutator model depicts a rapid increase in A3B-mediated mutations and APOBEC signature mutations after an A3B-activating event such as HPV-infection in cervical cancers or a currently unknown mechanism in breast cancers.

**c**, The continuous mutator plus activated (late) mutator model depicts the constant accumulation of APOBEC-signature mutations mediated by A3H-I as shown in panel a. For contrast, the distinct contribution from smoking-mediated mutagenesis (blue) is shown as an early finite time period. Late activation of A3B then leads to a more rapid accumulation of APOBEC signature mutations over time effectively eclipsing the A3H-I contribution.

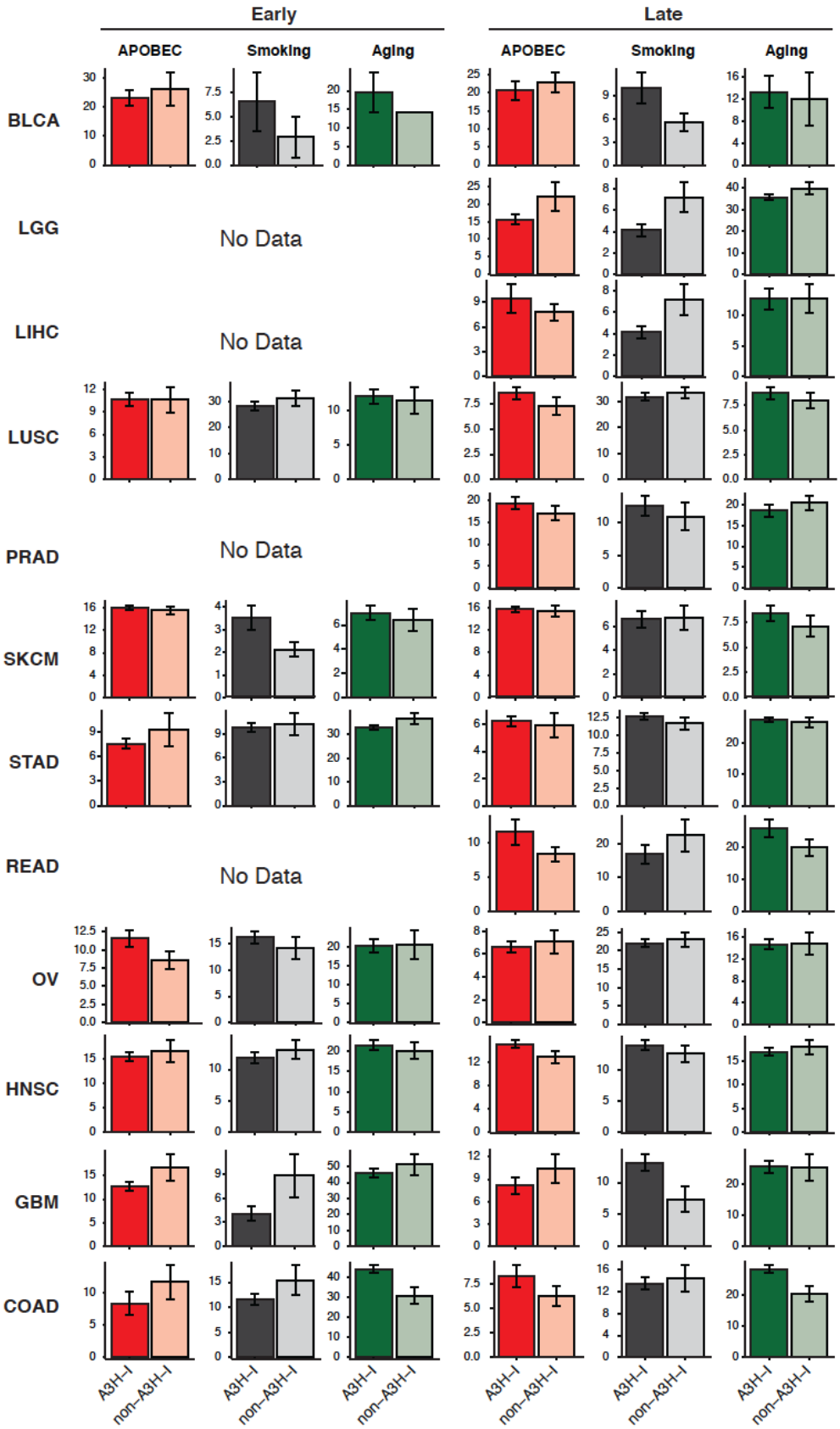
**d**, The activated (late) mutator model is nearly identical to the model shown in panel c**,** however the absence of A3H-I results in no early APOBEC-signature mutations as may be occurring in some lung adenocarcinomas.




**Supplementary Fig. S2.1. Evidence for cross-mapping of RNAseq reads from *A3B* to *A3A* in tumor samples impacting interpretations of mRNA expression.**

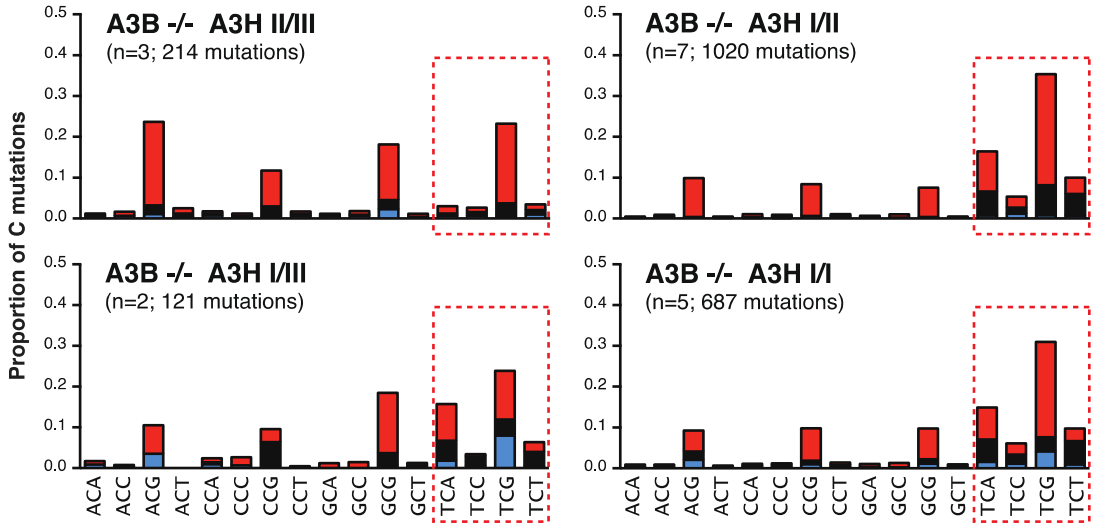
**a**, A significant positive correlation is evident between *A3B* and *A3A* mRNA expression by RNAseq from tumors expressing high levels of *A3B* and low levels of *A3A* (*e.g.*, BRCA, breast invasive ductal carcinoma; LUAD, lung adenocarcinoma). In contrast, consistent with physiological expression of *A3A* in myeloid lineage cell types, the positive correlation disappears in acute myeloid leukemia, where *A3A* appears to be expressed at higher levels (LAML, acute myeloid leukemia).

**b**, Alignment of synthetically generated Illumina-like reads from *A3B* mRNA sequence mapping to *A3B* (as expected) and mis-mapping to *A3A* (~25%).


Early
Late
APOBEC
Smoking
Aging
APOBEC
Smoking
Aging
BLCA
LGG
No Data
LIHC
No Data
LUSC
PRAD
No Data
SKCM
STAD
READ
No Data
OV
HNSC
GBM
COAD
A3H-I
non-A3H-I
A3H-I
non-A3H-I
A3H-I
non-A3H-I
A3H-I
non-A3H-I
A3H-I
non-A3H-I
A3H-I
non-A3H-I

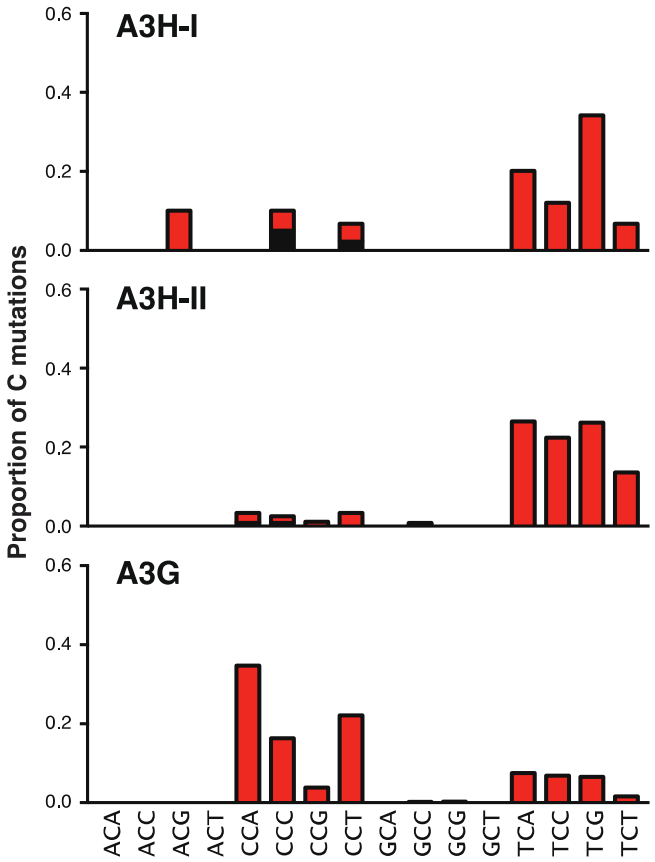
**Supplementary Fig. S2.2. A3H-I and temporal separation of mutation signature associations in select TCGA cohorts.**

Bar plots showing the frequency of clonal and subclonal mutations in the indicated cancer types attributable to APOBEC, smoking, and ageing (PRAD, prostate adenocarcinoma; BLCA, bladder urothelial carcinoma; SKCM, skin cutaneous melanoma; STAD, stomach adenocarcinoma; HNSC, head and neck squamous cell carcinoma; KIRC, kidney renal clear cell carcinoma; LGG, brain lower grade glioma; LUSC, lung squamous cell carcinoma; COAD, colon adenocarcinoma; READ, rectum adenocarcinoma; GBM, glioblastoma multiforme; THCA, thyroid carcinoma). Each bar represents the average proportion +/- SEM of signature mutations occurring within each *A3H* haplotype group (*i.e*., *A3H-I* versus non-*A3H-I*). The total number of tumors with 1 or 2 copies of G105 (A3H-I) or 2 copies of R105 (A3H-II and other haplotypes) is indicated within the “Late APOBEC” set of histogram bars. Welch’s two-tailed t-test for each category was used to calculate the p-values for all *A3H-I* and non-*A3H-I* pairings and none were significant for APOBEC mutation signatures. No data means either truly no available data in the indicated category or insufficient data for this analysis.



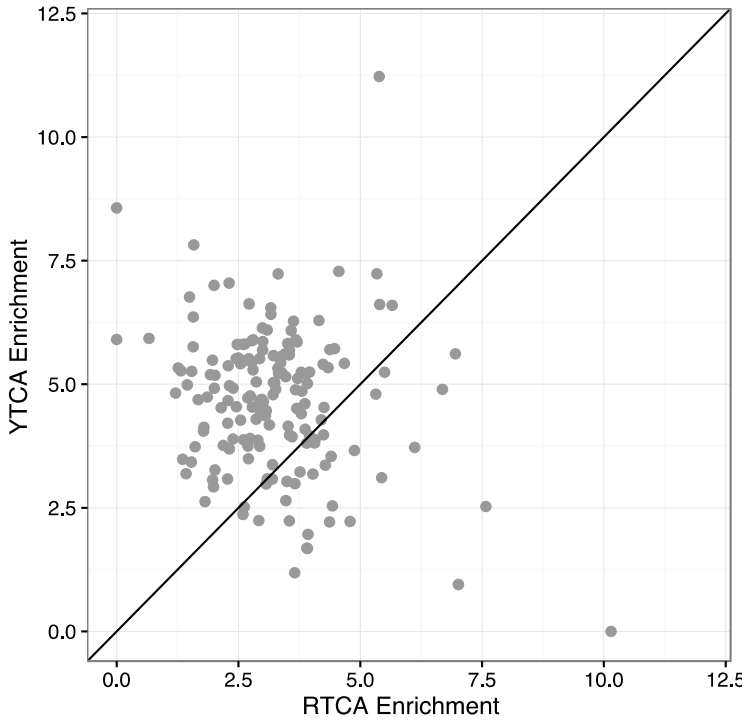

**Supplementary Fig. S2.3. Weighted mutation distributions by *A3H* haplotype from TCGA breast cancer data (additional analysis of data presented in Fig. 1d).**

Bar plots depicting the proportions of cytosine mutations occurring in the indicated trinucleotide motifs weighted by motif abundance in the human genome in *A3B*$^{-/-}$ breast cancers with the indicated *A3H* haplotype combinations (n-values and total mutation numbers in parentheses). C-to-T, C-to-G, and C-to-A are represented by red, black, and blue shading, respectively. Aging-related cytosine mutations in NCG motifs are prevalent in breast tumors regardless of *A3H* genotype, as expected. In contrast, APOBEC signature mutations are only evident in *A3B*-null tumors with *A3H-I* genotypes and, after correcting for motif abundance, TCG and TCA are mutated more than TCT, and TCC becomes the least mutated trinucleotide (analogous to biochemical data in Fig. 3e and HIV mutation data in Fig. 4c and weighted by motif abundance in Supplementary Fig.



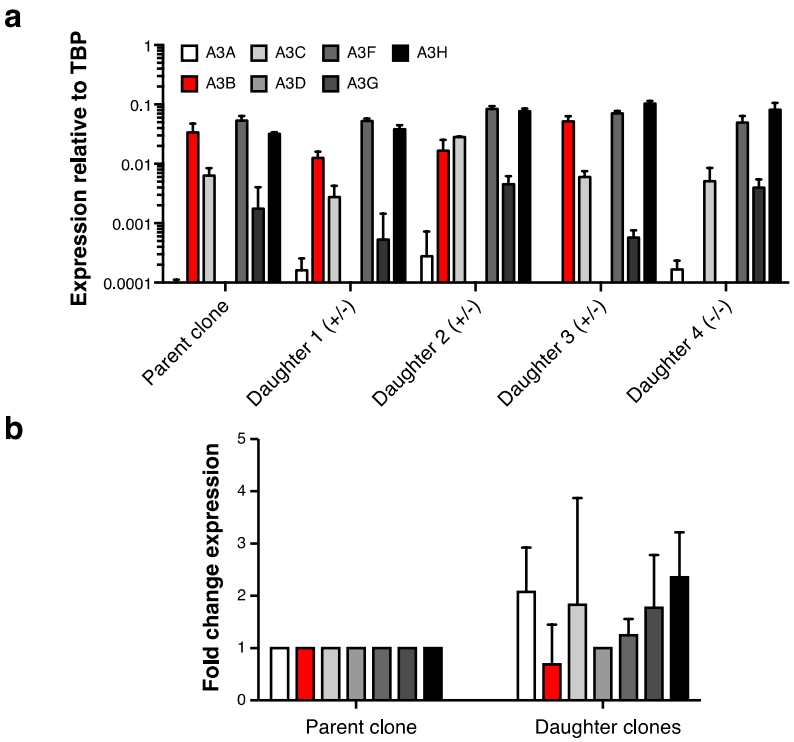

**Supplementary Fig. S2.4. Weighted mutation distribution of mutations in viral DNA sequences (additional analysis of data presented in Fig. 4c).**

C-to-T mutation distribution in viral DNA sequences weighted by motif abundance (NGN) in amplicon recovered from CEM-GFP reporter cells. C-to-T, C-to-G, and C-to-A are represented by red, black, and blue shading, respectively. The mutations are reported for the viral cDNA strand, rather than the conventional genomic strand to facilitate comparisons with tumor mutation data. After correcting for motif abundance, TCG and TCA are mutated more than TCT and TCC.




**Supplementary Fig. S2.5. Re-analysis of mutated YTCA vs RTCA motifs in breast tumors.**

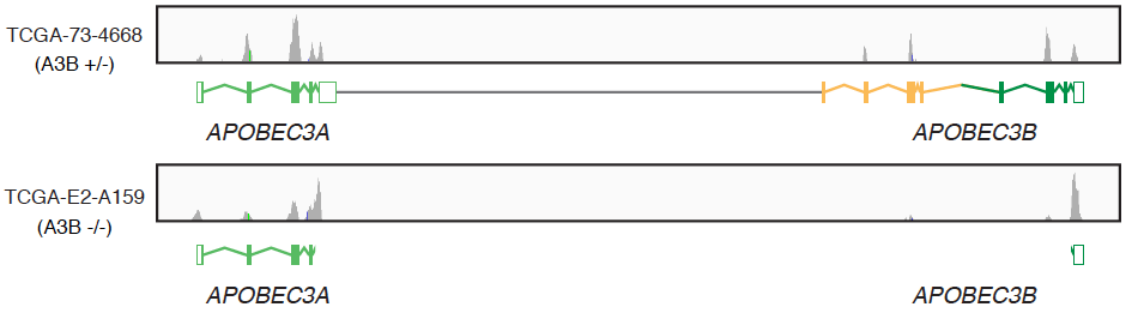
Scatterplot of the results of a re-analysis of data shown in Supplementary fig. 4g of Chan *et al*., 2016 (ref. #20). This plot shows a slight enrichment of C-to-T mutations in YTCA versus RTCA motifs in TCGA breast tumor exomes (each data point is derived from one tumor exome; n=164). This bias may be explained by A3H-I (tumors represented by red shaded symbols).




**Supplementary Fig. S2.6. *A3A* mRNA levels in MCF-7L cells and derivatives engineered by Cas9/CRISPR to have an *A3B* deletion identical to the naturally occurring deletion in humans.**

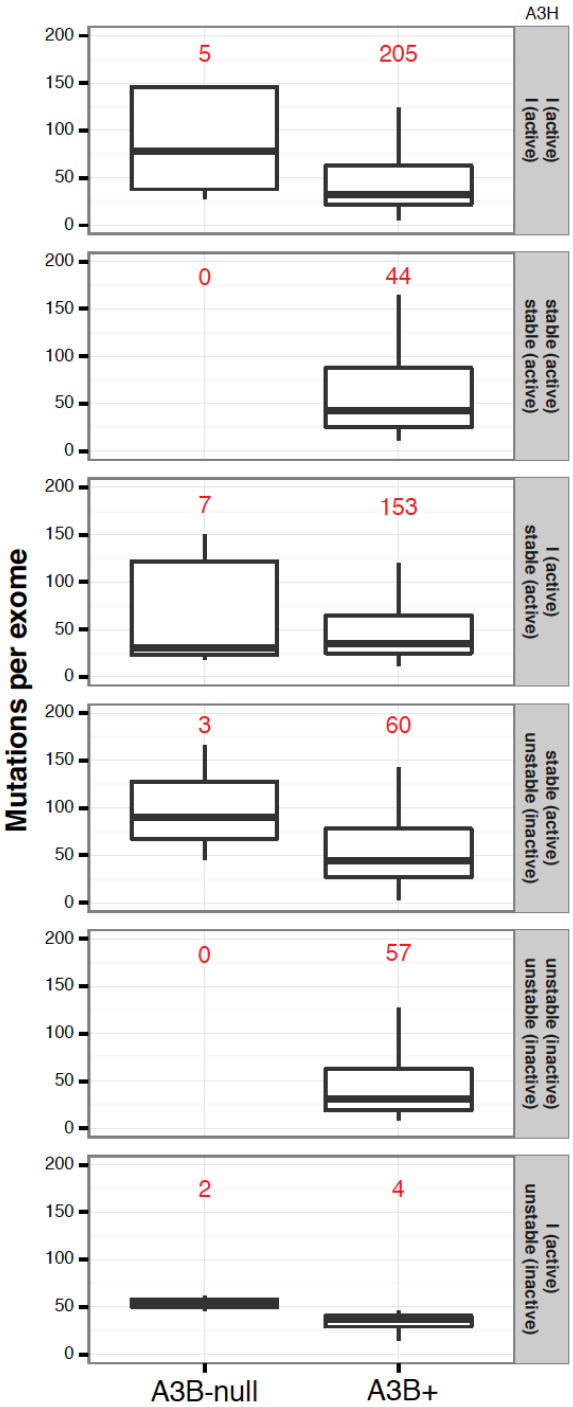
**a**, Expression of each *APOBEC3* family member relative to house-keeping gene, *TBP*, measured by RT-qPCR on a log10 axis for parent clone prior to engineering (*A3B* +/+) and 4 independent daughter clones after Cas9/CRISPR engineering (*A3B* +/- or -/-). Error bars show the standard error of 3 technical replicates.

**b**, Average fold change of *APOBEC* family member expression for daughter clones (right) targeted by Cas9/CRISPR relative to parent clone (left). Error bars show the standard error of the fold change in expression levels in the daughter clones (n=4; mean +/- SEM shown). The 1- to 2-fold differences seen here are not statistically significant and likely due to minor variations between clones.




**Supplementary Fig. S2.7. WES read coverage of *A3B* wild-type and *A3B*-null breast tumor specimens.**

Representative depth of coverage histograms from WES alignments used to positively identify a representative *A3B* null patient. The top histogram shows peaks mapping to the *A3B* gene that, in the bottom histogram, are not seen in truly *A3B* null individuals.




**Supplementary Fig. S2.8. Exome mutation loads as a function of *A3B* and *A3H* genotypes.**

Whisker plots quantifying the number of mutations per tumor grouped by *A3B* and *A3H* genotypes. The total number of tumors with each genotype is

represented by red numbers above each plots. The average is shown, the error box represents the first and third quartiles, and the whiskers extend to the highest value within 1.5x the interquartile range. Some genotypes are not represented in the entire TCGA breast cancer cohort.

**Supplementary Table S2.1. Fisher's exact test for cytosine mutation types occurring in *A3H-I* tumors versus *A3H-II/III* tumors.** This table is a statistical accompaniment to **Fig. 1d**. The proportion of each type of cytosine mutation in the A3H-II/III tumors was compared pairwise to the proportion of the same type of mutation in A3B-null breast tumors with the indicated A3H haplotypes.

| Tumor A3H haplotype | C-to-T transition mutation motif | | | | C-to-G transversion motif | | | |
|---|---|---|---|---|---|---|---|---|
| | **TCA** | **TCC** | **TCG** | **TCT** | **TCA** | **TCC** | **TCG** | **TCT** |
| **A3H-I/I** | 8.50 E-18 | 9.33 E-06 | 1.08 E-03 | 2.13E-07 | 1.14 E-17 | 5.47 E-05 | 3.18 E-01 | 4.42 E-17 |
| **A3H-II/III** | 1.80 E-22 | 2.50 E-05 | 3.39 E-04 | 6.03E-10 | 5.87 E-19 | 7.28 E-03 | 2.11 E-02 | 2.87 E-15 |
| **A3H-I/III** | 5.83 E-06 | 1.00 E+0 | 1.00 E+0 | 2.07E-01 | 3.68 E-05 | 9.45 E-02 | 1.00 E+0 | 7.10 E-03 |

**Supplementary Software S4.1. countSangerMuts.** Perl script to take fastq files produced from Sanger sequencing and a reference sequence and then output low frequency point mutations with context and whether or not the position or read is hypermutated. Available from: https://github.com/gjstarrett/countSangerMuts.

## CHAPTER 3:

# Functional upregulation of the DNA cytosine deaminase APOBEC3B by polyomaviruses

*This chapter was adapted with permission from: Verhalen and Starrett, Harris and Jiang. (2016) J Virol. 90(14):6379-6386*

**Authors:** Brandy Verhalen[1]*, Gabriel J. Starrett[2]*, Reuben S. Harris[2,3] and Mengxi Jiang[1]

*These authors contributed equally to this work

**Affiliations:** [1]Department of Microbiology, University of Alabama at Birmingham, Birmingham, Alabama, USA. [2]Biochemistry, Molecular Biology and Biophysics Department, Masonic Cancer Center, Institute for Molecular Virology, University of Minnesota, Minneapolis, Minnesota, USA. [3]Howard Hughes Medical Institute, University of Minnesota, Minneapolis, Minnesota, USA

**SUMMARY**

The APOBEC3 family of DNA cytosine deaminases has important roles in innate immunity and cancer. It is unclear how DNA tumor viruses regulate these enzymes and how these interactions, in turn, impact the integrity of both the viral and cellular genomes. Polyomavirus (PyVs) are small DNA pathogens that contain oncogenic potentials. In this study, we examined the effects of PyV infection on APOBEC3 expression and activity. We demonstrate that APOBEC3B is specifically upregulated by BK polyomavirus (BKPyV) infection in primary kidney cells and that the upregulated enzyme is active. We further show that the BKPyV large T antigen, as well as large T antigens from related polyomaviruses, is alone capable of upregulating APOBEC3B expression and activity. Furthermore, we assessed the impact of A3B on productive BKPyV infection and viral genome evolution. Although the specific knockdown of APOBEC3B has little short-term effect on productive BKPyV infection, our informatics analyses indicate that the preferred target sequences of APOBEC3B are depleted in BKPyV genomes and that this motif underrepresentation is enriched on the nontranscribed stand of the viral genome, which is also the lagging strand during viral DNA replication. Our results suggest that PyV infection upregulates APOBEC3B activity to influence virus sequence composition over longer evolutionary periods. These findings also imply that the increased activity of APOBEC3B may contribute to PyV-mediated tumorigenesis.

**IMPORTANCE**

Polyomaviruses (PyVs) are a group of emerging pathogens that can cause severe diseases, including cancers in immunosuppressed individuals. Here we describe the finding that PyV infection specifically induces the innate immune DNA cytosine deaminase APOBEC3B. The induced APOBEC3B enzyme is fully functional and therefore may exert mutational effects on both viral and host cell DNA. We provide bioinformatic evidence that, consistent with this idea, BK polyomavirus genomes are depleted of APOBEC3B-preferred target motifs and enriched for the corresponding predicted reaction products. These data imply that the interplay between PyV infection and APOBEC proteins may have significant impact on both viral evolution and virus-induced tumorigenesis.

## INTRODUCTION

Polyomaviruses (PyVs) are a family of small nonenveloped viruses containing an ~5-kb circular double-stranded DNA genome. Most human PyVs establish a subclinical persistent infection in healthy individuals (221). These viruses can reactivate under various immunosuppression conditions and cause a variety of severe diseases, including cancers (222). Among them, BK polyomavirus (BKPyV) reactivation is a major concern in kidney and bone marrow transplant patients due to the possibility of development of polyomavirus-associated nephropathy and hemorrhagic cystitis, respectively (223). Recently, there have also been increasing reports demonstrating an association between BKPyV infection and the occurrence of renourinary tumors (61). JC polyomavirus (JCPyV) reactivation can lead to progressive multifocal leukoencephalopathy

(PML), a serious demyelinating disease of the brain most prevalent in AIDS patients or associated with certain immunosuppressive or immunomodulatory treatments (224). Merkel cell polyomavirus (MCPyV) is so far the only human PyV directly linked to cancer, having been established to be the etiologic agent for Merkel cell carcinoma (MCC) (225). In most MCC cases, MCPyV is found integrated into the host DNA, leading to mutations that render the virus replication incompetent and simultaneously promoting tumorigenesis (226).

Even though PyVs have been studied since the 1950s, there are several knowledge gaps in PyV biology. First, innate immune responses to PyV infection are poorly understood. PyV large T antigens (TAgs) can induce interferon-stimulated genes (ISGs) in mouse embryonic fibroblasts (227). In contrast, studies with BKPyV infection of primary kidney cells have found little evidence of ISG activation (228). It is unclear how viral DNAs are recognized and reacted to by host immune DNA sensors. Second, there is limited knowledge with regard to how these DNA viruses evolve. For BKPyV, different subtypes of viral genomes have been demonstrated to evolve from distinct human populations, with the archetypal variants being the dominant circulating virus (229). The host cell factors that contribute to the molecular evolution of PyVs remain to be determined. Finally, although intensely studied, mechanistic details explaining how PyVs interact with cellular pathways to influence malignant transformation are still far from complete. In addition to the well-known functions of TAg to inactivate tumor suppressors such as retinoblastoma protein (Rb) and p53, novel

functions of PyV oncoproteins that impact cellular proliferation, transformation, and tumorigenesis continue to be discovered (230, 231).

One arm of the innate immune response in humans is comprised of the seven-membered APOBEC3 (A3) family of single-stranded DNA cytosine-to-uracil (C-to-U) deaminases: A3A, A3B, A3C, A3D, A3F, A3G, and A3H (183). Considerable evidence has shown that these enzymes combine to suppress the replication of many different DNA-based viruses (183, 232). Susceptible pathogens include retroviruses, parvoviruses, herpesviruses, papillomaviruses, and hepadnaviruses as well as endogenous retroelements. In contrast to these beneficial innate immune functions, recent studies have demonstrated that A3B and at least one other A3 enzyme deaminate genomic DNA and are responsible for cytosine-biased mutation patterns in many different human cancers (1, 2, 34, 81, 202, 233). Key evidence includes positive correlations between A3B expression levels and genomic mutation loads and the intrinsic biochemical preference of the enzyme closely matching the observed cytosine mutation biases (i.e., C-to-T and C-to-G mutations occurring within 5′-TCA, TCG, and TCT trinucleotide contexts). Moreover, these signature mutations are sometimes found in strand-coordinated clusters termed kataegis, which can be coincident with sites of DNA rearrangement (179). Finally, there appears to be a mechanistic relationship between the antiviral response and cancer mutagenesis, as human papillomavirus (HPV)-associated tumor types (cervical, head and neck, and bladder cancers) often manifest the highest levels of A3B expression and global DNA cytosine mutation biases, A3B expression is induced specifically

by HPV infection in relevant cell types (59), and the spectrum of PIK3CA activating mutations is biased toward an A3B deamination motif in HPV-positive head and neck tumors (60).

In this study, we set out to determine the relationship between PyV infection and the regulation of A3 family members. Using a relevant primary cell culture system, we demonstrated a unique and specific regulation of APOBEC3B by PyV and its encoded TAg. We also present evidence pointing to potential long-term effects of APOBEC-mediated deamination on PyV DNA genome composition and evolution.

