REGULATION OF APOBEC3B VIA INTERACTING PROTEINS

A DISSERTATION SUBMITTED TO THE FACULTY OF THE

UNIVERSITY OF MINNESOTA BY

Jennifer Lynn McCann

IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

Advisor: Reuben S. Harris

August 2019

Acknowledgement

I would like to thank former and present Harris lab members for their insights, discussions and overall support throughout my thesis studies. Additionally, I would like to thank my mentor, Reuben Harris, for providing both the resources and guidance needed to advance my scientific career. My committee members, Drs. Douglas Yee, David Largaespada, and Scott Dehm for your invaluable advice and support.

I would like to thank my many other mentors: Drs. DJ Daniel J Salamango (the man to always challenge me, my scientific thinking and the physiological relevance of everything), Michael Carpenter (a handyman for everything, including life itself), Nadine Shaban (99% god-given saint, 1% psychiatrist), Artur Serebrenick (baymate and graduate school lifeline) and Emily Law (my travel companion, and fellow sour-patch kid).

To my friends: Drs. Amber St. Martin, Amy Molan and soon to be real doctor Evelyn Leland without whom I would have had to trudge through this alone (and let's be real, I probably wouldn't have made it), thank you for your endless support.

My family: Michael, Syndee, Adam, Anna and Stephanie McCann, Ed Barrett, Linda and Richard Petersen for being everything and more.

Lastly, my husband Andrew for your enthusiasm and support on all days, good and bad.

Dedication

This thesis dedicated to Florence E. Chidester Petersen “Molly”.

For the world needs more women who raise strong women that drive change and ultimately never take no for an answer.

Thank you.

.

Abstract

Proteins play a major role in cellular processes; therefore, it is important to understand how they perform their biological functions. However, proteins rarely act independently, but rather they often create “molecular machines” and have intricate physical and chemical dynamic interactions to undertake biological functions at both cellular and system levels. A key step towards unraveling the complex molecular relationships in living systems lies in the mapping and characterization of protein-protein interactions (PPIs). These interactions form the backbone of signal transduction pathways and cellular networks in diverse physiological processes. Due to their critical roles in relaying key cellular signals in both normal and cancer cells; once deemed “undruggable” PPIs have led to the development of a potential new class of drug targets [for example, the interaction between p53 and MDM2 (1)].

In this thesis, we try to understand the interactome of APOBEC3B (A3B), a key molecular driver inducing mutations in multiple human cancers. We utilized an affinity purification mass spectrometry approach to identify cellular proteins that interact with A3B. Our results revealed a specific interaction with the cell cycle protein cyclin-dependent kinase 4 (CDK4). We validated and mapped this interaction by structure guided mutagenesis and co-immunoprecipitation experiments. Functional studies and immunofluorescence microscopy experiments in multiple cell lines revealed that A3B is not a substrate for CDK4-Cyclin D1 phosphorylation, nor is its deaminase activity modulated. Instead, we found that A3B is capable of disrupting the CDK4-dependent nuclear import of Cyclin D1. We propose that this interaction may favor a more potent antiviral response and simultaneously facilitate cancer mutagenesis.

In addition, our studies on the A3B “interactome” led to multiple advancements in genome engineering, specifically in cytosine base-editing (CBE). With the recent discovery of base editing systems in which a DNA editing enzyme, such as an APOBEC, is targeted to a specific base by covalent fusion to a Cas9 nickase (Cas9n) complex the prospect of precise genome engineering has become reality. However, current CBEs are prone to undesirable off-target mutations (most frequently occurring as target-adjacent mutations). Through the identification of novel A3B interacting proteins we developed a method, termed “MagnEdit”, in which we use these interacting proteins to serve as “magnets” fused to Cas9n in order to attract A3B to single-base editing hotspots. By untethering A3B from Cas9n and utilizing endogenous A3B “recruiters” we have dramatically reduced off target editing. These findings, while demonstrated here for A3B and its interacting proteins, have implications for other editing technologies and other recruiters such as antibodies, small molecules, and epitope tags.

Table of Contents

List of Figures

Chapter I.

Introduction – Regulation of APOBEC3B: Contributions to Cancer and Advancements in Base Editing

Authors: Jennifer L. McCann[1-4]

Affiliations: [1]Department of Biochemistry, Molecular Biology and Biophysics, [2]Institute for Molecular Virology, [3]Masonic Cancer Center, [4]Center for Genome Engineering, University of Minnesota, Minneapolis, MN 55455, USA.

APOBEC3B – A brief history

The apolipoprotein B mRNA-editing enzyme catalytic polypeptide-like (APOBEC) family of enzymes is a group of cytosine deaminases that have evolved in vertebrates, and particularly mammals, to mutate RNA and DNA at distinct preferred nucleotide hotspots on foreign genomes. This highly evolved family has gained a substantial amount of diversity through gene duplications and rearrangements over time (2,3). Phylogenetic analysis of the APOBEC family reveals its origin is from ancestral activation induced deaminase (*AID*) genes that were present in jawless fish over 500 million years ago and has now expanded to amphibians, birds and mammals (4). In higher primates and humans, the APOBEC family consists of eleven members, including APOBEC1 (A1), AID, APOBEC2 (A2), APOBEC3A (A3A), APOBEC3B (A3B), APOBEC3C (A3C), APOBEC3D (A3D), APOBEC3F (A3F), APOBEC3G (A3G), APOBEC3H (A3H), and APOBEC4 (A4) (2,5-9) (**Fig 1.1A**).

Initially, A1 and AID were the most well studied within the APOBEC superfamily. The founding member of the APOBEC family (and for which the family was named), APOBEC1, was first identified as the apolipoprotein B (*apoB*) mRNA editing enzyme in the 1990s (10). APOBEC1 was further characterized to bind A1CF (APOBEC1 complementation factor) and deaminate cytidine^{6666} to uridine on the *apoB* mRNA, thereby producing two isoforms of apoB proteins with individual and overlapping functions (11,12). Alternative to mRNA editing, AID plays an essential role in adaptive immunity by regulating antibody maturation and diversification in activated B cells via deamination of single-stranded (ss)DNA substrates (13). AID catalyzes the diversification in the sequence of the immunoglobulin genes through somatic hypermutation and class switch recombination

(14,15). The physiological functions of APOBEC2 and APOBEC4 remain largely unknown, as neither enzyme has been shown to be catalytically active (16-18).

The APOBEC3 (A3) subfamily of enzymes have gone through a series of duplications and rearrangements during the evolution of primates leading to seven family members in humans that are encoded in tandem on chromosome 22 (3-6,8). The A3 enzymes are composed of one or two catalytic zinc-coordinating domains that fall into three distinct phylogenetic groups, Z1, Z2 or Z3 (5). A3A (Z1), A3C (Z2) and A3H (Z3) all contain a single cytidine deaminase zinc-coordinating domain, while the remaining four A3s (A3B, A3D, A3F and A3G) contain both an N-terminal pseudocatalytic or inactive domain and a catalytically active C-terminal domain (2) (**Fig. 1.1*A***). Although the role for each A3 may be overlapping or unique within in the cell, the mechanism of deaminating cytosines to uracils in ssDNA is conserved (**Fig. 1.1*B***).

The APOBEC3 proteins, while distinctly different, are known to have three physiological functions in human cells. The most well-known function involves mechanisms of defense against a diverse array of viral pathogens, including retroviruses, hepatitis viruses, papillomaviruses, and others (19-26). For example, A3D, F, G and H are most highly expressed in T cells and have been widely studied for their role in retroviral restriction through hypermutation of viral ssDNA genome intermediates (27,28). An alternative function of A3s is inhibition of L1 and Alu transposable elements. Several A3s, including A3A, B and F have demonstrated some capacity to restrict retrotransposition via enzymatic and non-enzymatic mechanisms (29-38). Lastly, studies have shown that A3A and other family members have the potential to mediate clearance of foreign DNA via a deaminase dependent mechanism (25,39,40).

Although many A3 proteins seemingly have overlapping functions, these enzymes are differentially expressed in tissue and cell types (27,41). For example, different A3 proteins are expressed at various levels in immune cell populations, including $CD4^+$ and $CD8^+$ T cell subsets, B cells and myeloid cells and transcripts for A3 enzymes are also found in nonimmune tissues such as lung, ovary, breast and adipose tissue (27,28). In addition to cell type and expression differences, A3 proteins can have different subcellular localizations. A3D, F, G, and H are found to be cytoplasmic (H can also be nucleolar), whereas A3A, and C have been reported to localize cell-wide, and A3B is the only A3 enzyme that is constitutively nuclear (42-47). While cytoplasmic A3 enzymes are compartmentalized away from the genome, effective mechanisms likely exist within the cell to ensure proper discrimination of host vs non-host DNA for those A3 proteins with access to the nuclear compartment. However, there is evidence that when misregulated nuclear A3s, generally A3B (but also A3A and A3H), can act as genomic DNA mutators, and this poses an intrinsic hazard for the host if the genomic DNA is not properly regulated or repaired (28,48-60).

Contributions to cancer

In 2013, Burns *et al*. was the first to discover A3B as a dominant DNA mutating enzyme in breast cancer (48). Through the use of gene-specific knockdown, over-expression studies and somatic mutation signature analysis, data combined to show that A3B was driving mutations that diversified the genetic landscape in a large proportion of breast cancer cell lines. Additionally, bioinformatic and biochemical analysis of genomic substrates linked A3B deamination to a preferred sequence context, cytosines preceded by a 5'thymine and followed by a 3' adenosine or guanine (40,61-64). These deamination

events result in a genomic DNA uracil that when unfaithfully repaired and present through DNA replication will result in the transition of a cytosine to a thymine (C-to-T). Alternatively, this DNA uracil can undergo excision by uracil-DNA glycosylase (UDG) and repair of the remaining abasic site by REV1, commonly results in the transversion of a cytosine to a guanine (C-to-G) (65-69). These findings together led to a new class of mutation pattern deemed the “APOBEC signature”. Subsequently, A3B mediated mutations and elevated mRNA expression levels have been identified in many additional types of cancer and have been linked to poor clinical outcome (discussed later in *Clinical impact and association*) in a subset of them (70-78).

Next-generation sequencing (NGS) has revealed an APOBEC-mutation signature in >50% of human cancer types with variable impact within each tumor ranging from <5% to >90% of all base substitution mutations (28,49,51). In addition to single base substitutions, A3s can catalyze clustered hypermutation, or kataegis, in cancer cell genomes; however, these clusters often associate with recombination breakpoints, such as sites of chromosomal translocation, and only account for less than 10% of the total mutation load (52,79). Although a role for A3B in cancer mutagenesis and evolvability has been well documented, universally accepted genomic substrates have yet to be described. Conflicting studies show A3B can target transcription-associated substrates as well as replication and recombination-associated substrates (36,52,80-82). These APOBEC-mediated mutations are linked in part to R-loops and stalled replication forks, which can be associated with various types of DNA damage and ssDNA exposure (59,83,84). Despite lacking commonly accepted A3B genomic targets, two distinct APOBEC mutational hotspots resulting in PIK3CA activation have been reported (85,86). Altogether,

this supports the association seen between elevated expression of *APOBEC3B* and more aggressive phenotypes and recurrence in cancers (48,72,87,88).

Clinical impact and association with cancer

As studies continuously connect the DNA cytosine deaminase A3B to the genetic evolution of multiple cancers, association of its expression with clinical outcomes is largely in pursuit. It was recently shown in a large cohort of estrogen receptor-positive breast cancer patients that those with tumors expressing high levels of *A3B* were found to associate with shorter durations of both disease-free and overall survival (89). In addition to these findings, higher *APOBEC3B* expression levels are linked with increased resistance to the estrogen receptor agonist, tamoxifen, in murine models (70). Importantly, these studies strongly suggest *A3B* as a marker of pure prognosis and poor outcomes for estrogen receptor-positive breast cancer.

Broadening the initial breast cancer findings, Cescon *et al*. reported *A3B* expression is associated with clinicopathologic characteristic and unfavorable outcomes as well as a positive correlation with proliferative features in breast and 15 of 16 other solid tumor types (except lung squamous cell carcinoma) (73). These other solid tumor types include: bladder urothelial carcinoma, breast invasive carcinoma, cervical squamous cell carcinoma, colon adenocarcinoma, head and neck squamous cell carcinoma, kidney renal clear cell carcinoma, kidney renal papillary carcinoma, brain lower grade glioma, liver hepatocellular carcinoma, lung adenocarcinoma, lung squamous cell carcinoma, skin cutaneous melanoma, stomach adenocarcinoma, thyroid carcinoma and uterine corpus endometrial carcinoma. Together, genetic, cellular, and biochemical studies have demonstrated that A3B-catalyzed genomic uracil lesions are responsible for a large

proportion of both dispersed and clustered mutations in multiple distinct cancers (28,48,49,70,72,73,76,77,87,88,90-98). However, additional work and possibly prospective trials are needed to further evaluate the role of A3B as a prognostic marker and or contributor to drug resistance and metastasis.

Regulation and misregulation

Since A3B has been shown to play a large role in tumor survival and evolution, A3B regulation has now become an ever-growing area of cancer therapeutic research (**Fig. 1.2**). Both endogenous regulation and chemical modulation of A3B are novel strategies to rein this mutagenic enzyme. Recent reviews have suggested the development of small molecule and chemical inhibitors targeting the active site in the C-terminal domain of A3B; however, the crystal structure of full-length A3B remains to be solved and only recently did the first single domain crystal structures of A3B emerge (99-102). With these recent publications, structural-base design of A3B inhibitors can be pursued for the first time but concerns regarding specificity will remain a challenge. As one could imagine, non-specific A3B inhibitors targeting other A3s would leave patients susceptible to infection by viruses and/or retrotransposition.

An alternative approach focuses on targeting regulatory pathways of endogenous A3B, whether this be transcriptional or post-translational. Within the past three years numerous studies, including our own, have reported findings on transcriptional regulation of *A3B* (21,82,103-107). The primary goal of these studies is to better understand the mechanisms that regulate *A3B* expression, thereby providing insight into the processes driving acquisition of cancer mutations and tumor evolution (**Fig. 1.2**). These findings have linked the non-canonical NF-κB pathway, HPV E6 and E7 viral oncogenes, transcription

factors TEAD, B-Myb and p53 as well as the RB-E2F axis with *A3B* upregulation in various normal and cancer cell types (21,82,103-107). It's likely *A3B* is regulated via different pathways in a cell-type and stimulus dependent manner. Although it is in its early stages, these data will likely have important consequences for tumor development and tumor evolution, including response to therapies and regulation of A3B inflicted mutational damage.