## MATERIALS AND METHODS

*Cell culture, virus growth, and lentivirus production.*

Primary renal proximal tubule epithelial (RPTE) cells were expanded and maintained as described previously (234, 235). BKPyV Dunlop strain was grown in Vero cells, and titers were determined using an infectious-unit assay (236). RPTE cells were infected with BKPyV at a multiplicity of infection (MOI) of 0.5 infectious unit (IU)/cell as described previously (237). For TAg-expressing lentiviruses, JCPyV and MCPyV TAg cDNAs were subcloned from pBS-JT(Int-) (238) and pcDNA6.MCV.cLT206.V5 (Addgene), respectively, into lentivirus vector pLentiloxpuro (239). Lentiviruses expressing BKPyV TAg, JCPyV TAg, or MCPyV TAg were grown in 293T cells and transduced into RPTE cells as described previously (239).

Western blotting, RNA isolation, RT-qPCR, and single-stranded DNA (ssDNA) deaminase assays.

Protein lysates were harvested at one or 3 days postinfection (dpi) or posttransduction, quantified, and immunoblotted as described previously (236). The anti-A3B antibody 5210-87-13 is a rabbit monoclonal antibody used at a dilution of 1:50 (74). This antibody recognizes both A3B and A3G; however, the two proteins can be distinguished by differential gel migration ((74); this study). Mouse monoclonal antibody sc-136172 recognizing MCPyV-TAg was used at a dilution of 1:500 (Santa Cruz). JCPyV-TAg was recognized by pAb416 as described for BKPyV-TAg (236). Total RNA was harvested by removal of medium and resuspension in TRIzol (Thermo Fisher), and purification was done per the manufacturer's protocol. Reverse transcription-quantitative PCR (RT-qPCR) was used to quantify A3 transcripts as described previously (59). DNA deaminase activity assays were implemented as described previously (59).

*siRNA knockdown.*

Nontargeting control and A3B-targeting ON-TARGET plus small interfering RNAs (siRNAs) were purchased from Dharmacon (GE Healthcare). Each siRNA, at a concentration of 20 nM, was reverse transfected into RPTE cells as described previously (234). The cells were infected with BKPyV at 3 days posttransfection at an MOI of 0.5 IU/cell as described above. Total cell protein lysates were harvested, transferred, and probed for A3B, TAg, VP1 (236), and glyceraldehyde-3-phosphate dehydrogenase (GAPDH) as described above. For

viral DNA quantitation, low-molecular-weight DNA was purified and quantified using qPCR (29). Viral progeny were quantified using an infectious-unit assay (234).

*Bioinformatic analyses.*

Sequences from all available complete BK polyomavirus genomes (n = 302) were acquired from GenBank and aligned using Clustal Omega (240) to ensure that the homologous bases from each genome are in identical positions. Dinucleotide and trinucleotide motif enrichments for each aligned genome were calculated using Markov modeling as described by Ebrahimi and colleagues (119). Standard error was calculated and data were plotted using GraphPad Prism. Spearman's linear correlation coefficients of the enrichment between APOBEC target and product trinucleotides were calculated in the R statistical environment. Dinucleotide density was calculated across the BKPyV genome using 100-bp nonoverlapping windows. Smoothed fitted lines and 95% confidence intervals of these densities were calculated and plotted using the ggplot2 package in the R statistical environment.

## RESULTS

*BKPyV infection upregulates A3B specifically*.

We first asked whether polyomaviruses modulate cellular A3 levels and activity since these viruses have oncogenic potential (241). Using BKPyV and an established primary renal proximal tubule epithelial (RPTE) cell culture system

(234, 235), we determined whether virus infection alters A3 gene expression (Fig. 1). RPTE cells were infected with BKPyV at an MOI of 0.5 IU/cell, and total RNA was extracted 1 and 3 days postinfection (dpi) and subjected to RT-qPCR analyses (59). One day postinfection is an early time point when the input virus has just entered the nucleus (236), whereas 3 dpi represents a time when viral replication is robust (234). Among all A3 family members, only A3B expression changed significantly by BKPyV infection at 3 dpi (**Fig. 3.1**).

We then asked whether the observed increase in A3B mRNA also occurs at the protein level (**Fig. 3.2A**). Total cell lysates were immunoblotted for A3B, TAg, and GAPDH. Although the anti-A3B MAb recognized both A3B (lower band) and A3G (top band) (74), the immunoblots clearly show A3B protein upregulation by 3 dpi (**Fig. 3.2A**). It is notable that 3 dpi is around the time point that viral TAg becomes highly expressed (239). A3G protein levels are also upregulated at 3 dpi in both mock-infected and virus-infected cells, which could be due to a posttranscriptional regulatory process triggered by cellular growth.

To determine whether increased A3B protein levels also correlate with elevated activity, we performed a series of DNA deaminase activity assays using whole-cell extracts and an ssDNA oligonucleotide containing a single TCA target motif (59). All APOBEC family members, except A3G, prefer this trinucleotide as a deamination target. A representative gel and quantification of 3 independent experiments are shown in Fig. 2B and C. C-to-U deamination, uracil excision, and mild hydroxide treatment combine to break the ssDNA at the site of deamination and result in a smaller, faster-migrating fragment. The positive

control, purified A3A, converted nearly all of the substrate to product. In comparison, all mock-infected or 1-dpi RPTE cell extracts yielded only 5 to 10% product accumulation, whereas 3-dpi BKPyV-infected RPTE cell extracts showed ~50% product accumulation. Thus, the increase in A3B mRNA and protein levels correlated with increases in measurable enzymatic activity that were similar and of significant magnitude in RPTE cell extracts.

*Polyomavirus TAg is sufficient for A3B upregulation.*

For high-risk HPV, the viral E6 oncoprotein alone is sufficient to trigger A3B upregulation (59). Therefore, we hypothesized that polyomavirus TAg may trigger a similarly specific response, since TAg and E6 share some overlapping functions. To test this idea, we expressed BKPyV TAg in RPTE cells using a previously reported lentivirus delivery system (237) and measured the A3 mRNA levels as for Fig. 1. Consistent with the BKPyV infection data, the expression of TAg increased A3B transcript levels specifically (**Fig. 3.3A**). As with the full virus, A3B upregulation also manifested at the protein level (**Fig. 3.3B**). Interestingly, as previously shown for HIV-1 infection of T lymphocytes (25), the lentiviral vector itself caused A3G induction, which was evident at both the mRNA and protein levels (**Fig. 3.3**).

We then investigated whether A3B upregulation might be a general feature of the polyomavirus family. To address this point, we subcloned JCPyV and MCPyV TAg cDNA into the lentiviral expression vector pLentiloxpuro as was done for BKPyV TAg (239). Similar to the case with BKPyV TAg, we found by

both RT-qPCR (**Fig. 3.4A**) and immunoblotting (**Fig. 3.4B and C**) that both JCPyV and MCPyV TAg induced A3B levels. Interestingly, the magnitudes of A3B induction at the mRNA level were similar, whereas the magnitudes at the protein level were considerably more variable, with JCPyV causing the highest level of induction (**Fig. 3.2 to to 3.4**). The molecular explanation for this difference is not clear at this time, but the data suggest that the different PyV TAg proteins may impact A3B regulation differentially at multiple levels.

*A3B is not required for productive BKPyV infection.*

Next, we investigated whether upregulated A3B impacts BKPyV productive infection. To do this, we used siRNA to specifically knock down A3B in RPTE cells and assessed the effects of the knockdown on both cellular deaminase activity and viral life cycle. The knockdown of A3B was robust and specific, as evidenced by RT-qPCR and immunoblot data (**Fig. 3.5A and B**, respectively). A3B siRNA completely abolished the cellular deaminase activity, further confirming that the increased deaminase activity observed in BKPyV-infected cells originated solely from A3B upregulation (**Fig. 3.5C**). The specific depletion of A3B, however, did not affect viral TAg expression (**Fig. 3.5B**), viral DNA replication (**Fig. 3.5D**), or production of infectious viral progeny (**Fig. 3.5E**). Collectively, these data suggest that A3B is not necessary for short-term acute BKPyV productive infection.

*BKPyV genomes display A3B-mediated mutation signatures.*

To address whether A3B upregulation by BKPyV TAg has had an impact on BK polyomavirus genome composition over an evolutionary time frame, we acquired and analyzed all available BKPyV complete genomes from GenBank (n = 302). We then calculated the enrichment of all dinucleotide motifs using the ratio of observed dinucleotides versus expected determined by Markov modeling and then focused on dinucleotides containing 3′ cytosines on the positive or negative strand (NC or GN, respectively) that can be targeted by APOBEC family member deamination and the dinucleotide product that would be produced upon deamination and replication of the viral genome (NT or AN). This analysis revealed an underrepresentation of TC:GA dinucleotides and an overrepresentation of TT:AA dinucleotides, which reflects the preferred deamination target of most APOBEC3 family members, including A3B, and the products of deamination without repair, respectively (**Fig. 3.6A and B**). Targets and their respective products of deamination by other APOBEC family members, such as A3G (CC:GG) or AID (RC:GY), were not observed to be distinctly depleted or enriched.

We next looked at the linear correlation between the A3B strongly preferred TCA:TGA trinucleotide to see if the reduction of specific target motifs in each genome directly correlated with the enrichment of product trinucleotides independently of spontaneous deamination of methylcytosines that would reduce NCG motifs, which are known to be underrepresented in polyomavirus genomes (35). We observed strong negative linear correlations for both TCA-TTA and TGA-TAA trinucleotide pairs, indicating that as one motif is depleted the other

increases, strongly supporting a model in which deamination and mutation by A3B has shaped BKPyV evolution (**Fig. 3.6C and D**). Finally, we observed a clear inverse relationship between TC and GA dinucleotide density that corresponds to either the deamination of the nontranscribed strand of the viral genome during transcription or the lagging strand during replication in the BKPyV genome (**Fig. 3.6E**). The latter is supported by several recent studies investigating the nature of APOBEC-mediated mutations in various cancers (15, 44, 66, 67, 242). However, considering the gene organization of the BKPyV genome and the bidirectional nature of its genome replication, we cannot at this time definitively determine whether transcription or lagging-strand DNA replication predisposes to more viral genomic mutation.

## DISCUSSION

In summary, we report here that polyomavirus infection specifically upregulates A3B expression and activity and that the viral TAg is sufficient to mediate this response. To our knowledge, this is the first work to show a direct link between polyomavirus infection and A3B upregulation. Our results also reveal that TAgs from multiple human PyVs are capable of upregulating the expression and activity of A3B. Given the fact that other small DNA viruses, such as HPV, also specifically increase A3B (59), these findings suggest that A3B upregulation may be a conserved response to small DNA tumor virus infections.

Using siRNA knockdown, we showed that A3B is indeed responsible for the increased cellular deaminase activity in BKPyV-infected cells (**Fig. 3.5C**).

However, A3B expression does not facilitate or impair productive BKPyV infection, as neither viral DNA replication nor viral progeny production was affected by A3B depletion (**Fig. 3.5D and E**). It is possible that since our assays only measured initial rounds of viral infection, any subtle restrictive effects of A3B may have been missed. Additional investigation of long-term A3B depletion will be necessary to assess the functional significance of A3B upregulation over an evolutionary time span.

Although we did not detect immediate viral infection or replication defects associated with A3B knockdown, our computational analysis of all publicly available BKPyV genomes revealed that BKPyV genomes are depleted of the A3B preferred target dinucleotides (TCA:TGA), with a concomitant overrepresentation of the corresponding deamination products (TTA:TAA) (**Fig. 3.6A to D**). These results suggest that the functional increase in A3B activity has exerted long-term evolutionary pressure on the nucleotide composition of the viral genomic DNA. It is unclear whether the virus directly induces the A3B response to such a level as to avoid restriction while generating additional genetic diversity for better fitness as has been observed in HIV-1 (243, 244). Our data are also consistent with recent reports that point to possible roles of A3 enzymes during HPV evolution (143). Interestingly, we also identified a strong asymmetry in the abundance TC or its reverse complement GA that corresponds to the direction of both transcription and viral genomic DNA lagging-strand replication. In support of this potential mechanistic relationship, several groups recently reported that APOBEC-mediated mutation of cancer (and model)

genomes is preferentially linked to the lagging strand of DNA replication, where more ssDNA intermediates would be expected to occur (15, 44, 66, 67).

A range of A3B transcriptional regulation mechanisms in both viral and nonviral cancers have been reported (59, 74). It has been postulated that E6 protein from high-risk HPV derepresses A3B gene transcription via p53 inactivation (59). From our results, this is an unlikely mechanism for polyomavirus-mediated upregulation of A3B, since MCPyV TAg does not bind p53 (226) yet it clearly triggers an increase in A3B level (**Fig. 3.4C**). Another study reported that HPV E7 protein, which is known to inhibit Rb function, is required for upregulation of both A3A and A3B in HPV-positive normal human immortalized keratinocyte cells (140). In multiple cancer cell lines not associated with virus infection and for which A3B overexpression is observed, the protein kinase C (PKC)/noncanonical NF-κB signaling pathway has been shown to be necessary for A3B induction (74). It is possible that each of these studies reveals separate pathways regulating the expression and activity of A3B or that may be connected through yet-to-be-determined mechanisms. Future investigation on polyomavirus TAg will be valuable to determine if a direct or indirect function of the TAg is responsible for A3B upregulation.

Finally, it is possible that TAg-mediated upregulation of A3B may also lead to host genomic DNA mutations that contribute to carcinogenesis. Several recent studies have implicated BKPyV seropositivity with the development of urothelial cancers in immunocompromised patients (61). Polyomavirus-mediated

upregulation of A3B may contribute to carcinogenesis in these patients much as HPV-mediated upregulation of A3B likely contributes to the tumor evolution and the dominant APOBEC signature mutation spectra observed in cervical and HPV-positive head and neck cancers (59, 60). Future studies on the host genomic DNA mutations occurring in cells infected with BKPyV and tumors expressing TAg will address this issue. If this contribution is confirmed, it will add to the already diverse functions of TAg for its oncogenic potential.

**ACKNOWLEDGMENTS**

We thank Richard Frisque for the JCPyV pBS-JCT(Int-) plasmid, Ellen Cahir-McFarland for thoughtful comments, and James DeCaprio for MCPyV reagents.

Salary support for G.J.S. was provided by a National Science Foundation Graduate Research Fellowship (grant 00039202). R.S.H. is an Investigator of Howard Hughes Medical Institute. Jiang lab contributions were supported in part by the UAB Department of Microbiology start-up fund, AHA Scientist Development grant 15SDG25680061, and NIH grant R01AI123162.

R.S.H. is a cofounder, shareholder, and consultant of ApoGen Biotechnologies Inc. A research project in the Harris laboratory is supported by Biogen.

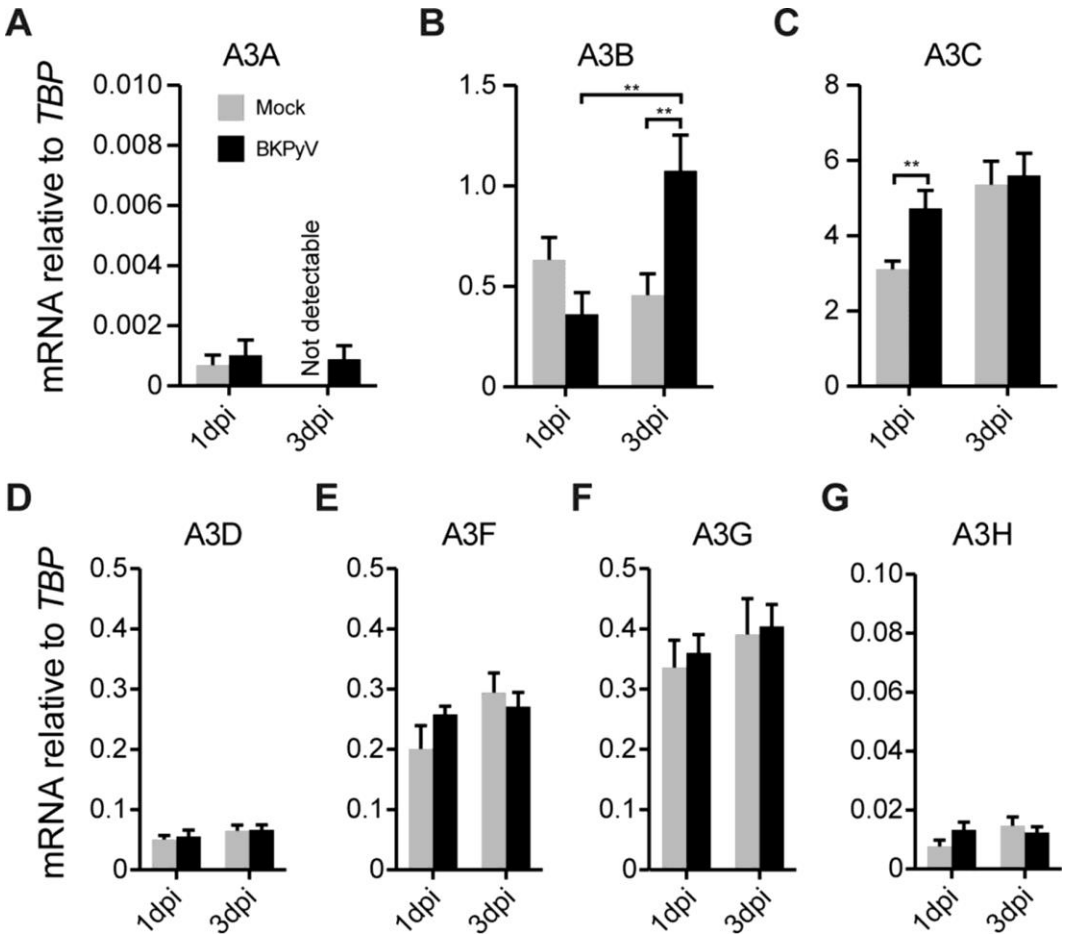



**Figure 3.1. BKPyV infection upregulates *A3B* transcripts.**

RNA was harvested and purified from mock-infected or BKPyV-infected cells, and *A3* mRNA levels were quantified by RT-qPCR. Histograms report mean *A3* mRNA expression levels in mock- or BKPyV-infected RPTE cells in three independent experiments normalized to those of the housekeeping gene *TBP*. Error bars show standard deviations, and Student's t test was used to assess significance (**, $P < 0.01$). *APOBEC1*, *APOBEC2*, *APOBEC4*, and *AID* mRNAs were not detectable.

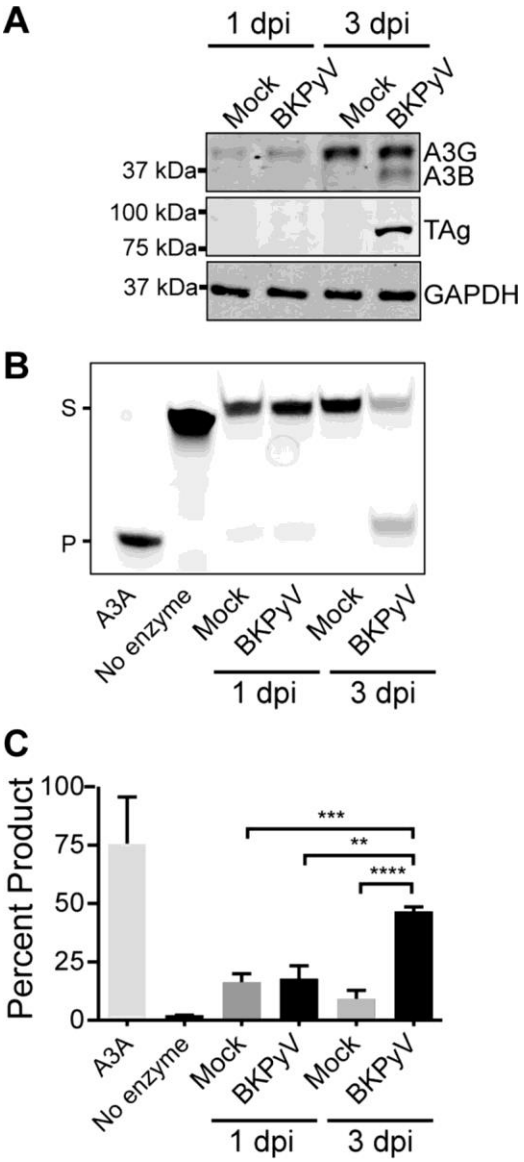



**Figure 3.2. BKPyV infection increases A3B protein and cellular deaminase activity.**

(A) Total protein lysates from representative cells in Fig. 1 were immunoblotted for A3B (the antibody also recognizes A3G), TAg, and GAPDH. Shown are representative blots from three independent repeats. (B) Representative gel image of a DNA cytosine deaminase assay performed with cell extracts from mock- or BKPyV-infected cells as in panel A. Recombinant purified APOBEC3A

was used as a positive control, and reaction buffer alone was used as a negative control. S, substrate; P, product. (C) The deaminated products were quantified by the Fiji gel analysis tool (http://fiji.sc/). Shown are combined results from three independent repeats. **, $P < 0.01$; ***, $P < 0.001$; ****, $P < 0.0001$.

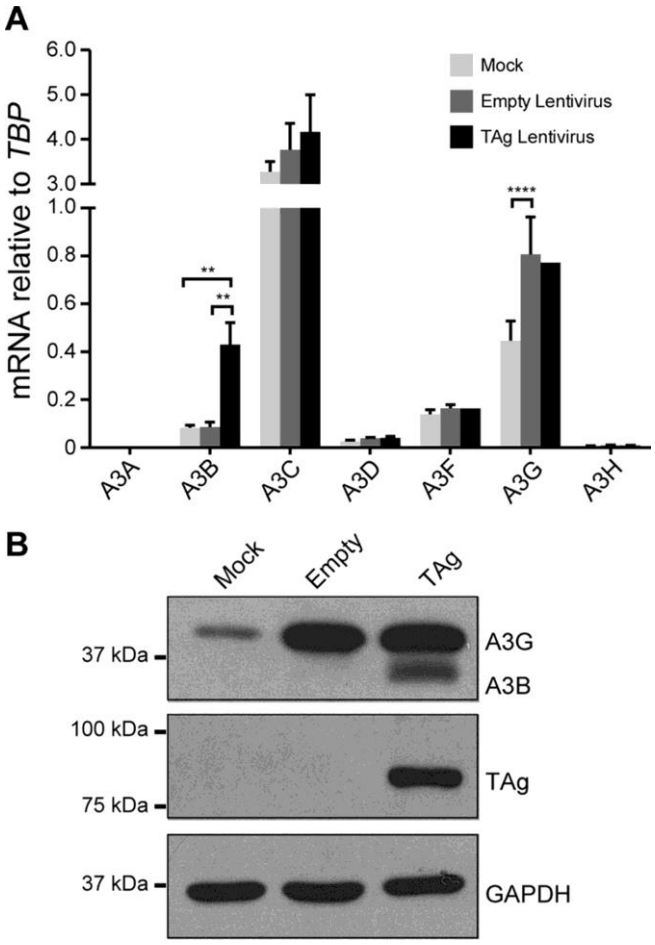



**Figure 3.3. PyV large T antigen is sufficient for A3B upregulation.**

RPTE cells were mock transduced, transduced with an empty lentivirus control, or transduced with a lentivirus expressing BKPyV TAg. A3 family member mRNA levels (A) and A3B protein levels (B) were examined as for Fig. 3.1 and 3.2, respectively. Note that A3A mRNA was detected but at too low a level to be graphed in panel A. **, $P < 0.01$; ****, $P < 0.0001$.

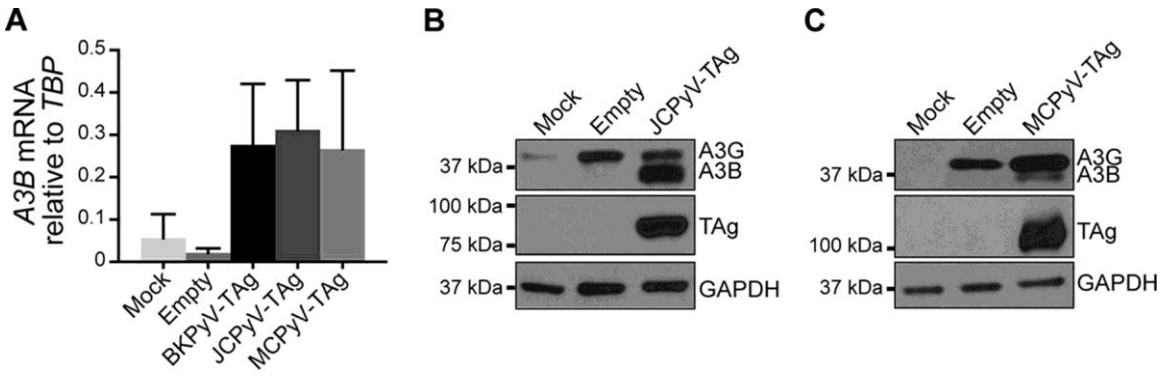



**Figure 3.4. Other human PyV large T antigens also upregulate A3B.**

RPTE cells were transduced with lentivirus expressing JVPyV or MCPyV Tag, and A3B mRNA and protein levels were determined by RT-qPCR (A) and immunoblotting (B and C), respectively, as in Fig. 3.3.

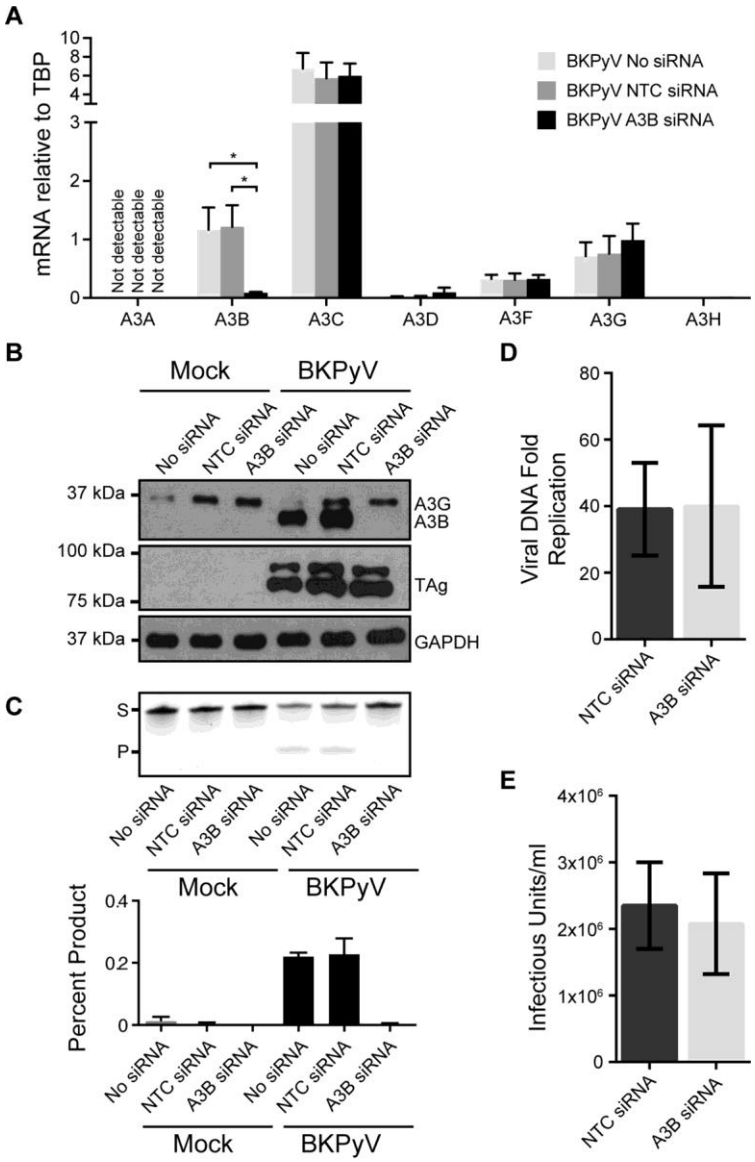



**Figure 3.5. Knockdown of A3B eliminates deaminase activity during BKPyV infection but does not affect BKPyV lytic infection**.

Cells were left untreated or treated with nontargeting control (NTC) siRNA or A3B-targeting siRNA for 3 days, followed by mock infection or infection with BKPyV for 3 days. (A) A3 family member mRNA levels were determined as for Fig. 1. Shown are combined results from three independent experiments with infected cells. (B) Total protein lysates were immunoblotted for A3B, TAg, and

GAPDH. Shown is a representative result from three independent repeats. (C) Representative gel image and quantitation of a DNA cytosine deaminase assay (as for Fig. 2) performed with cell extracts as for panel B. (D and E) Bar graphs of BKPyV viral DNA fold replication (D) and viral infectious units (E) produced from infected cells treated with NTC or A3B-targeting siRNA. Shown are combined results from three independent repeats.

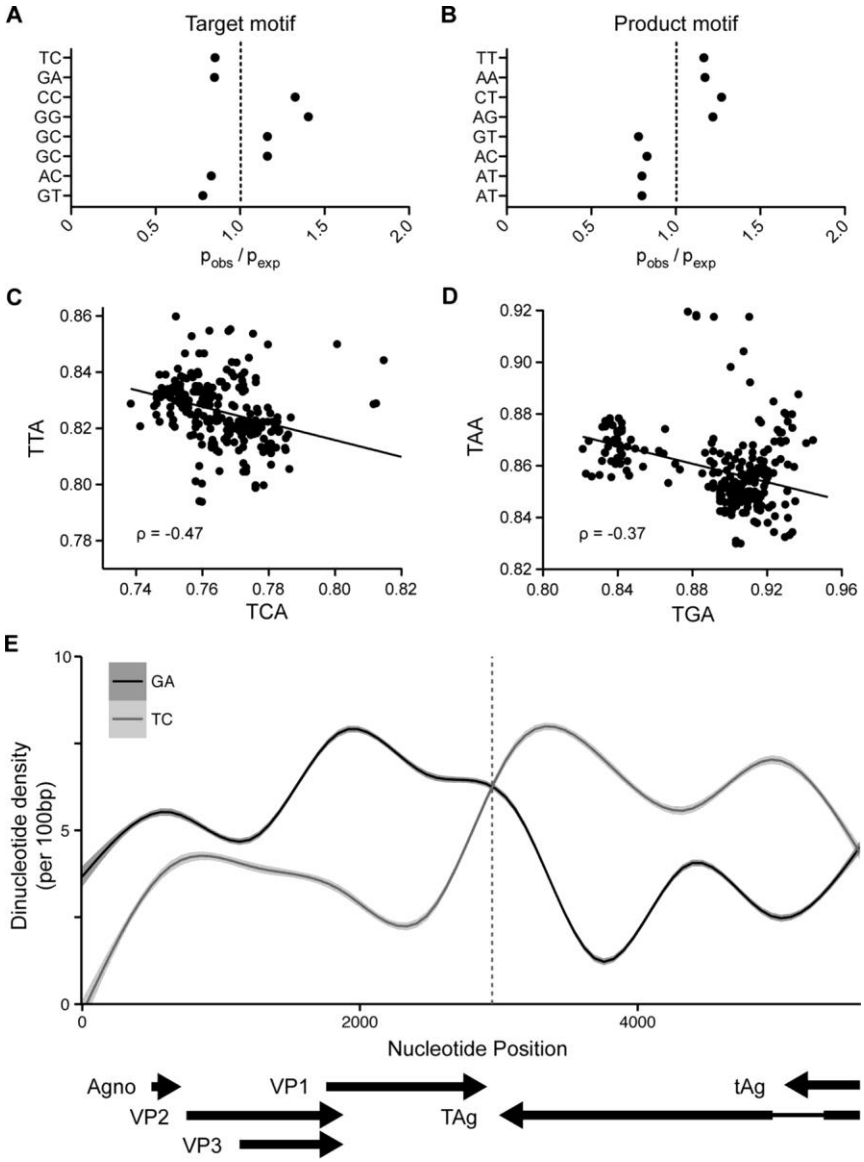



**Figure 3.6. Evidence for APOBEC mutagenesis on the BKPyV polyomavirus genome.**

(A) Over- and underrepresentation with standard errors of potential APOBEC-targeted cytosine dinucleotides in the BKPyV genome. (B) Over- and underrepresentation of dinucleotides produced by the deamination of the target dinucleotides in panel A. (C and D) Scatter plots and best-fit lines representing the linear correlations between A3B preferred target and the respective product

trinucleotides TCA/TTA (C) and TGA/TAA (D). (E) Line graph of TC/GA dinucleotide density across the BKPyV genome with viral genes and transcription direction annotated at the bottom.

## CHAPTER 4:

# Merkel cell polyomavirus exhibits dominant control of the tumor genome and transcriptome of Merkel cell carcinoma

*This chapter was adapted with permission from: Starrett et al. (2017) mBio. 1(8): e02079-16.*

**Authors:** Gabriel J. Starrett[1], Christina Marcelus[2], Paul G. Cantalupo[3], Joshua P. Katz[3], Jingwei Cheng[2,4], Keiko Akagi[6,7], Manisha Thakuria[5], Guilherme Rabinowits[2,4], Linda C. Wang[5], David E. Symer[6,7,8], James M. Pipas[3], Reuben S. Harris[1,9], James A. DeCaprio[2,4]

**Affiliations:** [1]Department of Biochemistry, Molecular Biology and Biophysics, Masonic Cancer Center, Institute for Molecular Virology, University of Minnesota, Minneapolis, Minnesota, USA. [2]Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, Massachusetts, USA. [3]Department of Biological Sciences, University of Pittsburgh, Pittsburgh, Pennsylvania, USA. [4]Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts, USA. [5]Department of Dermatology, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts, USA. [6]Human Cancer Genetics Program, The Ohio State University Comprehensive Cancer Center, Columbus, Ohio, USA. [7]Department of Cancer Biology and Genetics, The Ohio State University, Columbus, Ohio, USA. [8]Department of Biomedical Informatics,

The Ohio State University, Columbus, Ohio, USA. [9]Howard Hughes Medical Institute, University of Minnesota, Minneapolis, Minnesota, USA

## SUMMARY

Merkel cell polyomavirus is the primary etiological agent of the aggressive skin cancer, Merkel cell carcinoma (MCC). Recent studies have revealed that UV radiation is the primary mechanism for somatic mutagenesis in non-viral forms of MCC. Here we analyze the whole transcriptomes and genomes of primary MCC tumors. Our reveal that virus-associated tumors have minimally altered genomes compared to non-virus-associated tumors, which are dominated by UV-mediated mutations. Although virus-associated tumors contain relatively small mutation burdens, they exhibit a distinct mutation signature with observable transcriptionally biased kataegic events. In addition, viral integration sites overlap focal genome amplifications in virus-associated tumors suggesting a potential mechanism for these events. Collectively, our studies indicate that Merkel cell polyomavirus is capable of hijacking cellular processes and driving tumorigenesis to the same severity as tens of thousands of somatic genome alterations.

## IMPORTANCE

A variety of mutagenic processes that shape the evolution of tumors are critical determinants of disease outcome. Here we sequenced the entire genome of virus-positive and virus-negative primary Merkel cell carcinomas (MCC) revealing distinct mutation spectra and corresponding expression profiles. Our studies highlight the strong effect that Merkel cell polyomavirus has on the divergent development of viral MCC compared to the somatic alterations that

typically drive non-viral tumorigenesis. A more comprehensive understanding of the distinct mutagenic processes operative in viral and non-viral MCC has implications for the effective treatment of these tumors.

## INTRODUCTION

Merkel cell carcinoma (MCC) is an aggressive skin cancer, associated with advanced age, excessive UV exposure, immune deficiencies and presence of the human virus, Merkel cell polyomavirus (MCPyV) (155, 245). MCPyV DNA is clonally integrated in approximately 80% of MCC and the expression of viral T antigens is required for driving tumor cell proliferation (155, 246, 247). Deletion mutations of the C-terminus of the viral large T antigen are common in MCC tumors, rendering the viral genome replication-deficient (248). The effect of this integration event and the constitutive expression of viral proteins on the host genome structure and somatic mutation profile of the tumor genome have not been studied in depth. Using *in vitro* models, it has been suggested that expression of full length MCPyV large T antigen is able to disrupt the stability of the host genome and upregulate the mutagenic enzyme APOBEC3B (192, 249). Another small DNA tumor virus, human papillomavirus (HPV), also triggers the upregulation of the DNA cytosine deaminase, APOBEC3B, and is likely responsible for the majority of mutations observed in HPV-positive cervical, head and neck squamous cell, and bladder carcinomas (2, 34, 59, 138, 250). To date, there has been no high-coverage whole genome sequencing performed in MCC.

High-throughput sequencing has been highly beneficial to many fields including cancer biology and virology. Sequencing shows individual mutations as well as mutation patterns or signatures, which implicate distinct mutational processes acting within tumors over time. These processes are responsible for intratumoral genetic heterogeneity and provide the necessary substrate for evolution, survival, and metastasis (4, 16, 179, 180). Additionally, these studies have been critical for our understanding of cancer subtypes and how to better improve targeted therapies. However, to date, deep sequencing projects have been restricted to more common cancers studied such as breast cancer by large consortia. Now due to decreased sequencing costs, rare cancers can be sequenced to expand our knowledge on how these tumors arise. In fact, early next-generation sequencing was used to discover MCPyV from primary human tumors in 2008 (157). Here we leverage modern sequencing platforms to sequence the RNA and DNA from six primary MCC tumors and analyzed both the mutation spectra and corresponding transcriptome characteristics based on detectable Merkel cell polyomavirus transcripts.

## RESULTS

*Virus-negative MCC tumor genomic DNA is heavily mutagenized.*

To determine if a tumor expressed viral genes, RNAseq reads obtained from 6 MCC tumor specimens were aligned to a reference containing both the human (hg19) and MCPyV (NCBI) genomes (**Table 4.1**). Merkel cell polyomavirus T antigen transcripts were readily detected in 4 out of 6 tumors. Tumors with viral

transcripts were defined as virus-positive, whereas those without were defined as virus-negative.

We performed high-coverage (~100X) whole genome sequencing of two virus-positive MCC tumors and one virus-negative MCC tumor and analyzed somatic mutations, copy number variants and structural rearrangements compared to the normal somatic DNA isolated from PBMC isolated from the corresponding patients. What was exceptionally striking was that the virus-negative MCC tumor had over 30 fold more somatic mutations with a total of 127,236 mutations in addition to many copy number alterations and inter-chromosomal translocations compared to the two virus-positive tumors (**Fig. 4.1A-B, Table 4.1**). This mutation load is consistent with recent reports from targeted sequencing (159, 251, 252). Within all tumors, the majority of mutations fell into intergenic regions, but a large fraction of these mutations (37.7%) occurred near or within genes that did not significantly differ based on tumor type (**Table 4.2**). Within the intergenic regions, virus-positive tumors did show greater than two-fold enrichment of mutations in both HERVK and simple repeat regions of the genome, but not in any other type of mobile element compared to the virus-negative tumor.

By functionally annotating the mutations overlapping genes, the virus-positive tumors had a cumulative total of 12 missense and nonsense mutations targeting genes implicated in cancer as annotated by COSMIC; whereas, the virus-negative tumor harbored 51 missense and nonsense mutations targeting COSMIC annotated genes. Of these 51 mutations, 34 were predicted by either

SIFT or PROVEAN to be deleterious to the function of the primary protein product. The effects of these mutations on all potential protein products are detailed in **Table 4.S1**. Of note, there are damaging mutations predicted to occur in *CBFA2T3*, *CHEK2, FANCC*, *FLI1*, *ITPR1*, *MUC16*, *NF1, NUTM*, *PTPRB*, *PTPRR*, *SETX*, and *STK11IP* in the virus-negative tumor, which may further promote tumor survival and evolution.

The relative abundance of structural variants in each tumor genome mimicked the profiles of the aforementioned somatic mutations. Tumor-088 had no amplifications or deletions corresponding to known copy number variations in cancer. The other virus-positive tumor-076 had a single copy amplification of *MDM4* and single copy deletions of *PTEN* and *SUFU.* In stark contrast to the virus-positive tumors, the virus-negative tumor-050 had single copy amplifications of *EGFR* and *JUN* and single copy deletions of *APC, ATM, BIRC*, *BRCA1, BRCA2, FANCA, FANCD2*, *CDKN2A, MLH1, PAX5, PBRM1, RB1* and *VHL*. *RB1* function may be absent in sample-050 as there was a somatic G-to-A transition mutation in the remaining allele at position chr13:49047495. This base substitution is predicted to interfere with the splice acceptor for the adjacent exon 20 with the potential to produce a non-functional protein. Of the detected inter-chromosomal translocations in these tumors, none of them reflected known annotated translocations in cancer. To further define and consolidate the impact of the sheer number of somatic alterations in the non-viral tumor we performed pathway analysis on the aforementioned variants. This analysis predicted significant inhibition of p53, ATM, and BRCA1 signaling and inhibition of DNA

damage checkpoint regulation all of which would contribute to the observed severe genome instability and the ability of the tumor to survive the corresponding stress (**Fig. 4.1C**). Activation of pathways observed in other cancers such as glioma, glioblastoma, and metastasis in colorectal cancer were also predicted and were commonly linked by the inactivation of *ATM* and *CDKN2A* and amplification of *EGFR*. Although EGF signaling was also predicted to be significantly impacted, it was neither activated nor inhibited due to inactivation of *ATM* and *ITPR1* and amplification of *EGFR* and *JUN*.

*Different mutation signatures occur in virus-positive and virus-negative tumors.*

Recent studies have highlighted and classified the multitude of mutation processes critical for shaping tumor development and evolution across cancers (176, 215, 253). Upon subdividing the detected somatic mutations in these MCC tumors by base change and trinucleotide context to visualize the overall mutation landscapes of these tumors, even more differences were revealed between virus-positive and virus-negative tumors. The MCPyV-positive tumors were highly similar to each other and showed mutation profiles that were modestly enriched for both C-to-T and T-to-C mutations (**Fig. 4.2A**). In contrast, the MCPyV-negative tumor showed a dominant proportion of C-to-T mutations in both TCN and CCN trinucleotides, corresponding to cross-linked pyrimidine dimers induced by UV-radiation and subsequent error-prone repair (8, 19). Using somatic signature prediction software, we modeled three signatures from these samples and determined their relative contribution to each tumor indicating that the virus-

associated tumors may represent a mixture of mutational processes including a small proportion of UV-mediated mutations evident through a slight enrichment for C-to-T mutations in dipyrimidine contexts (**Fig. 4.2A** and data not shown). Hierarchical clustering revealed that these mutation profiles are most similar to signatures 5 and 16 for the MCPyV-positive tumors and signature 7 for the MCPyV-negative tumor as defined by Alexandrov and colleagues (4) (**Fig. 4.2B**).

At this time, signatures 5 and 16 currently have no known etiologies. However, it was recently reported that the mutations commonly observed in liver cancer, corresponding to signature 5, exhibit a bias for mutations on the non-transcribed strand of genic DNA, termed transcription-coupled damage (TCD) (15). To address whether the observed signatures from MCC exhibit similar mutation asymmetries, we analyzed the replication and transcription strand biases of the somatic mutations for the MCPyV-negative and MCPyV-positive tumors using methods published by Haradhvala and colleagues (15). Consistent with observations in liver cancer, the T-to-C base substitutions in the virus-positive MCC tumors had a clear preference to accumulate on the non-transcribed strand of genes. The overall mutation density remained constant or increased and the bias became more pronounced as the expression of the gene increased strongly indicating that these were mediated by TCD (**Fig. 4.2C**). Additionally, the C-to-T mutations in the virus-negative tumor, corresponding to signature 7 and attributable to UV-mediated DNA damage, exhibited a strong bias for the non-transcribed strand, with mutation density decreasing as expression of the gene increased. Signature 7 mutations are also dominant in

other forms of skin cancer such as basal cell carcinoma, squamous cell carcinoma, and melanoma (254–256). The transcription-biased mutation asymmetry we observed is also consistent to that observed in melanoma and transcription-coupled repair of UV-mediated damage. No significant replication-biased mutation asymmetries were observed in MCPyV-positive or negative tumors (**Fig. 4.2C**). Interestingly there was also no strong evidence for signature 1 mutations in either tumor type, which are the most common mutations detected in cancer and are associated with aging and the spontaneous deamination of 5-methycytosines in CpG motifs. Furthermore, there was no similarity to signatures 2 or 13 attributed to APOBEC-mediated mutation and have been observed in many cancers and are especially prominent in HPV-associated cancers (2, 34).

To further characterize the mutational processes in MCC, we evaluated the density and distribution of mutations across the entire genome by calculating intermutational distance for each somatic base substitution and plotting the values by position (**Fig. 4.3A**). As anticipated from other UV-mutated tumors and the high mutation burden the virus-negative tumor had a generally dense distribution of mutations across the genome with any clusters or other patterns occluded by the UV-attributable C-to-T transitions. The virus-positive tumors had a sparser distribution of mutations across the genome, but this highlighted several unique mutation clusters or kataegis events in tumor-076, but none in tumor-088 (**Fig. 4.3A**). Several non-specific clusters were observed in more than one sample and are likely due to errors from the sequencing platform. The clusters observed on chromosome 10 for tumor-076 appear to correspond to

several copy number alteration events that were observed and this is consistent with kataegis events typically being associated with DNA double stranded breaks (**Fig. 4.3B**). The minimal amount of copy number alterations in tumor-088 further supports that kataegis events in viral MCC correspond to DNA breaks. By plotting the abundance and context of each base substitution located at these kataegis events reveals a mutation profile similar to both APOBEC and the recently identified non-APOBEC mediated kataegis events as observed in breast cancer whole genome sequencing implicating multiple sources of DNA damage (**Fig. 4.3C**) (181). Evaluation of more viral MCC tumors at the whole genome level will reveal whether kataegis is a common characteristic of the mutation profile and the mechanism by which they occur.

*Whole transcriptome analysis reflects the state of the genome*.

To further delineate the differences between virus-positive and virus-negative tumors and establish potential mechanistic effects of the previously reported somatic genome alterations, we used RNAseq to analyze the full transcriptomes of the 4 virus-positive and 2 virus-negative MCC tumors. At the transcriptome level, all MCPyV-positive tumors formed a discrete cluster when analyzed by principal component analysis indicating a high level of homogeneity while the MCPyV-negative tumors were highly divergent and did not form a cluster (**Fig. 4.4A**). There were over 1100 significantly differentially expressed genes between MCPyV-positive and -negative tumors (**Table 4.S2**). Notably, the MCPyV-negative tumors expressed significantly reduced levels of DNA damage response

genes, such as *MSH2*, *MLH1* and Fanconi Anemia family genes, *FANCA* and *FANCC*, suggestive for a potential mechanism for the accumulation of the large amount of somatic mutations identified in the MCPyV-negative genome and the low number of somatic mutations in the MCPyV-positive tumors (**Table 4.S2**). Many of these relative decreases in gene expression levels, such as *MLH1*, correspond to our previously described alterations to genomic DNA indicating functional implications of these variants. Of particular interest, the *P16INK4A* isoform of the tumor suppressor *CDKN2A* shows a significant decrease in the virus-negative tumors compared to the virus-positive tumors. This alteration suggests that a common mechanism may promote tumor development, potentially mediated by the aforementioned single copy deletion of the *CDKN2A* locus observed in our virus-negative whole genome sequencing data (**Fig. 4.4B**).

We used Ingenuity Pathway Analysis software (Qiagen) to study differentially expressed genes. These results indicated that virus-negative MCC tumors were significantly enriched for genes associated with basal cell carcinoma signaling pathways and many of these genes are associated with the WNT signaling pathway, which is consistent with results inferred previously for MCC from microarray data (**Fig. 4.4C**) (257). Many of these pathways were also predicted by the pathway analysis of somatic variants shown in **Fig. 4.1C** and highlighted in bold. In contrast, virus-positive tumors show significant upregulation of the GABA receptor signaling pathway, commonly associated with neuronal development, estrogen-mediated S-phase entry, and a mild, positive enrichment for WNT signaling. GABA receptor signaling pathway enrichment was defined by

elevated expression of *GABRB3* and potassium channel genes, *KCNN1*, *KCNN2*, and *KCNQ3*, and decreased expression of *ADCY2*, which have all been indicated as important in tumor growth (258–260). Interestingly the pathways enriched in our MCPyV-positive tumors also included cell cycle genes, cyclin A1 and cyclin D1, which are detailed in **Fig. 4.4C**. However, in contrast with a previous publication, we did not observe a difference between tumor types in regards to the expression of genes associated with tumor-infiltrating lymphocytes, such as *CD3D* (257).

*Virus integration sites result in focal host genome amplifications and fusion transcripts.*

Virus integration has the potential to disrupt or alter the function of genes as well as produce novel fusion transcripts. To identify the integration sites in our virus positive whole genome alignments, we used a custom pipeline to discover reads that map to both the host and viral genomes. Tumor-076 revealed two integration sites on chromosome 1 that are approximately 40kb apart. Discordant read pairs show that these insertional breakpoints are linked to the C-terminus end of Large T antigen (**Fig. 4.5A**). Tumor-088 had one integration site detected on chromosome 6, which mapped to the N and C termini of Large T antigen with a proportion of reads supporting the deletion of the DNA binding domain (**Fig. 4.5B**).

A previous publication reported that HPV integrants frequently coincide with focal copy number alterations in cancer cell lines and HNSCC primary

tumors (261). To determine if this was also a characteristic of MCPyV integration events, we examined relative copy number of the host genomes across these regions. The location of all integration events in each patient overlapped with single copy amplifications of the host genome (**Fig. 4.5C, 4.5D**). The integrants in tumor-076 flank a tandem duplication indicating that this amplification and these copies of the viral genome were mediated by the same viral integration event (**Fig. 4.5C**). The insertion site in tumor-088 is located near the 3' edge of and within a tandem duplication event amplifying chr6:20646000-20768000 (**Fig 4.5D**).

To resolve the insertion sites between viral and host DNA, we de novo assembled all of the read pairs that mapped to the host and viral genomes into fusion contigs. The reads were remapped to these contigs to identify their original positions from the viral and host genomes (**Fig. 4.5G, 4.5H**). We do not detect host-virus fusion contigs that fully explain the integration events in tumor-076 and instead have numerous contigs comprised of only viral sequences (**Fig. 4.5E, 4.5G & Fig. 4.S1**). This analysis indicates that the integration event likely has complex rearrangements and potential amplifications of the MCPyV genome at the 5' end of the amplification. Generally, these data do support a common breakpoint in the C-terminus of Large T antigen and a DNA level deletion of the DNA binding domain of Large T antigen that was observed in the RNAseq data (**Fig 4.5A, 4.6A**). For tumor-088, we assembled two contigs containing host and viral sequences that support the junctions of a single identical integrant and the deletion of the DNA binding domain of Large T antigen (**Fig. 4.5F, 4.5H**).

As was proposed for HPV integration, our data support a similar looping model for the focal amplifications observed near the MCPyV integration sites in MCC (**Fig. 4.5I, 4.5J**) (261). This model proposes that after MCPyV integration, transiently circular DNA is formed and activation of the viral origin of replication amplifies neighboring regions of the host genome. Dissociation of this transiently circular DNA then is followed by recombination of the newly amplified regions and subsequent repair. Depending on the location of recombination and repair, the amplified regions can result in multiple virus-host concatemers as observed in tumor-076 or can appear as a single virus integration event within a tandem duplication as observed in tumor-088 (**Fig. 4.5I, 4.5J**).

To further characterize the integration sites of MCPyV and address whether these are affecting host genes, we aligned RNAseq reads from the virus positive tumors to the viral genome and assembled viral contigs from these reads using a custom analysis pipeline (262). Each of these tumors expressed at least part of the viral early region and in each of these cases the Large T antigen was truncated and nearby host gene expression was unaffected. Of the two tumors that we also had WGS data for, sample-088 (**Fig. 4.6B**) contained a single chimeric junction within two overlapping genes *RP3-348I23.2* and *CDKAL1* (at chr6: 20,757,000). The observed contig indicates a deletion between coordinates 1,560-2,754 of the viral genome, causing a frameshift after V311 that results in a 321aa amino-terminal truncation of Large T antigen, which are also supported by the integration analysis from the WGS reads. Analysis of tumor-076 (**Fig. 4.6A**) resulted in one MCPyV contig, which aligns to positions 146-429, 861-1,580 and

2,254-3,096. The deletion causes codon D318 (positions 1,578-1,580) to be placed immediately in frame with a stop codon (positions 2,254-2,256) encoding a 318aa amino-terminal truncation of Large T antigen. However, no chimeric junctions were detected in this sample.

For tumor-090, the read depth graph shows expression of the full-length Large T antigen transcript and truncated variants (**Fig. 4.6C**). One chimeric junction was mapped at chr20:32,132,694 within *CBFA2T2*. Another chimeric junction was mapped between exon 1 of *CBFA2T2* and the large T antigen splice acceptor at position 861 in MCPyV. This analysis also suggests additional copies of the viral genome, either integrated or episomal, are present in this sample. There are two C-to-T nonsense mutations that change Q432 and Q504 to stop codons that we predict to encode for a 432aa C-terminal truncated large T antigen. We mapped four chimeric junctions in tumor-146 (**Fig. 4.6D**). One chimeric junction was mapped within an intron of *FLJ46066* (approximately at chr3: 182,180,601) indicating a chimeric transcript anti-sense to the FLJ46066 gene. The other three chimeric junctions mapped within the *ECH1* (at host positions chr19: 39,307,703 and chr19: 39,307,378). The viral detection pipeline resulted in two MCPyV contigs. One contig shows a deletion occurring between coordinates 1,330-1,877 of the viral genome. This results in a frameshift after codon P234 resulting in a 240aa amino-terminal truncation of Large T antigen. The other contig aligns to VP1 and VP2.

## DISCUSSION

By combining the analyses of point mutations, copy number alterations, structural variants, and viral integration sites of primary MCC tumors at the single nucleotide resolution level for both the transcriptome and whole genome, we identified numerous common features and pathways manipulated in virus-associated MCC and distinct features from non-viral associated MCC. First, the distinct dichotomy between the number of mutations and the mutation signatures of the virus-positive and the virus-negative tumors is surprising, since UV-damage has generally been thought to be a significant contributing factor to both types of MCC (245, 263). Although only one virus-negative MCC tumor was subjected to WGS here, recent targeted sequencing studies support the likelihood that this tumor type is likely to have a high UV mutation burden (159, 251).

Conversely, viral MCC has a low mutation load and is enriched for signature 5, which has been identified previously in many cancers, but most well defined in hepatocellular carcinomas (15). Although signature 5 does not yet have a accepted mechanism, it is linked to the recently identified process of transcription-coupled damage, which results in an enrichment of T-to-C mutations on the transcribed strand (15). Liver tumors harboring this signature were not enriched for hepatitis virus infection, indicating that, at this time, this is not a common virus-mediated mutation process (4, 6). Our work also identified kataegis in one virus-positive tumor overlapping with apparent sites of DNA breaks, which previously had been associated primarily with APOBEC-mediated

cancers, with non-APOBEC related events only recently identified in a large study of 560 breast cancer genomes (181). The similarity of these events in MCC to both types of kataegis has the potential to better characterize the mutagenic processes active in virus-associated MCC and how this contributes to tumorigenesis.

Nearly all cervical and a growing proportion of head and neck carcinomas are caused by the similar small dsDNA virus, HPV, and exhibit high *APOBEC3B* expression and a dominant proportion of APOBEC-signature mutations (34, 59, 60, 131, 264). Considering this, it is unusual that there is no strong evidence of APOBEC3 family upregulation or activity in MCPyV-positive tumors. A recent study also demonstrated that another human polyomavirus, BK polyomavirus, is able to upregulate APOBEC3B in infections of primary renal tubule epithelial cells and this is at least partially mediated by large T antigen expression (192). This same study also demonstrated that MCPyV Large T antigen is able to upregulate APOBEC3B in this cell culture system. Possible explanations for this paradox is that upregulation of APOBEC3B in the cell of origin for viral MCC is not possible due to chromatin-mediated gene silencing or, since only Large T antigen was tested, another protein involved in MCPyV infection prevents TAg-mediated activation of APOBEC3B.

It is curious that the continued expression of viral genes in patients appears to associate with the maintenance of the host genome integrity compared to virus-free tumors considering the ability of MCPyV to integrate into the host genome and the apparent necessity of this event to establish cancer. From the

standpoint of the virus, less DNA damage is beneficial for continued proliferation of the host cell and the virus as integration is not part of the normal viral life cycle and results in a replication deficient virus. This provides a potential explanation as to why MCC is an exceptionally rare cancer despite upwards of a 90% prevalence of MCPyV infection in the human population (265). As seen in this study, when integration does occur, these events coincide with host genome amplifications. These amplifications flanking MCPyV integrants are consistent with observations of CNVs flanking HPV and HBV integrants in HNSCC and HCC, respectively (261, 266–268). Additionally, integration of MCPyV into chromosomes 1 and 6 have both been previously observed at elevated frequencies with breakpoints primarily occurring in the second exon of Large T antigen (269). This suggests potential integration hotspots, yet all observed integration sites have been unique. Compared to HPV-positive tumors and cell lines, we observed less complex integration events in each tumor and these events overlapped with single copy amplifications, whereas, HPV integrants have been shown to flank amplifications up to 90-fold.

Generally, our data support the hypothesis that oncogenic viruses, including HPV, HBV and MCPyV, are able to induce focal genomic CNVs and potentially greater genomic instability through the activation of their origin of replication after integration into the host genome. Despite CNVs being infrequent in virus-associated MCC, there are several recurrent copy number alterations that have been observed between studies that may be initiated by virus mediated genome instability. For example, *SUFU* in our virus-positive tumor-076, which mirrors a

recent report that identified an inactivating mutation of *SUFU* in another MCC tumor. This particular tumor was characterized by an absence of mutations in any of more than 300 cancer related genes sequenced, which based on our results and others suggests that it was also a virus-associated MCC (270). Analysis of the host-virus DNA junctions were limited in this study by the insert size and the 20bp mappable length of the reads, but can be improved in future studies using different sequencing technologies. It would also be interesting to test whether MCPyV can seed recurrent CNV. Expanded genome-wide studies of virus-associated MCC will also reveal if the observed copy number alterations, structural variants, and integration sites are common characteristics and mechanisms of virus-associated MCC. The non-virus-associated tumor in this study exhibited many more somatic alterations compared to the virus-associated tumors that frequently observed in other skin tumors (10). Many of these alterations affected the DNA damage response in the cell, which have important implications for treatment and the evolution rate of the tumor. Ultimately our study highlights the overwhelming ability of Merkel cell polyomavirus to hijack specific cellular processes and produce a tumorigenic phenotype without necessitating the accumulation of hundreds or thousands of somatic mutations and may have important implications for how these tumors progress.

## MATERIALS AND METHODS

*Sample collection.*

Primary tumor tissue and whole blood were collected from a cohort of six individuals summarized in **Table 1**. Patients ranged from 64 to 82 years of age, five white males and one white female. Most shared a previous medical history of non-melanoma skin cancer and actinic keratosis. Other medical history includes coronary artery disease, gout and rheumatoid arthritis. Primary tumor sites were variable for each patient, although most tumors were found in areas of the skin susceptible to increased sun exposure including the forehead, arm and ear.

*DNA sequencing, alignment and analysis.*

Tumor and normal (PBMC) DNA preparations (10 μg) were sequenced by the Beijing Genome Institute (BGI) on the Complete Genomics platform to an average of 100x depth (271). Alignment of reads and calling of somatic mutation, copy number, structural variants and annotation of repetitive elements were performed by BGI using their analysis pipeline. Somatic mutations were filtered out if they did not score as SQHIGH as defined by BGI analysis workflow. Additionally, somatic mutations that had identical 41mer flanking sequences were removed. Only mutations occurring in genes implicated in cancer by the COSMIC cancer gene census were further characterized (272). Functional implications for missense mutations were determined using the SIFT and PROVEAN v1.1.3 protein batch analysis tool submitted through the J. Craig Venter website (273–276). COSMIC annotations were further used to annotate copy number alterations and structural variants for genes commonly altered in cancer. Pathway analysis was conducted using the core analysis pipeline of the Ingenuity

Pathway Analysis software (IPA) (Qiagen) and pathways were only further analyzed if an enrichment z-score was able to be calculated and only pathways with an enrichment p-value less than 0.05 were considered statistically significant. Z-scores are a measure of the relative enrichment or depletion of a pathway in the dataset.