Post-translational regulation (including protein-protein interactions and post-translational modifications)

While research on transcriptional regulation of *A3B* has "taken-off", understanding post-translational regulation has been far more difficult due to the limited A3B-specific reagents in addition to the insoluble properties of A3B. However, post-translational properties such as localization, modifications and interacting proteins have been recently reported. As mentioned above, A3B is the only family member shown to be constitutively nuclear (42-45). After mitosis and nuclear membrane reconstitution, A3B is rapidly reimported into the nuclear compartment where it remains constitutively present until the next mitotic cell division (42,43,46). Although the mechanism of nuclear import has yet to be fully elucidated, the identification of A3B N-terminal domain *cis* elements required for active nuclear import has shed light on this mechanism of regulation (43-46). Although, A3B may still remain an active deaminase, re-localization of A3B to the cytoplasm, and excluding it from the nuclear compartment where in deaminates the host genome, using inhibitors of these newly identified residues may serve as a novel therapeutic approach in future studies.

Alternatively, post-translational modifications of A3B have just recently been reported as a potential mechanism of A3B inhibition. Matsumoto *et al*., published findings of protein kinase A (PKA)-mediated phosphorylation at Thr214 inhibits A3B mutagenic activity (108). While this *in vitro* data is convincing for A3Bs cell-type specific HIV-1 viral restriction role, it is far from complete and provides minimal, if any, cancer/mutation-based studies. The only other modification studies performed on A3B demonstrate that lysine residues are unlikely targets of A3B regulation as they are completely dispensable for A3B mutagenic activity and localization (109). Together, these data provide insights into A3B post-translational regulation, but further advancements need to be made to better understand the mechanisms governing the activity, localization and overall regulation of A3B.

As is known, proteins rarely act alone, but rather they often create "molecular machines" and have intricate physicochemical dynamic interactions to perform biological functions at both cellular and system levels. A key step towards unraveling the complex molecular relationships in living systems is the mapping of protein-protein interactions (PPIs). These interactions form the backbone of signal transduction pathways and networks in diverse physiological processes. That being said, few A3B interacting proteins have been identified and only one to-date has shown a direct mechanism of A3B-inhibition. The first A3B-interactor identified was hnRNPK in 2008 (110). These findings showed no form of A3B regulation but rather that A3B was inhibiting hnRNPK-mediated transcription and expression of HBV genes. Eight years later in 2015, simian immunodeficiency virus (SIV) Vif proteins were shown to degrade A3B, with SIV_{mac239} being the most potent (111). This mechanism of degradation is likely to occur through the canonical polyubiquitination mechanism as A3B protein levels are restored with treatment

of MG132 (111). This was the first report to suggest development of known interacting proteins, like SIV_{mac239}, as molecular tools for future studies focused at neutralized A3B in cancer to halt tumor mutagenesis, prevent therapy resistance, and improve the treatment and prognosis of cancer patients. Although these studies are interesting and potentially translational, this is not physiological as SIV (monkey) Vif is not of direct relevance to human cells. Just two years later in 2017, hnRNPA3 was reported to bind A3B, but again no findings of A3B regulation (112). Lastly, we published this year (2019) that the Epstein-Barr virus protein, BORF2, can inhibit cellular A3B through binding of the A3B C-terminal catalytic domain in addition to re-localization of nuclear A3B to perinuclear bodies (19). This was the first report to show interaction and inhibition of the A3B catalytic domain providing a novel break-through in A3B research. In addition to these findings, later discussed in **Chapter 2**, we also published in 2019 the interaction between A3B and the cell cycle protein CDK4 (113). Although we did not find a mechanism of A3B inhibition, we do propose that this interaction may favor a more potent antiviral response and simultaneously facilitate cancer mutagenesis through disruption of the canonical G1/S phase transition. As summarized above, research on post-translational regulation of A3B are clearly on-going, but only recently did we make great strides towards addressing mechanisms of A3B inhibition and suppression of cancer progression.

Advancements in base editing

In parallel to rapidly advancing research on APOBEC3s as antivirals and as cancer mutagens, as described above, the APOBEC enzymes have also taken center stage in the area of genome engineering. CRISPR/Cas9 as a genome engineering tool has revolutionized basic, biomedical, and agricultural research by enabling simple targeting and manipulation of user-defined locations in the genome. In its simplest form, CRISPR is an RNA-guided DNA endonuclease that catalyzes DNA double-stranded breaks (DSBs) that often leads to error-prone repair and permanent genetic scars. Broad application and innovation of CRIPSR tools has now made it relatively straight-forward to knock out a gene of interest in a variety of different cell types. Though targeting specific genome sites has now become relatively simple, the introduction of pre-defined, precise alterations at these sites has remained a challenge.

In 2016, two groups simultaneously described the development of base editing (BE) enzymes that couple the genome-targeting features of Cas9, with the direct DNA modifying activity of cytidine deaminases, such as AID and APOBEC1 (114,115). In this adapted technology, base editors utilize covalent fusions between a catalytically impaired Cas9 nuclease, deemed Cas9 nickase (Cas9n or $Cas9n_{D10A}$), and a base-modifying enzyme that targets the single-stranded (ss)DNA. Upon binding to its target locus in the DNA, base pairing between the guide (g)RNA and the target DNA strand leads to an RNA-DNA hybrid and the formation of an R-loop where a segment of ssDNA is displaced (**Fig. 1.3**) (116). The DNA strand separation positions the deaminase enzyme to induce cytosine to thymine (C-to-T) transitions within a small nucleotide window at the 5' end of the sgRNA target sequence (**Fig. 1.3**). Importantly, the nickase activity of the Cas9 generates a nick in the non-edited DNA strand to induce cells to repair the non-edited

strand using the edited strand as a template (**Fig. 1.3**) (114,115,117). Further, the recent evolution of non-natural DNA editing enzymes has broadened the scope of base editing to include adenine base editors (ABEs) that convert adenine to guanine (A-to-G transition mutations) (117). Collectively, cytosine base editors (CBEs) and ABEs can mediate four types of transition mutations (C-to-T and G-to-A; A-to-G and T-to-C). The already broad therapeutic applications for CRIPSR/Cas9 now includes base editing approaches that can be used to target the many genetic diseases that can be corrected through a single base mutation rather than the total disruption of the gene.

Even before the development of BE technology, and as mentioned above, it was known that APOBECs induce C-to-T mutations in both the cellular genome as well as in complementary DNA that is reverse transcribed from viral RNA genome (16,28,48,118). Indeed, association between APOBECs and CRISPR/Cas9-mediated gene editing was discovered early on when trying to understand off-target effects. Tsai *et al*. found off target editing of Cas9n and most editing events fell in a C-to-T context, manifesting a typical APOBEC mutational signature (119). In addition, APOBEC/AID has been largely reported in targeting of R-loops (which are essentially ssDNA loops created by RNA/DNA hybrids) (120-126). However, the real landmark series of experiments began with the development of CBEs. These initiated with the development of the first-generation base editor (BE1), a construct consisting of rat APOBEC1 fused to catalytically inactive Cas9 (dCas9) through an XTEN linker, that was able to deaminase cytosines between positions 4 and 8 within the target sequence at rates of up 7% in cellular based systems (114). Following BE1 was BE2, a second-generation base editor aimed at inhibiting uracil DNA glycosylase (UNG) through the fusing of uracil DNA glycosylase inhibitor (UGI), a small protein from bacteriophage PBS, to the C-terminus of BE1 (114). This addition increased editing

efficiencies in bacterial cells but only moderately increase editing in mammalian-based systems (114,127). Shortly after BE2 came BE3, which changed the dead Cas9 (dCas9) to a Cas9 nickase (Cas9n) (114). As mentioned above, the nickase activity of the Cas9 generates a nick in the non-edited DNA strand to induce cells to repair the non-edited strand using the editing strand as a template (114,115,117). This dramatically increased editing efficiencies in mammalian cells, up to 37% across six loci in the initial report (114). Lastly, BE4 and BE4max were recently published and only differs from BE3 in that it has an additional copy of UGI and BE4max is codon optimized (**Fig. 1.4**). Notably, base editing technology did not ablate the off target indels but it did dampen their frequency compared to those induced by Cas-mediated DSBs (114).

Base-editing limitations

With these recent technological advancements comes a better understanding of limitations, off-target frequencies, and remaining challenges. One major limitation to this technology is the inability to catalyze transversion mutations, purine to pyrimidine and vice versa. Overcoming this chemistry and protein engineering roadblock would significantly broaden the applications of this method and provide even more flexibility in genome engineering. Going hand-in-hand with the first limitation is the need for protein engineering of Cas9 enzymes that broaden the sequence specificity of a protospacer adjacent motif (PAM). While this restriction is common to all CRISPR-based tools, it is particularly important for BE, as effective editing occurs within a small nucleotide window of the R-loop. Consequently, moving the gRNA target to the nearest PAM can shift the target nucleotide outside the editing window. Although other species and variants of Cas9 with alternative PAM sequences exists, current studies are attempting to circumvent this

limitation by manipulating the window of editing (128). Most importantly, like all genome engineering systems, BE carries the possibility of off-target effects.

Base-editing precision

To date, fusion of DNA cytosine modifying enzymes (i.e APOBEC) has resulted in a substantial amount of off-target editing. These editing events can be random genomic DNA mutations (129-132), RNA editing events (133-135), and most frequently target-adjacent editing (114,117,129,136-139). Surprisingly, recent whole-genome sequencing CBE studies in both mice and rice detected significant amounts of off-target C-to-T editing in transcribed regions, even in the absence of a gRNA (131,132). These results suggest that the presence of a cytosine deaminase tethered to a Cas9n complex alone are sufficient for a primed encounter with exposed ssDNA. Nevertheless, the off-targeting outcomes for BEs and Cas9 alone are not always the same, raising the concern for separate evaluation methods. In comparison with CBEs, whole-genome sequencing of ABEs revealed no off-target base editing (above spontaneous mutation rates) in mice studies (140,141). Together these data largely support a need for a more accurate and advanced method of targeting and editing specific cytosine bases.

In light of this off target limitation and need for optimized cytosine base editing, we have devised a method, deemed “MagnEdit”, to attract an endogenous DNA editor to a target base for efficient and specific editing without gratuitous off-target adjacent damage (manuscript in review). In this method, later discussed in **Chapter 3**, we demonstrate the feasibility of this method using previously validated A3B-interacting proteins to serve as “magnets” fused to Cas9n and UGI, in order to attract A3B to single-base editing hotspots. The key to this system is removal of the covalent fusion of the deaminase to the Cas9n

complex, which traps the tethered deaminase locally and inextricable links both on-target and target-adjacent cytosine deamination events. We demonstrate that current CBE methodology generates adjacent off-target editing at a frequency of >25%, while MagnEdit catalyzes an on-target efficiency of >95% and no significant off-target events mutations. Overall, our advancements to the genome engineering field demonstrate a modular system that can be adapted for any recruiter (interacting proteins, antibodies, epitopes etc.) and base editor system with current studies producing substantially lower off-target editing than current base editing methodology.

Concluding introductory remarks and chapter prelude

The APOBEC3 family of DNA cytosine deaminases in important for mutagenesis in cancer and viral evolution (where viral restriction is merely the outcome for viruses that fail to evolve and adapt to the host APOBECs). Importantly, A3B has been identified as a major source of somatic mutagenesis in cancer cells that drive tumor evolution such as metastasis, and/or therapy resistance. Although research in this field is in its early stages, A3B regulation has now become an ever-growing area of cancer therapeutic interest. Both endogenous regulation and chemical modulation of A3B are novel strategies to rein this mutagenic enzyme but remain largely unknown. To address this aforementioned "black-box" of A3B endogenous regulation, we investigated the interactome of proteins that bind and potentially regulate this mutagenic enzyme. These proteomic efforts led to two different discoveries, the first being a novel role for A3B and its ability to bind CDK4 and disrupt CDK4-dependent nuclear import of Cyclin D1 (**Chapter 2**). In summary, we propose that this interaction may alter canonical cell cycle progression and favor a more potent antiviral response and simultaneously facilitate cancer mutagenesis.

Additionally, our work on A3B interacting proteins led us to develop a novel technology that merges the current CRISPR/Cas9 cytosine base editing (CBE) technology with our new understanding of the A3B proteome. This system stems from the current base editing technology in which a DNA editing enzyme, such as an APOBEC, is targeted to a specific base by covalent fusion to a Cas9 nickase (Cas9n) complex to thereby provide precision genome engineering. However, current CBE methods are fraught with undesirable target-adjacent and off-target mutations. Through the identification of novel A3B interacting proteins we developed a method, termed "MagnEdit", in which we use these interacting proteins to serve as "magnets" fused to Cas9n in order to attract A3B to

single-base editing hotspots (**Chapter 3**). By untethering A3B from Cas9n and utilizing endogenous A3B “recruiters” we have dramatically reduced off-target editing events. These findings, while demonstrated here for A3B and its interacting proteins, have implications for other editing technologies and other recruiters such as antibodies, small molecules, and epitope tags.

.

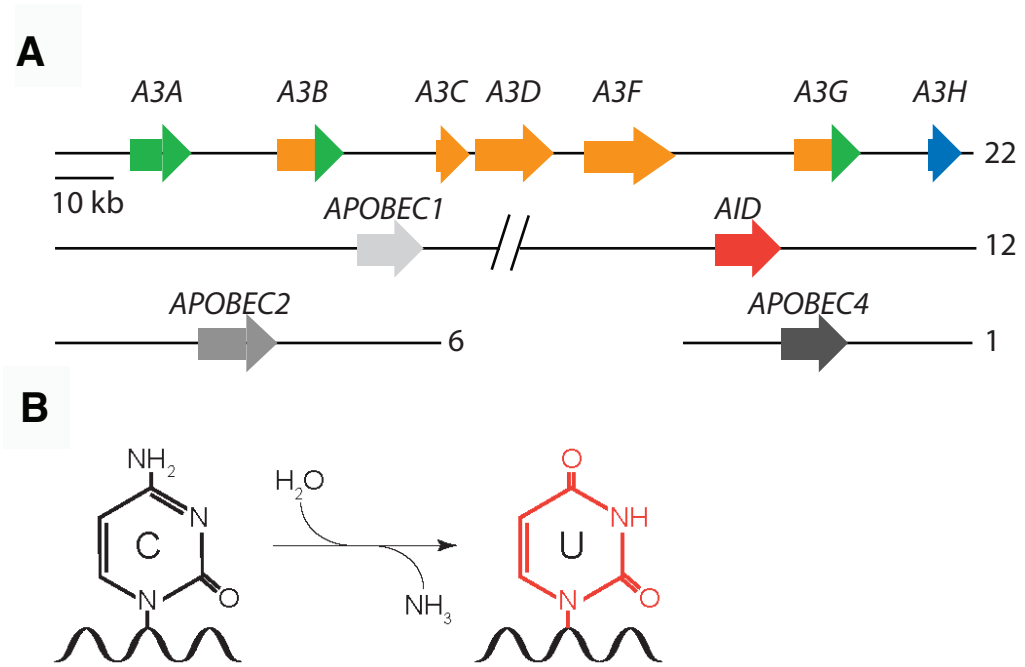


Figure 1.1. The APOBEC family of polynucleotide cytosine deaminases

A, Diagram of the eleven APOBEC family members. Colored arrows represent the evolutionary conservation of the three zinc-coordinating deaminase domains; green depicts Z1, orange depicts Z2, and blue depicts Z3. All other colors represent unique regions to that enzyme.

***B*,** Schematic of APOBEC-catalyzed deamination of a cytosine to uracil in single stranded (ss)DNA.

This figure was drafted by J. McCann and adapted from (142) .

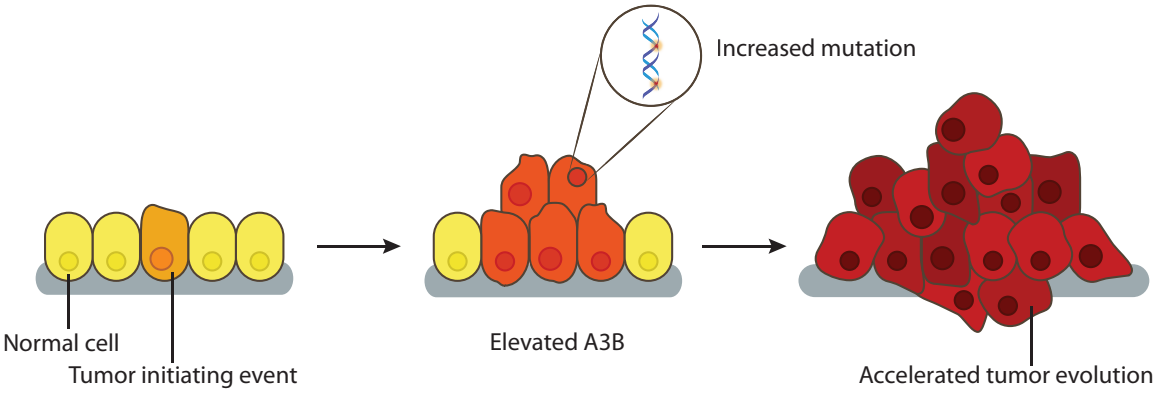


Figure 1.2. Model for APOBEC3B driven tumor evolution

Normal, healthy cells are depicted in yellow while cells in the early stages of APOBEC3B dysregulation are colored in light orange. Cells undergoing high APOBEC3B expression and mutagenesis are in dark orange and red. Together this model depicts upregulation of APOBEC3B in nascent cancer cells or during cancer development and thereby increases mutation rates and drives tumor evolution.

This figure was drafted by Harris Lab artist Amanda Miller and adapted from (142).

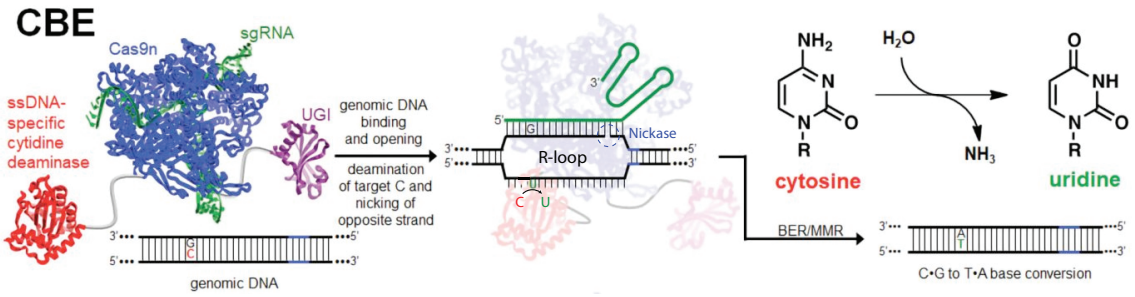


Figure 1.3. Molecular reactions that underlie cytosine base editing (CBE)

Schematic of a cytosine base editor where the cytidine deaminase is covalently tethered to Cas9 nickase (Cas9n) and the uracil glycosylase inhibitor (UGI). The target cytosine, in red, is exposed as ssDNA once the Cas9n gRNA binds to the target DNA strand resulting in the formation of an R-loop. Deamination of the target C occurs, Cas9n generates a nick on the non-edited strand (circled in blue) and UGI prevents the U:G mismatch from being repaired back to a C:G resulting in C-to-T mutation.

This figure was adjusted by J. McCann and adapted from (114).

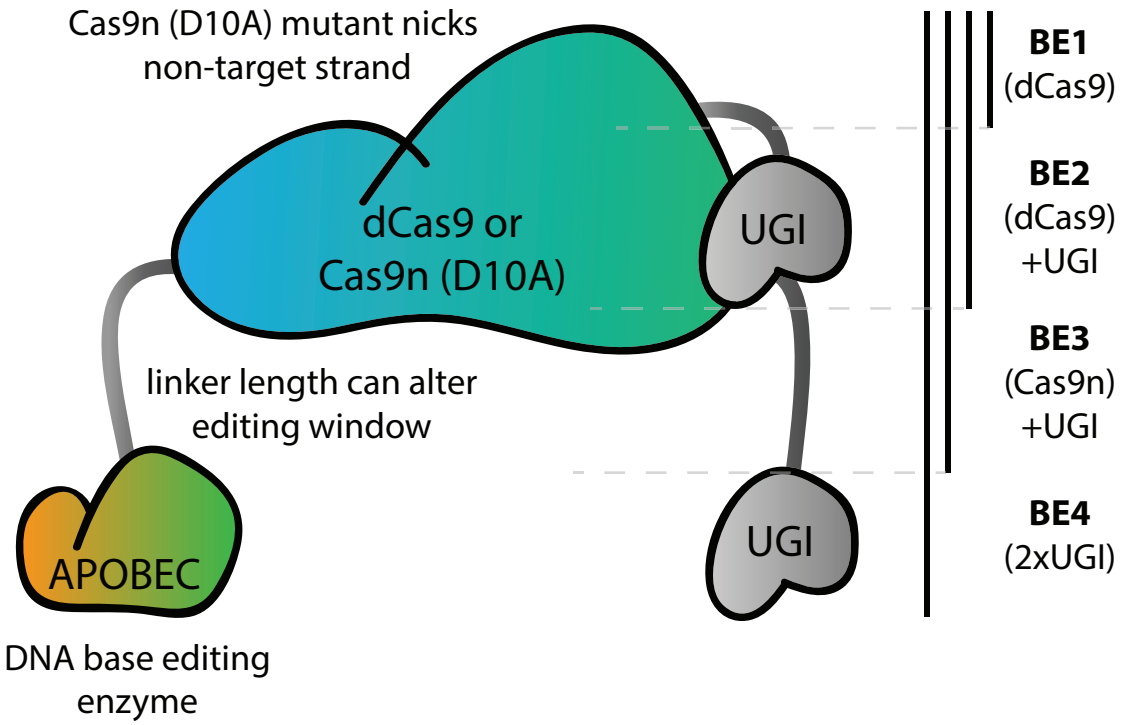


Figure 1.4. Model of cytosine base editing advancements

Schematic of the core components of current cytosine base editing technology, including Cas9 nickase (Cas9n), APOBEC and uracil glycosylase inhibitor (UGI); alongside modifications that have been introduced to improve editing efficiency and accuracy.

This figure was drafted by J. McCann and adapted from (143).

Chapter II.

The DNA Deaminase APOBEC3B Interacts with the Cell Cycle Protein CDK4 and Disrupts CDK4-Mediated Nuclear Import of Cyclin D1

This chapter was adapted with permission from: McCann et al. (2019) Journal of Biological Chemistry

Authors: Jennifer L. McCann[1-4], Madeline M. Klein[1], Evelyn M. Leland[1], Emily K. Law[1-5], William L. Brown[1-4], Daniel J. Salamango[1-4], and Reuben S. Harris[1-5]

Affiliations: [1]Department of Biochemistry, Molecular Biology and Biophysics, [2]Institute for Molecular Virology, [3]Masonic Cancer Center, [4]Center for Genome Engineering, [5]Howard Hughes Medical Institute, University of Minnesota, Minneapolis, MN 55455, USA.

Summary

Apolipoprotein B mRNA editing enzyme catalytic subunit-like protein 3B (APOBEC3B or A3B), as other APOBEC3 members, is a single-stranded (ss)DNA cytosine deaminase with antiviral activity. A3B is also overexpressed in multiple tumor types such as carcinomas of the bladder, cervix, lung, head/neck, and breast. A3B generates both dispersed and clustered C-to-T and C-to-G mutations in intrinsically preferred trinucleotide motifs (TCA/TCG/TCT). A3B-catalyzed mutations are likely to promote tumor evolution and cancer progression and, as such, are associated with poor clinical outcomes. However, little is known about cellular processes that regulate A3B. Here, we used a proteomics approach involving affinity purification coupled to mass spectrometry with human 293T cells to identify cellular proteins that interact with A3B. This approach revealed a specific interaction with cyclin-dependent kinase 4 (CDK4). We validated and mapped this interaction by co-immunoprecipitation experiments. Functional studies and immunofluorescence microscopy experiments in multiple cell lines revealed that A3B is not a substrate for CDK4-Cyclin D1 phosphorylation nor is its deaminase activity modulated. Instead, we found that A3B is capable of disrupting the CDK4-dependent nuclear import of Cyclin D1. We propose that this interaction may favor a more potent antiviral response and simultaneously facilitate cancer mutagenesis.

Introduction

Human apolipoprotein B mRNA editing enzyme catalytic subunit-like protein 3B (APOBEC3B or A3B)[6] is a member of a larger family of zinc-dependent cytosine deaminases that convert cytosine bases to uracils in single-stranded (ss)DNA (6,26,48,54,144,145). APOBEC3 family members have overlapping roles in restricting the replication of a variety of DNA-based viruses. For instance, several APOBEC3 family members including A3B inflict DNA C-to-U lesions in reverse-transcribing retroviruses and retrotransposons (23-26). In addition, A3B has elicited restriction activity against EBV, HBV, papillomaviruses, and polyomaviruses (19-23).

Recently, A3B has also been implicated as a source of mutation in various cancer types, with major influences in primary breast tumors (comprising 20% of base substitutions) and in metastatic disease (mutational contribution greater than 50%) (28,48-58). Additionally, A3B has been correlated with poor clinical outcomes including drug resistance (70-78). Efforts are being focused on developing therapies to inhibit A3B-mediated mutagenesis in cancer as an adjuvant to current treatment strategies (146). However, post-translational regulation of A3B enzymatic activity, the molecular mechanisms of how it gains access to ssDNA, and how it enters the nuclear compartment have remained elusive “black-boxes”.

Interestingly, A3B is the only family member shown to be constitutively nuclear (42-45). After mitosis and nuclear membrane reconstitution, A3B is rapidly reimported into the nuclear compartment where it remains constitutively present until the next mitotic cell division (42,43,46). Although the mechanism of nuclear import has yet to be fully elucidated, the identification of A3B-interacting factors may shed light in this process.

The cell cycle is comprised of a complex set of interacting proteins that tightly regulate progression through each phase. A key component of cell cycle regulation is the

family of serine/threonine cyclin-dependent kinases (CDK) (147-151). CDKs cooperate with various Cyclin proteins to regulate cell-cycle checkpoints (149). Cells must progress through four phases of the cell cycle to divide and replicate: G1, S phase (DNA synthesis), G2, and M phase (mitosis). The key regulator of the G1/S transition is CDK4, which forms a complex with Cyclin D1 (encoded by *CCND1*) and inactivates the retinoblastoma protein RB1 through phosphorylation of residue S780 (152,153). This relieves RB1-mediated inhibition of the transcription factor E2F, which commits the cell to progression through the cell cycle (151,154-156). Although there are many CDK-dependent points of cell cycle regulation, the regulation of the G1/S transition via CDK4-Cyclin D1 is perturbed in a large proportion of human cancers (157-160).

Our studies here began with the goal of identifying direct protein-protein interactors that post-translationally regulate A3B. We used affinity purification coupled to mass spectrometry (AP-MS) to identify CDK4 as a high-confidence A3B interactor. AP-MS results were validated by showing that A3B interacts directly with CDK4 but not with the closely related family members CDK2 or CDK6. Additionally, we use structure-guided mutagenesis to define the regions of both A3B and CDK4 required for interaction. We could also engraft key residues from CDK4 into CDK2 and endow it with the ability to interact with A3B. Finally, we demonstrated that both endogenous and exogenous expression of A3B could disrupt CDK4-dependent import of Cyclin D1, thereby providing a novel post-translational function for A3B.

Results

AP-MS revealed a specific interaction between A3B and the cell cycle regulator CDK4

To identify direct cellular regulators of A3B, we utilized AP-MS to identify A3B-interacting proteins. We expressed an A3B-2xStrep-3xFlag-tagged (hereafter referred to as A3B-SF) construct in 293T cells followed by affinity purification via anti-Strep resin and stringent high salt washes to enrich for tightly bound proteins. Immunoblot analysis, Coomassie gels, and activity assays using ssDNA oligonucleotide substrates were used to validate the presence, enrichment, and activity of affinity-purified A3B (schematic in **Fig. 2.1*A*** and data not shown). As controls, eGFP-SF, A3A-SF, and A3G-SF constructs were expressed in 293T cells and affinity purified in parallel. Co-immunoprecipitated samples were digested with trypsin and subjected to analysis by mass-spectrometry (**Fig. 2.1*A***). To our surprise, we identified CDK4 as a putative A3B interacting protein based on the recovery of multiple unique peptides across six independent AP-MS experiments (cumulatively spanning 24.7% of CDK4 amino acid sequence; **Fig. 2.1*B-C***). Additionally, CDK4 was not identified in the eGFP-SF, A3A-SF, or A3G-SF control experiments. Interestingly, even though CDK4 shares high homology with closely related family members (70% identity with CDK6 and 42% with CDK2), it was the only CDK and the only cell cycle protein identified in our AP-MS datasets. Furthermore, Cyclin D1, the canonical binding partner of CDK4, was not present in any of the A3B-SF AP-MS datasets.