*RNA sequencing, alignment and analysis.*

RNA was purified using Ultra RNA Library Prep Kit (New England BioLabs) and was sequenced (0.1 μg total) on the Illumina HiSeq 2500 platform with paired end flowcells and 50 cycles in each direction. Sequences were aligned to a combination of the hg19 and MCPyV genomes (NCBI) using TopHat2 (277). Differential expression analysis was performed with CuffLinks and DESeq2 (277, 278). Only genes with a differential expression q-value of less than 0.05 and a greater than or equal to 3-fold change in expression in virus-positive vs. virus-negative samples were considered significant. To focus on relevant genes, only genes implicated in cancer by the COSMIC cancer gene census, E2F-regulated genes, and leukocyte-related genes were further characterized (272). Pathway analysis was completed by submitting the log-base-2-fold change of the top differentially expressed genes between MCPyV positive and negative tumors into the core analysis pipeline of IPA (Qiagen). The nature of this analysis indicated that pathways enriched for virus-positive tumors were pathways with the highest positive z-scores as calculated by IPA and virus-negative tumors were pathways with the lowest negative z-score, only pathways

with an enrichment p-value less than 0.05 were considered statistically significant. Principal component analysis was performed using the R statistical package with all annotated genes.

*Mutation profile analysis.*

Flanking 5’ and 3’ bases at the site of each somatic mutation were collected from the hg19 reference genome. Proportion of each mutation in its trinucleotide context was calculated in respect to the total number of somatic mutations. Mutation profiles were plotted using the R statistical software with the “SomaticSignatures” package (279). This package was also used to predict mutation signatures from the somatic mutations of each tumor genome. To determine mutation strand asymmetries and produce subsequent plots, we input somatic mutations grouped by MCPyV status into the AsymTools MatLab script (15).

Mutation clusters (kataegis) was evaluated by taking the distance in base pairs from one somatic single base substitution to the previous mutation or intermutational distance (IMD). The genomic distributions of mutations were plotted using ggplot2. Clusters of mutations were determined by the same method as Alexandrov and colleagues (4), which they defined as at least six concurrent mutations with an average intermutational distance less than 1000bp. Unique events did not overlap clusters observed in other samples, which are likely a byproduct of sequencing errors.

*Virus integration site identification pipeline.*

Half mapped read pairs were extracted from the whole genome alignments using a custom script. Due to the variable, gapped structure of Complete Genomics reads we only used the 20bp continuous segment located at the beginning of the read (271). These read pairs were then mapped back to the reference genome using Bowtie2 and the virus-host reference genome used for the RNAseq analysis (280). After determining their mapping coordinates from the Bowtie2 alignment, discordant read pairs were extracted and de novo assembled using velvet with a word size of 11bp (281). The discordant reads were then remapped to the new contigs using nucleotide BLAST with “short” settings and a word size set to 9bp (282). Using these BLAST results, the de novo assembled contigs were filtered to identify those that contained reads that initially mapped to the human and viral genomes. The resulting junctions were visualized by plotting out the mapped starting positions of the reads fitting the aforementioned criteria (according to the BLAST alignment) and coloring them by origin (viral or host) with ggplot2.

*Virus-host fusion transcript identification pipeline.*

Identification of viral integration sites follows the pipeline suggested by the SummonChimera (http://www.biomedcentral.com/1471-2105/15/348) software. Raw fastq paired-end reads were mapped with default Bowtie2 parameters to a database composed of the Merkel Cell Polyomavirus (HM355825.1) and Human hg19 genomes. Next, all unmapped reads are input into BLASTN with

parameters ‘–word_size 16’ and ‘-outfmt 6’ and compared against the Merkel Cell Polyomavirus genome. Then, all reads with a BLASTN hit to the viral genome are run through BLASTN against the hg19 genome, using the same parameters. Finally, SummonChimera was run with the BLASTN and SAM report files and generated a report containing all detected chimeric junctions.

*Virus identification pipeline.*

Raw fastq reads were mapped to the hg19 version of the human genome and a human mRNA database (to remove spliced reads) with Bowtie2 using default parameters (280). Then unmapped reads were extracted and low quality reads were removed and poor quality ends were trimmed with Prinseq (http://prinseq.sourceforge.net/). High quality reads were assembled with CLC Assembler. Contigs >= 500bp were masked with Repeat Masker and filtered as described(283). Then, high quality contigs were annotated by a computation subtraction pipeline: (i) the human genome using BLASTN; (ii) GenBank nt database using BLASTN; (iii) GenBank nr database using BLASTX; and (iv) the NCBI viral refseq genome database using TBLASTX. Minimal E-value cutoff of 1e-5 for all steps were applied. Additionally, a minimal query coverage of 50% and minimal percent identity of 80% were applied to the BLASTN steps.

## ACKNOWLEDGEMENTS

We are grateful for sequencing and bioinformatic support by the DFCI Center for Cancer Computational Biology. We would also like to thank Dr. Birgit Crain of Omicia Inc. for providing us with scripts to process Complete Genomics reads.

## FUNDING INFORMATION

This work was supported in part by US Public Health Service grants R01CA63113, R01CA173023, and P01CA050661 and the DFCI Helen Pappas Merkel Cell Research Fund and the Claudia Adams Barr Program in Cancer Research to JAD, and R21 CA170248 to JMP. Salary support for G.J.S. was provided by the National Science Foundation Graduate Research Fellowship (Grant 00039202). R.S.H. in an investigator of the Howard Hughes Medical Institute.

## SUPPLEMENTAL DATA

Supplemental data are available from http://mbio.asm.org/content/8/1/e02079-16.full

## Tables

**Table 4.1**. Summary of patients and tumors used in this study.

| ID | Sex | Age at Diagnosis | Past Medical History | Primary Tumor Site | Viral Reads | WGS | Somatic mutations |
|---|---|---|---|---|---|---|---|
| 09156-050 | M | 76 | Actinic keratosis, basal cell carcinoma, squamous cell carcinoma | Right forehead | 0 | Yes | 127236 |
| 09156-076 | M | 82 | Hypothyroidism, Diabetes mellitus, Hypoaldosteronism | Left third finger | 10441 | Yes | 4132 |
| 09156-088 | M | 64 | Rheumatoid arthritis | Right upper medial thigh | 5449 | Yes | 3397 |
| 09156-090 | F | 77 | Basal cell carcinoma, hypothyroidism | Right dorsal foot | 26289 | No | NA |
| 09156-142 | M | 79 | Actinic keratosis, Basal cell carcinoma, Squamous cell carcinoma, Polymyalgia rheumatica | Right postauricular | 0 | No | NA |
| 09156-146 | M | 77 | Actinic keratosis, Basal cell carcinoma, polymyalgia rheumatica | Left upper arm | 18947 | No | NA |

**Table 4.2**. Annotation of somatic point mutations in MCC tumors.

| | 09156-050 | | 09156-076 | | 09156-088 | | Log2 fold change |
|---|---|---|---|---|---|---|---|
| **Upstream** | 4129 | 3.25% | 170 | 3.99% | 169 | 4.79% | 0.43 |
| **CDS** | 1107 | 0.87% | 44 | 1.03% | 35 | 0.99% | 0.22 |
| Synonymous | 1087 | 0.85% | 43 | 1.01% | 35 | 0.99% | 0.23 |
| Missense | 658 | 0.52% | 28 | 0.66% | 21 | 0.59% | 0.27 |
| Nonsense | 47 | 0.04% | 5 | 0.12% | 2 | 0.06% | 1.23 |
| Stop loss | 1 | 0.00% | 0 | 0.00% | 0 | 0.00% | 0.00 |
| **Intron** | 40809 | 32.07% | 1417 | 33.22% | 1195 | 33.84% | 0.06 |
| **UTR3** | 612 | 0.48% | 33 | 0.77% | 27 | 0.76% | 0.68 |
| **UTR5** | 155 | 0.12% | 7 | 0.16% | 6 | 0.17% | 0.46 |
| **Total in genes** | 45603 | 35.84% | 1620 | 37.98% | 1389 | 39.34% | 0.11 |
| | | | | | | | |
| **Alu** | 12406 | 9.75% | 612 | 14.35% | 483 | 13.68% | 0.52 |
| **ERV1** | 4964 | 3.90% | 210 | 4.92% | 181 | 5.13% | 0.37 |
| **ERVK** | 469 | 0.37% | 45 | 1.06% | 25 | 0.71% | 1.26 |
| **ERVL** | 10045 | 7.89% | 249 | 5.84% | 178 | 5.04% | -0.54 |
| **hAT** | 2293 | 1.80% | 47 | 1.10% | 49 | 1.39% | -0.53 |
| **L1** | 23438 | 18.42% | 948 | 22.23% | 756 | 21.41% | 0.24 |
| **L2** | 4786 | 3.76% | 104 | 2.44% | 104 | 2.95% | -0.48 |
| **MIR** | 3267 | 2.57% | 85 | 1.99% | 69 | 1.95% | -0.38 |
| **RTE** | 129 | 0.10% | 8 | 0.19% | 3 | 0.08% | 0.43 |
| | | | | | | | |
| **Low complexity** | 551 | 0.43% | 31 | 0.73% | 29 | 0.82% | 0.84 |
| **Simple repeat** | 361 | 0.28% | 78 | 1.83% | 69 | 1.95% | 2.74 |
| | | | | | | | |
| **Total** | **127236** | | **4132** | | **3397** | | |

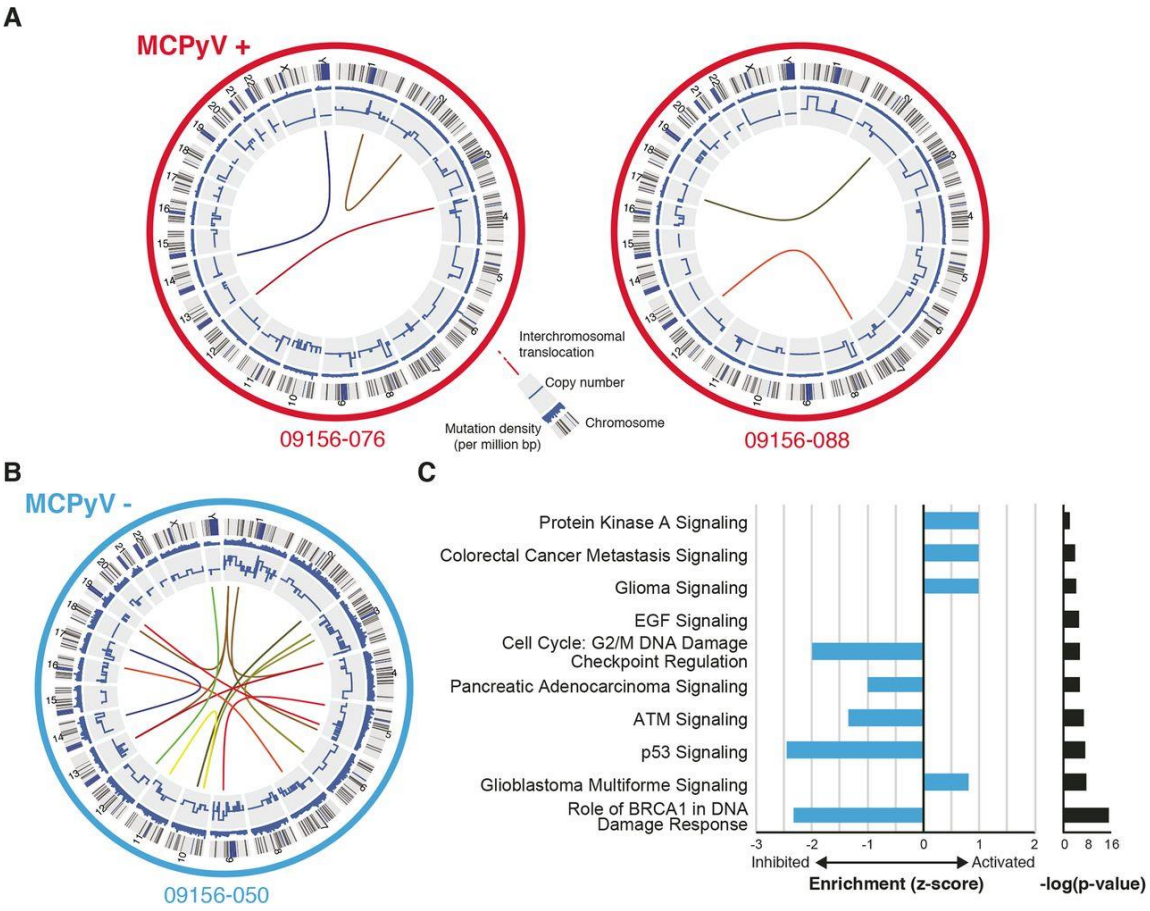



**Figure 4.1. Circos plots and functional annotation of genomic alterations in MCC tumors**

(A-B) The MCPyV-positive tumors are highlighted in red and the MCPyV-negative tumor is highlighted in blue. The outermost ring represents each chromosome. The next ring represents the density of somatic mutations calculated in 1 Mbp regions. The inner most ring represents the copy number alterations of each chromosome. The colored lines in the inner circle represent inter-chromosomal translocations. (C) Bar plot of the enrichment z-score (blue) and p-values (black) of pathways predicted from somatic variants in tumor-050.

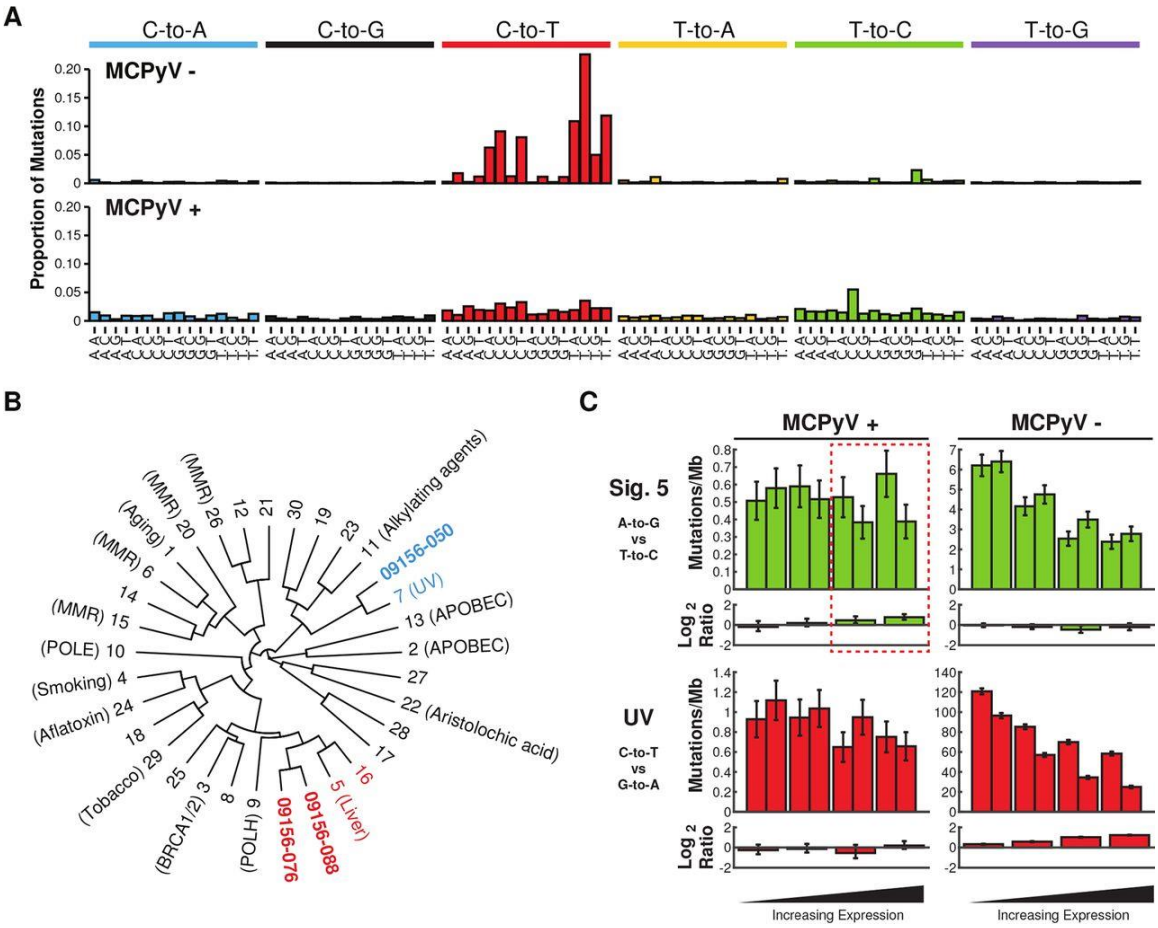



**Figure 4.2. Summary of mutation signatures detected in MCPyV positive and negative MCC.**

(A) Bar plot of the average contribution of each base substitution at each possible trinucleotide context across the genome in MCPyV-positive and MCPyV-negative tumors. (B) Dendrogram representing the similarity of mutation signatures detected in each tumor to known mutation signatures in cancer. MCPyV-positive tumors are highlighted in red and -negative tumors are in blue. (C) Bar plots of the transcriptional strand asymmetry measured for “UV-signature”, C-to-T, and “signature 5”, T-to-C, over four quartiles of expression and divided by MCPyV-positive and MCPyV-negative tumors. Upper plots show

mutation density by each base substitution and its complement substitution. The lower plots show the log2 ratio representing the degree of transcriptional asymmetry. More positive denotes enrichment for the non-transcribed strand; more negative denotes enrichment for the transcribed strand.

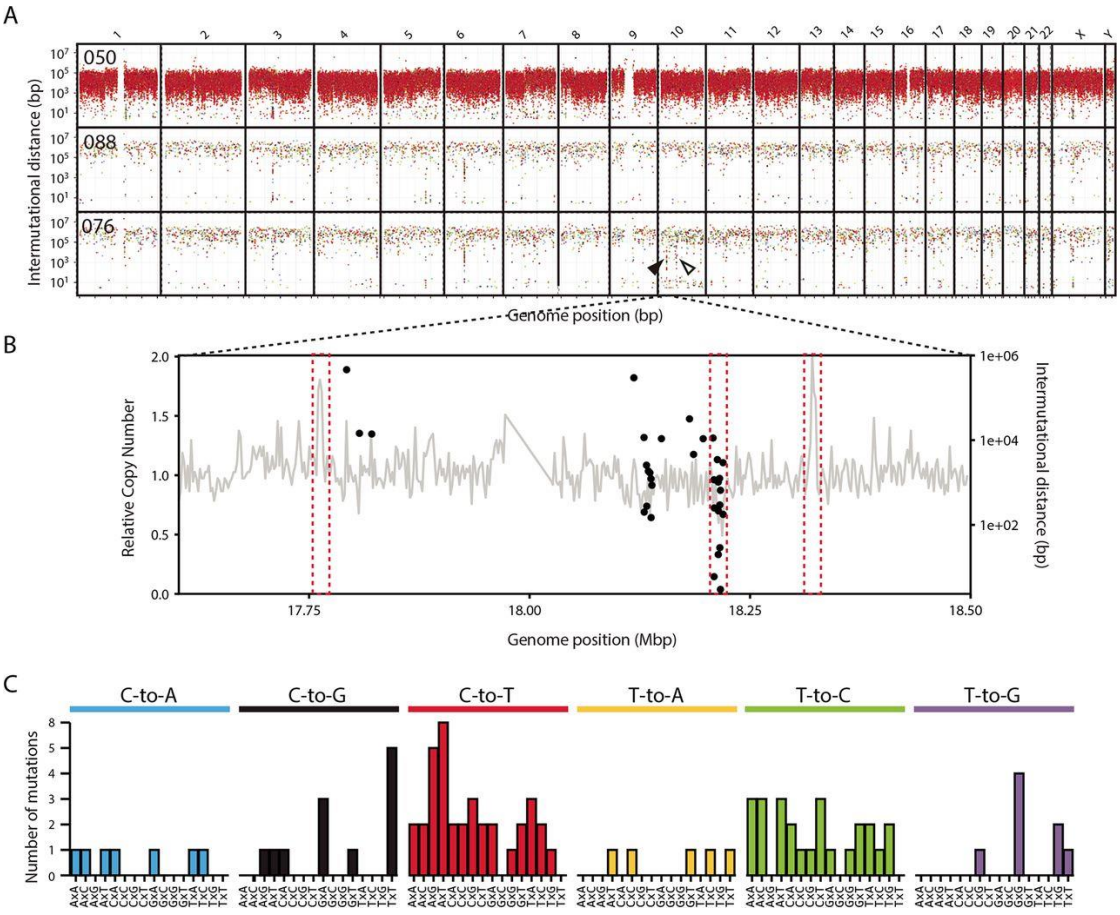



**Figure 4.3. Summary of mutation clusters observed in MCC.**

(A) Rainfall plot of intermutational distances between somatic single base substitutions by genome position. Colors for each type of base substitution are the same as in Figure 2A. Unique kataegis events are indicated by arrows. The kataegis event expanded in B is indicated by a filled in black arrow (B) Coincidence of kataegis events and structural alterations along the X axis representing genomic position, which are indicative of DNA breaks are highlighted by dotted red boxes. Relative DNA copy number is shown as a grey line graph using the left Y axis for its scale. The intermutational distance for each point mutation is shown by a black dot using the right Y axis. (C) Bar graph of the

number of mutations observed by base substitution and context in the kataegis events in tumor-076.

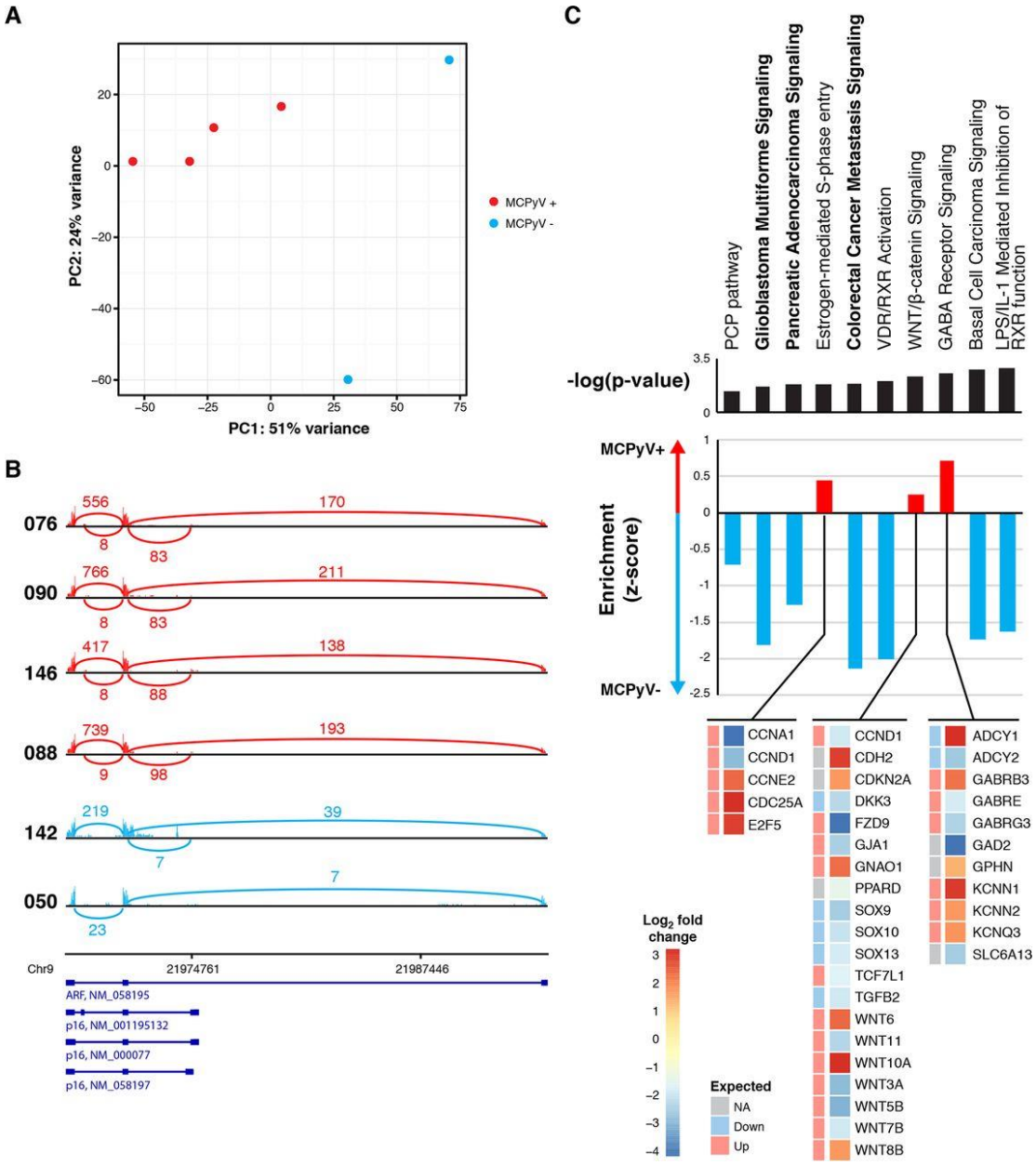



**Figure 4.4. Summary of MCPyV positive and negative MCC transcriptome.**

(A) Plot of the first two principal components from PCA of the transcriptomes of each MCC tumor colored by the presence of MCPyV, red for positive, blue for negative. (B) Sashimi plots showing the number of reads spanning the exon junctions of the CDKN2A locus for each tumor samples, labeled on the left. Virus-positive samples are in red; virus-negative tumors are in blue. Known

transcript variants of CDKN2A are shown below the sashimi plots. (C) Bar plot of the enrichment z-score (red for MCPyV-associated and blue for non-MCPyV-associated) and p-values (black) of pathways predicted from significantly differentially expressed genes. Pathways supported by somatic variants are bold. Below the bar plots, details of the log2 fold change and predicted direction of expression of genes significantly differentially regulated and associated with the pathways attributed to MCPyV-positive tumors. For both observed and expected expression changes, increased fold change is in red and decreased is in blue

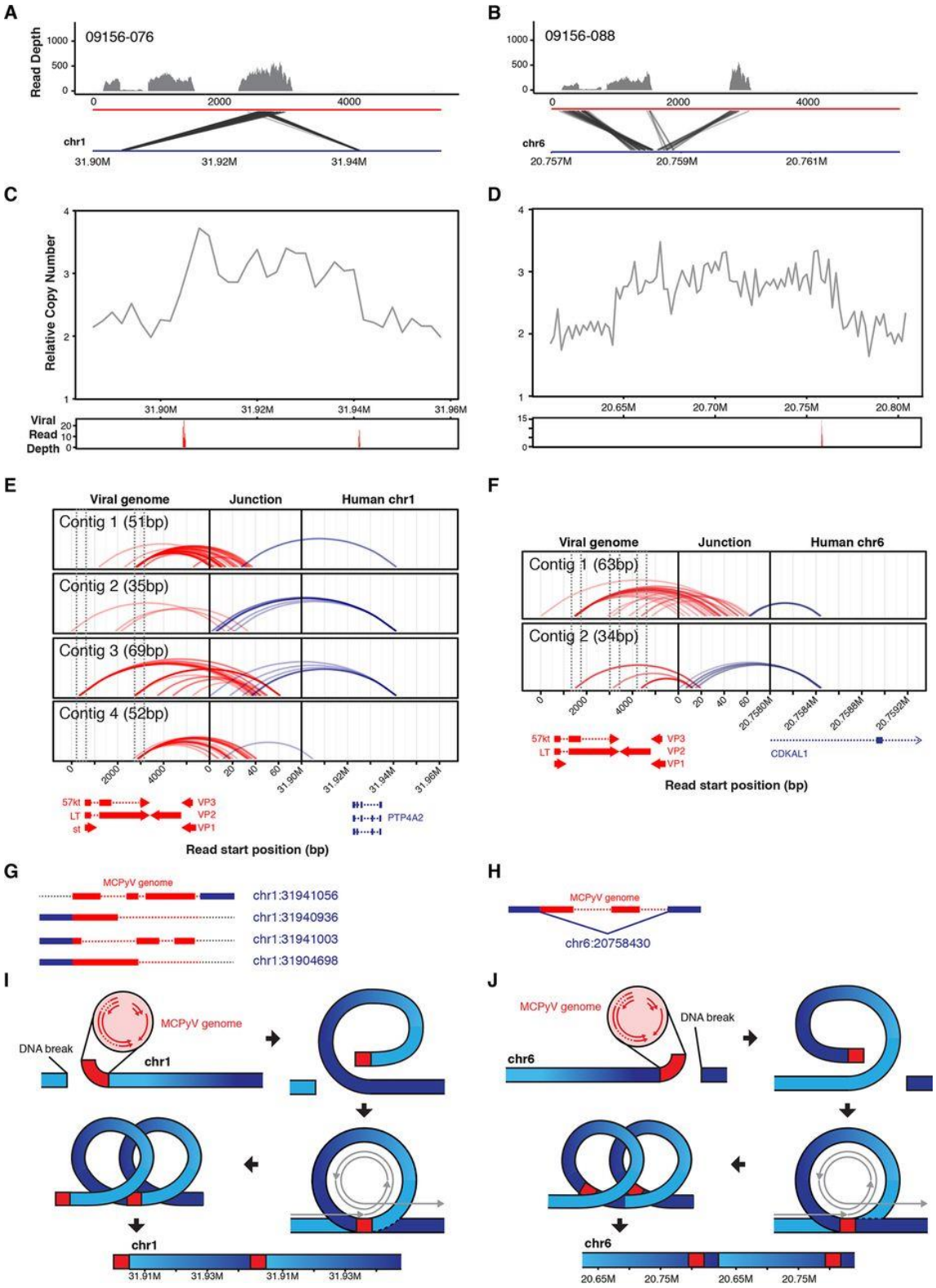

A
09156-076
Read Depth
chr1
31.90M
31.92M
31.94M
B
09156-088
chr6
20.757M
20.759M
20.761M
C
Relative Copy Number
Viral Read Depth
31.90M
31.92M
31.94M
31.96M
D
20.65M
20.70M
20.75M
20.80M
E
Viral genome
Junction
Human chr1
Contig 1 (51bp)
Contig 2 (35bp)
Contig 3 (69bp)
Contig 4 (52bp)
57kt
LT
st
VP3
VP2
VP1
PTP4A2
Read start position (bp)
F
Viral genome
Junction
Human chr6
Contig 1 (63bp)
Contig 2 (34bp)
57kt
LT
VP3
VP2
VP1
CDKAL1
Read start position (bp)
G
MCPyV genome
chr1:31941056
chr1:31940936
chr1:31941003
chr1:31904698
H
MCPyV genome
chr6:20758430
I
MCPyV genome
DNA break
chr1
chr1
31.91M
31.93M
31.91M
31.93M
J
MCPyV genome
DNA break
chr6
chr6
20.65M
20.75M
20.65M
20.75M

**Figure 4.5. Detailed evaluation of MCPyV insertion sites in MCC tumors.** A,B. Diagrams of discordant read pairs and association with observed RNAseq coverage from MCC tumors. Depth of coverage histograms for RNAseq reads across the MCPyV genome are shown in the upper section. Discordant read pairs are shown in the lower section as transparent black lines linking the MCPyV genome to putative insertion sites in the human genome. C,D. Relative copy numbers from each patient near the detected viral integration sites are shown in the upper panel. Depth of coverage of read pairs that map to the host and viral genomes are shown in red in the lower panel. E,F. Diagram of the de novo assembled virus-host fusion contigs. The start positions of each read are connected from the viral genome (red, left) and the host genome (blue, right) to the de novo assembled fusion contig representing the integration event (center) via colored arches. Virus and host genes are shown below the arch diagram. G,H. Simplified schematic of the integration events interpreted from the corresponding data in E and F. The viral and host genomes are shown in red and blue respectively. Deletions in the viral genome are represented by red dotted lines and junctions without support from the de novo assembled contigs are shown in grey dotted lines. Host chromosome positions are in blue adjacent to the schematic. I,J. Model for MCPyV-insertion-mediated host structural variants. DNA double strand break initiates insertion of the linearized MCPyV genome into the host genome. After insertion DNA loops over, forms transiently circular DNA, and allows for rolling circle DNA replication initiated from the viral origin of

replication. Separation of the transiently circular DNA results in a focal amplification of the host genome flanked by viral DNA.