Experimental validation of A3B-CDK4 interaction

To validate our AP-MS results implicating an A3B-CDK4 interaction, we performed an immunoprecipitation (IP) of eGFP-SF and A3B-SF in 293T cells. Immunoblot analysis confirmed that endogenous CDK4 could IP with A3B-SF but not with eGFP-SF (**Fig. 2.2*A***). A similar approach was taken to IP endogenous A3B; however, despite multiple attempts

in various cell lines using several commercial and custom in-house antibodies, immunoprecipitation of endogenous A3B was unsuccessful due to an overt failure to bind A3B, cross-reactivity to related APOBEC3s, and/or inferred low affinity (*i.e.*, all were non-IP grade reagents; data not shown).

To overcome these technical limitations, we further verified the interaction in co-immunoprecipitation (co-IP) experiments using two different sets of epitope-tagged A3B and CDK4 constructs. We expressed either Strep- or Flag-tagged A3B and CDK4 constructs in 293T cells and performed co-IPs probing for A3B first, CDK4 second, and *vice versa.* Immunoblot analysis of both Strep and Flag co-IPs revealed that, regardless of resin or epitope-tag, these two proteins reproducibly immunoprecipitated CDK4 (**Fig. 2.2*B***).

While our AP-MS approach was performed in the presence of both DNA and RNA nucleases, we sought to determine if the A3B-CDK4 interaction could occur in a heterologous system, which would indicate a direct protein-protein interaction. Utilizing a His-tagged full-length A3B catalytically dead construct ($A3B_{E255A}$) expressed in *E. coli*, we were able to successfully purify co-expressed CDK4 (**Fig. 2.2*C***). These results showed that no other human protein or specific RNA/DNA intermediate is required for the A3B-CDK4 interaction, such as the canonical CDK4 binding partner Cyclin D1.

Although CDK4 was not identified in our A3A or A3G AP-MS data sets, sequence homology between A3B and these APOBEC3 family members is high (89% homology between A3A and A3B C-terminal domain and 57% homology between A3G and A3B). Therefore, we sought to reconfirm our AP-MS results by directly examining the ability of A3A and A3G to interact with CDK4. As shown in **Fig. 2.2*D***, A3B is the only deaminase able to co-IP CDK4.

A3B is comprised of two deaminase domains, an N-terminal domain (NTD) and a C-terminal domain (CTD). The CTD is catalytically active and responsible for deaminating cytosines in ssDNA. The N-terminal domain (NTD) is catalytically inactive and implicated in protein oligomerization and mediating interactions with nucleic acids (81,161). To begin to determine the binding interface for CDK4, we performed comparative co-IPs with full-length Strep-tagged A3B ($A3B_{FL}$), $A3B_{NTD}$, and $A3B_{CTD}$ constructs. These experiments revealed that CDK4 could IP with $A3B_{FL}$ and $A3B_{NTD}$, but not with $A3B_{CTD}$ (**Fig. 2.2*D***).

To address whether this interaction was occurring endogenously within a cell, a proximity ligation assay (PLA) was performed using A3B, CDK4, and Cyclin D1 antibodies to visualize these interactions in MCF10A cells treated with 25 ng/ml phorbol 12-myristate 13-acetate (PMA) to induce A3B expression. Notably, A3B and CDK4 showed PLA foci in the nuclear compartment, whereas CDK4 and Cyclin D1 PLA foci were cell-wide (**Fig. 2.2*E***). In further support of our proteomics results, A3B and Cyclin D1 did not form PLA foci.

CDK4 does not affect A3B phosphorylation, localization or DNA deaminase activity

Interestingly, we noticed that $A3B_{FL}$ and $A3B_{NTD}$ displayed altered mobilities in co-IP experiments with overexpressed CDK4 (seen throughout **Fig. 2.2*A-D***). It is well known that altered mobility of protein species in SDS-PAGE can be attributed to post-translational modifications such as phosphorylation (162,163). This raised the possibility that A3B may be phosphorylated by CDK4.

We used kinase prediction algorithms [GPS 2.0 and Scansite 2.0 (164,165)] to identify putative phosphorylation sites in $A3B_{NTD}$ that could be targeted by CDK4. These programs identified a ^{95}TPxP^{98} motif within $A3B_{NTD}$ that closely resembles the S/TPxK consensus motif recognized by CDK4 (166-170). To determine if CDK4 was capable of

phosphorylating A3B, we performed an *in vitro* kinase assay by incubating purified RB1-3xHA (positive control), A3G-3xHA (negative control), A3B-3xHA, and two phospho-mutants at amino acid T95 (phospho-null $A3B_{T95A}$ and phospho-mimetic $A3B_{T95D}$) with active recombinant CDK4-Cyclin D1 in the presence of [γ-^{32}P]-ATP. Autoradiograph and immunoblot analysis revealed [γ-^{32}P] incorporation into RB1 but not into A3G or A3B, suggesting that the altered migration of A3B observed in **Fig. 2.2** is not due to CDK4-mediated phosphorylation (**Fig. 2.3*A***).

Several studies have demonstrated that $A3B_{NTD}$ contains *cis* elements that are required for active nuclear import (43-46). Since CDK4 is also actively imported into the nuclear compartment (153), we were curious if it is required for nuclear import of A3B. To address this idea, we challenged A3B nuclear import with knockdown of CDK4. Despite binding the localization domain of A3B, CDK4 depletion did not seem to alter the steady-state accumulation of A3B within the nuclear compartment (**Fig. 2.3*B***).

Next, because DNA deamination is the canonical activity of APOBEC3 enzymes, we asked whether CDK4 alters A3B deaminase activity. U2OS cells were engineered to stably express either an *A3B*-specific short hairpin (sh)RNA construct (shA3B) or a nonspecific shRNA construct as a control (shCTRL), and they were then transfected with either a *CDK4*-specific small interfering RNA (siRNA) or scrambled siRNA as a control. Using an established ssDNA deamination assay and immunoblots as additional controls, we observed that CDK4 depletion does not cause a major alteration in the enzymatic activity of endogenous A3B (**Fig. 2.3*C***).

$A3B_{NTD}$ residues required for CDK4 binding

We next sought to map the surface(s) of the NTD that mediate the interaction with CDK4. As illustrated in **Fig. 2.4*A-B***, structural knowledge of $A3B_{NTD}$ and A3A was used to inform the construction of a series of chimeric proteins to map the CDK4 interaction

(101,102,171). A3A was chosen as a chimeric partner because it displays cell-wide localization in many cell types, it is represented by multiple high-resolution crystal structures, and it does not interact with CDK4 (43,101,171,172) (**Fig. 2.2*D***). A comparison between A3A and $A3B_{NTD}$ structures (PDB 4XXO and 5TKM respectively) led us to focus on four surface exposed regions with major differences in amino acid composition (sections 1-4 in **Fig. 2.4*A***). Co-IPs and immunoblot analysis were performed with cells co-expressing CDK4 and $A3A/B_{1-4}$ chimeric constructs. Interestingly, $A3A/B_{1-3}$ proteins showed wild-type binding to CDK4, whereas $A3A/B_4$ had a compromised capacity to bind (**Fig. 2.4*C***). These results suggested that amino acid residues 80-85 may be required for mediating the interaction with CDK4 (highlighted in blue in **Fig. 2.4*B***). Because the $A3B_{NTD}$ nuclear localization surfaces defined by residues D19, E24, and T95-A102 do not overlap with these CDK4 binding residues, our results implicate a distinct A3B protein surface in the interaction with CDK4.

N-terminal region of CDK4 is required for A3B binding

As described above, CDK4 is unlikely to be using obvious mechanisms to regulate A3B activities. We therefore hypothesized instead that A3B could be regulating CDK4. As CDK4 family members serve different functions, we initially wanted to establish which family members bind to A3B. Similar to APOBEC3s, the CDK family shares high homology in protein sequence and function (*e.g.*, CDK4 and CDK2 alignment in **Fig. 2.5*A***). CDK4 is one of four classical cell cycle CDKs, sharing the most sequence homology with CDK6 and CDK2 (70% and 42% identity respectively), both of which regulate cell cycle progression (149,150). Co-IP of CDK4, 2 and 6 was performed in the presence of A3B followed by immunoblot analysis. As shown in **Fig. 2.5*B***, A3B bound to CDK4, but not to CDK6 or CDK2, further indicating a specific interaction.

Next, we wanted to delineate the CDK4 interaction surface to better understand what biological role the complex might play. We generated several reciprocal chimeras that exchanged surface exposed regions of CDK4 with CDK2. CDK2 was used as a chimeric partner for several reasons. First, CDK2 is negative for A3B binding. Second, available high-resolution crystal structures allow for structure-guided mutagenesis (147,173). Third, CDK2 binds an alternative cyclin (Cyclin E) (150) thereby providing clarity for lack of Cyclin D1 in proteomics and A3B-CDK4 binding. Chimeric junctions were guided by available crystal structures as well as by amino acid conservation between the protein coding sequences (**Fig. 2.5*A***). Co-IPs and immunoblot analysis were performed on cells co-expressing A3B and CDK4/$2_{A\text{-}C}$ chimeras. As described above, CDK2 serves as a negative control and does not bind A3B. Interestingly, all CDK4/$2_{A\text{-}C}$ chimeric constructs had the ability to bind A3B, indicating that region "A" (aa 1-96) is sufficient for binding (**Fig. 2.5*C***). To further support this finding, a reciprocal CDK2/4_A chimera was constructed and used for co-IP. This construct lost the ability to bind A3B, confirming that region "A" of CDK4 is necessary for binding to A3B (**Fig. 2.5*D***). Having identified a variant of CDK4 (CDK2/4_A) defective for binding A3B, we also sought to reconfirm that the ssDNA deaminase activity of A3B is not altered by this interaction. HeLa cells were transfected with either wild-type A3B-HA or the catalytic mutant ($A3B_{CM}$-HA) in the presence of either empty vector (ev) or Flag-tagged CDK constructs. As above, DNA deaminase activity assays and control immunoblots combined to show that neither CDK4 nor the A3B-binding mutant (CDK2/4_A) alters A3B deaminase activity (data not shown).

A3B disrupts CDK4-dependent nuclear import of Cyclin D1

Interestingly, the N-terminal region of CDK4 identified for A3B binding also contains the known CDK4-Cyclin D1 binding interface (152), which together with the aforementioned data suggests that these interactions may be mutually exclusive. This

observation may also explain why Cyclin D1 was not identified in our proteomics analysis, why no exogenous Cyclin D1 is needed for co-IP validations, and why $CDK4_A$/CDK2 (68% CDK2 residues) retains the ability to bind to A3B [even though CDK2 normally binds a different Cyclin (166)]. These insights, along with domain mapping of the A3B-CDK4 interaction above, support a model in which A3B and Cyclin D1 share an overlapping binding interface on CDK4.

To ask whether A3B binding to CDK4 could disrupt the nuclear import of Cyclin D1, A3B-eGFP was expressed in HeLa cells and immunofluorescence microscopy was used to examine Cyclin D1 localization. We observed that CDK4 and Cyclin D1 are both predominantly nuclear in the absence of exogenous A3B, as anticipated from prior studies (153) (see untransfected GFP-negative cells in **Fig. 2.6*A*** and quantification in **Fig. 2.6*B***). However, in the presence of A3B-eGFP, A3B and CDK4 were found to co-localize in the nuclear compartment, whereas Cyclin D1 became more cell-wide (see transfected A3B-eGFP-positive cells in **Fig. 2.6*A*** and quantification in **Fig. 2.6*B***). Additionally, A3B-eGFP expression caused significant accumulation of Cyclin D1 in foci that appeared to form within the nuclear compartment (**Fig. 2.6*C***). In support of these observations, endogenous A3B depletion in U2OS cells caused significant Cyclin D1 relocalization (increased nuclear to cytoplasmic expression ratios) and a trend toward fewer Cyclin D1 foci, in comparison to control short-hairpin RNA transduced cells (shA3B versus shCTRL in **Fig. 2.6*D-G***).

Discussion

A3B is unique among the human APOBEC3 family of DNA cytosine deaminase enzymes in that it constitutively localizes to the nucleus and is upregulated in several different cancer types. Here, we use a combination of AP-MS, structural-guided mutagenesis, genetic, and immunofluorescence microscopy approaches to demonstrate that CDK4 is a cellular A3B interacting protein. We show that A3B, and not homologous family members A3A and A3G, can specifically co-IP with CDK4. In addition, we demonstrate that A3B does not co-IP with other closely related CDK family members such as CDK2 and CDK6. Using structure-guided analysis of recently solved crystal structures and co-IP techniques we show that the interaction between A3B and CDK4 is mediated through the N-terminal domain of A3B (utilizing but not limited to aa 80-85) and the N-terminal region of CDK4 (aa 1-96). Additionally, we conclude that CDK4 is not likely to be regulating A3B via established mechanisms (CDK4-Cyclin D1 phosphorylation or CDK4-dependent nuclear import) or a potentially novel mechanism (DNA deamination inhibition). Lastly, PLA and immunofluorescence microscopy studies show A3B and CDK4 co-localization in the nuclear compartment, as well as aberrant Cyclin D1 localization.

Our results favor a model in which A3B may regulate CDK4-Cyclin D1 by sequestering CDK4 in the nuclear compartment, which in turn disrupts Cyclin D1 localization, likely via the secondary consequence of less available CDK4 for Cyclin D1 interaction (**Fig. 2.7**). A3B binds the N-terminal region of CDK4 and thereby disrupts CDK4-dependent import of Cyclin D1, likely due to overlapping and therefore competing binding surfaces. This is supported by studies that mapped the CDK4-Cyclin D1 binding interface and identified the N-terminal domain, specifically CDK4 residues 50-56 (PISTVRE), as the Cyclin D1 binding surface (152). Additional evidence supporting a mutually exclusive interaction for A3B-CDK4 is as follows: Cyclin D1 was not identified in

proteomics analysis; exogenous Cyclin D1 is not needed for co-IP validations; and $CDK4_A$/CDK2 (68% CDK2 residues) retained the ability to bind to A3B, even though CDK2 normally binds to a different Cyclin (166). Additionally, aberrant localization of Cyclin D1 is only observed in cells expressing A3B-eGFP as neighboring untransfected cells show normal nuclear enrichment for Cyclin D1. Our studies thereby demonstrate that A3B disrupts CDK4-dependent import of Cyclin D1, which may perturb cell cycle regulation and stall cells in G1/S.

It is possible that the A3B-CDK4 interaction benefits the innate antiviral response of cells. The APOBEC3s can restrict a wide variety of DNA-based pathogens and have been shown to be induced by viral infection (7,19,21,47,82,174-176). Divergent viral families manipulate the cell cycle in an attempt to optimize viral growth conditions and/or viral fitness. Manipulation of the cell cycle for optimized viral fitness may aid in overcoming mechanisms of APOBEC3-mediated restriction. For example, the EBV protein BORF2 causes cell cycle arrest of host cells undergoing lytic viral DNA replication, and it also binds and inhibits A3B in order to prevent hypermutation of the EBV genome (19,177). As a second example, the HIV-1 Vif protein has been reported to cause host cells to arrest in the G2 phase of the cell cycle, and it too binds APOBEC3 proteins and targets them for proteasomal degradation in order to escape restriction (178,179). Additionally, irregular cell cycle phenotypes such as cell cycle arrest, cell cycle delay, and aborted cell division (multi-nucleation) have been reported in the literature for A3B but remain poorly understood (43,180-182).

Interestingly, many viruses that are susceptible to DNA deamination by A3B, including EBV and HPV, also subvert the cell cycle to their advantage. For instance, EBNA3C (encoded by EBV) and E7 (encoded by HPV) are able to perturb the control of G1/S phase transition in order to replicate their genomes at the same time as cellular DNA

replication (183-185). Other viruses can induce a G2/M arrest to provide an optimized cellular environment for maximal levels of viral replication (186-188). One speculation for how this might facilitate an antiviral response is by altering the cell cycle to expose more viral ssDNA replication intermediates to A3B mutagenesis. One example for this is the recent study on A3B and the herpesvirus EBV (and KSHV) where hypermutation of the viral genome is driven by A3B during lytic replication (19). In addition to EBV, studies on BK and JC polyomaviruses have indicated that the preferred trinucleotide motifs of A3B are depleted from viral genomes, specifically on the DNA strand that is most likely to be single-stranded [*i.e*. the nontranscribed strand or the DNA replication lagging strand template (20)]. Despite these and other possible models, full elucidation of the molecular mechanism and potential benefits of the A3B-CDK4 interaction to the cell will require systematic studies with different viruses in pathologically relevant systems.

It is also possible that the A3B-CDK4 interaction plays a role in healthy, uninfected cells. However, *A3B* expression appears to be kept at very low levels in normal cell types, and it is more likely that the cell-cycle phenotype described here could become exacerbated upon A3B overexpression in cancer (potentially benefitting the evolvability of different tumor cell types). One canonical role of CDK4 is to control the nuclear import of Cyclin D1, which facilitates the downstream progression of the G1/S phase transition (153,189-194). Several different cancer types have been shown to upregulate A3B, which may lead to G1/S phase stalling via its interaction with CDK4. As the G1/S phase transition involves numerous points of exposed ssDNA (DNA replication, transcription, and DNA repair), this may result in increased substrate availability and could lead to increased genomic DNA deamination, non-clustered and clustered mutation, and overall genomic instability. Although a role for A3B in cancer mutagenesis and evolvability has been well documented, universally accepted genomic substrates have yet to be described.

Conflicting studies show A3B can target transcription-associated substrates as well as replication and recombination-associated substrates (36,52,80-82). Therefore, it is plausible that A3B-mediated mutagenesis can be exacerbated during G1/S stalling, which may create more exposed ssDNA substrates.

Methods

Constructs – Strep and Flag epitope tagged constructs were cloned into pcDNA4/TO using standard PCR cloning techniques. C-terminal 2xStrep and/or 3xFlag tags were cloned into pcDNA4/TO using *Xho*I-*Apa*I. N-terminal 3xFlag tag was cloned into pcDNA4/TO using *Hind*III-*Not*I. The full set of pcDNA3.1 (+) human APOBEC-HA expression constructs have been described (7) [A3A (GenBank accession NM_145699), A3B (NM_004900), A3G (NM_021822)]. CDK2 (BC003065.2), CDK4(CR542247.1), and CDK6 (BC052264) were cloned into pcDNA4/TO-2xStrep using *Hind*III-*Not*I and N-terminally tagged pcDNA4/TO-3xFlag using *Not*I-*Xba*I. RB1 (NM_000321.2) was cloned in the pcDNA3.1 (+) – 3xHA using *Hind*III-*Kpn*I. A3B phosphorylation mutants were generated using standard PCR based site-directed mutagenesis techniques. The A3Bi-GFP construct has been described (43). A3B-eGFP-pQCXIH was made by cloning A3B-eGFP into pQCXIH using *Not*I-*Age*I. Chimeric A3A/$B_{1\text{-}4}$, CDK4/$2_{A\text{-}C}$, and CDK2/4_{A} constructs were cloned using standard overlap PCR methods. Bacterial expression constructs were cloned into pRSF-Duet using *Sbf*I and *Hind*III for A3B and *Nae*I and *Pac*I for CDK4. *A3B* knockdown and control shRNA constructs were validated previously (48,98).

Cells – 293T and HeLa cells were cultured in RPMI 1640 supplemented with 10% fetal bovine serum (FBS) and penicillin-streptomycin. 293T cells stably expressing A3B-eGFP were made by transducing 293T cells with A3B-eGFP-pQCXIH virus following an established protocol **(195)**. MCF10A cells were cultured in DMEM/F12 supplemented with 5% equine serum, EGF (20 ng/ml), insulin (10 μg/ml), hydrocortisone (0.5 mg/ml), cholera toxin (100 ng/ml), and penicillin-streptomycin. U2OS cells were cultured in McCoy's 5A medium supplements with 10% FBS and penicillin-streptomycin.

Affinity purification - mass spectrometry– 293T cells were transfected with pcDNA4/TO-A3B-2xStrep-3xFlag, A3A-2xStrep-3xFlag, A3G-2xStrep-3xFlag or eGFP-2xStrep-3xFlag using Transit LT1 (Mirus). Cells were harvested in 1xPBS 48 hours post-transfection. Cells were washed two times in 1xPBS followed by lysis [50 mM Tris-HCl pH 8.0, 1% Tergitol, 150 mM NaCl, 0.5% sodium deoxycholate, 0.1% SDS, 1 mM DTT, 1x Protease Inhibitor (Roche), RNase and DNase]. Lysates were subjected to sonication prior to clearing by centrifugation. Cleared lysates were then added to Strep-Tactin Superflow resin (IBA) followed by end-over-end rotation for 2 hours at 4°C. Following IP, the anti-Strep resin was washed three times in high salt wash buffer [20 mM Tris-HCl pH 8.0, 1.5 mM $MgCl_2$, 1 M NaCl, 0.2% Tergitol, 0.5 mM DTT and 5% glycerol] followed by three washes in low salt wash buffer [same as high salt but with 150 mM NaCl]. To remove detergents for proteomics submission, samples were subjected to three washes of no-detergent wash buffer [20 mM Tris-HCl pH 8.0, 1.5 mM $MgCl_2$, 150 mM NaCl, 0.5 mM DTT and 5% glycerol]. Protein was eluted from the resin in elution buffer [100 mM Tris-HCl pH 8.0, 150 mM NaCl and 2.5 mM desthiobiotin]. Samples were validated using immunoblotting, DNA deaminase activity assays (discussed below), and Coomassie staining. In-solution samples were analyzed by LC-MS/MS at the Harvard Proteomic Core.

Co-immunoprecipitation experiments and immunoblotting – Semi-confluent 293T cells were transfected with plasmids using TransIT-LT1 (Mirus) per manufacturer's protocol. Cells were harvested in 1xPBS 48 hours post-transfection. Cells were washed two times in 1xPBS followed by lysis [150 mM NaCl, 50 mM Tris-HCl pH 8.0, 0.5% Tergitol, 1x Protease inhibitor (Roche), RNase and DNase]. Cells were vortexed vigorously and incubated at 4°C for 30 minutes prior to clearing by centrifugation. Cleared lysates were then added to Strep-Tactin Superflow resin (IBA) or anti-Flag M2 Magnetic Beads (Sigma M8823) followed by end-over-end rotation overnight at 4°C. Beads were then washed

three times in lysis buffer followed by elution in elution buffer [lysis buffer + 0.15 mg/mL Flag Peptide (Sigma) or 2.5 mM desthiobiotin]. Protein was analyzed by immunoblot. Antibodies used include mouse anti-Flag 1:10,000 (Sigma F1804), rabbit anti-HA 1:5000 (Cell Signaling C29F4), rabbit anti-APOBEC3B 1:2000 [5210-87-13 (196)], rabbit anti-Strep Tag II 1:5000 (Abcam, ab76949), and mouse anti-CDK4 1:2000 (Santa Cruz, sc-23896).

DNA deaminase activity assays – Deamination reactions were performed at 37°C for 2 hours using purified protein, 4 pmol of oligonucleotide(5′-ATTATTATTATTCAAATGGATTTATTTATTTATTTATTTATTT-fluorescein), 0.025 units uracil DNA glycosylase (UDG), 1x UDG buffer (NEB), and 1.75 units RNase A, as described in **(70)**. Reaction mixtures were treated with 100 mM NaOH at 95°C for 10 min to achieve complete backbone breakage. Reaction mixtures were separated on 15% Tris-borate-EDTA (TBE)-urea gels to separate substrate from product. Gels were scanned using a Fujifilm FLA-7000 image reader.

***Protein purification from* E. coli** – pRSF-Duet-A3B$_{E255A}$ and pRSF-Duet-A3B$_{E255A}$-CDK4 were transformed into BL21 $CaCl_2$ *E. coli* and colonies were used to inoculate 1L 2xYT media. Cultures were grown to an OD of 0.5 and 50 μM Zn sulfate was added 30 minutes prior to induction with ITPG to a final concentration of 0.5 mM. Induced cultures were grown overnight at 16°C and pelleted. Pellets were resuspended in lysis buffer [1% TritonX-100, 50 mM HEPES (pH 7.4), 150 mM NaCl, 1x Protease Inhibitor (Roche), 0.1 mg/ml RNase A, 0.1 mg/ml DNase, 100 μg/ml lysozyme, 5 mM imidazole and 1.5 mM $MgCl_2$] incubated on ice for 10 minutes and sonicated (Branson Sonfier). Lysate was cleared with centrifugation and insoluble pellet and input lysate were collected and boiled for immunoblot. Cleared lysate was subjected to protein purification using Ni-NTA Agarose (Qiagen). Bound protein was washed in lysis buffer (without RNase A, DNase and

lysozyme) and purified protein was eluted by boiling. Samples were run for immunoblot analysis for confirmation of expression and purification.

mRNA quantification – *CDK4* mRNA was quantified by RT-qPCR relative to the stable housekeeping transcript, *TBP,* using specific primers and probed (Roche Lightcycler). RNA isolation and RT-qPCR methods have been described **(27)**.

siRNA knockdown – siRNA targeting CDK4 (Dharmacon, L-003238-00) and control siRNA (Dharmacon, D-001810-10-05) were purchased and diluted to a working concentration of 20 μM. A final concentration of 20 nM was used in 293T cells stably expressing A3B-eGFP. siRNAs were delivered to cells using Lipofectamine RNAiMAX (Thermo Fisher Scientific) per manufacturer's protocol.

Proximity ligation assay – MCF10A cells were treated with 25 ng/ml of PMA for 18 hours. Cells were fixed with 4% paraformaldehyde in phosphate buffered saline (PFA in PBS) for 15 minutes at room temperature. Cells were permeabilized using 0.5% Triton X-100 for 10 minutes at 4°C. The remainder of proximity ligation assay was performed using Duolink In Situ Red Starter Kit Mouse/Rabbit (Sigma, DUO92101). Antibodies were mouse anti-CDK4 1:50 (Santa Cruz, sc-23896), rabbit anti-Cyclin D1 1:100 (Abcam, ab134175), rabbit anti-APOBEC3B (1:50) [5210-87-13 **(196)**], and mouse anti-Cyclin D1 1:100 (Santa Cruz, sc-20044).

In vitro *kinase assay* – Purified RB1-3xHA and A3-3xHA proteins were produced in in 293T cells, purified using Anti-HA magnetic beads (ThermoFisher 88836), and validated using immunoblot techniques. Recombinant CDK4-Cyclin D1 proteins were produced in Baculovirus-infected Sf9 cells (Abcam, ab55695). Purified proteins were incubated with 0.1 μg of recombinant CDK4-Cyclin D1 in kinase reaction buffer [50 mM HEPES (pH 7.5), 10 mM $MgCl_2$, 2.5 mM EGTA, 20 μM DTT, 0.1 mM ATP and 5.4 μCi [γ-^{32}P]-ATP

(PerkinElmer) at 30°C for 2 hours. Reactants were subjected to SDS-PAGE and proteins were detected by immunoblot while phosphorylated proteins were detected by autoradiography.

Immunofluorescence microscopy – Microscopy procedures have been described **(197)**. HeLa cells were seeded on glass coverslips in 6 well plates and were transiently transfected with 500 ng each of A3B-eGFP or empty vector (Mirus, Transit-LT1) and incubated for 48 hours. The cells were fixed with 4% PFA in PBS for 15 minutes at room temperature. Cells were permeabilized using 0.5% Triton X-100 for 10 minutes at 4°C. Slides were blocked in 5% normal goat serum, 4% bovine serum albumin in PBS for 1 hour at room temperature. Primary antibodies were diluted in blocking buffer and slides were incubated in primary antibody overnight at 4°C. Antibodies were mouse anti-CDK4 1:50 (Santa Cruz, sc-23896), rabbit anti-Cyclin D1 1:100 (Abcam, ab134175), anti-rabbit Cy5 (Abcam, ab6564), and anti-mouse AlexaFluor 594 (Invitrogen, A11032). All slides were treated with 0.1% Hoechst dye to stain nuclei. The slides were mounted with 50% glycerol and imaged using a Nikon inverted Ti-E deconvolution microscope. All images were analyzed for foci and nuclear-to-cytoplasmic ratio using ImageJ-Fiji Software and P values were calculated using an unpaired student's *t*-test.