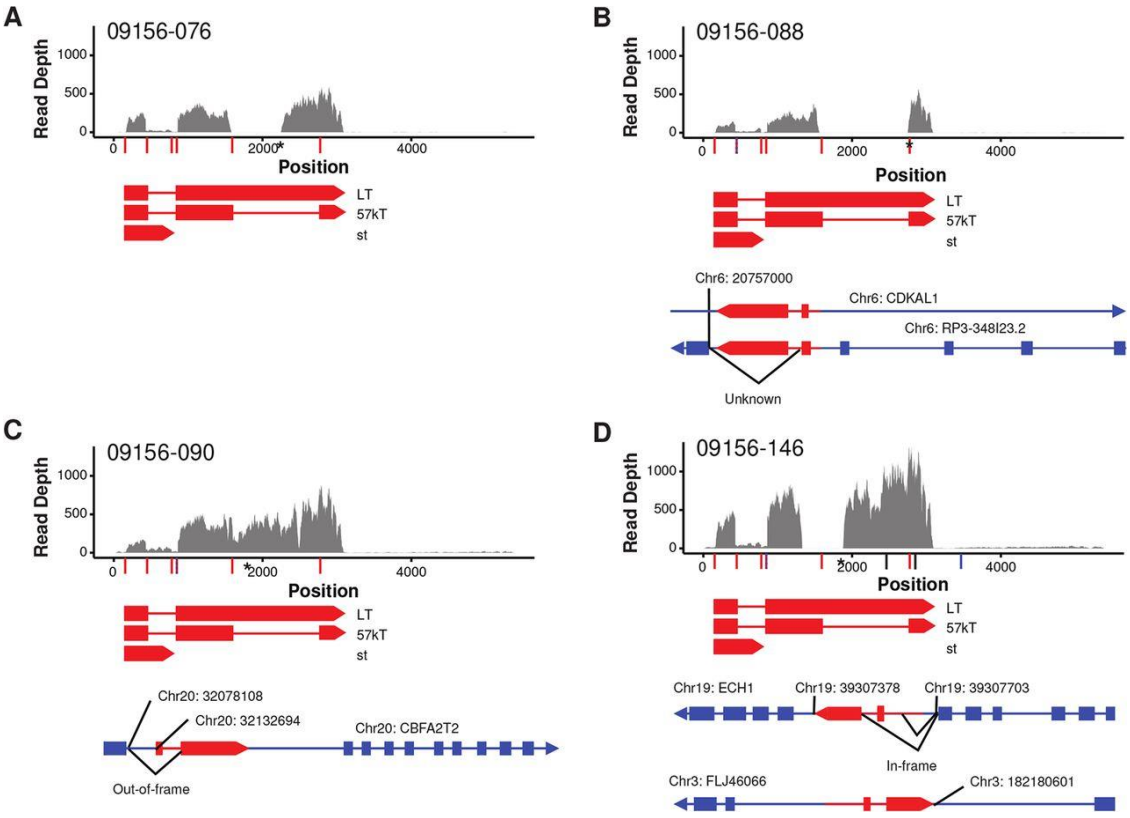



**Figure 4.6. MCPyV genome coverage and diagrams of the detected viral-host transcript chimeras.**

(A-D) Plot of depth of coverage over MCPyV genome for each patient tumor. Known T antigen isoforms are represented below with known splice junctions, virus host splice fusions, and potential DNA chimeric junctions indicated by red, blue and black vertical lines on the X axis, respectively. Overlapping junctions are represented by dashed lines. Asterisks on the x-axis represent stop codons introduced by mutation within the Large T antigen coding region. Diagrams of the detected viral-host fusion transcripts for the corresponding patient are below the depth of coverage plots. Arrows indicate the direction of transcription. Human genes are represented in blue and the MCPyV genome is represented in red, only Large T antigen exons are diagramed with red boxes. Chromosome and

position of each DNA junction is labeled above the diagrams. (D) Tumor corresponding to 09156-076 had no detected integration or fusion transcripts.

# CHAPTER 5:

# Conclusions and Discussion

**Author:** Gabriel Starrett

## CONCLUSIONS

*Chapter 2: APOBEC3H haplotype I is a likely contributor to breast and lung cancer mutagenesis* (90)

Since the discovery of APOBEC3B as a major contributor to mutagenesis in breast and others cancer types (1, 34), the geographically distributed germline deletion of APOBEC3B (82) has proved to be a difficult paradox to address. Many studies have shown that this germline deletion corresponds to increased risk (not a decreased risk as expected from aforementioned discoveries) for the development of breast cancer. Further complicating the story these tumors still bear a strong APOBEC-mediated mutation signature (83, 84, 86, 202). Several other APOBEC3s have been implicated as mutagens in the absence of APOBEC3B. APOBEC3A has been the leading candidate for the source of these mutations as it is the most active of the human APOBEC3s and the transcript is modified with the deletion allele (89). However, APOBEC3A is expressed at nearly undetectable levels in most tumors and there has not been consistent evidence that the APOBEC3A is differentially regulated by the deletion allele (34, 90).

We mined the Cancer Genome Atlas (TCGA) to identify individuals homozygous for the *APOBEC3B* deletion. We then determined the APOBEC3H haplotype for each of these tumors. *APOBEC3H* is the most polymorphic of the human *APOBEC3* genes. This surprisingly revealed that only individuals with at least one copy of APOBEC3H-I haplotype had significant APOBEC-associated mutation signatures. APOBEC3H-I has traditionally been considered partially

unstable and inactive. We demonstrated that APOBECH-I is equally active to APOBEC3H-II, which is normally considered the HIV-1 restrictive haplotype, in virus restriction and deaminase assays when normalized to equal protein levels. In these assays we were also able to show that APOBEC3H is able to produce a mutation profile that is strikingly similar to that observed in APOBEC3B-null cancers. Lastly, we were able to show that APOBEC3H-I has greater nuclear localization in several cell lines than APOBEC3H-II. Providing further evidence for APOBEC3H-I involvement in cancer, early clonal mutations were significantly enriched for the APOBEC signature only in lung adenocarcinoma, but with a strong trend in breast cancer. We concluded that the combination of APOBEC3B and APOBEC3H-I accounts for the bulk of, and perhaps all, APOBEC signature mutations in cancer.

*Chapter 3: Functional upregulation of APOBEC3B by polyomaviruses* (192)

Previously the E6 oncoprotein of high-risk HPV has been shown to specifically upregulate APOBEC3B expression in keratinocyte cell lines and is supported by increased expression of APOBEC3B in HPV+ head and neck squamous cell carcinomas over those that are HPV- (59). This ultimately led to the hypothesis that, generally, small DNA tumor viruses are capable of upregulating APOBEC3B and this is part of their mechanism of tumorigenesis. To test this, we utilized BK polyomavirus and the primary renal proximal tubule epithelial cell culture system (RPTE). BK polyomavirus is the leading cause of kidney transplant rejection and has a growing association with bladder cancer (61, 284, 285). Half the polyomavirus genome is dedicated to encoding the large

tumor antigen (LTAg) which overlaps both the functions of HPV E6 and E7, inhibiting p53 and pRB1, respectively (135, 136, 286, 287).

At one-day post infection, the viral capsid has yet to reach the nucleus and release the viral genome. At three days post infection, BKPyV infected cells express high levels of LTAg and actively replicate the viral genome (236). Also at this timepoint, APOBEC3B is the only APOBEC family member that is significantly upregulated. This upregulation corresponds to increased deaminase activity within the infected RPTEs. Knockdown of APOBEC3B does not affect BKPyV replication dynamics, indicating that APOBEC3B does not overtly promote or inhibit BKPyV infection. Nucleotide motif analysis of the publicly available BKPyV sequences from Genbank revealed the BKPyV genome is depleted for APOBEC-target TC motifs and enriched for the product TT motifs. Additionally, these motifs are depleted in a strand-biased manner that is consistent with the preferred function of APOBEC3B to deaminate the lagging strand of replicating DNA. We concluded that the Large T antigen of multiple different human polyomaviruses is able to upregulate APOBEC3B and this upregulation has likely influenced the nucleotide motif abundance in genomes of polyomaviruses.

*Chapter 4: Merkel cell polyomavirus has dominant control over the tumor genome and transcriptome of Merkel cell carcinoma* (161)

HPV-mediated cancers consistently show some of the highest APOBEC3B expression and strongest APOBEC-mediated mutation signatures of any type of cancer (2, 34, 60). This is likely due to the active expression of E6

from the integrated high-risk HPV genome (59). In the previous chapter, we showed that the large T antigen of human polyomaviruses, such as Merkel cell polyomavirus (MCPyV), is able to specifically and significantly upregulate APOBEC3B protein and activity. Merkel cell carcinoma is a rare skin cancer and approximately 80% of cases have copies of the MCPyV genome integrated into chromosomal DNA (155, 157). We performed RNA sequencing on the tumors of six Merkel cell carcinoma patients and whole genome sequencing (WGS) of three patients.

Splitting the patients by detection of MCPyV transcripts initially revealed a striking difference in the abundance and context of mutations in virus-associated versus non-associated tumors. Tumors without MCPyV had abundant UV-signature mutations and chromosomal translocations that interrupted genes associated with DNA damage recognition and repair. MCPyV-associated tumors had ~20-fold fewer mutations that were primarily T-to-C with a transcriptional strand bias consistent with a previously defined mutation signature. When further investigating the integration site of the virus in the WGS data, we identified that these sites overlap focal amplifications of the host genome. These amplifications are very similar to those observed in HPV-associated cancers and support a looping model where integration of the viral genome forms transiently circular DNA that facilitates the firing of the viral origin to amplify the viral genome as well as flanking host chromosomal DNA. In conclusion, we identified that MCPyV drives carcinogenesis in Merkel cell carcinoma in the absence of abundant point

mutations and copy number alterations observed in most tumor types and integration of the virus seeds focal amplifications in the tumor genome.

## DISCUSSION

*The antiviral enzyme APOBEC3H-I contributes to cancer mutagenesis*

Through carefully controlled bioinformatic and biochemical experiments, we identified in the absence of APOBEC3B that APOBEC3H-I is a significant contributor of APOBEC signature mutations in breast and lung cancers. These results advance multiple models in which APOBEC3H-I causes a continuous smoldering mutator phenotype that is eclipsed upon activation of the highly active mutator APOBEC3B by a tumor-initiating event. This explains why it the effect could only be observed in the absence of APOBEC3B. Collectively expression and activity of APOBEC3B and APOBEC3H-I can now account for all APOBEC-signature mutations in cancer, apart from some possibly coming from APOBEC3A in some myeloid lineage cancers (90). However, many questions still remain in regards to APOBEC-mediated mutagenesis in cancer. Specifically, known upregulation of APOBEC3H by HIV-1 infection opens the potential for flares of cancer mutagenesis initiated by viral infection (288).

One of the more interesting phenotypes that doesn’t have a complete mechanistic explanation is the cellular localization of APBOEC3H-I vs APOBEC3H-II (48). This is a fundamental difference between the two haplotypes that facilitates APOBEC3H-I to act as a genomic DNA mutator. Differential localization may be due to a yet-to-be discovered cytoplasmic retention factor

that requires the R105 amino acid to bind to APOBEC3H. Despite the number of new studies on APOBEC3s and their roles in cancer, we still understand very little about the initiation and early stages of tumor development with or without APOBEC3B. Continued mechanistic studies in APOBEC3B-null primary tumors will facilitate these insights.

*The cancer mutagenic enzyme APOBEC3B is induced by polyomavirus infection and contributes to virus evolution*

Studies into the mutation signatures in cancer gave the first indications of a connection between tumor viruses and APOBEC-mediated cancer mutagenesis. Continued studies into the oncoproteins of HPV confirmed that these viruses manipulate the expression of *APOBEC* family members through direct binding of the *APOBEC3B* promoter and activation of transcription factors such as TEADs (59, 137, 138). Other studies into the regulation of APOBEC3B in cancer has further revealed the complexity of its regulatory mechanisms and the need for novel molecular methods to interrogate it (74, 166, 195, 197). Tumor viruses continue to act as valuable molecular probes to understand the regulation of oncogenes and tumor suppressors, highlighting the value of studying small DNA viruses such as BK polyomavirus. Our study has revealed that T-antigen, which shares some overlapping function with high-risk HPV E6, is able to specifically upregulate APOBEC3B at the transcript, protein and enzymatic activity levels.

Human papillomaviruses were the first to be shown to be depleted for APOBEC-target motifs and enriched for product motifs and this corresponded to

the ability of that virus to upregulate APOBEC3B (143). We were able to show that the BKPyV genome has similar motif distributions while also showing a strong replication-based strand biased depletion not observed in HPV. Additional studies into the structure of small DNA viruses may also provide valuable insights into how APOBEC3B is able to mutate host genomic DNA.

*Copy number variants in polyomavirus-mediated Merkel cell carcinoma are induced by viral integration*

Although the previous chapter showed that transduction of MCPyV LTAg in primary RPTEs is able to upregulate APOBEC3B, there is no evidence of elevated expression of any APOBEC3 family member in our study of Merkel cell carcinoma (MCC) and others (159, 160, 251). Additionally, virus-positive MCC tumors lacked any significant observable APOBEC-signature mutations, which is in stark contrast to HPV-associated tumors, which have some of the highest APOBEC-mediated mutation loads (34). Additional studies in other cell lines, such as the novel primary cell line for MCPyV replication (289), will be required to determine if this is cell type specific phenomenon. It is also possible that MCPyV has additional mechanisms to prevent or counteract the expression and activities of APOBEC3 enzymes since we know that they can act on polyomavirus genomes.

However, like HPV-associated tumors, we identified that MCPyV integration sites are able to seed focal amplifications in the tumor genome (261). Additional WGS studies on MCC will determine the degree of complexity that these rearrangements can take and what initiates the integration. A recent study

showed that increased expression of APOBEC3A corresponds with integrated HPV in tumors (290). It could be that MCPyV initially upregulates APOBEC3B and this increases DNA damage in the host and viral genomes. This damage then promotes double-strand breaks and facilitates the integration of the viral genome. After integration, large sections of the viral T antigen are frequently deleted (155, 226). This could be an explanation as to why MCPyV-positive tumors don't show APOBEC3B expression or mutation signature.

## CLOSING REMARKS

Historically, the study of viruses has proven to be integral in regards to understanding oncoproteins, tumor suppressors, and the biology of DNA replication. Additionally both viruses and cancers are similar in the characteristic that they require constant new genetic diversity to survive and outpace the host immune system, but are also susceptible to hypermutation and genome instability. The primary goal of my thesis has been to better understand how the antiviral APOBEC3 family contributes to the genetic diversity of cancer and tumor-associated viruses and how in turn tumors and viruses affect the regulation of APOBEC3 enzymes. Here I have shown by investigating tumors with a homozygous germline deletion of APOBEC3B, a specific haplotype of the normally virus-associated APOBEC, APOBEC3H, can contribute to the APOBEC signature mutations observed in breast and lung cancers (**Chapter 2**). The normally cancer associated enzyme, APOBEC3B, is induced by the Large T antigen of human polyomaviruses, much like what is observed in HPV infections

and HPV-associated cancers (**Chapter 3**). Paradoxically when investigating Merkel cell carcinoma, which is caused by Merkel cell polyomavirus, we found that these tumors do not have a significant proportion of APOBEC-signature mutations. Rather integration of the virus seeds focal amplifications of the host genome, in which Large T antigen expression is lost (**Chapter 4**). These studies combine to provide novel insights into the complex interplay between oncoviruses and APOBEC regulation and activity across multiple tissue types and cancers.

## REFERENCES


1. Burns MB, Lackey L, Carpenter M a, Rathore A, Land AM, Leonard B, Refsland EW, Kotandeniya D, Tretyakova N, Nikas JB, Yee D, Temiz N a, Donohue DE, McDougle RM, Brown WL, Law EK, Harris RS. 2013. APOBEC3B is an enzymatic source of mutation in breast cancer. Nature 494:366–70.
2. Roberts S a, Lawrence MS, Klimczak LJ, Grimm S a, Fargo D, Stojanov P, Kiezun A, Kryukov G V, Carter SL, Saksena G, Harris S, Shah RR, Resnick M a, Getz G, Gordenin D a. 2013. An APOBEC cytidine deaminase mutagenesis pattern is widespread in human cancers. Nat Genet 45:970–6.
3. Nik-Zainal S, Van Loo P, Wedge DC, Alexandrov LB, Greenman CD, Lau KW, Raine K, Jones D, Marshall J, Ramakrishna M, Shlien A, Cooke SL, Hinton J, Menzies A, Stebbings L a, Leroy C, Jia M, Rance R, Mudie LJ, Gamble SJ, Stephens PJ, McLaren S, Tarpey PS, Papaemmanuil E, Davies HR, Varela I, McBride DJ, Bignell GR, Leung K, Butler AP, Teague JW, Martin S, Jönsson G, Mariani O, Boyault S, Miron P, Fatima A, Langerød A, Aparicio S a JR, Tutt A, Sieuwerts AM, Borg Å, Thomas G, Salomon AV, Richardson AL, Børresen-Dale A-L, Futreal PA, Stratton MR, Campbell PJ. 2012. The life history of 21 breast cancers. Cell 149:994–1007.
4. Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio S a JR, Behjati S, Biankin A V, Bignell GR, Bolli N, Borg A, Børresen-Dale A-L, Boyault S,

Burkhardt B, Butler AP, Caldas C, Davies HR, Desmedt C, Eils R, Eyfjörd JE, Foekens J a, Greaves M, Hosoda F, Hutter B, Ilicic T, Imbeaud S, Imielinski M, Imielinsk M, Jäger N, Jones DTW, Jones D, Knappskog S, Kool M, Lakhani SR, López-Otín C, Martin S, Munshi NC, Nakamura H, Northcott P a, Pajic M, Papaemmanuil E, Paradiso A, Pearson J V, Puente XS, Raine K, Ramakrishna M, Richardson AL, Richter J, Rosenstiel P, Schlesner M, Schumacher TN, Span PN, Teague JW, Totoki Y, Tutt ANJ, Valdés-Mas R, van Buuren MM, van 't Veer L, Vincent-Salomon A, Waddell N, Yates LR, Zucman-Rossi J, Futreal PA, McDermott U, Lichter P, Meyerson M, Grimmond SM, Siebert R, Campo E, Shibata T, Pfister SM, Campbell PJ, Stratton MR. 2013. Signatures of mutational processes in human cancer. Nature 500:415–21.

5. Poon SL, Huang MN, Choo Y, McPherson JR, Yu W, Heng HL, Gan A, Myint SS, Siew EY, Ler LD, Ng LG, Weng W-H, Chuang C-K, Yuen JS, Pang S-T, Tan P, Teh BT, Rozen SG. 2015. Mutation signatures implicate aristolochic acid in bladder cancer development. Genome Med 7:38.

6. Schulze K, Imbeaud S, Letouzé E, Alexandrov LB, Calderaro J, Rebouissou S, Couchy G, Meiller C, Shinde J, Soysouvanh F, Calatayud A-L, Pinyol R, Pelletier L, Balabaud C, Laurent A, Blanc J-F, Mazzaferro V, Calvo F, Villanueva A, Nault J-C, Bioulac-Sage P, Stratton MR, Llovet JM, Zucman-Rossi J. 2015. Exome sequencing of hepatocellular carcinomas identifies new mutational signatures and potential therapeutic targets. Nat Genet 47:505–511.

7. Pleasance ED, Stephens PJ, O'Meara S, McBride DJ, Meynert A, Jones D, Lin M-L, Beare D, Lau KW, Greenman C, Varela I, Nik-Zainal S, Davies HR, Ordoñez GR, Mudie LJ, Latimer C, Edkins S, Stebbings L, Chen L, Jia M, Leroy C, Marshall J, Menzies A, Butler A, Teague JW, Mangion J, Sun Y a., McLaughlin SF, Peckham HE, Tsung EF, Costa GL, Lee CC, Minna JD, Gazdar A, Birney E, Rhodes MD, McKernan KJ, Stratton MR, Futreal PA, Campbell PJ. 2010. A small-cell lung cancer genome with complex signatures of tobacco exposure. Nature 463:184–190.
8. Brash DE, Haseltine WA. 1982. UV-induced mutation hotspots occur at DNA damage hotspots. Nature 298:189–92.
9. Wei Q, Lee JE, Gershenwald JE, Ross MI, Mansfield PF, Strom SS, Wang L-E, Guo Z, Qiao Y, Amos CI, Spitz MR, Duvic M. 2003. Repair of UV light-induced DNA damage and risk of cutaneous malignant melanoma. J Natl Cancer Inst 95:308–315.
10. De Gruijl FR, Van Kranen HJ, Mullenders LHF. 2001. UV-induced DNA damage, repair, mutations and oncogenic pathways in skin cancer. J Photochem Photobiol B Biol 63:19–27.
11. Franklin A, Milburn PJ, Blanden R V, Steele EJ. 2004. Human DNA polymerase-eta, an A-T mutator in somatic hypermutation of rearranged immunoglobulin genes, is a reverse transcriptase. Immunol Cell Biol 82:219–25.
12. Wang Y, Woodgate R, McManus TP, Mead S, McCormick JJ, Maher VM. 2007. Evidence that in xeroderma pigmentosum variant cells, which lack

DNA polymerase eta, DNA polymerase iota causes the very high frequency and unique spectrum of UV-induced mutations. Cancer Res 67:3018–26.

13. Hemnani T, Parihar MS. 1998. Reactive oxygen species and oxidative DNA damage. Indian J Physiol Pharmacol 42:440–452.
14. Jena NR. 2012. DNA damage by reactive species: Mechanisms, mutation and repair. J Biosci 37:503–507.
15. Haradhvala NJ, Polak P, Stojanov P, Covington KR, Shinbrot E, Hess JM, Rheinbay E, Kim J, Maruvka YE, Braunstein LZ, Kamburov A, Hanawalt PC, Wheeler DA, Koren A, Lawrence MS, Getz G. 2016. Mutational strand asymmetries in cancer genomes reveal mechanisms of DNA damage and repair. Cell 1–12.
16. Alexandrov LB, Nik-Zainal S, Wedge DC, Campbell PJ, Stratton MR. 2013. Deciphering signatures of mutational processes operative in human cancer. Cell Rep 3:246–259.
17. Spivak G. 2016. Transcription-coupled repair: an update. Arch Toxicol 90:2583–2594.
18. Rayner E, van Gool IC, Palles C, Kearsey SE, Bosse T, Tomlinson I, Church DN. 2016. A panoply of errors: polymerase proofreading domain mutations in cancer. Nat Rev Cancer 16:71–81.
19. Strauss B, Rabkin S, Sagher D, Moore P. 1982. The role of DNA polymerase in base substitution mutagenesis on non-instructional templates. Biochimie 64:829–838.

20. Duffy S, Shackelton LA, Holmes EC. 2008. Rates of evolutionary change in viruses : patterns and determinants 9.
21. Harris RS, Liddament MT. 2004. Retroviral restriction by APOBEC proteins. Nat Rev Immunol 4:868–77.
22. Münk C, Willemsen A, Bravo IG. 2012. An ancient history of gene duplications, fusions and losses in the evolution of APOBEC3 mutators in mammals. BMC Evol Biol 12:71.
23. LaRue RS, Andrésdóttir V, Blanchard Y, Conticello SG, Derse D, Emerman M, Greene WC, Jónsson SR, Landau NR, Löchelt M, Malik HS, Malim MH, Münk C, O'Brien SJ, Pathak VK, Strebel K, Wain-Hobson S, Yu X-F, Yuhki N, Harris RS. 2009. Guidelines for naming nonprimate APOBEC3 genes and proteins. J Virol 83:494–7.
24. Conticello SG. 2008. The AID/APOBEC family of nucleic acid mutators. Genome Biol 9:229.
25. Hultquist JF, Lengyel JA, Refsland EW, LaRue RS, Lackey L, Brown WL, Harris RS. 2011. Human and rhesus APOBEC3D, APOBEC3F, APOBEC3G, and APOBEC3H demonstrate a conserved capacity to restrict Vif-deficient HIV-1. J Virol 85:11220–34.
26. 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 79:285–296.
27. Rogozin IB, Basu MK, Jordan IK, Pavlov YI, Koonin E V. 2005. APOBEC4, a new member of the AID/APOBEC family of polynucleotide

(deoxy)cytidine deaminases predicted by computational analysis. Cell Cycle 4:1281–1285.

28. Liao W, Hong S-H, Chan BH-J, Rudolph FB, Clark SC, 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.

29. Greeve J, Navaratnam N, Scott J. 1991. Characterization of the apolipoprotein B mRNA editing enzyme: no similarity to the proposed mechanism of RNA editing in kinetoplastid protozoa. Nucleic Acids Res 19:3569–76.

30. Teng B, Burant CF, Davidson NO. 1993. Molecular cloning of an apolipoprotein B messenger RNA editing protein. Science 260:1816–9.

31. Navaratnam N, Morrison JR, Bhattacharya S, Patel D, Funahashi T, Giannoni F, Teng BB, Davidson NO, Scott J. 1993. The p27 catalytic subunit of the apolipoprotein B mRNA editing enzyme is a cytidine deaminase. J Biol Chem 268:20709–12.

32. Prochnow C, Bransteitter R, Klein MG, Goodman MF, Chen XS. 2007. The APOBEC-2 crystal structure and functional implications for the deaminase AID 445:447–451.

33. Refsland EW, Stenglein MD, Shindo K, Albin JS, Brown WL, Harris RS. 2010. Quantitative profiling of the full APOBEC3 mRNA repertoire in lymphocytes and tissues: Implications for HIV-1 restriction. Nucleic Acids Res 38:4274–4284.

34. Burns MB, Temiz N a, Harris RS. 2013. Evidence for APOBEC3B mutagenesis in multiple human cancers. Nat Genet 45:977–83.

35. Schumann GG. 2007. APOBEC3 proteins: major players in intracellular defence against LINE-1-mediated retrotransposition. Biochem Soc Trans 35:637–42.

36. Horn A V., Klawitter S, Held U, Berger A, Jaguva Vasudevan AA, Bock A, Hofmann H, Hanschmann KMO, Trösemeier JH, Flory E, Jabulowsky R a., Han JS, Löwer J, Löwer R, Münk C, Schumann GG. 2014. Human LINE-1 restriction by APOBEC3C is deaminase independent and mediated by an ORF1p interaction that affects LINE reverse transcriptase activity. Nucleic Acids Res 42:396–416.

37. Muckenfuss H, Hamdorf M, Held U, Perkovic M, Löwer J, Cichutek K, Flory E, Schumann GG, Münk C. 2006. APOBEC3 proteins inhibit human LINE-1 retrotransposition. J Biol Chem 281:22161–72.

38. Heidmann O. 2005. APOBEC3G cytidine deaminase inhibits retrotransposition of endogenous retroviruses 433:1–4.

39. Wissing S, Montano M, Garcia-Perez JL, Moran J V., Greene WC. 2011. Endogenous APOBEC3B restricts LINE-1 retrotransposition in transformed cells and human embryonic stem cells. J Biol Chem 286:36427–36437.

40. Dutko J a, Schäfer A, Kenny AE, Cullen BR, Curcio MJ. 2005. Inhibition of a yeast LTR retrotransposon by human APOBEC3 cytidine deaminases. Curr Biol 15:661–6.

41. 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.