Additional contributions

We thank Chuanmao Zhang for CDK4 constructs and Ivan D'Orso for pcDNA4/TO Strep and Flag tagged base vectors. This work was supported by the University of Minnesota Masonic Cancer Center, Academic Health Center, and College of Biological Sciences. Salary support for JLM was provided in part by the NSF Graduate Research Fellowship (Grant Number 00039202). We thank the University of Minnesota Microbiology, Immunology and Cancer Biology Graduate Program for academic support. RSH is the

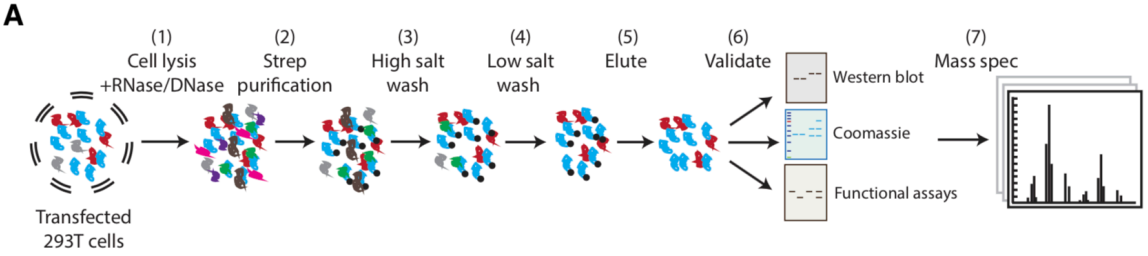
Margaret Harvey Schering Land Grant Chair for Cancer Research, a Distinguished University McKnight Professor, and an Investigator of the Howard Hughes Medical Institute. A subset of the imaging studies was done with a Nikon inverted Ti-E deconvolution microscope provided by the University of Minnesota - University Imaging Centers.

Conflicts of interest

RSH is a co-founder, shareholder, and consultant for ApoGen Biotechnologies Inc. The other authors have no conflicts of interest to declare.

Figures

A



B

Protein Name	kDa	Accession	SC	Coverage
APOBEC3B	46	Q9UH17	161	59.6%
CDK4	34	P11802	15	24.7%

C

```
MATSRYEPVA EIGVGAYGTV YKARDPHSGH FVALKSVRVP NGGGGGGGLP  50
ISTVREVALL RRLEAFEHPN VVRLMDVCAT SRTDREIKVT LVFEHVDQDL 100
RTYLDKAPPP GLPAETIKDL MRQFLRGLDF LHANCIVHRD LKPENILVTS 150
GGTVKLADFG LARIYSYQMA LTPVVVTLWY RAPEVLLQST YATPVDMWSV 200
GCIFAEMFRR KPLFCGNSEA DQLGKIFDLI GLPPEDDWPR DVSLPRGAFP 250
PRGPRPVQSV VPEMEESGAQ LLLEMLTFNP HKRISAFRAL QHSYLHKDEG 300
NPE
```

Figure 2.1. AP-MS reveals an interaction between A3B and CDK4

A, Schematic of the AP-MS workflow used to identify A3B interacting proteins.

***B*,** Summary of the total spectral counts (SC) and percent coverage for A3B and CDK4.

C, Amino acid sequence of CDK4 with AP-MS peptides highlighted in red.

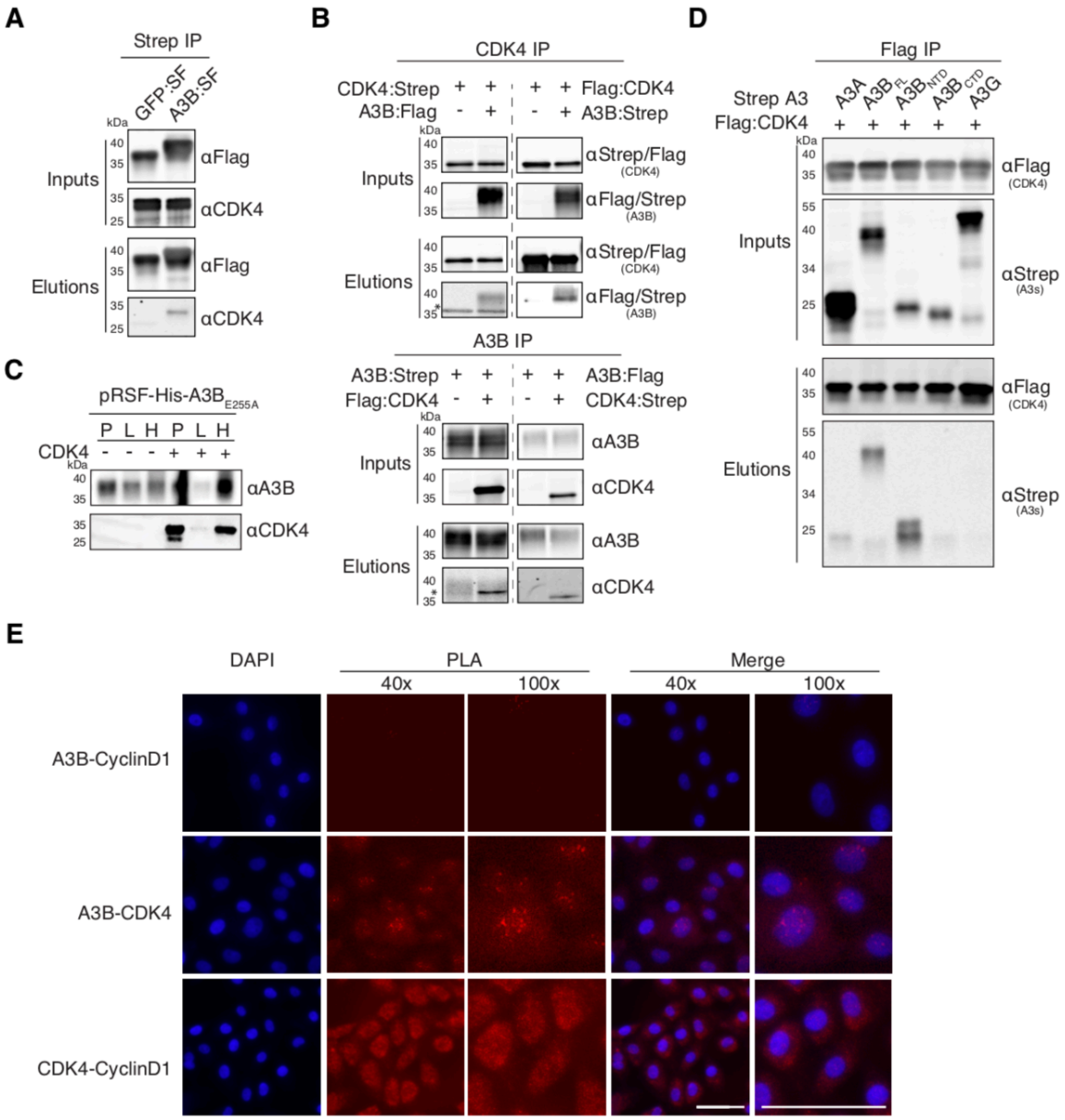


Figure 2.2. Confirmation of the A3B-CDK4 interaction through IP experiments

A, Co-IP of endogenous CDK4 in 293T cells transfected with A3B-SF. Parallel reactions with eGFP-SF serve as a negative control. Upper two immunoblots show the indicated proteins in whole cell lysates (input), and the bottom two immunoblots show the

immunoprecipitated samples (elution). kDa markers are shown to the left of each blot and the primary antibody used for detection is shown to the right.

B, Co-IP results for Strep- and Flag-tagged CDK4 (upper blots) and Strep- and Flag-tagged A3B (bottom blots) from 293T cells. The asterisk denotes residual signal from the IP antibody due to insufficient stripping.

C, Immunoblots showing an interaction between *E. coli* expressed His-tagged $A3B_{E255A}$ and untagged CDK4. Immunoblot labels are as follows: insoluble pellet (P), lysate (L) and His-purification (H).

D, Co-IP of the indicated Flag-tagged CDK4 constructs with Strep-tagged A3 constructs in 293T cells.

E, Fluorescence microscopy images of PLA results showing the interaction between A3B and CDK4 as well as CDK4 and Cyclin D1 in MCF10A cells treated with PMA to induce A3B expression (scale bar, 50 μm).

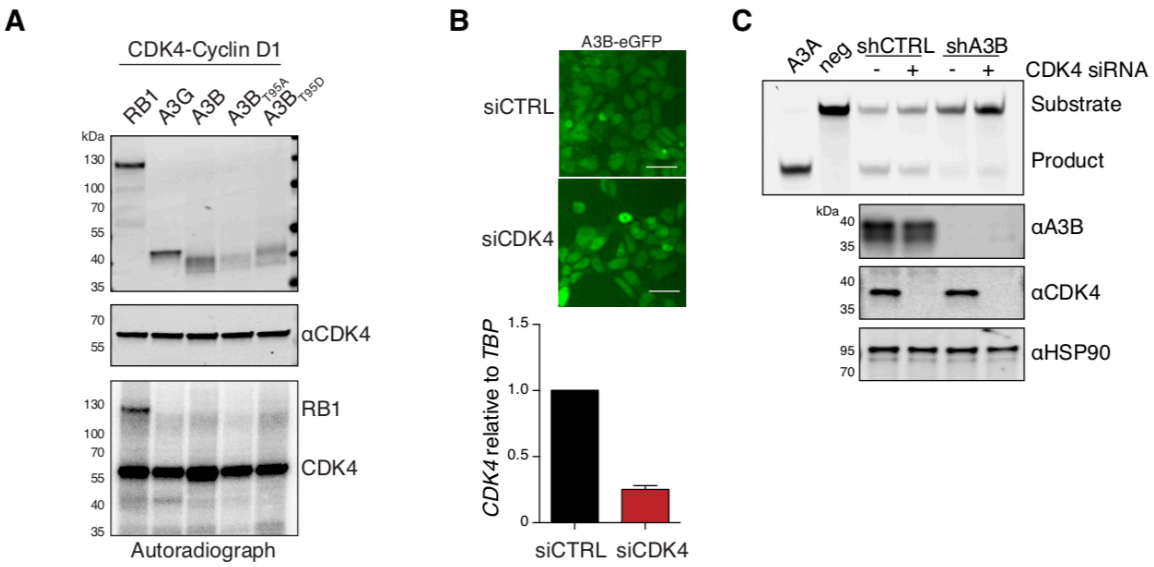


Figure 2.3. Multiple A3B properties are not altered by CDK4

A, *In vitro* kinase assay of indicated HA-tagged proteins in the presence of recombinant CDK4-Cyclin D1 and [γ-^{32}P]-ATP. The top two panels show immunoblots of the indicated proteins, and the bottom panel shows a representative autoradiograph of the same reaction following separation by SDS-PAGE.

B, Fluorescence microscopy of 293T cells stably expressing A3B-eGFP following transfection with siCtrl or siCDK4 RNAs (scale bar, 20 μm). Histogram showing *CDK4* mRNA levels from RT-qPCR relative to *TBP* mRNA (TATA-binding protein).

C, PAGE analysis of DNA deamination reaction products following incubation of ssDNA substrates with U2OS whole cell extracts plus/minus A3B and CDK4 depletion as indicated (see text and methods for details). Recombinant A3A and no APOBEC3 provide positive and negative controls, respectively. Immunoblot controls are shown below.

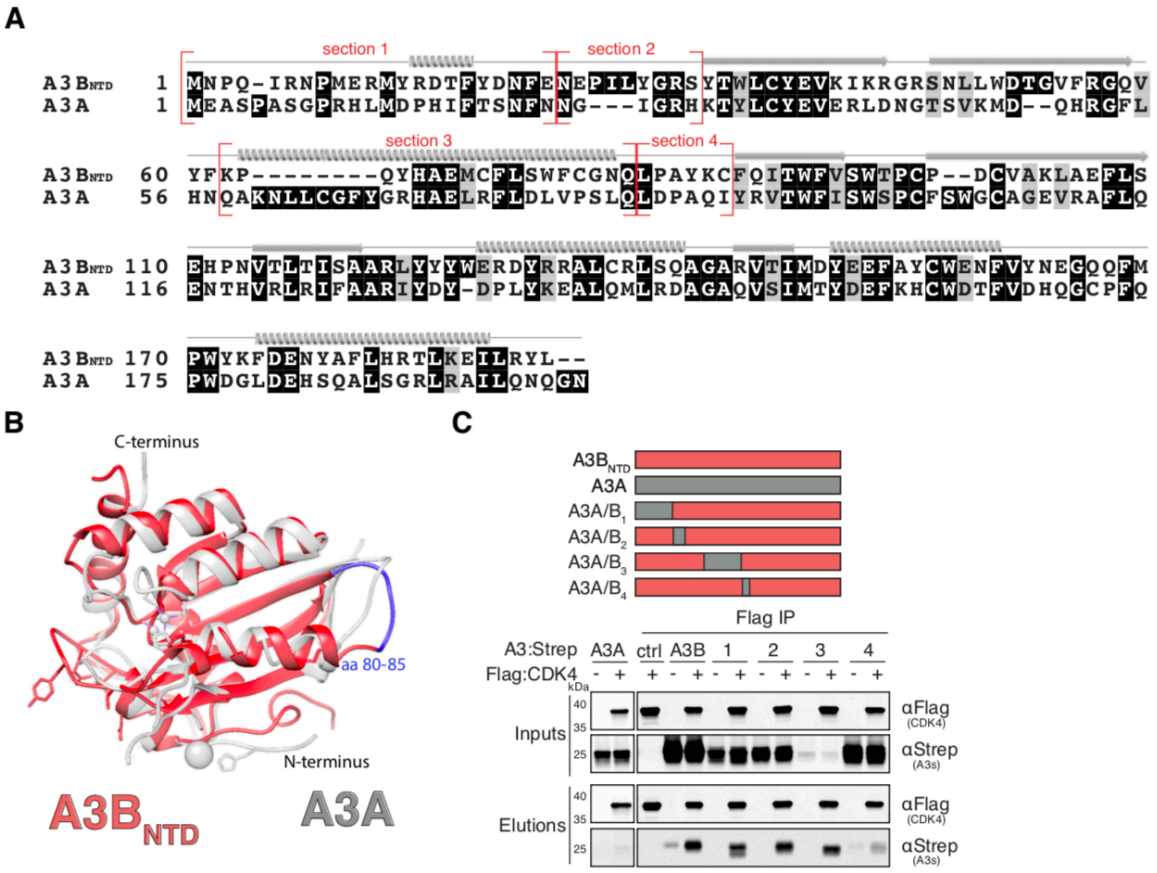


Figure 2.4. CDK4 interacts with a N-terminal region of A3B

A, Amino acid alignment of $A3B_{NTD}$ and A3A with chimeric junctions indicated in red and structural elements shown above the alignment (alpha helices, beta strands, and loop regions).

B, Ribbon schematic overlay of the crystal structures of $A3B_{NTD}$ (red, PDB 5TKM) and A3A (grey, PDB 4XXO). $A3B_{NTD}$ residues 80-85 are highlighted in blue (see text for details).

C, Anti-Flag (CDK4) co-IP of indicated Strep-tagged $A3B_{NTD}$ constructs from 293T cells. Parallel reactions were done with empty vector and A3A-S as negative controls. See Fig. 2*A* legend for a description of the IP labeling scheme.

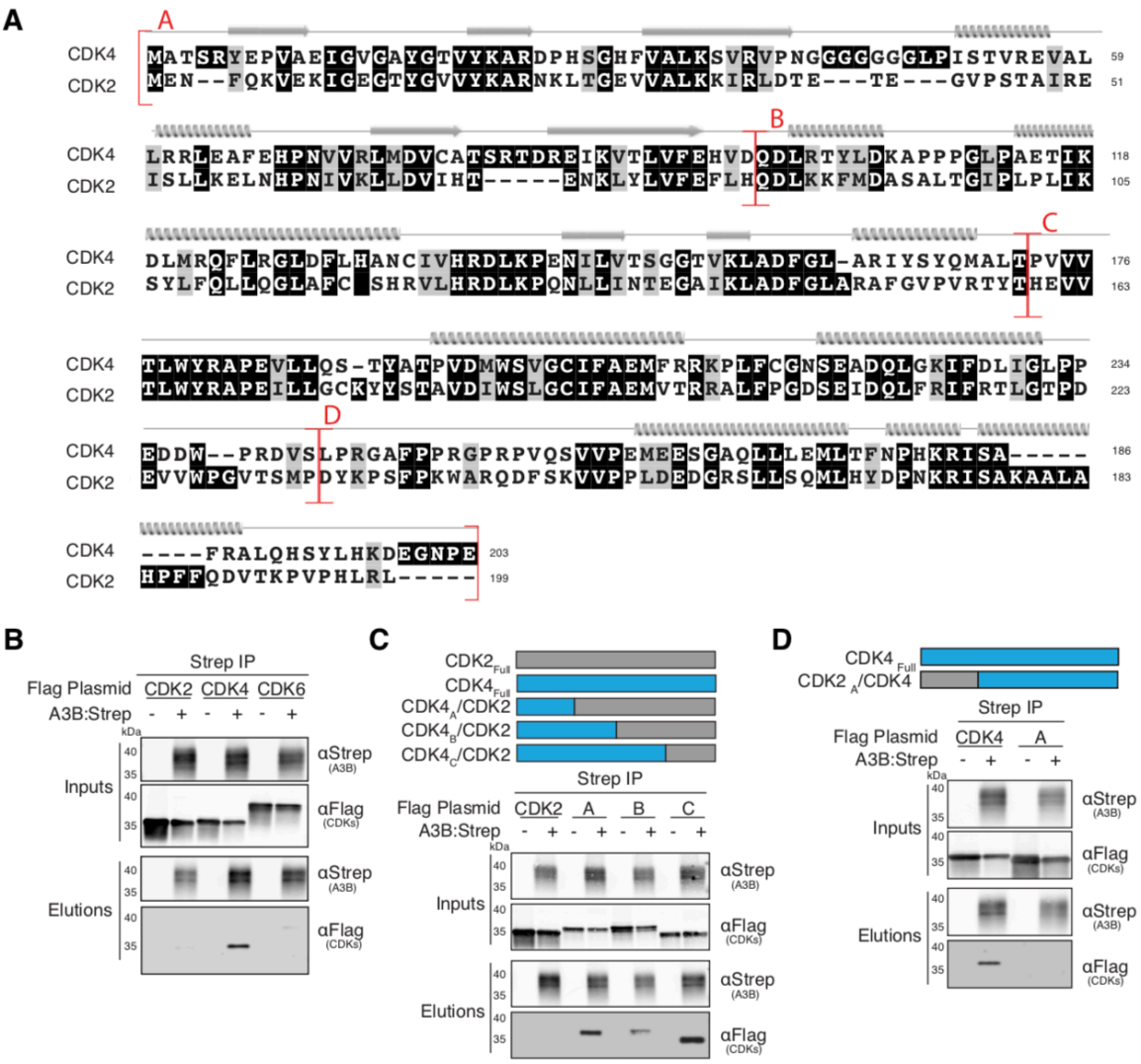


Figure 2.5. N-terminal region of CDK4 is required for A3B binding

A, Amino acid alignment of CDK4 and CDK2 with chimeric junctions indicated in red and structural elements shown above the alignment (alpha helices, beta strands, and loop regions).

B-D, Anti-Strep (A3B) co-IPs of the indicated Flag-tagged CDK constructs from 293T cells. See Fig. 2*A* legend for a description of the IP labeling scheme.

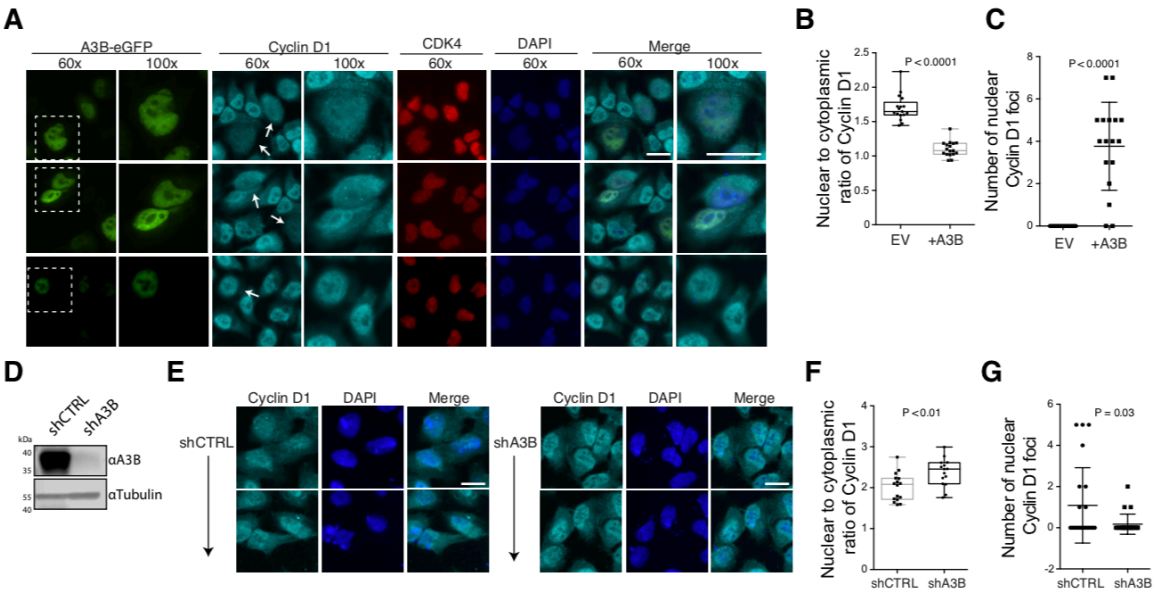


Figure 2.6. A3B disrupts CDK4-dependent nuclear import of Cyclin D1

A, Three sets of representative immunofluorescence microscopy images of HeLa cells transfected with A3B-eGFP and stained for endogenous Cyclin D1, endogenous CDK4, and DAPI (60x magnification image scale bar, 20 μm). A 100x magnification of boxed regions is also provided for Cyclin D1-stained, A3B-eGFP expressing cells for better visualization of Cyclin D1 redistribution and foci (white arrows).

B-C, Whiskered dot plots for quantification of nuclear to cytoplasmic ratio of Cyclin D1 as well as number of Cyclin D1 foci, respectively (n=20 cells; p-values calculated using an unpaired student's *t*-test).

D, Immunoblots showing A3B levels in U2OS cells stably expressing an *A3B*-specific short hairpin RNA construct (shA3B) or a nonspecific shRNA construct as a control (shCTRL).

E, Two sets of representative immunofluorescence microscopy images of U2OS cells stably expressing shCTRL or shA3B constructs and stained for endogenous Cyclin D1 and DAPI (scale bar, 10 μm).

***F-G*,** Whiskered dot plots for quantification of nuclear to cytoplasmic ratio of Cyclin D1 as well as number of Cyclin D1 foci, respectively (n=20 cells; p-values calculated using an unpaired student's *t*-test).

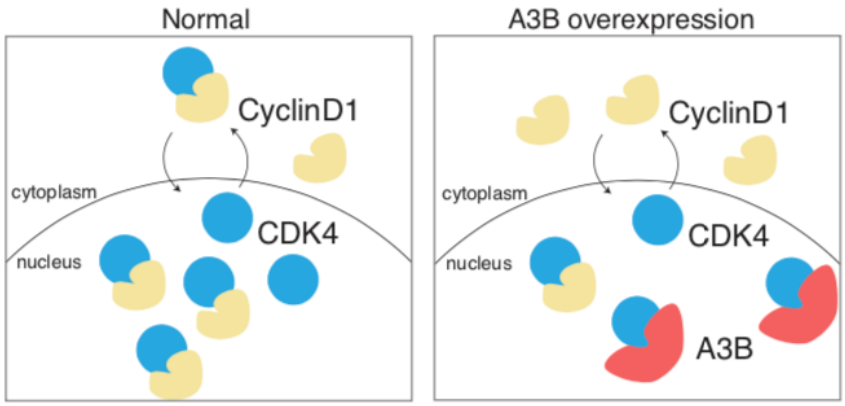


Figure 2.7. Working model for A3B-CDK4 interaction

A cartoon depicting the normal nucleo-cytoplasmic shuttling of CDK4 and CDK4-Cyclin D1 complexes (left panel). In comparison, A3B-mediated sequestration of CDK4 in the nuclear compartment causes a cell-wide distribution of Cyclin D1 (right panel).

Chapter III.

MagnEdit – Interacting Factors that Recruit DNA Editing Enzymes to Single Base Targets

This chapter was adapted with permission from: McCann et al. in preparation for Nature Biotechnology

Authors: Jennifer L. McCann[1-4], Daniel J. Salamango[1-4], Emily K. Law[1-5], William L. Brown[1-4], , and Reuben S. Harris[1-5]

Affiliations: [1]Department of Biochemistry, Molecular Biology and Biophysics, [2]Institute for Molecular Virology, [3]Masonic Cancer Center, [4]Center for Genome Engineering, [5]Howard Hughes Medical Institute, University of Minnesota, Minneapolis, MN 55455, USA.

Summary

Although CRISPR/Cas9 technology has created a renaissance in genome engineering, particularly for gene knockout generation, methods to introduce precise single base changes are also highly desirable (139,198). The covalent fusion of a DNA editing enzyme such as APOBEC to a Cas9 nickase (Cas9n) complex has heighted hopes for such precision genome engineering (114,117). However, current cytosine base editors (CBEs) are prone to undesirable off-target mutations (most frequently target-adjacent mutations)(114,117,138,139). Here, we report a method to “attract” an endogenous DNA deaminase, APOBEC3B (A3B), to a target cytosine base for specific editing with minimal damage to the rest of the cellular genome. The key to this system is fusing an APOBEC-interacting protein (not APOBEC itself) to Cas9n, which attracts nuclear A3B transiently to the target site for editing. Several A3B interactors were tested and one, hnRNPUL1, demonstrated proof-of-concept with successful C-to-T editing of episomal and chromosomal substrates with negligible target-adjacent affects.

Article

The original CBE (BE3) is comprised of the rat APOBEC1 deaminase fused to the N-terminal end of a Cas9 nickase [Cas9n (D10A)](114). Appropriate guide (g)RNAs are able to target this assembly to genomic cytosine bases and facilitate high frequency editing (10-90% depending on a number of variables including distance between target C and PAM) (114,117). However, this technology is prone to a number of off-target effects including RNA editing (135,199), random genomic DNA editing (129-132), and most frequently target-adjacent editing (114,129,136,137). The latter problem is due to deamination of single-stranded (ss)DNA cytosines located adjacent to the desired target cytosine in the same gRNA-displaced R-loop. This issue has been diminished – but not eliminated – by mutating APOBEC1 (137,200), trying different DNA deaminase family members (133,134,201,202), mutating Cas9 (200,203,204), and trying different Cas enzymes/complexes(200,205-207). However, an invariant feature of all current designs is *covalent* fusion of the deaminase to the Cas9 complex, which traps the tethered deaminase locally and inextricably links both on-target and target-adjacent cytosine deamination events (schematic in **Fig. 3.1*A***).

We hypothesize that *non-covalent* methods to "attract" a DNA cytosine deaminase to a particular genomic cytosine target will decouple the fates of on-target and target-adjacent editing events and thereby enhance the likelihood of achieving precise single base substitution mutations. A key to implementing this non-covalent strategy is identifying appropriate APOBEC-interacting proteins, which bind the deaminase without blocking access to the active site. Such interacting proteins can then be tethered to a Cas9n/gRNA complex and used to "attract" a co-expressed APOBEC enzyme (exogenous or

endogenous) to edit a particular genomic target cytosine. Inspired by an analogy to magnetism, this novel system is called MagnEdit (schematic in **Fig. 3.1*B***).

As an initial test of MagnEdit, we fused A3B-interacting proteins from the literature [SIV Vif (111), hnRNPK (110)] and proteomic screens [CDK4 (113) and hnRNPUL1 (McCann *et al*., unpublished)] to the N-terminal end of Cas9n and asked whether these complexes are able to recruit A3B to edit an episomal *eGFP* reporter (134) in 293T cells (TC to TT schematic in **Fig. 3.1*B*** and actual *eGFP* gRNA target sequence in **Fig. 3.1*C*** inset). Due to simultaneous overexpression of reaction components following co-transfection, including A3B, a low level of eGFP-positive cells (~1-2%) was observed in the absence of a gRNA and a candidate interacting protein (reactions represented by gray and black bars in **Fig. 3.1*C***). Interestingly, addition an eGFP Leu202-targeting gRNA (again without an interactor) enabled higher levels of eGFP editing by A3B (~5-7%; empty Cas9n plus gRNA reaction in **Fig. 3.1*C***). Unfortunately, most MagnEdit complexes failed to stimulate editing beyond these background levels or those caused by a non-interacting blue fluorescent protein (BFP)-Cas9n control (**Fig. 3.1*C***). SIV Vif (SLQ-AAA)-Cas9n even yielded lower overall frequencies of background editing, likely due to poorer expression relative to other MagnEdit constructs [the SLQ-AAA was necessary to prevent Vif from binding ELOC and triggering A3B degradation (111)]. However, one MagnEdit construct, hnRNPUL1-Cas9n, was clearly capable of recruiting A3B in a dose-dependent manner to catalyze editing and activation of the eGFP reporter (**Fig. 3.1*C***). Editing frequencies due to hnRNPUL1-Cas9n were at least 2-fold higher than the BFP-Cas9n/gRNA-induced background in these transient transfection experiments ($p<0.0001$ by unpaired student's t-test).