42. Stenglein MD, Harris RS. 2006. APOBEC3B and APOBEC3F inhibit L1 retrotransposition by a DNA deamination-independent mechanism. J Biol Chem 281:16837–41.

43. Bogerd HP, Wiegand HL, Doehle BP, Lueders KK, Cullen BR. 2006. APOBEC3A and APOBEC3B are potent inhibitors of LTR-retrotransposon function in human cells. Nucleic Acids Res 34:89–95.

44. Seplyarskiy VB, Soldatov RA, Popadin KY, Antonarakis SE, Bazykin GA, Nikolaev SI. 2016. APOBEC-induced mutations in human cancers are strongly enriched on the lagging DNA strand during replication 1–9.

45. Knisbacher BA, Levanon EY. 2015. DNA editing of LTR retrotransposons reveals the impact of APOBECs on vertebrate genomes. Mol Biol Evol 33:msv239.

46. Koning F a, Newman ENC, Kim E-Y, Kunstman KJ, Wolinsky SM, Malim MH. 2009. Defining APOBEC3 expression patterns in human tissues and hematopoietic cell subsets. J Virol 83:9474–85.

47. Lackey L, Law EK, Brown WL, Harris RS. 2013. Subcellular localization of the APOBEC3 proteins during mitosis and implications for genomic DNA deamination 12:762–772.

48. Li MMH, Emerman M. 2011. Polymorphism in human APOBEC3H affects a phenotype dominant for subcellular localization and antiviral activity. J Virol 85:8197–207.

49. Lackey L, Demorest ZL, Land AM, Hultquist JF, Brown WL, Harris RS. 2012. APOBEC3B and AID have similar nuclear import mechanisms. J Mol Biol 419:301–14.

50. Chen J, MacCarthy T, Smith M, Zhang B, Pan X, Ferguson-Smith A. 2017. The preferred nucleotide contexts of the AID/APOBEC cytidine deaminases have differential effects when mutating retrotransposon and virus sequences compared to host genesPLOS Computational Biology.

51. Carpenter MA, Li M, Rathore A, Lackey L, Law EK, Land AM, Leonard B, Shandilya SMD, Bohn M-F, Schiffer CA, Brown WL, Harris RS. 2012. Methylcytosine and normal cytosine deamination by the foreign DNA restriction enzyme APOBEC3A. J Biol Chem 287:34801–34808.

52. Nabel C. 2013. Nucleic acid determinants of cytosine deamination by AID / APOBEC enzymes in immunity and epigenetics.

53. Kohli RM, Maul RW, Guminski AF, McClure RL, Gajula KS, Saribasak H, McMahon MA, Siliciano RF, Gearhart PJ, Stivers JT. 2010. Local sequence targeting in the AID/APOBEC family differentially impacts retroviral restriction and antibody diversification. J Biol Chem 285:40956–64.

54. Prasad R, Poltoratsky V, Hou EW, Wilson SH. 2016. Rev1 is a base excision repair enzyme with 5′-deoxyribose phosphate lyase activity. Nucleic Acids Res gkw869.

55. Chan K, Resnick MA, Gordenin DA. 2013. The choice of nucleotide inserted opposite abasic sites formed within chromosomal DNA reveals the

polymerase activities participating in translesion DNA synthesis. DNA Repair (Amst) 12:878–889.

56. Kim N, Mudrak S V., Jinks-Robertson S. 2011. The dCMP transferase activity of yeast Rev1 is biologically relevant during the bypass of endogenously generated AP sites. DNA Repair (Amst) 10:1262–1271.

57. Doseth B, Ekre C, Slupphaug G, Krokan HE, Kavli B. 2012. Strikingly different properties of uracil-DNA glycosylases UNG2 and SMUG1 may explain divergent roles in processing of genomic uracilDNA Repair.

58. Pulukuri SMK, Knost JA, Estes N, Rao JS. 2009. Small interfering RNA-directed knockdown of uracil DNA glycosylase induces apoptosis and sensitizes human prostate cancer cells to genotoxic stress. Mol Cancer Res 7:1285–1293.

59. Vieira VC, Leonard B, White EA, Starrett GJ, Temiz N a, Lorenz LD, Lee D, Soares MA, Lambert PF, Howley PM, Harris RS, Harris S. 2014. Human papillomavirus E6 triggers upregulation of the antiviral and cancer genomic DNA deaminase APOBEC3B. MBio 5:1–8.

60. Henderson S, Chakravarthy A, Su X, Boshoff C, Fenton TR. 2014. APOBEC-mediated cytosine deamination links PIK3CA helical domain mutations to human papillomavirus-driven tumor development. Cell Rep 7:1833–1841.

61. Papadimitriou JC, Randhawa P, Rinaldo CH, Drachenberg CB, Alexiev B, Hirsch HH. 2016. BK polyomavirus infection and renourinary tumorigenesis. Am J Transplant 16:398–406.

62. Pino L, Rijo E, Nohales G, Frances A, Ubre A, Arango O. 2013. Bladder transitional cell carcinoma and BK virus in a young kidney transplant recipient. Transpl Infect Dis 15:E25-7.

63. Cancer T, Atlas G, Weinstein JN, Akbani R, Broom BM, Wang W, Verhaak RGW, McConkey D, Lerner S, Morgan M, Creighton CJ, Smith C, Cherniack AD, Kim J, Pedamallu CS, Noble MS, Al-Ahmadie HA, Reuter VE, Rosenberg JE, F.Bajorin D, Bochner BH, Solit DB, Koppie T, Robinson B, Gordenin DA, Fargo D, Klimczak LJ, Roberts SA, Au J, Laird PW, Hinoue T, Schultz N, Ramirez R, Hansel D, Hoadley KA, Kim WY, Damrauer JS, Baylin SB, Mungall AJ, Robertson AG, Chu A, Kwiatkowski DJ, Sougnez C, Cibulskis K, Lichtenstein L, Sivachenko A, Stewart C, Lawrence MS, Getz G, Lander E, Gabrie SB, Donehower L, Carter SL, Saksena G, Schumacher SE, Freeman SS, Jung J, Bhatt AS, Pugh T, Beroukhim R, Meyerson M, Ally A, Balasundaram M, Butterfield YSN, Dhalla N, Hirst C, Holt RA, Jones SJM, Lee D, Li HI, Marra MA, Mayo M, Moore RA, Schein JE, Sipahimalani P, Tam A, Thiessen N, Wong T, Wye N, Bowlby R, Chuah E, Guin R, Shen H, Bootwalla MS, Triche Jr T, Lai PH, Van Den Berg DJ, Weisenberger DJ, Balu S, Bodenheimer T, Hoyle AP, Jefferys SR, Meng S, Mose LE, Simons J V, Soloway MG, Wu J, Parker JS, Hayes DN, Roach J, Buda E, Jones CD, Mieczkowski PA, Tan D, Veluvolu U, Waring S, Auman JT, Perou CM, Wilkerson MD, Santoso N, Parfenov M, Ren X, Pantazi A, Hadjipanayis A, Seidman J, Kucherlapati R, Lee S, Yang L, Park PJ, Xu AW, Protopopov A, Zhang J, Bristow C,

Mahadeshwar HS, Seth S, Song X, Tang J, Zeng D, Chin L, Guo C, Casasent TD, Liu W, Ju Z, Motter T, Peng B, Ryan M, Su X, Yang JY, Lorenzi PL, Yao H, Zhang N, Mills GB, Cho J, DiCara D, Frazer S, Gehlenborg N, Heiman DI, Lin P, Liu Y, Stojanov P, Voet D, Zhang H, Zou L, Bernard B, Kreisberg D, Reynolds S, Rovira H, Shmulevich I, Gao J, Jacobsen A, Aksoy BA, Antipin Y, Ciriello G, Dresdner G, Gross B, Lee W, Reva B, Shen R, Sinha R, Sumer SO, Weinhold N, Ladanyi M, Sander C, Benz C, Carlin D, Haussler D, Ng S, Paull E, Stuart J, Zhu J, Zhang W, Taylor BS, Lichtenberg TM, Zmuda E, Barr T, Black AD, George M, Hanf B, Helsel C, McAllister C, Ramirez NC, Tabler TR, Weaver S, Wise L, Bowen J, Gastier-Foster JM, Jian W, Tello S, Ittman M, Castro P, McClenden WD, Gibbs R, Saller C, Tarvin K, DiPiero JM, Owens J, Bollag R, Li Q, Weinberger P, Czerwinski C, Huelsenbeck-Dill L, Iacocca M, Petrelli N, Rabeno B, Swanson P, Shelton T, Curley E, Gardner J, Mallery D, Penny R, Van Bang N, Hanh PT, Kohl B, Van Le X, Phu BD, Thorp R, Tien N V, Vinh LQ, Sandusky G, Burks E, Christ K, Gee J, Holway A, Moinzadeh A, Sorcini A, Sullivan T, Garcia-Grossman IR, Regazzi AM, Boice L, Rathmell WK, Thorne L, Bastacky S, Davies B, Dhir R, Gingrich J, Hrebinko R, Maranchie J, Nelson J, Parwani A, Bshara W, Gaudioso C, Morrison C, Alexopoulou V, Bartlett J, Engel J, Kodeeswaran S, Antic T, O'Donnell PH, Smith ND, Steinberg GD, Egea S, Gomez-Fernandez C, Herbert L, Jorda M, Soloway MG, Beaver A, Carter SL, Kapur P, Lewis C, Lotan Y, Bondaruk J, Czerniak B, Skinner E, Aldape K, Jensen MA, Kahn

AB, Pihl TD, Pot DA, Srinivasan D, Wan Y, Ferguson ML, Zenklusen JC, Davidsen T, Demchok JA, Shaw KRM, Sheth M, Tarnuzzer R, Wang Z, Hutter C, Ozenberger BA, Sofia HJ, Eley G. 2014. Comprehensive molecular characterization of urothelial bladder carcinoma. Nature 507:315–22.

64. Nordentoft I, Lamy P, Birkenkamp-Demtröder K, Shumansky K, Vang S, Hornshøj H, Juul M, Villesen P, Hedegaard J, Roth A, Thorsen K, Høyer S, Borre M, Reinert T, Fristrup N, Dyrskjøt L, Shah S, Pedersen JS, Orntoft TF, Ørntoft TF. 2014. Mutational context and diverse clonal development in early and late bladder cancer. Cell Rep 7:1649–63.

65. Liu S, Chaudhry MR, Berrebi AA, Papadimitriou JC, Drachenberg CB, Haririan A, Alexiev BA. 2016. Polyomavirus replication and smoking are independent risk factors for bladder cancer after renal transplantation. Transplantation 0:1.

66. Hoopes JI, Cortez LM, Mertz TM, Malc EP, Mieczkowski PA, Roberts SA. 2016. APOBEC3A and APOBEC3B preferentially deaminate the lagging strand template during DNA replication. Cell Rep 1–10.

67. Bhagwat AS, Hao W, Townes JP, Lee H, Tang H, Foster PL. 2016. Strand-biased cytosine deamination at the replication fork causes cytosine to thymine mutations in Escherichia coli. Proc Natl Acad Sci U S A 113:2176–81.

68. Green AM, Landry S, Budagyan K, Avgousti DC, Shalhout S, Bhagwat AS, Weitzman MD. 2016. APOBEC3A damages the cellular genome during

DNA replication. Cell Cycle 15:998–1008.

69. Seplyarskiy VB, Andrianova MA, Bazykin GA. 2016. APOBEC3A / B-induced mutagenesis is responsible for 20 % of heritable mutations in the TpCpW context.

70. Yan S, He F, Gao B, Wu H, Li M, Huang L, Liang J. 2016. Increased APOBEC3B predicts worse outcomes in lung cancer: a comprehensive retrospective study. J Cancer 7:618–625.

71. Tsuboi M, Yamane A, Horiguchi J, Yokobori T, Kawabata-Iwakawa R, Yoshiyama S, Rokudai S, Odawara H, Tokiniwa H, Oyama T, Takeyoshi I, Nishiyama M. 2015. APOBEC3B high expression status is associated with aggressive phenotype in Japanese breast cancers. Breast Cancer.

72. Xu L, Chang Y, An H, Zhu Y, Yang Y, Xu J. 2015. High APOBEC3B expression is a predictor of recurrence in patients with low-risk clear cell renal cell carcinoma. Urol Oncol 33:1–8.

73. Pim D, Massimi P, Dilworth SM, Banks L. 2005. Activation of the protein kinase B pathway by the HPV-16 E7 oncoprotein occurs through a mechanism involving interaction with PP2A. Oncogene 24:7830–8.

74. Leonard B, McCann JL, Starrett GJ, Kosyakovsky L, Luengas EM, Molan AM, Burns MB, McDougle RM, Parker PJ, Brown WL, Harris RS. 2015. The PKC/NF-kB signaling pathway induces APOBEC3B expression in multiple human cancers. Cancer Res 75:4538–4547.

75. Lucifora J, Xia Y, Reisinger F, Zhang K, Stadler D, Cheng X, Sprinzl MF, Koppensteiner H, Makowska Z, Volz T, Remouchamps C, Chou W-M,

Thasler WE, Hüser N, Durantel D, Liang TJ, Münk C, Heim MH, Browning JL, Dejardin E, Dandri M, Schindler M, Heikenwalder M, Protzer U. 2014. Specific and nonhepatotoxic degradation of nuclear hepatitis B virus cccDNA. Science 343:1221–8.

76. Siriwardena SU, Guruge TA, Bhagwat AS. 2015. Characterization of the catalytic domain of human APOBEC3B and the critical structural role for a conserved methionine. J Mol Biol 427:3042–55.

77. Shi K, Carpenter M a., Kurahashi K, Harris RS, Aihara H. 2015. Crystal structure of the DNA deaminase APOBEC3B catalytic domain. J Biol Chem 290:28120–28130.

78. Shi K, Carpenter MA, Banerjee S, Shaban NM, Kurahashi K, Salamango DJ, McCann JL, Starrett GJ, Duffy J V, Demir Ö, Amaro RE, Harki DA, Harris RS, Aihara H. 2016. Structural basis for targeted DNA cytosine deamination and mutagenesis by APOBEC3A and APOBEC3B. Nat Struct Mol Biol 24:131–139.

79. Byeon I-JL, Byeon C-H, Wu T, Mitra M, Singer D, Levin JG, Gronenborn AM. 2016. Nuclear magnetic resonance structure of the APOBEC3B catalytic domain: structural basis for substrate binding and DNA deaminase activity. Biochemistry 55:2944–59.

80. Kouno T, Silvas T V, Hilbert BJ, Shandilya SMD, Bohn MF, Kelch BA, Royer WE, Somasundaran M, Kurt Yilmaz N, Matsuo H, Schiffer CA. 2017. Crystal structure of APOBEC3A bound to single-stranded DNA reveals structural basis for cytidine deamination and specificity. Nat Commun

8:15024.

81. Chan K, Roberts S a, Klimczak LJ, Sterling JF, Saini N, Malc EP, Kim J, Kwiatkowski DJ, Fargo DC, Mieczkowski P a, Getz G, Gordenin D a. 2015. An APOBEC3A hypermutation signature is distinguishable from the signature of background mutagenesis by APOBEC3B in human cancers. Nat Genet.

82. Kidd JM, Newman TL, Tuzun E, Kaul R, Eichler EE. 2007. Population stratification of a common APOBEC gene deletion polymorphism. PLoS Genet 3:e63.

83. Xuan D, Li G, Cai Q, Deming-Halverson S, Shrubsole MJ, Shu X-O, Kelley MC, Zheng W, Long J. 2013. APOBEC3 deletion polymorphism is associated with breast cancer risk among women of European ancestry. Carcinogenesis 34:2240–3.

84. Wen WX, Soo JS-S, Kwan PY, Hong E, Khang TF, Mariapun S, Lee CS-M, Hasan SN, Rajadurai P, Yip CH, Mohd Taib NA, Teo SH. 2016. Germline APOBEC3B deletion is associated with breast cancer risk in an Asian multi-ethnic cohort and with immune cell presentation. Breast Cancer Res 18:56.

85. Revathidevi S, Manikandan M, Rao AKDM, Vinothkumar V, Arunkumar G, Rajkumar KS, Ramani R, Rajaraman R, Ajay C, Munirajan AK. 2016. Analysis of APOBEC3A/3B germline deletion polymorphism in breast, cervical and oral cancers from South India and its impact on miRNA regulation. Tumor Biol.

86. Long J, Delahanty RJ, Li G, Gao Y-T, Lu W, Cai Q, Xiang Y-BY-B, Li C, Ji B-T, Zheng Y, Ali S, Shu X-OX-O, Zheng W. 2013. A Common Deletion in the APOBEC3 Genes and Breast Cancer Risk. JNCI J Natl Cancer Inst 105:573–579.

87. Qi G, Xiong H, Zhou C. 2014. APOBEC3 deletion polymorphism is associated with epithelial ovarian cancer risk among Chinese women. Tumour Biol 35:5723–6.

88. Rezaei M, Hashemi M, Hashemi SM, Mashhadi MA, Taheri M. 2015. APOBEC3 deletion is associated with breast cancer risk in a sample of Southeast Iranian population. Int J Mol Cell Med 4:4–9.

89. Caval V, Suspène R, Shapira M, Vartanian J-P, Wain-Hobson S. 2014. A prevalent cancer susceptibility APOBEC3A hybrid allele bearing APOBEC3B 3′UTR enhances chromosomal DNA damage. Nat Commun 5:5129.

90. Starrett GJ, Luengas EM, McCann JL, Ebrahimi D, Temiz NA, Love RP, Feng Y, Adolph MB, Chelico L, Law EK, Carpenter MA, Harris RS. 2016. The DNA cytosine deaminase APOBEC3H haplotype I likely contributes to breast and lung cancer mutagenesis. Nat Commun 7:12918.

91. Ezzikouri S, Kitab B, Rebbani K, Marchio a, Wain-Hobson S, Dejean a, Vartanian J-P, Pineau P, Benjelloun S. 2013. Polymorphic APOBEC3 modulates chronic hepatitis B in Moroccan population. J Viral Hepat 20:678–86.

92. Prasetyo AA, Sariyatun R, Sari Y, Haryati S, Adnan ZA, Kageyama S.

2015. The APOBEC3B deletion polymorphism is associated with prevalence of hepatitis B virus, hepatitis C virus, Torque Teno virus, and Toxoplasma gondii co-infection among HIV-infected individuals. J Clin Virol 70:67–71.

93. He X, Xu H, Wang X, He X, Gao P, Niu J. 2016. Association between APOBEC3B deletion polymorphism and susceptibility to chronic hepatitis B infection and outcomes of hepatocellular carcinoma in Chinese Han population. Int J Clin Exp Pathol 9:9520–9528.

94. Abe H, Ochi H, Maekawa T, Hatakeyama T, Tsuge M, Kitamura S, Kimura T, Miki D, Mitsui F, Hiraga N, Imamura M, Fujimoto Y, Takahashi S, Nakamura Y, Kumada H, Chayama K. 2009. Effects of structural variations of APOBEC3A and APOBEC3B genes in chronic hepatitis B virus infection. Hepatol Res 39:1159–1168.

95. Vartanian JP, Henry M, Marchio A, Suspène R, Aynaud MM, Guétard D, Cervantes-Gonzalez M, Battiston C, Mazzaferro V, Pineau P, Dejean A, Wain-Hobson S. 2010. Massive APOBEC3 editing of hepatitis B viral DNA in cirrhosis. PLoS Pathog 6:1–9.

96. Xu R, Zhang X, Zhang W, Fang Y, Zheng S, Yu XF. 2007. Association of human APOBEC3 cytidine deaminases with the generation of hepatitis virus B x antigen mutants and hepatocellular carcinoma. Hepatology 46:1810–1820.

97. Jha P, Sinha S, Kanchan K, Qidwai T, Narang A, Singh PK, Pati SS, Mohanty S, Mishra SK, Sharma SK, Awasthi S, Venkatesh V, Jain S, Basu

A, Xu S, Mukerji M, Habib S. 2012. Deletion of the APOBEC3B gene strongly impacts susceptibility to falciparum malaria. Infect Genet Evol 12:142–148.

98. Middlebrooks CD, Banday AR, Matsuda K, Udquim K-I, Onabajo OO, Paquin A, Figueroa JD, Zhu B, Koutros S, Kubo M, Shuin T, Freedman ND, Kogevinas M, Malats N, Chanock SJ, Garcia-Closas M, Silverman DT, Rothman N, Prokunina-Olsson L. 2016. Association of germline variants in the APOBEC3 region with cancer risk and enrichment with APOBEC-signature mutations in tumors. Nat Genet.

99. Wittkopp CJ, Adolph MB, Wu LI, Chelico L, Emerman M. 2016. A single nucleotide polymorphism in human APOBEC3C enhances restriction of lentiviruses. PLoS Pathog 12:e1005865.

100. An P, Bleiber G, Duggal P, Nelson G, May M, Mangeat B, Alobwede I, Trono D, Vlahov D, Donfield S, Goedert JJ, Phair J, Buchbinder S, O'Brien SJ, Telenti A, Winkler CA. 2004. APOBEC3G genetic variants and their influence on the progression to AIDS. J Virol 78:11070–6.

101. An P, An P, Bleiber G, Bleiber G, Duggal P, Duggal P, Nelson G, Nelson G, May M, May M, Mangeat B, Mangeat B, Winkler CA, Winkler CA, Alobwede I, Trono D, Vlahov D, Donfield S, Goedert JJ, Phair J, Buchbinder S, O'Brien SJ, Telenti A, Winkler CA. 2004. APOBEC3G genetic variants and their infuence on the progression to AIDS. J Virol 78:11070–11076.

102. Singh KK, Wang Y, Gray KP, Farhad M, Brummel S, Fenton T, Trout R,

Spector SA. 2013. Genetic variants in the host restriction factor APOBEC3G are associated with HIV-1-related disease progression and central nervous system impairment in children. J Acquir Immune Defic Syndr 62:197–203.

103. Auton A, Abecasis GR, Altshuler DM, Durbin RM, Abecasis GR, Bentley DR, Chakravarti A, Clark AG, Donnelly P, Eichler EE, Flicek P, Gabriel SB, Gibbs RA, Green ED, Hurles ME, Knoppers BM, Korbel JO, Lander ES, Lee C, Lehrach H, Mardis ER, Marth GT, McVean GA, Nickerson DA, Schmidt JP, Sherry ST, Wang J, Wilson RK, Gibbs RA, Boerwinkle E, Doddapaneni H, Han Y, Korchina V, Kovar C, Lee S, Muzny D, Reid JG, Zhu Y, Wang J, Chang Y, Feng Q, Fang X, Guo X, Jian M, Jiang H, Jin X, Lan T, Li G, Li J, Li Y, Liu S, Liu X, Lu Y, Ma X, Tang M, Wang B, Wang G, Wu H, Wu R, Xu X, Yin Y, Zhang D, Zhang W, Zhao J, Zhao M, Zheng X, Lander ES, Altshuler DM, Gabriel SB, Gupta N, Gharani N, Toji LH, Gerry NP, Resch AM, Flicek P, Barker J, Clarke L, Gil L, Hunt SE, Kelman G, Kulesha E, Leinonen R, McLaren WM, Radhakrishnan R, Roa A, Smirnov D, Smith RE, Streeter I, Thormann A, Toneva I, Vaughan B, Zheng-Bradley X, Bentley DR, Grocock R, Humphray S, James T, Kingsbury Z, Lehrach H, Sudbrak R, Albrecht MW, Amstislavskiy VS, Borodina TA, Lienhard M, Mertes F, Sultan M, Timmermann B, Yaspo M-L, Mardis ER, Wilson RK, Fulton L, Fulton R, Sherry ST, Ananiev V, Belaia Z, Beloslyudtsev D, Bouk N, Chen C, Church D, Cohen R, Cook C, Garner J, Hefferon T, Kimelman M, Liu C, Lopez J, Meric P, O'Sullivan C, Ostapchuk Y, Phan L,

Ponomarov S, Schneider V, Shekhtman E, Sirotkin K, Slotta D, Zhang H, McVean GA, Durbin RM, Balasubramaniam S, Burton J, Danecek P, Keane TM, Kolb-Kokocinski A, McCarthy S, Stalker J, Quail M, Schmidt JP, Davies CJ, Gollub J, Webster T, Wong B, Zhan Y, Auton A, Campbell CL, Kong Y, Marcketta A, Gibbs RA, Yu F, Antunes L, Bainbridge M, Muzny D, Sabo A, Huang Z, Wang J, Coin LJM, Fang L, Guo X, Jin X, Li G, Li Q, Li Y, Li Z, Lin H, Liu B, Luo R, Shao H, Xie Y, Ye C, Yu C, Zhang F, Zheng H, Zhu H, Alkan C, Dal E, Kahveci F, Marth GT, Garrison EP, Kural D, Lee W-P, Fung Leong W, Stromberg M, Ward AN, Wu J, Zhang M, Daly MJ, DePristo MA, Handsaker RE, Altshuler DM, Banks E, Bhatia G, del Angel G, Gabriel SB, Genovese G, Gupta N, Li H, Kashin S, Lander ES, McCarroll SA, Nemesh JC, Poplin RE, Yoon SC, Lihm J, Makarov V, Clark AG, Gottipati S, Keinan A, Rodriguez-Flores JL, Korbel JO, Rausch T, Fritz MH, Stütz AM, Flicek P, Beal K, Clarke L, Datta A, Herrero J, McLaren WM, Ritchie GRS, Smith RE, Zerbino D, Zheng-Bradley X, Sabeti PC, Shlyakhter I, Schaffner SF, Vitti J, Cooper DN, Ball E V., Stenson PD, Bentley DR, Barnes B, Bauer M, Keira Cheetham R, Cox A, Eberle M, Humphray S, Kahn S, Murray L, Peden J, Shaw R, Kenny EE, Batzer MA, Konkel MK, Walker JA, MacArthur DG, Lek M, Sudbrak R, Amstislavskiy VS, Herwig R, Mardis ER, Ding L, Koboldt DC, Larson D, Ye K, Gravel S, Swaroop A, Chew E, Lappalainen T, Erlich Y, Gymrek M, Frederick Willems T, Simpson JT, Shriver MD, Rosenfeld JA, Bustamante CD, Montgomery SB, De La Vega FM, Byrnes JK, Carroll AW, DeGorter MK,

Lacroute P, Maples BK, Martin AR, Moreno-Estrada A, Shringarpure SS, Zakharia F, Halperin E, Baran Y, Lee C, Cerveira E, Hwang J, Malhotra A, Plewczynski D, Radew K, Romanovitch M, Zhang C, Hyland FCL, Craig DW, Christoforides A, Homer N, Izatt T, Kurdoglu AA, Sinari SA, Squire K, Sherry ST, Xiao C, Sebat J, Antaki D, Gujral M, Noor A, Ye K, Burchard EG, Hernandez RD, Gignoux CR, Haussler D, Katzman SJ, James Kent W, Howie B, Ruiz-Linares A, Dermitzakis ET, Devine SE, Abecasis GR, Min Kang H, Kidd JM, Blackwell T, Caron S, Chen W, Emery S, Fritsche L, Fuchsberger C, Jun G, Li B, Lyons R, Scheller C, Sidore C, Song S, Sliwerska E, Taliun D, Tan A, Welch R, Kate Wing M, Zhan X, Awadalla P, Hodgkinson A, Li Y, Shi X, Quitadamo A, Lunter G, McVean GA, Marchini JL, Myers S, Churchhouse C, Delaneau O, Gupta-Hinch A, Kretzschmar W, Iqbal Z, Mathieson I, Menelaou A, Rimmer A, Xifara DK, Oleksyk TK, Fu Y, Liu X, Xiong M, Jorde L, Witherspoon D, Xing J, Eichler EE, Browning BL, Browning SR, Hormozdiari F, Sudmant PH, Khurana E, Durbin RM, Hurles ME, Tyler-Smith C, Albers CA, Ayub Q, Balasubramaniam S, Chen Y, Colonna V, Danecek P, Jostins L, Keane TM, McCarthy S, Walter K, Xue Y, Gerstein MB, Abyzov A, Balasubramanian S, Chen J, Clarke D, Fu Y, Harmanci AO, Jin M, Lee D, Liu J, Jasmine Mu X, Zhang J, Zhang Y, Li Y, Luo R, Zhu H, Alkan C, Dal E, Kahveci F, Marth GT, Garrison EP, Kural D, Lee W-P, Ward AN, Wu J, Zhang M, McCarroll SA, Handsaker RE, Altshuler DM, Banks E, del Angel G, Genovese G, Hartl C, Li H, Kashin S, Nemesh JC, Shakir K, Yoon SC,

Lihm J, Makarov V, Degenhardt J, Korbel JO, Fritz MH, Meiers S, Raeder B, Rausch T, Stütz AM, Flicek P, Paolo Casale F, Clarke L, Smith RE, Stegle O, Zheng-Bradley X, Bentley DR, Barnes B, Keira Cheetham R, Eberle M, Humphray S, Kahn S, Murray L, Shaw R, Lameijer E-W, Batzer MA, Konkel MK, Walker JA, Ding L, Hall I, Ye K, Lacroute P, Lee C, Cerveira E, Malhotra A, Hwang J, Plewczynski D, Radew K, Romanovitch M, Zhang C, Craig DW, Homer N, Church D, Xiao C, Sebat J, Antaki D, Bafna V, Michaelson J, Ye K, Devine SE, Gardner EJ, Abecasis GR, Kidd JM, Mills RE, Dayama G, Emery S, Jun G, Shi X, Quitadamo A, Lunter G, McVean GA, Chen K, Fan X, Chong Z, Chen T, Witherspoon D, Xing J, Eichler EE, Chaisson MJ, Hormozdiari F, Huddleston J, Malig M, Nelson BJ, Sudmant PH, Parrish NF, Khurana E, Hurles ME, Blackburne B, Lindsay SJ, Ning Z, Walter K, Zhang Y, Gerstein MB, Abyzov A, Chen J, Clarke D, Lam H, Jasmine Mu X, Sisu C, Zhang J, Zhang Y, Gibbs RA, Yu F, Bainbridge M, Challis D, Evani US, Kovar C, Lu J, Muzny D, Nagaswamy U, Reid JG, Sabo A, Yu J, Guo X, Li W, Li Y, Wu R, Marth GT, Garrison EP, Fung Leong W, Ward AN, del Angel G, DePristo MA, Gabriel SB, Gupta N, Hartl C, Poplin RE, Clark AG, Rodriguez-Flores JL, Flicek P, Clarke L, Smith RE, Zheng-Bradley X, MacArthur DG, Mardis ER, Fulton R, Koboldt DC, Gravel S, Bustamante CD, Craig DW, Christoforides A, Homer N, Izatt T, Sherry ST, Xiao C, Dermitzakis ET, Abecasis GR, Min Kang H, McVean GA, Gerstein MB, Balasubramanian S, Habegger L, Yu H, Flicek P, Clarke L, Cunningham F, Dunham I, Zerbino D, Zheng-Bradley

X, Lage K, Berg Jespersen J, Horn H, Montgomery SB, DeGorter MK, Khurana E, Tyler-Smith C, Chen Y, Colonna V, Xue Y, Gerstein MB, Balasubramanian S, Fu Y, Kim D, Auton A, Marcketta A, Desalle R, Narechania A, Wilson Sayres MA, Garrison EP, Handsaker RE, Kashin S, McCarroll SA, Rodriguez-Flores JL, Flicek P, Clarke L, Zheng-Bradley X, Erlich Y, Gymrek M, Frederick Willems T, Bustamante CD, Mendez FL, David Poznik G, Underhill PA, Lee C, Cerveira E, Malhotra A, Romanovitch M, Zhang C, Abecasis GR, Coin L, Shao H, Mittelman D, Tyler-Smith C, Ayub Q, Banerjee R, Cerezo M, Chen Y, Fitzgerald TW, Louzada S, Massaia A, McCarthy S, Ritchie GR, Xue Y, Yang F, Gibbs RA, Kovar C, Kalra D, Hale W, Muzny D, Reid JG, Wang J, Dan X, Guo X, Li G, Li Y, Ye C, Zheng X, Altshuler DM, Flicek P, Clarke L, Zheng-Bradley X, Bentley DR, Cox A, Humphray S, Kahn S, Sudbrak R, Albrecht MW, Lienhard M, Larson D, Craig DW, Izatt T, Kurdoglu AA, Sherry ST, Xiao C, Haussler D, Abecasis GR, McVean GA, Durbin RM, Balasubramaniam S, Keane TM, McCarthy S, Stalker J, Chakravarti A, Knoppers BM, Abecasis GR, Barnes KC, Beiswanger C, Burchard EG, Bustamante CD, Cai H, Cao H, Durbin RM, Gerry NP, Gharani N, Gibbs RA, Gignoux CR, Gravel S, Henn B, Jones D, Jorde L, Kaye JS, Keinan A, Kent A, Kerasidou A, Li Y, Mathias R, McVean GA, Moreno-Estrada A, Ossorio PN, Parker M, Resch AM, Rotimi CN, Royal CD, Sandoval K, Su Y, Sudbrak R, Tian Z, Tishkoff S, Toji LH, Tyler-Smith C, Via M, Wang Y, Yang H, Yang L, Zhu J, Bodmer W, Bedoya G, Ruiz-Linares A, Cai Z, Gao Y, Chu J, Peltonen L, Garcia-

Montero A, Orfao A, Dutil J, Martinez-Cruzado JC, Oleksyk TK, Barnes KC, Mathias RA, Hennis A, Watson H, McKenzie C, Qadri F, LaRocque R, Sabeti PC, Zhu J, Deng X, Sabeti PC, Asogun D, Folarin O, Happi C, Omoniwa O, Stremlau M, Tariyal R, Jallow M, Sisay Joof F, Corrah T, Rockett K, Kwiatkowski D, Kooner J, Tịnh Hiê`n T, Dunstan SJ, Thuy Hang N, Fonnie R, Garry R, Kanneh L, Moses L, Sabeti PC, Schieffelin J, Grant DS, Gallo C, Poletti G, Saleheen D, Rasheed A, Brooks LD, Felsenfeld AL, McEwen JE, Vaydylevich Y, Green ED, Duncanson A, Dunn M, Schloss JA, Wang J, Yang H, Auton A, Brooks LD, Durbin RM, Garrison EP, Min Kang H, Korbel JO, Marchini JL, McCarthy S, McVean GA, Abecasis GR. 2015. A global reference for human genetic variation. Nature 526:68–74.

104. Pace C, Keller J, Nolan D, James I, Gaudieri S, Moore C, Mallal S. 2006. Population level analysis of human immunodeficiency virus type 1 hypermutation and its relationship with APOBEC3G and vif genetic variation. J Virol 80:9259–69.

105. Compaore TR, Soubeiga ST, Ouattara AK, Obiri-Yeboah D, Tchelougou D, Maiga M, Assih M, Bisseye C, Bakouan D, Compaore IP, Dembele A, Martinson J, Simpore J. 2016. APOBEC3G variants and protection against HIV-1 infection in Burkina Faso. PLoS One 11:e0146386.

106. Harari A, Ooms M, Mulder LCF, Simon V. 2009. Polymorphisms and splice variants influence the antiretroviral activity of human APOBEC3H. J Virol 83:295–303.

107. Wang X, Abudu A, Son S, Dang Y, Venta PJ, Zheng Y-H. 2011. Analysis of

human APOBEC3H haplotypes and anti-human immunodeficiency virus type 1 activity. J Virol 85:3142–3152.

108. OhAinle M, Kerns J a., Li MMH, Malik HS, Emerman M. 2008. Antiretroelement activity of APOBEC3H was lost twice in recent human evolution. Cell Host Microbe 4:249–259.

109. Zhen A, Du J, Zhou X, Xiong Y, Yu X-F. 2012. Reduced APOBEC3H variant anti-viral activities are associated with altered RNA binding activities. PLoS One 7:e38771.

110. Zhu M, Wang Y, Wang C, Shen W, Liu J, Geng L, Cheng Y, Dai J, Jin G, Ma H, Hu Z, Shen H. 2015. The eQTL-missense polymorphisms of APOBEC3H are associated with lung cancer risk in a Han Chinese population. Sci Rep 5:14969.

111. Dang Y, Siew LM, Wang X, Han Y, Lampen R, Zheng Y-H. 2008. Human cytidine deaminase APOBEC3H restricts HIV-1 replication. J Biol Chem 283:11606–14.

112. Kim E-Y, Lorenzo-Redondo R, Little SJ, Chung Y-S, Phalora PK, Maljkovic Berry I, Archer J, Penugonda S, Fischer W, Richman DD, Bhattacharya T, Malim MH, Wolinsky SM. 2014. Human APOBEC3 induced mutation of human immunodeficiency virus type-1 contributes to adaptation and evolution in natural infection. PLoS Pathog 10:e1004281.

113. Beral V, Peterman TA, Berkelman RL, Jaffe HW. 1990. Kaposi's sarcoma among persons with AIDS: a sexually transmitted infection? Lancet 335:123–128.

114. Chang Y, Cesarman E, Pessin MS, Lee F, Culpepper J, Knowles DM, Moore PS. 1994. Identification of herpesvirus-like DNA sequences in AIDS-associated Kaposi's sarcoma. Science 266:1865–9.

115. Thel E, Esarman C, Uan Y, Hang C, Oore ASM, Onathan J, Aid WS, Nowles AMK. 1995. Kaposi's sarcoma–associated herpesvirus-like dna sequences in aids-related body-cavity–based lymphomas. N Engl J Med 332:1186–91.

116. Cuevas JM, Geller R, Garijo R, López-Aldeguer J, Sanjuán R. 2015. Extremely High Mutation Rate of HIV-1 In Vivo. PLOS Biol 13:e1002251.

117. Rawson JMO, Landman SR, Reilly CS, Mansky LM. 2015. HIV-1 and HIV-2 exhibit similar mutation frequencies and spectra in the absence of G-to-A hypermutation. Retrovirology 12:60.

118. Koning FA, Goujon C, Bauby H, Malim MH. 2011. Target cell-mediated editing of HIV-1 cDNA by APOBEC3 proteins in human macrophages. J Virol 85:13448–13452.

119. Ebrahimi D, Anwar F, Davenport MP. 2011. APOBEC3 has not left an evolutionary footprint on the HIV-1 genome. J Virol 85:9139–9146.

120. Delviks-Frankenberry KA, Nikolaitchik OA, Burdick RC, Gorelick RJ, Keele BF, Hu W-S, Pathak VK. 2016. Minimal contribution of APOBEC3-induced G-to-A hypermutation to HIV-1 recombination and genetic variation. PLOS Pathog 12:e1005646.

121. Gonçalves DU, Proietti FA, Ribas JGR, Araújo MG, Pinheiro SR, Guedes AC, Carneiro-Proietti ABF. 2010. Epidemiology, treatment, and prevention

of human T-cell leukemia virus type 1-associated diseases. Clin Microbiol Rev 23:577–89.

122. Ooms M, Krikoni A, Kress AK, Simon V, Münk C. 2012. APOBEC3A, APOBEC3B, and APOBEC3H haplotype 2 restrict human T-lymphotropic virus type 1. J Virol 86:6097–108.

123. 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 86:2489–2494.

124. Fan J, Ma G, Nosaka K, Tanabe J, Satou Y, Koito A, Wain-Hobson S, Vartanian J-P, Matsuoka M. 2010. APOBEC3G generates nonsense mutations in human T-cell leukemia virus type 1 proviral genomes in vivo. J Virol 84:7278–87.

125. Derse D, Hill SA, Princler G, Lloyd P, Heidecker G. 2007. Resistance of human T cell leukemia virus type 1 to APOBEC3G restriction is mediated by elements in nucleocapsid. Proc Natl Acad Sci U S A 104:2915–20.

126. Jern P, Coffin JM. 2008. Effects of Retroviruses on Host Genome Function. Annu Rev Genet 42:709–732.

127. Maksakova IA, Romanish MT, Gagnier L, Dunn CA, van de Lagemaat LN, Mager DL. 2006. Retroviral elements and their hosts: insertional mutagenesis in the mouse germ line. PLoS Genet 2:e2.

128. Chen H, Lilley CE, Yu Q, Lee D V., Chou J, Narvaiza I, Landau NR, Weitzman MD. 2006. APOBEC3A is a potent inhibitor of adeno-associated

virus and retrotransposons. Curr Biol 16:480–485.

129. Speek M. 2001. Antisense promoter of human L1 retrotransposon drives transcription of adjacent cellular genes. Mol Cell Biol 21:1973–85.

130. Duggal NK, Fu W, Akey JM, Emerman M. 2013. Identification and antiviral activity of common polymorphisms in the APOBEC3 locus in human populations. Virology 443:329–37.

131. D'Souza G, Dempsey A. 2011. The role of HPV in head and neck cancer and review of the HPV vaccine. Prev Med (Baltim) 53 Suppl 1:S5–S11.

132. zur Hausen H. 2009. Papillomaviruses in the causation of human cancers - a brief historical account. Virology 384:260–265.

133. zur Hausen H. 1996. Papillomavirus infections--a major cause of human cancers. Biochim Biophys Acta 1288:F55-78.

134. Kessis TD, Connolly DC, Hedrick L, Cho KR. 1996. Expression of HPV16 E6 or E7 increases integration of foreign DNA. Oncogene 13:427–31.

135. Barbosa MS, Edmonds C, Fisher C, Schiller JT, Lowy DR, Vousden KH. 1990. The region of the HPV E7 oncoprotein homologous to adenovirus E1a and Sv40 large T antigen contains separate domains for Rb binding and casein kinase II phosphorylation. EMBO J 9:153–60.

136. Crook T, Tidy JA, Vousden KH. 1991. Degradation of p53 can be targeted by HPV E6 sequences distinct from those required. Cell 67:547–556.

137. Mori S, Takeuchi T, Ishii Y, Yugawa T, Kiyono T, Nishina H, Kukimoto I. 2017. Human papillomavirus 16 E6 up-regulates APOBEC3B via the TEAD transcription factor. J Virol JVI.02413-16.

138. Mori S, Takeuchi T, Ishii Y, Kukimoto I. 2015. Identification of APOBEC3B promoter elements responsible for activation by human papillomavirus type 16 E6. Biochem Biophys Res Commun 460:555–560.

139. den Boon J a., Pyeon D, Wang SS, Horswill M, Schiffman M, Sherman M, Zuna RE, Wang Z, Hewitt SM, Pearson R, Schott M, Chung L, He Q, Lambert P, Walker J, Newton M a., Wentzensen N, Ahlquist P. 2015. Molecular transitions from papillomavirus infection to cervical precancer and cancer: Role of stromal estrogen receptor signaling. Proc Natl Acad Sci 112:E3255–E3264.

140. Warren CJ, Xu T, Guo K, Griffin LM, Westrich J a, Lee D, Lambert PF, Santiago ML, Pyeon D. 2014. APOBEC3A functions as a restriction factor of human papillomavirus. J Virol 89:688–702.

141. Ahasan MM, Wakae K, Wang Z, Kitamura K, Liu G, Koura M, Imayasu M, Sakamoto N, Hanaoka K, Nakamura M, Kyo S, Kondo S, Fujiwara H, Yoshizaki T, Mori S, Kukimoto I, Muramatsu M. 2015. APOBEC3A and 3C decrease human papillomavirus 16 pseudovirion infectivity. Biochem Biophys Res Commun 457:295–9.

142. Wang Z, Wakae K, Kitamura K, Aoyama S, Liu G, Koura M, Monjurul a. M, Kukimoto I, Muramatsu M. 2013. APOBEC3 Deaminases induce hypermutation in human papillomavirus 16 DNA upon beta interferon stimulation. J Virol 88:1308–1317.

143. Warren CJ, Van Doorslaer K, Pandey A, Espinosa JM, Pyeon D. 2015. Role of the host restriction factor APOBEC3 on papillomavirus evolution.

Virus Evol 1:vev015.

144. Gosert R, Kardas P, Major EO, Hirsch HH. 2010. Rearranged JC virus noncoding control regions found in progressive multifocal leukoencephalopathy patient samples increase virus early gene expression and replication rate. J Virol 84:10448–10456.

145. Gorelik L, Reid C, Testa M, Brickelmaier M, Bossolasco S, Pazzi A, Bestetti A, Carmillo P, Wilson E, McAuliffe M, Tonkin C, Carulli JP, Lugovskoy A, Lazzarin A, Sunyaev S, Simon K, Cinque P. 2011. Progressive multifocal leukoencephalopathy (PML) development is associated with mutations in JC Virus capsid protein VP1 that change its receptor specificity. J Infect Dis 204:103–114.

146. Padgett BL, Rogers CM. 1977. JC virus , a human polyomavirus associated with progressive multifocal leukoencephalopathy: Additional biological characteristics and antigenic relationships. Microbiology 15:656–662.

147. Saundh BK, Tibble S, Baker R, Sasnauskas K, Harris M, Hale A. 2010. Different patterns of BK and JC polyomavirus reactivation following renal transplantation.

148. Gosert R, Rinaldo CH, Funk GA, Egli A, Ramos E, Drachenberg CB, Hirsch HH. 2008. Polyomavirus BK with rearranged noncoding control region emerge in vivo in renal transplant patients and increase viral replication and cytopathology. J Exp Med 205:841–52.

149. Sullivan CS, Tremblay JD, Fewell SW, Lewis JA, Brodsky JL, Pipas JM.

2000. Species-specific elements in the large T-antigen J domain are required for cellular transformation and DNA replication by simian virus 40. Mol Cell Biol 20:5749–57.

150. Kwun HJ, Guastafierro A, Shuda M, Meinke G, Bohm A, Moore PS, Chang Y. 2009. The minimum replication origin of merkel cell polyomavirus has a unique large T-antigen loading architecture and requires small T-antigen expression for optimal replication. J Virol 83:12118–12128.

151. Gai D, Wang D, Li S-X, Chen XS. 2016. The structure of SV40 large T hexameric helicase in complex with AT-rich origin DNA. Elife 5:1–18.

152. Caracciolo V, Reiss K, Khalili K, De Falco G, Giordano a. 2006. Role of the interaction between large T antigen and Rb family members in the oncogenicity of JC virus. Oncogene 25:5294–5301.

153. Abend JR, Joseph AE, Das D, Campbell-Cecen DB, Imperiale MJ. 2009. A truncated T antigen expressed from an alternatively spliced BK virus early mRNA. J Gen Virol 90:1238–1245.

154. Nguyen KD, Lee EE, Yue Y, Stork J, Pock L, North JP, Vandergriff T, Cockerell C, Hosler GA, Pastrana D V., Buck CB, Wang RC. 2016. Human polyomavirus 6 and 7 are associated with pruritic and dyskeratotic dermatoses. J Am Acad Dermatol 76:932–940.e3.

155. Kassem A, Schöpflin A, Diaz C, Weyers W, Stickeler E, Werner M, Zur Hausen A. 2008. Frequent detection of merkel cell polyomavirus in human merkel cell carcinomas and identification of a unique deletion in the VP1 gene. Cancer Res 68:5009–5013.

156. Bialasiewicz S, Whiley DM, Lambert SB, Nissen MD, Sloots TP. 2009. Detection of BK, JC, WU, or KI polyomaviruses in faecal, urine, blood, cerebrospinal fluid and respiratory samples. J Clin Virol 45:249–254.
157. Feng H, Shuda M, Chang Y, Moore PS. 2008. Clonal integration of a polyomavirus in human Merkel cell carcinoma. Science 319:1096–100.
158. Buck CB, Phan GQ, Raiji MT, Murphy PM, McDermott DH, McBride AA. 2012. Complete genome sequence of a tenth human polyomavirus. J Virol 86:10887.
159. Harms PW, Vats P, Verhaegen ME, Robinson DR, Wu Y-M, Dhanasekaran SM, Palanisamy N, Siddiqui J, Cao X, Su F, Wang R, Xiao H, Kunju LP, Mehra R, Tomlins S a., Fullen DR, Bichakjian CK, Johnson TM, Dlugosz a. a., Chinnaiyan a. M. 2015. The distinctive mutational spectra of polyomavirus-negative Merkel cell carcinoma. Cancer Res 75:3720–3728.
160. Goh G, Walradt T, Markarov V, Blom A, Doumani R, Stafstrom K, Moshiri A, Yelistratova L. 2015. Mutational landscape of MCPyV-positive and MCPyV-negative merkel cell carcinomas with implications for immunotherapy 7.
161. Starrett GJ, Marcelus C, Cantalupo PG, Katz JP, Cheng J, Akagi K, Thakuria M, Rabinowits G, Wang LC, Symer DE, Pipas JM, Harris RS, DeCaprio JA. 2017. Merkel cell polyomavirus exhibits dominant control of the tumor genome and transcriptome in virus-associated merkel cell carcinoma. MBio 8:1–14.
162. Wen Y-M, Wang Y-X. 2009. Biological features of hepatitis B virus isolates

from patients based on full-length genomic analysis. Rev Med Virol 19:57–64.

163. Turelli P, Mangeat B, Jost S, Vianin S, Trono D. 2004. Inhibition of hepatitis B virus replication by APOBEC3G. Science (80- ) 303:1829.

164. Suspène R, Guétard D, Henry M, Sommer P, Wain-Hobson S, Vartanian J-P. 2005. Extensive editing of both hepatitis B virus DNA strands by APOBEC3 cytidine deaminases in vitro and in vivo. Proc Natl Acad Sci U S A 102:8321–6.

165. Deng Y, Du Y, Zhang Q, Han X, Cao G. 2014. Human cytidine deaminases facilitate hepatitis B virus evolution and link inflammation and hepatocellular carcinoma. Cancer Lett 343:161–171.

166. Gao Y, Feng J, Guang Y, Zhang S, Liu Y, Bu Y, Xun M, Zhao M, Chen F, Zhang W, Ye L, Zhang X. 2017. HBx-elevated MSL2 modulates HBV cccDNA through inducing degradation of APOBEC3B to enhance hepatocarcinogenesis. Hepatology 10.1002/hep.29316.

167. Bouzidi MS, Caval V, Suspène R, Hallez C, Pineau P, Wain-Hobson S, Vartanian J-P. 2016. APOBEC3DE Antagonizes Hepatitis B Virus Restriction Factors APOBEC3F and APOBEC3G. J Mol Biol 428:3514–3528.

168. Suspene R, Aynaud M-MM-M, Koch S, Pasdeloup D, Labetoulle M, Gaertner B, Vartanian J-PJ-P, Meyerhans A, Wain-Hobson S, Suspène R, Aynaud M-MM-M, Koch S, Pasdeloup D, Labetoulle M, Gaertner B, Vartanian J-PJ-P, Meyerhans A, Wain-Hobson S. 2011. Genetic editing of

herpes simplex virus 1 and Epstein-Barr herpesvirus genomes by human APOBEC3 cytidine deaminases in culture and in vivo. J Virol 85:7594–602.

169. Minkah N, Chavez K, Shah P, Maccarthy T, Chen H, Landau N, Krug LT. 2014. Host restriction of murine gammaherpesvirus 68 replication by human APOBEC3 cytidine deaminases but not murine APOBEC3. Virology 454–455:215–26.

170. Bekerman E, Jeon D, Ardolino M, Coscoy L. 2013. A role for host activation-induced cytidine deaminase in innate immune defense against KSHV. PLoS Pathog 9:e1003748.

171. Wagener R, Alexandrov LB, Montesinos-Rongen M, Schlesner M, Haake A, Drexler HG, Richter J, Bignell GR, McDermott U, Siebert R. 2015. Analysis of mutational signatures in exomes from B-cell lymphoma cell lines suggest APOBEC3 family members to be involved in the pathogenesis of primary effusion lymphoma. Leukemia 29:1612–1615.

172. Bass AJ, Thorsson V, Shmulevich I, Reynolds SM, Miller M, Bernard B, Hinoue T, Laird PW, Curtis C, Shen H, Weisenberger DJ, Schultz N, Shen R, Weinhold N, Kelsen DP, Bowlby R, Chu A, Kasaian K, Mungall AJ, Gordon Robertson A, Sipahimalani P, Cherniack A, Getz G, Liu Y, Noble MS, Pedamallu C, Sougnez C, Taylor-Weiner A, Akbani R, Lee J-S, Liu W, Mills GB, Yang D, Zhang W, Pantazi A, Parfenov M, Gulley M, Blanca Piazuelo M, Schneider BG, Kim J, Boussioutas A, Sheth M, Demchok JA, Rabkin CS, Willis JE, Ng S, Garman K, Beer DG, Pennathur A, Raphael BJ, Wu H-T, Odze R, Kim HK, Bowen J, Leraas KM, Lichtenberg TM,

Weaver S, McLellan M, Wiznerowicz M, Sakai R, Getz G, Sougnez C, Lawrence MS, Cibulskis K, Lichtenstein L, Fisher S, Gabriel SB, Lander ES, Ding L, Niu B, Ally A, Balasundaram M, Birol I, Bowlby R, Brooks D, Butterfield YSN, Carlsen R, Chu A, Chu J, Chuah E, Chun H-JE, Clarke A, Dhalla N, Guin R, Holt RA, Jones SJM, Kasaian K, Lee D, Li HA, Lim E, Ma Y, Marra MA, Mayo M, Moore RA, Mungall AJ, Mungall KL, Ming Nip K, Gordon Robertson A, Schein JE, Sipahimalani P, Tam A, Thiessen N, Beroukhim R, Carter SL, Cherniack AD, Cho J, Cibulskis K, DiCara D, Frazer S, Fisher S, Gabriel SB, Gehlenborg N, Heiman DI, Jung J, Kim J, Lander ES, Lawrence MS, Lichtenstein L, Lin P, Meyerson M, Ojesina AI, Sekhar Pedamallu C, Saksena G, Schumacher SE, Sougnez C, Stojanov P, Tabak B, Taylor-Weiner A, Voet D, Rosenberg M, Zack TI, Zhang H, Zou L, Protopopov A, Santoso N, Parfenov M, Lee S, Zhang J, Mahadeshwar HS, Tang J, Ren X, Seth S, Yang L, Xu AW, Song X, Pantazi A, Xi R, Bristow CA, Hadjipanayis A, Seidman J, Chin L, Park PJ, Kucherlapati R, Akbani R, Ling S, Liu W, Rao A, Weinstein JN, Kim S-B, Lee J-S, Lu Y, Mills G, Laird PW, Hinoue T, Weisenberger DJ, Bootwalla MS, Lai PH, Shen H, Triche Jr T, Van Den Berg DJ, Baylin SB, Herman JG, Getz G, Chin L, Liu Y, Murray BA, Noble MS, Arman Askoy r B, Ciriello G, Dresdner G, Gao J, Gross B, Jacobsen A, Lee W, Ramirez R, Sander C, Schultz N, Senbabaoglu Y, Sinha R, Onur Sumer S, Sun Y, Weinhold N, Thorsson V, Bernard B, Iype L, Kramer RW, Kreisberg R, Miller M, Reynolds SM, Rovira H, Tasman N, Shmulevich I, Ng S, Haussler

D, Stuart JM, Akbani R, Ling S, Liu W, Rao A, Weinstein JN, Verhaak RGW, Mills GB, Leiserson MDM, Raphael BJ, Wu H-T, Taylor BS, Black AD, Bowen J, Ann Carney J, Gastier-Foster JM, Helsel C, Leraas KM, Lichtenberg TM, McAllister C, Ramirez NC, Tabler TR, Wise L, Zmuda E, Penny R, Crain D, Gardner J, Lau K, Curely E, Mallery D, Morris S, Paulauskis J, Shelton T, Shelton C, Sherman M, Benz C, Lee J-H, Fedosenko K, Manikhas G, Potapova O, Voronina O, Belyaev D, Dolzhansky O, Kimryn Rathmell W, Brzezinski J, Ibbs M, Korski K, Kycler W, Łaźniak R, Leporowska E, Mackiewicz A, Murawa D, Murawa P, Spychała A, Suchorska WM, Tatka H, Teresiak M, Wiznerowicz M, Abdel-Misih R, Bennett J, Brown J, Iacocca M, Rabeno B, Kwon S-Y, Penny R, Gardner J, Kemkes A, Mallery D, Morris S, Shelton T, Shelton C, Curley E, Alexopoulou I, Engel J, Bartlett J, Albert M, Park D-Y, Dhir R, Luketich J, Landreneau R, Janjigian YY, Kelsen DP, Cho E, Ladanyi M, Tang L, McCall SJ, Park YS, Cheong J-H, Ajani J, Constanza Camargo M, Alonso S, Ayala B, Jensen MA, Pihl T, Raman R, Walton J, Wan Y, Demchok JA, Eley G, Mills Shaw KR, Sheth M, Tarnuzzer R, Wang Z, Yang L, Claude Zenklusen J, Davidsen T, Hutter CM, Sofia HJ, Burton R, Chudamani S, Liu J. 2014. Comprehensive molecular characterization of gastric adenocarcinoma. Nature 513:202–209.

173. Zhang J, Wei W, Jin H, Ying R, Zhu A, Zhang F. 2015. The roles of APOBEC3B in gastric cancer. Int J Clin Exp Pathol 8:5089–5096.

174. Zheng H, Dai W, Cheung AKL, Ko JMY, Kan R, Wong BWY, Leong MML,

Deng M, Kwok TCT, Chan JY-W, Kwong DL-W, Lee AW-M, Ng WT, Ngan RKC, Yau CC, Tung S, Lee VH-F, Lam K-O, Kwan CK, Li WS, Yau S, Chan K-W, Lung ML. 2016. Whole-exome sequencing identifies multiple loss-of-function mutations of NF-κB pathway regulators in nasopharyngeal carcinoma. Proc Natl Acad Sci U S A 113:11283–11288.

175. Kuong KJ, Loeb LA. 2013. APOBEC3B mutagenesis in cancer. Nat Genet 45:964–5.

176. Helleday T, Eshtad S, Nik-Zainal S. 2014. Mechanisms underlying mutational signatures in human cancers. Nat Rev Genet 1–14.

177. McGranahan N, Swanton C. 2015. Biological and Therapeutic Impact of Intratumor Heterogeneity in Cancer Evolution. Cancer Cell 27:15–26.

178. Roberts SA, Gordenin DA. 2014. Hypermutation in human cancer genomes: footprints and mechanisms. Nat Rev Cancer 14:786–800.

179. Nik-Zainal S, Alexandrov LB, Wedge DC, Van Loo P, Greenman CD, 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 KW, Mudie LJ, Varela I, McBride DJ, Bignell GR, Cooke SL, Shlien A, Gamble J, Whitmore I, Maddison M, Tarpey PS, Davies HR, Papaemmanuil E, Stephens PJ, McLaren S, Butler AP, Teague JW, Jönsson G, Garber JE, Silver D, Miron P, Fatima A, Boyault S, Langerød A, Tutt A, Martens JWM, Aparicio S a JR, Borg Å, Salomon AV, Thomas G, Børresen-Dale A-L, Richardson AL, Neuberger MS, Futreal PA, Campbell PJ, Stratton MR. 2012. Mutational processes molding the genomes of 21 breast cancers.

Cell 149:979–93.

180. de Bruin EC, McGranahan N, Mitter R, Salm M, Wedge DC, Yates L, Jamal-Hanjani M, Shafi S, Murugaesu N, Rowan AJ, Gronroos E, Muhammad MA, Horswell S, Gerlinger M, Varela I, Jones D, Marshall J, Voet T, Van Loo P, Rassl DM, Rintoul RC, Janes SM, Lee S-M, Forster M, Ahmad T, Lawrence D, Falzon M, Capitanio A, Harkins TT, Lee CC, Tom W, Teefe E, Chen S-C, Begum S, Rabinowitz A, Phillimore B, Spencer-Dene B, Stamp G, Szallasi Z, Matthews N, Stewart A, Campbell P, Swanton C. 2014. Spatial and temporal diversity in genomic instability processes defines lung cancer evolution. Science (80- ) 346:251–256.

181. Nik-Zainal S, Davies H, Staaf J, Ramakrishna M, Glodzik D, Zou X, Martincorena I, Alexandrov LB, Martin S, Wedge DC, Van Loo P, Ju YS, Smid M, Brinkman AB, Morganella S, Aure MR, Lingjærde OC, Langerød A, Ringnér M, Ahn S-M, Boyault S, Brock JE, Broeks A, Butler A, Desmedt C, Dirix L, Dronov S, Fatima A, Foekens JA, Gerstung M, Hooijer GKJ, Jang SJ, Jones DR, Kim H-Y, King TA, Krishnamurthy S, Lee HJ, Lee J-Y, Li Y, McLaren S, Menzies A, Mustonen V, O'Meara S, Pauporté I, Pivot X, Purdie CA, Raine K, Ramakrishnan K, Rodríguez-González FG, Romieu G, Sieuwerts AM, Simpson PT, Shepherd R, Stebbings L, Stefansson OA, Teague J, Tommasi S, Treilleux I, Van den Eynden GG, Vermeulen P, Vincent-Salomon A, Yates L, Caldas C, Veer L van't, Tutt A, Knappskog S, Tan BKT, Jonkers J, Borg Å, Ueno NT, Sotiriou C, Viari A, Futreal PA, Campbell PJ, Span PN, Van Laere S, Lakhani SR, Eyfjord JE, Thompson

AM, Birney E, Stunnenberg HG, van de Vijver MJ, Martens JWM, Børresen-Dale A-L, Richardson AL, Kong G, Thomas G, Stratton MR. 2016. Landscape of somatic mutations in 560 breast cancer whole-genome sequences. Nature 1–20.

182. Morganella S, Alexandrov LB, Glodzik D, Zou X, Davies H, Staaf J, Sieuwerts AM, Brinkman AB, Martin S, Ramakrishna M, Butler A, Kim H-Y, Borg Å, Sotiriou C, Futreal PA, Campbell PJ, Span PN, Van Laere S, Lakhani SR, Eyfjord JE, Thompson AM, Stunnenberg HG, van de Vijver MJ, Martens JWM, Børresen-Dale A-L, Richardson AL, Kong G, Thomas G, Sale J, Rada C, Stratton MR, Birney E, Nik-Zainal S. 2016. The topography of mutational processes in breast cancer genomes. Nat Commun 7:11383.

183. Harris RS, Dudley JP. 2015. APOBECs and virus restriction. Virology 479–480:1–15.

184. Simon V, Bloch N, Landau NR. 2015. Intrinsic host restrictions to HIV-1 and mechanisms of viral escape. Nat Immunol 16:546–553.

185. Landry S, Narvaiza I, Linfesty DC, Weitzman MD. 2011. APOBEC3A can activate the DNA damage response and cause cell-cycle arrest. EMBO Rep 12:444–50.

186. 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.

187. Caval V, Bouzidi MS, Suspène R, Laude H, Dumargne M-C, Bashamboo A, Krey T, Vartanian J-P, Wain-Hobson S. 2015. Molecular basis of the attenuated phenotype of human APOBEC3B DNA mutator enzyme. Nucleic Acids Res 43:9340–9.

188. Shee C, Cox BD, Gu F, Luengas EM, Joshi MC, Chiu L-Y, Magnan D, Halliday JA, Frisch RL, Gibson JL, Nehring RB, Do HG, Hernandez M, Li L, Herman C, Hastings PJ, Bates D, Harris RS, Miller KM, Rosenberg SM. 2013. Engineered proteins detect spontaneous DNA breakage in human and bacterial cells. Elife 2:e01222.

189. Stenglein MD, Burns MB, Li M, Lengyel J, Harris RS. 2010. APOBEC3 proteins mediate the clearance of foreign DNA from human cells. Nat Struct Mol Biol 17:222–9.

190. Land AM, Law EK, Carpenter M a, Lackey L, Brown WL, Harris RS. 2013. Endogenous APOBEC3A DNA cytosine deaminase is cytoplasmic and nongenotoxic. J Biol Chem 288:17253–60.

191. Thielen BK, McNevin JP, McElrath MJ, Hunt BVS, Klein KC, Lingappa JR. 2010. Innate immune signaling induces high levels of TC-specific deaminase activity in primary monocyte-derived cells through expression of APOBEC3A isoforms. J Biol Chem 285:27753–27766.

192. Verhalen B, Starrett GJ, Harris RS, Jiang M. 2016. Functional Upregulation of the DNA Cytosine Deaminase APOBEC3B by Polyomaviruses. J Virol 90:6379–6386.

193. Bogerd HP, Wiegand HL, Hulme AE, Garcia-Perez JL, O'Shea KS, Moran J V, Cullen BR. 2006. Cellular inhibitors of long interspersed element 1 and Alu retrotransposition. Proc Natl Acad Sci U S A 103:8780–5.

194. Leonard B, Hart SN, Burns MB, Carpenter M a., Temiz N a., Rathore A, Vogel RI, Nikas JB, Law EK, Brown WL, Li Y, Zhang Y, Maurer MJ, Oberg a. L, Cunningham JM, Shridhar V, Bell D a., April C, Bentley D, Bibikova M, Cheetham RK, Fan J-B, Grocock R, Humphray S, Kingsbury Z, Peden J, Chien J, Swisher EM, Hartmann LC, Kalli KR, Goode EL, Sicotte H, Kaufmann SH, Harris RS. 2013. APOBEC3B upregulation and genomic mutation patterns in serous ovarian carcinoma. Cancer Res 73:7222–7231.

195. Cescon DW, Haibe-Kains B, Mak TW. 2015. APOBEC3B expression in breast cancer reflects cellular proliferation, while a deletion polymorphism is associated with immune activation. Proc Natl Acad Sci 201424869.

196. Sieuwerts AM, Willis S, Burns MB, Look MP, Gelder MEM Van, Schlicker A, Heideman MR, Jacobs H, Wessels L, Leyland-Jones B, Gray KP, Foekens J a., Harris RS, Martens JWM. 2014. Elevated APOBEC3B correlates with poor outcomes for estrogen-receptor-positive breast cancers. Horm Cancer 5:405–413.

197. Periyasamy M, Patel H, Lai C-F, Nguyen VTM, Nevedomskaya E, Harrod A, Russell R, Remenyi J, Ochocka AM, Thomas RS, Fuller-Pace F, Győrffy B, Caldas C, Navaratnam N, Carroll JS, Zwart W, Coombes RC, Magnani L, Buluwela L, Ali S. 2015. APOBEC3B-mediated cytidine deamination is

required for estrogen receptor action in breast cancer. Cell Rep 13:1–14.

198. Walker B a., Wardell CP, Murison A, Boyle EM, Begum DB, Dahir NM, Proszek PZ, Melchor L, Pawlyn C, Kaiser MF, Johnson DC, Qiang Y-W, Jones JR, Cairns D a., Gregory WM, Owen RG, Cook G, Drayson MT, Jackson GH, Davies FE, Morgan GJ. 2015. APOBEC family mutational signatures are associated with poor prognosis translocations in multiple myeloma. Nat Commun 6:6997.

199. Burns MB, Leonard B, Harris RS. 2015. APOBEC3B: Pathological consequences of an innate immune DNA mutator. Biomed J 38:102.

200. Harris RS. 2015. Molecular mechanism and clinical impact of APOBEC3B-catalyzed mutagenesis in breast cancer. Breast Cancer Res 17:8.

201. Swanton C, McGranahan N, Starrett GJ, Harris RS. 2015. APOBEC enzymes: Mutagenic fuel for cancer evolution and heterogeneity. Cancer Discov 5:704–12.

202. Nik-Zainal S, Wedge DC, Alexandrov LB, Petljak M, Butler AP, Bolli N, Davies HR, Knappskog S, Martin S, Papaemmanuil E, Ramakrishna M, Shlien A, Simonic I, Xue Y, Tyler-Smith C, Campbell PJ, Stratton MR. 2014. Association of a germline copy number polymorphism of APOBEC3A and APOBEC3B with burden of putative APOBEC-dependent mutations in breast cancer. Nat Genet 46:487–91.

203. Refsland EW, Hultquist JF, Luengas EM, Ikeda T, Shaban NM, Law EK, Brown WL, Reilly C, Emerman M, Harris RS. 2014. Natural polymorphisms in human APOBEC3H and HIV-1 Vif combine in primary T lymphocytes to

affect viral G-to-A mutation levels and infectivity. PLoS Genet 10:e1004761.

204. Ooms M, Majdak S, Seibert CW, Harari A, Simon V. 2010. The localization of APOBEC3H variants in HIV-1 virions determines their antiviral activity. J Virol 84:7961–9.

205. Harris RS, Bishop KN, Sheehy AM, Craig HM, Petersen-mahrt SK, Watt IN, Neuberger MS, Malim MH. 2003. DNA Deamination Mediates Innate Immunity to Retroviral Infection. Cell 113:803–809.

206. Yu Q, K?nig R, Pillai S, Chiles K, Kearney M, Palmer S, Richman D, Coffin JM, Landau NR. 2004. Single-strand specificity of APOBEC3G accounts for minus-strand deamination of the HIV genome. Nat Struct Mol Biol 11:435–442.

207. Chan K, Roberts SA, Klimczak LJ, Sterling JF, Saini N, Malc EP, Kim J, Kwiatkowski DJ, Fargo DC, Mieczkowski PA, Getz G, Gordenin DA. 2015. An APOBEC3A hypermutation signature is distinguishable from the signature of background mutagenesis by APOBEC3B in human cancers. Nat Genet 47:1067–1072.

208. Lawrence MS, Stojanov P, Polak P, Kryukov G V, Cibulskis K, Sivachenko A, Carter SL, Stewart C, Mermel CH, Roberts SA, Kiezun A, Hammerman PS, McKenna A, Drier Y, Zou L, Ramos AH, Pugh TJ, Stransky N, Helman E, Kim J, Sougnez C, Ambrogio L, Nickerson E, Shefler E, Cortés ML, Auclair D, Saksena G, Voet D, Noble M, DiCara D, Lin P, Lichtenstein L, Heiman DI, Fennell T, Imielinski M, Hernandez B, Hodis E, Baca S, Dulak

AM, Lohr J, Landau D-A, Wu CJ, Melendez-Zajgla J, Hidalgo-Miranda A, Koren A, McCarroll SA, Mora J, Lee RS, Crompton B, Onofrio R, Parkin M, Winckler W, Ardlie K, Gabriel SB, Roberts CWM, Biegel JA, Stegmaier K, Bass AJ, Garraway LA, Meyerson M, Golub TR, Gordenin DA, Sunyaev S, Lander ES, Getz G. 2013. Mutational heterogeneity in cancer and the search for new cancer-associated genes. Nature 499:214–8.

209. Rathore A, Carpenter MA, Demir Ö, Ikeda T, Li M, Shaban NM, Law EK, Anokhin D, Brown WL, Amaro RE, Harris RS. 2013. The local dinucleotide preference of APOBEC3G can be altered from 5'-CC to 5'-TC by a single amino acid substitution. J Mol Biol 425:4442–54.
210. Eftedal I, Guddal PH, Slupphaug G, Volden G, Krokan HE. 1993. Consensus sequences for good and poor removal of uracil from double stranded DNA by uracil-DNA glycosylase. Nucleic Acids Res 21:2095–101.
211. Makridakis NM, Reichardt JK V. 2012. Translesion DNA Polymerases and Cancer. Front Genet 3:174.
212. Nabel CS, Jia H, Ye Y, Shen L, Goldschmidt HL, Stivers JT, Zhang Y, Kohli RM. 2012. AID/APOBEC deaminases disfavor modified cytosines implicated in DNA demethylation. Nat Chem Biol 8:751–8.
213. Ehrlich M, Norris KF, Wang RY, Kuo KC, Gehrke CW. 1986. DNA cytosine methylation and heat-induced deamination. Biosci Rep 6.
214. Leonard B, Starrett GJ, Maurer MJ, Oberg A, Van Bockstal M, Van Dorpe J, De Wever O, Helleman J, Sieuwerts AM, Berns EMJJ, Martens JWM, Anderson B, Brown WL, Kalli KR, Kaufmann SH, Harris RS. 2016.

APOBEC3G expression correlates with T cell infiltration and improved clinical outcomes in high-grade serous ovarian carcinoma. Clin Cancer Res 1078-0432.CCR-15-2910-.

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

216. Yoshihara K, Shahmoradgoli M, Martínez E, Vegesna R, Kim H, Torres-Garcia W, Treviño V, Shen H, Laird PW, Levine D a, Carter SL, Getz G, Stemke-Hale K, Mills GB, Verhaak RGW. 2013. Inferring tumour purity and stromal and immune cell admixture from expression data. Nat Commun 4:2612.

217. Larue RS, Lengyel J, Jónsson SR, Andrésdóttir V, Harris RS. 2010. Lentiviral Vif degrades the APOBEC3Z3/APOBEC3H protein of its mammalian host and is capable of cross-species activity. J Virol 84:8193–201.

218. Chelico L, Prochnow C, Erie DA, Chen XS, Goodman MF. 2010. Structural model for deoxycytidine deamination mechanisms of the HIV-1 inactivation enzyme APOBEC3G. J Biol Chem 285:16195–205.

219. Chelico L, Pham P, Calabrese P, Goodman MF. 2006. APOBEC3G DNA deaminase acts processively 3′ → 5′ on single-stranded DNA. Nat Struct Mol Biol 13:392–399.

220. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. 2009. The Sequence Alignment/Map format and

SAMtools. Bioinformatics 25:2078–9.

221. Doerries K. 2006. Human polyomavirus JC and BK persistent infection. Adv Exp Med Biol. Springer New York, New York, NY.

222. DeCaprio JA, Garcea RL. 2013. A cornucopia of human polyomaviruses. Nat Rev Microbiol 11:264–76.

223. Bennett SM, Broekema NM, Imperiale MJ. 2012. BK polyomavirus: Emerging pathogen. Microbes Infect 14:672–683.

224. Wollebo HS, White MK, Gordon J, Berger JR, Khalili K. 2015. Persistence and pathogenesis of the neurotropic polyomavirus JC. Ann Neurol 77:560–70.

225. Chang Y, Moore PS. 2012. Merkel cell carcinoma: a virus-induced human cancer. Annu Rev Pathol 7:123–44.

226. Cheng J, Rozenblatt-Rosen O, Paulson KG, Nghiem P, DeCaprio JA. 2013. Merkel cell polyomavirus large T antigen has growth-promoting and inhibitory activities. J Virol 87:6118–26.

227. Giacobbi NS, Gupta T, Coxon AT, Pipas JM. 2015. Polyomavirus T antigens activate an antiviral state. Virology 476:377–385.

228. Abend JR, Low JA, Imperiale MJ. 2010. Global effects of BKV infection on gene expression in human primary kidney epithelial cells. Virology 397:73–9.

229. Yogo Y, Sugimoto C, Zhong S, Homma Y. 2009. Evolution of the BK polyomavirus: epidemiological, anthropological and clinical implications. Rev Med Virol 19:185–199.

230. Gjoerup O, Chang Y. 2010. Update on human polyomaviruses and cancer. Adv Cancer Res 106:1–51.

231. Wendzicki JA, Moore PS, Chang Y. 2015. Large T and small T antigens of Merkel cell polyomavirus. Curr Opin Virol 11:38–43.

232. Vieira VC, Soares M a. 2013. The role of cytidine deaminases on innate immune responses against human viral infections. Biomed Res Int 2013.

233. Shinohara M, Io K, Shindo K, Matsui M, Sakamoto T, Tada K, Kobayashi M, Kadowaki N, Takaori-Kondo A. 2012. APOBEC3B can impair genomic stability by inducing base substitutions in genomic DNA in human cells. Sci Rep 2:806.

234. Jiang M, Zhao L, Gamez M, Imperiale MJ. 2012. Roles of ATM and ATR-mediated DNA damage responses during lytic BK polyomavirus infection. PLoS Pathog 8.

235. Low J, Humes HD, Szczypka M, Imperiale M. 2004. BKV and SV40 infection of human kidney tubular epithelial cells in vitro. Virology 323:182–188.

236. Jiang M, Abend JR, Tsai B, Imperiale MJ. 2009. Early events during BK virus entry and disassembly. J Virol 83:1350–8.

237. Verhalen B, Justice JL, Imperiale MJ, Jiang M. 2015. Viral DNA replication-dependent DNA damage response activation during BK polyomavirus infection. J Virol 89:JVI.03650-14.

238. Bollag B, Mackeen PC, Frisque RJ. 1996. Purified JC virus T antigen derived from insect cells preferentially interacts with binding site II of the

viral core origin under replication conditions. Virology 218:81–93.

239. Jiang M, Entezami P, Gamez M. 2011. Functional Reorganization of Promyelocytic Leukemia Nuclear Bodies during BK Virus Infection. MBio 2:e00281-10.

240. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Söding J, Thompson JD, Higgins DG. 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol 7:539.

241. Moens U, Rasheed K, Abdulsalam I, Sveinbjørnsson B. 2015. The role of Merkel cell polyomavirus and other human polyomaviruses in emerging hallmarks of cancer. Viruses 7:1871–901.

242. Hoelzer K, Shackelton LA, Parrish CR. 2008. Presence and role of cytosine methylation in DNA viruses of animals. Nucleic Acids Res 36:2825–37.

243. Monajemi M, Woodworth CF, Benkaroun J, Grant M, Larijani M. 2012. Emerging complexities of APOBEC3G action on immunity and viral fitness during HIV infection and treatment. Retrovirology 9:35.

244. Haché G, Mansky LM, Harris RS. Human APOBEC3 proteins, retrovirus restriction, and HIV drug resistance. AIDS Rev 8:148–57.

245. Heath M, Jaimes N, Lemos B, Mostaghimi A, Wang LC, Peñas PF, Nghiem P. 2008. Clinical characteristics of Merkel cell carcinoma at diagnosis in 195 patients: the AEIOU features. J Am Acad Dermatol 58:375–81.

246. Houben R, Shuda M, Weinkam R, Schrama D, Feng H, Chang Y, Moore

PS, Becker JC. 2010. Merkel cell polyomavirus-infected Merkel cell carcinoma cells require expression of viral T antigens. J Virol 84:7064–7072.

247. Andres C, Belloni B, Puchta U, Sander C a, Flaig MJ. 2010. Prevalence of MCPyV in Merkel cell carcinoma and non-MCC tumors. J Cutan Pathol 37:28–34.

248. Shuda M, Feng H, Kwun HJ, Rosen ST, Gjoerup O, Moore PS, Chang Y. 2008. T antigen mutations are a human tumor-specific signature for Merkel cell polyomavirus. Proc Natl Acad Sci U S A 105:16272–16277.

249. Li J, Wang X, Diaz J, Tsang SH, Buck CB, You J. 2013. Merkel cell polyomavirus large T antigen disrupts host genomic integrity and inhibits cellular proliferation. J Virol 87:9173–88.

250. Ohba K, Ichiyama K, Yajima M, Gemma N, Nikaido M, Wu Q, Chong P, Mori S, Yamamoto R, Wong JEL, Yamamoto N. 2014. In vivo and in vitro studies suggest a possible involvement of HPV infection in the early stage of breast carcinogenesis via APOBEC3B induction. PLoS One 9:e97787.

251. Wong SQ, Waldeck K, Vergara IA, Schroder J, Madore J, Wilmott JS, Colebatch AJ, De Paoli-Iseppi R, Li J, Lupat R, Semple T, Arnau GM, Fellowes A, Leonard JH, Hruby G, Mann GJ, Thompson JF, Cullinane C, Johnston M, Shackleton M, Sandhu S, Bowtell DDL, Johnstone RW, Fox SB, McArthur GA, Papenfuss AT, Scolyer RA, Gill AJ, Hicks RJ, Tothill RW. 2015. UV-Associated mutations underlie the etiology of MCV-negative Merkel cell carcinomas. Cancer Res 75:5228–5234.

252. Cohen PR, Tomson BN, Elkin SK, Marchlik E, Jennifer L, Kurzrock R. 2016. Genomic portfolio of Merkel cell carcinoma as determined by comprehensive genomic profiling: implications for targeted therapeutics. Oncotarget 7.

253. Alexandrov LB, Stratton MR. 2014. Mutational signatures: The patterns of somatic mutations hidden in cancer genomes. Curr Opin Genet Dev 24:52–60.

254. Jayaraman SS, Rayhan DJ, Hazany S, Kolodney MS. 2014. Mutational landscape of basal cell carcinomas by whole-exome sequencing. J Invest Dermatol 134:213–20.

255. Boukamp P. 2005. Non-melanoma skin cancer: what drives tumor development and progression? Carcinogenesis 26:1657–1667.

256. Rass K, Reichrath J. 2008. UV damage and DNA repair in malignant melanoma and nonmelanoma skin cancer. Adv Exp Med Biol 624:162–178.

257. Harms PW, Patel RM, Verhaegen ME, Giordano TJ, Nash KT, Johnson CN, Daignault S, Thomas DG, Gudjonsson JE, Elder JT, Dlugosz A a, Johnson TM, Fullen DR, Bichakjian CK. 2013. Distinct gene expression profiles of viral- and nonviral-associated merkel cell carcinoma revealed by transcriptome analysis. J Invest Dermatol 133:936–45.

258. Pardo L a, Stühmer W. 2014. The roles of K(+) channels in cancer. Nat Rev Cancer 14:39–48.

259. Schuller HM, Al-Wadei HAN, Majidi M. 2008. GABAB receptor is a novel

drug target for pancreatic cancer. Cancer 112:767–778.

260. Miao Y, Zhang Y, Wan H, Chen L, Wang F. 2010. GABA-receptor agonist, propofol inhibits invasion of colon carcinoma cells. Biomed Pharmacother 64:583–588.

261. Akagi K, Li J, Broutian TR, Padilla-Nash H, Xiao W, Jiang B, Rocco JW, Teknos TN, Kumar B, Wangsa D, He D, Ried T, Symer DE, Gillison ML. 2014. Genome-wide analysis of HPV integration in human cancers reveals recurrent, focal genomic instability. Genome Res 24:185–99.

262. Katz JP, Pipas JM. 2014. SummonChimera infers integrated viral genomes with nucleotide precision from NGS data 15:1–6.

263. Agelli M, Clegg LX. 2003. Epidemiology of primary Merkel cell carcinoma in the United States. J Am Acad Dermatol 49:832–841.

264. Ojesina AI, Lichtenstein L, Freeman SS, Pedamallu CS, Imaz-Rosshandler I, Pugh TJ, Cherniack AD, Ambrogio L, Cibulskis K, Bertelsen B, Romero-Cordoba S, Treviño V, Vazquez-Santillan K, Guadarrama AS, Wright A a, Rosenberg MW, Duke F, Kaplan B, Wang R, Nickerson E, Walline HM, Lawrence MS, Stewart C, Carter SL, McKenna A, Rodriguez-Sanchez IP, Espinosa-Castilla M, Woie K, Bjorge L, Wik E, Halle MK, Hoivik E a, Krakstad C, Gabiño NB, Gómez-Macías GS, Valdez-Chapa LD, Garza-Rodríguez ML, Maytorena G, Vazquez J, Rodea C, Cravioto A, Cortes ML, Greulich H, Crum CP, Neuberg DS, Hidalgo-Miranda A, Escareno CR, Akslen L a, Carey TE, Vintermyr OK, Gabriel SB, Barrera-Saldaña H a, Melendez-Zajgla J, Getz G, Salvesen HB, Meyerson M. 2014. Landscape

of genomic alterations in cervical carcinomas. Nature 506:371–5.

265. Schowalter RM, Pastrana D V., Pumphrey KA, Moyer AL, Buck CB. 2010. Merkel cell polyomavirus and two previously unknown polyomaviruses are chronically shed from human skin. Cell Host Microbe 7:509–515.

266. Toh ST, Jin Y, Liu L, Wang J, Babrzadeh F, Gharizadeh B, Ronaghi M, Toh HC, Chow PKH, Chung AYF, Ooi LLPJ, Lee CGL. 2013. Deep sequencing of the hepatitis B virus in hepatocellular carcinoma patients reveals enriched integration events, structural alterations and sequence variations. Carcinogenesis 34:787–798.

267. Jiang Z, Jhunjhunwala S, Liu J, Haverty PM, Kennemer MI, Guan Y, Lee W, Carnevali P, Stinson J, Johnson S, Diao J, Yeung S, Jubb A, Ye W, Wu TD, Kapadia SB, Sauvage FJ De, Gentleman RC, Stern HM, Seshagiri S, Pant KP, Modrusan Z, Ballinger DG, Zhang Z, de Sauvage FJ, Gentleman RC, Stern HM, Seshagiri S, Pant KP, Modrusan Z, Ballinger DG, Zhang Z. 2012. The effects of hepatitis B virus integration into the genomes of hepatocellular carcinoma patients. Genome Res 22:593–601.

268. Sung W-K, Zheng H, Li S, Chen R, Liu X, Li Y, Lee NP, Lee WH, Ariyaratne PN, Tennakoon C, Mulawadi FH, Wong KF, Liu AM, Poon RT, Fan ST, Chan KL, Gong Z, Hu Y, Lin Z, Wang G, Zhang Q, Barber TD, Chou W-C, Aggarwal A, Hao K, Zhou W, Zhang C, Hardwick J, Buser C, Xu J, Kan Z, Dai H, Mao M, Reinhard C, Wang J, Luk JM. 2012. Genome-wide survey of recurrent HBV integration in hepatocellular carcinoma. Nat Genet 44:765–9.

269. Martel-Jantin C, Filippone C, Cassar O, Peter M, Tomasic G, Vielh P, Brière J, Petrella T, Aubriot-Lorton MH, Mortier L, Jouvion G, Sastre-Garau X, Robert C, Gessain A. 2012. Genetic variability and integration of Merkel cell polyomavirus in Merkel cell carcinoma. Virology 426:134–142.

270. Cohen PR, Kurzrock R. 2015. Merkel cell carcinoma with a suppressor of fused (SUFU) mutation: Case report and potential therapeutic implications. Dermatol Ther (Heidelb) 129–143.

271. Drmanac R, Sparks AB, Callow MJ, Halpern AL, Burns NL, Kermani BG, Carnevali P, Nazarenko I, Nilsen GB, Yeung G, Dahl F, Fernandez A, Staker B, Pant KP, Baccash J, Borcherding AP, Brownley A, Cedeno R, Chen L, Chernikoff D, Cheung A, Chirita R, Curson B, Ebert JC, Hacker CR, Hartlage R, Hauser B, Huang S, Jiang Y, Karpinchyk V, Koenig M, Kong C, Landers T, Le C, Liu J, McBride CE, Morenzoni M, Morey RE, Mutch K, Perazich H, Perry K, Peters BA, Peterson J, Pethiyagoda CL, Pothuraju K, Richter C, Rosenbaum AM, Roy S, Shafto J, Sharanhovich U, Shannon KW, Sheppy CG, Sun M, Thakuria J V, Tran A, Vu D, Zaranek AW, Wu X, Drmanac S, Oliphant AR, Banyai WC, Martin B, Ballinger DG, Church GM, Reid CA. 2010. Human genome sequencing using unchained base reads on self-assembling DNA nanoarrays. Science (80- ) 327:78–81.

272. Futreal PA, Coin L, Marshall M, Down T, Hubbard T, Wooster R, Rahman N, Stratton MR. 2004. A census of human cancer genes. Nat Rev Cancer 4:177–183.

273. Choi Y, Chan AP. 2015. PROVEAN web server: a tool to predict the

functional effect of amino acid substitutions and indels. Bioinformatics 31:2745–2747.

274. Choi Y. 2012. A fast computation of pairwise sequence alignment scores between a protein and a set of single-locus variants of another protein, p. 414–417. *In* Proceedings of the ACM Conference on Bioinformatics, Computational Biology and Biomedicine - BCB ’12.

275. Choi Y, Sims GE, Murphy S, Miller JR, Chan AP. 2012. Predicting the functional effect of amino acid substitutions and indels. PLoS One 7:e46688.

276. Ng PC. 2003. SIFT: predicting amino acid changes that affect protein function. Nucleic Acids Res 31:3812–3814.

277. Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, Pimentel H, Salzberg SL, Rinn JL, Pachter L. 2012. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat Protoc 7:562–78.

278. Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 15:1–21.

279. Gehring JS, Fischer B, Lawrence M, Huber W. 2015. SomaticSignatures: inferring mutational signatures from single-nucleotide variants. Bioinformatics 31:3673–5.

280. Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. Nat Methods 9:357–9.

281. Zerbino DR, Birney E. 2008. Velvet: Algorithms for de novo short read

assembly using de Bruijn graphs. Genome Res 18:821–829.

282. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+: architecture and applications. BMC Bioinformatics 10:1–9.

283. Cantalupo PG, Calgua B, Zhao G. 2011. Raw sewage harbors diverse viral populations. MBio 2:e00180-11.

284. Dalianis T, Hirsch HH. 2013. Human polyomaviruses in disease and cancer. Virology 437:63–72.

285. Jiang M, Abend JR, Johnson SF, Imperiale MJ. 2009. The role of polyomaviruses in human disease. Virology 384:266–273.

286. Harris KF, Christensen JB, Imperiale MJ. 1996. BK virus large T antigen: interactions with the retinoblastoma family of tumor suppressor proteins and effects on cellular growth control. J Virol 70:2378–86.

287. Pipas JM, Levine AJ. 2001. Role of T antigen interactions with p53 in tumorigenesis. Semin Cancer Biol 11:23–30.

288. Mehta H V, Jones PH, Weiss JP, Okeoma CM. 2012. IFN-α and lipopolysaccharide upregulate APOBEC3 mRNA through different signaling pathways. J Immunol 189:4088–103.

289. Liu W, Yang R, Payne AS, Schowalter RM, Spurgeon ME, Lambert PF, Xu X, Buck CB, You J. 2016. Identifying the Target Cells and Mechanisms of Merkel Cell Polyomavirus Infection. Cell Host Microbe 1–13.

290. Kondo S, Wakae K, Wakisaka N, Nakanishi Y, Ishikawa K, Komori T, Moriyama-Kita M, Endo K, Murono S, Wang Z, Kitamura K, Nishiyama T,

Yamaguchi K, Shigenobu S, Muramatsu M, Yoshizaki T. 2016. APOBEC3A associates with human papillomavirus genome integration in oropharyngeal cancers. Oncogene 36:1–11.