Next, we analyzed chromosomal DNA editing by MagnEdit. The same eGFP Leu202 reporter was integrated into the genome of 293T cells by low MOI lentiviral transduction followed by hygromycin selection to ensure that every cell has one editing target (uniform mCherry-positive population confirmed by flow cytometry). This pool was then transfected, as above, with the panel of A3B interactor-Cas9n complexes with or without the Leu202 targeting gRNA in the presence or absence of exogenous A3B. Also, as above, empty-Cas9n and BFP-Cas9n were used as negative controls, and most MagnEdit complexes showed no activity above background levels. Flow cytometry noise was the likely source of these low background levels of eGFP positivity because no difference was seen here with/without the eGFP Leu202 targeting gRNA or different amounts of A3B. However, in agreement with episomal editing data, hnRNPUL1 MagnEdit complexes yielded a dose-dependent increase in A3B editing (quantification and representative immunoblots in **Fig. 3.2*A***; $p<0.0009$ by unpaired student's t-test). As expected, all components of the MagnEdit reaction were required for chromosomal DNA editing (hnRNPUL1-Cas9n complex, Leu202 gRNA, and A3B-HA; **Fig. 3.2*B***).

To further investigate the mechanistic requirements for MagnEdit, we asked whether the nuclear import activity of A3B is required. A3B is the only constitutively nuclear human APOBEC family member (42,43,46) and nuclear localization is predicted to be essential for MagnEdit. Recent studies have combined to delineate a non-canonical nuclear import mechanism involving multiple A3B surface residues in two distinct patches (46). Indeed, two previously characterized import-defective mutants, Val54Asp (42) and chim 22-32 (46), were no longer capable of editing the chromosomal eGFP Leu202 reporter (**Fig. 3.2*C***). These amino acid substitutions localize to the N-terminal regulatory domain of A3B and the editing phenotype is indistinguishable from that of a C-terminal

domain catalytic mutant (CM in **Fig. 3.2*C***). Additionally, the chromosomal DNA editing reaction can be suppressed in a dose-dependent manner by BORF2, a recently discovered A3B antagonist encoded by Epstein-Barr virus (19) (**Fig. 3.2*D***).

Last, DNA sequencing was used to compare the ratios of on-target and target-adjacent editing by a current CBE (A3B-Cas9n)(134) and the MagnEdit complex described here (A3B plus hnRNPUL1-Cas9n). A3B-Cas9n was used for these comparisons because its catalytic domain is less promiscuous than BE3 (134) and it provides an isogenic comparison for covalent versus non-covalent editing reactions catalyzed by A3B. As above, chromosomal DNA editing was performed by transfecting Cherry-positive 293T pools with the eGFP Leu202 gRNA expression vector and plasmids encoding either A3B-Cas9n or hnRNPUL1-Cas9n with a separate vector for A3B. FACS was used 72 hours post-transfection to isolate eGFP-positive positive pools for target recovery and deep sequencing. As indicated by bright eGFP-positive signals in each editing reaction 72 hours post-transfection, both editing technologies activated the reporter with the A3B CBE appearing only 4-fold more efficient (6.1% for A3B-Cas9n and 1.5% for A3B plus hnRNPUL1-Cas9n; **Fig. 3.3*A***). In each instance, FACS resulted in enrichment of similar numbers of eGFP-positive cells for deep sequencing (98% for A3B-Cas9n and 99% for A3B plus hnRNPUL1-Cas9n).

As negative controls, parallel reactions without gRNAs were directly converted to genomic DNA for deep sequencing and no target cytosine mutations were observed. In contrast, as anticipated above and from prior studies (134), the inclusion of a gRNA enabled both technologies to restore functionality to eGFP codon 202 [TCA (Ser) to TTA (Leu); represented by a red T and normalized to 1 for comparisons in **Fig. 3.3*B***]. However,

target-adjacent editing frequencies were clearly different for these two different base editing technologies. The covalently tethered A3B-Cas9n CBE caused high frequencies of target-adjacent editing within the R-loop created by gRNA-binding region (27% at the -5 position and 16% at the -7 position in **Fig. 3.3*B***). In contract, the hnRNPUL1-Cas9n MagnEdit system showed much lower target-adjacent editing within the gRNA-binding region (0.9% at the -5 position and 3.6% at the -7 position in **Fig. 3.3*B***). Thus, our results combine to demonstrate that MagnEdit is capable of yielding high frequencies of on-target editing with significantly lower frequencies of target-adjacent editing events.

This study is the first to describe a fundamentally different approach to single base editing through the use of non-covalent interactions to "attract" a DNA cytosine deaminase to a single target cytosine. A3B is particularly attractive for this application because it is normally nuclear (not shuttling or cytoplasmic like related family members)(42,43,46,208,209) and, due to active site structural constraints (100,146), unlikely to elicit RNA level off-target editing events as documented recently for BE3 and A3A CBEs (135,199). A variety of techniques may be utilized in the future to identify additional APOBEC-interacting "baits" for the MagnEdit system (proteomic, genetic, directed-evolution, *etc.*) or different interactors for the adenosine base editing systems. In general, proteins such as hnRNPUL1 that interact with the non-catalytic N-terminal domain of A3B may be more effective than those that bind the catalytic C-terminal domain simply because they are less likely to interfere with catalytic activity. For instance, EBV BORF2 is the only A3B catalytic domain interactor described to-date (19) and, as shown here, it potently blocks editing in the MagnEdit system. However, not all A3B-interacting proteins are likely to be effective in the MagnEdit system because affinities may be too low, nuclear access may not be allowed, and/or binding confirmations may be

unproductive (*e.g.*, CDK4, hnRNPK, and SIV Vif). Nevertheless, the single “hit-and-run” kinetics of the non-covalent MagnEdit system is attractive for helping to minimize off-target effects and ultimately enable true single base editing.

Methods

Cell lines – 293T and 293T-Leu202 cells were cultured in RPMI 1640 supplemented with 10% fetal bovine serum (FBS) and penicillin-streptomycin. Chromosomal 293T-Leu202 reporter line was constructed using viral transduction followed by hygromycin selection (detailed below).

Constructs – The rat APOBEC1-Cas9n-UGI-NLS construct (BE3) was provided by David Liu (114). Interactor cDNA sequences were amplified using standard PCR techniques and cloned into the BE3 vector in place of APOBEC1 using standard PCR subcloning methods. GenBank accession numbers for BFP (MK178577.1), CDK4 (NM_000075.4), hnRNPK (NM_031263.4), and hnRNPUL1 (EU831487.1). SIV-Vif was subcloned from previously reported construct (111,210) using standard PCR techniques. Leu202 gRNA, NS gRNA, empty-Cas9n-UGI-NLS and Leu202 reporter (pLenti-CMV-mCherry-T2A-eGFP) have been reported (134).

pcDNA3.1-3xHA, A3Bi-3xHA and A3Bi$_{V54D}$-3xHA have been reported (42), and A3B$_{chim22-32}$-3xHA was subcloned from a previously reported construct (46) using standard PCR cloning techniques. BORF2-3xFlag has also been reported (19).

Episomal base editing experiments – Semi-confluent 293T cells in a 6-well plate format were transfected with 200 ng gRNA, 400 ng reporter, 600 ng Cas9n-UGI-NLS, and either 600 ng pcDNA3.1-3xHA, 300ng pcDNA3.1-3xHA and 300 ng A3B-3xHA or 600 ng A3B-3xHA [25 min at RT with 3:1 ratio of TransIT LT1 (Mirus) and 250 µl of serum-free RPMI 1640 (Hyclone)]. Cells were harvested following 72 hrs incubation for editing quantification by flow cytometry.

Chromosomal base editing experiments – Semi-confluent 10 cm plates of 293 T cells were transfected with 8 µg of an HIV-1 Gag-Pol packaging plasmid, 1.5 µg of a VSV-G expression plasmid, and 3 µg of pLenti-CMV-mCherry-T2A-eGFP$_{Leu202}$-IRES-Hygro. Viruses were harvested 48 hrs post-transfection and used to transduce target cells. 48 hrs post-transduction cells were selected using 250 µg/ml Hygromycin. Transduced, mCherry-positive cells were transfected with 600 ng Cas9n-UGI editor, 200 ng of Leu202 or NS-gRNA and either 600 ng pcDNA3.1-3xHA, 300ng pcDNA3.1-3xHA and 300 ng A3B-3xHA or 600 ng A3B-3xHA. Cells were harvested 72 hrs post-transfection and editing was quantified by flow cytometry (fraction of eGFP and mCherry double-positive cells in the total mCherry-positive population).

MiSeq – eGFP target sequences were amplified using Phusion high-fidelity DNA polymerase (NEB) and previously reported primers (134). To add diversity to the sequence library, zero, one, or two extra cytosine bases were added to forward and reverse primers for each amplicon. Barcodes were added to generate full-length Illumina amplicons. Samples were analyzed using Illumina MiSeq 2 × 75-nucleotide paired-end reads (University of Minnesota Genomics Center). Reads were paired using FLASh (211). Data processing was performed using a locally installed FASTX-Toolkit. Fastx-clipper was used to trim the 3′ constant adapter region from sequences, and a stand-alone script was used to trim 5′ constant regions. Trimmed sequences were then filtered for high-quality reads using the Fastx-quality filter. Sequences with a Phred quality score less than 30 (99.9% base calling accuracy) at any position were eliminated. Preprocessed sequences were then further analyzed using the FASTAptamer toolkit (212). FASTAptamer-Count was used to determine the number of times each sequence was sampled from the population. Each sequence was then ranked and sorted based on overall abundance,

normalized to the total number of reads in each population, and directed into FASTAptamer-Enrich. FASTAptamer-Enrich calculates the fold enrichment ratios from a starting population to a selected population by using the normalized reads-per-million (RPM) values for each sequence. Sequences at abundances lower than 5 RPM in the A3-editosome samples were discarded. For reporter and A3-editosome comparisons, sequences that appeared only in the A3-containing samples (with an RPM value over 5), or, sequences that occurred at a frequency below 5 RPM in the no-gRNA controls were included for analysis.

Immunoblots – 1×10^6 cells were lysed directly into 2.5x Laemmli sample buffer, separated by 4-20% SDS-PAGE, and transferred to PVDF-FL membranes (Millipore). Membranes were blocked in 5% milk in PBS and incubated with primary antibody diluted in 5% milk in PBS supplemented with 0.1% Tween20. Secondary antibodies were diluted in 5% milk in PBS supplemented with 0.1% Tween20 and 0.01% SDS. Membranes were imaged with a LI-COR Odyssey instrument. Primary antibodies used in these experiments were rabbit anti-Cas9 (Abcam ab189380), mouse anti-Tubulin (Sigma T5168), rabbit anti-HA (Cell Signaling 3724S) and mouse anti-Flag (Sigma F1804). Secondary antibodies used were goat anti-rabbit IRdye 800CW (Licor 827-08365) and goat anti-mouse Alexa Fluor 680 (Molecular Probes A-21057).

Additional contributions

We thank David Liu for providing BE3. These studies were supported in part by a grant from the NCI (P01-CA234228 to RSH). Salary support for JLM was provided in part by the NSF Graduate Research Fellowship (Grant Number 00039202) and for DJS from the University of Minnesota Craniofacial Research Training (MinnCResT) program (NIH

T90DE022732). RSH is the Margaret Harvey Schering Land Grant Chair for Cancer Research, a Distinguished University McKnight Professor, and an Investigator of the Howard Hughes Medical Institute.

Conflicts of interest

RSH is a co-founder, shareholder, and consultant for ApoGen Biotechnologies Inc. The other authors have no conflicts of interest to declare.

Figures

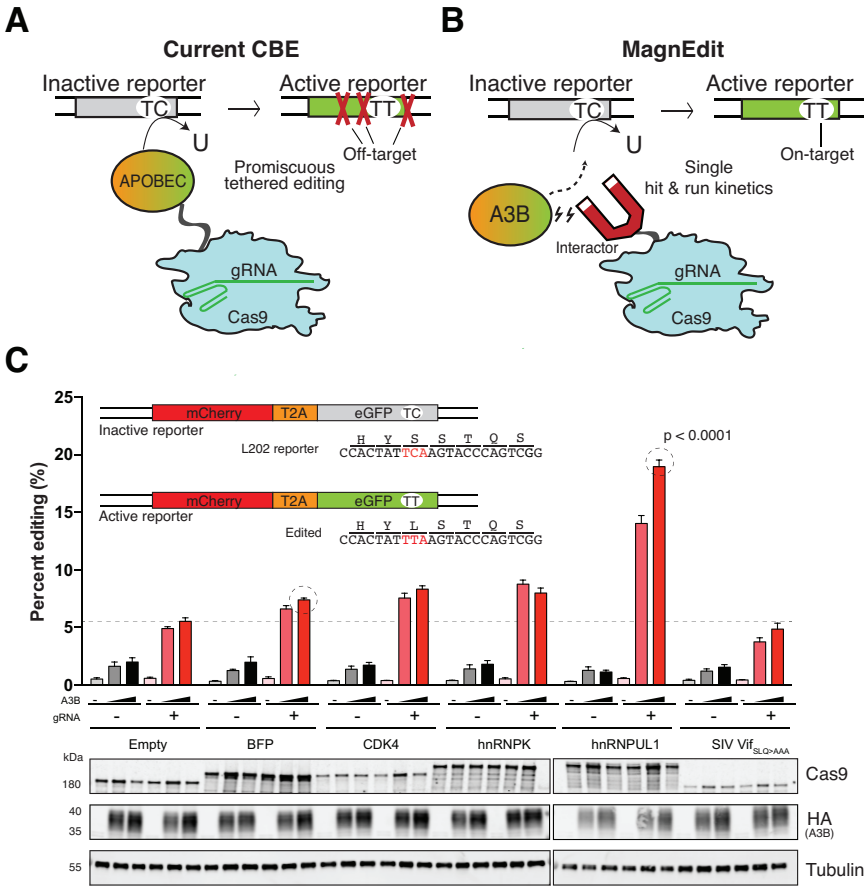


Figure 3.1. Covalent CBE versus non-covalent MagnEdit technology for DNA cytosine base editing

A, Schematic of current CBE methodology with APOBEC-Cas9n/gRNA editosome engaging the eGFP Leu202 reporter. Target-adjacent mutations are indicated by red X's. ***B***, Schematic of MagnEdit with interactor-Cas9n/gRNA complex recruiting untethered A3B to the eGFP Leu202 reporter.

C, Quantification of episomal eGFP reporter editing activity of the indicated MagnEdit complexes in 293T cells (n = 3 biologically independent experiments, average ± SD, p<0.0001 by unpaired student t-test for circled reactions). The immunoblots below are

from one of these experiments. The inset schematic shows the eGFP Leu202 reporter, the DNA region matching the gRNA, and the target cytosine in red.

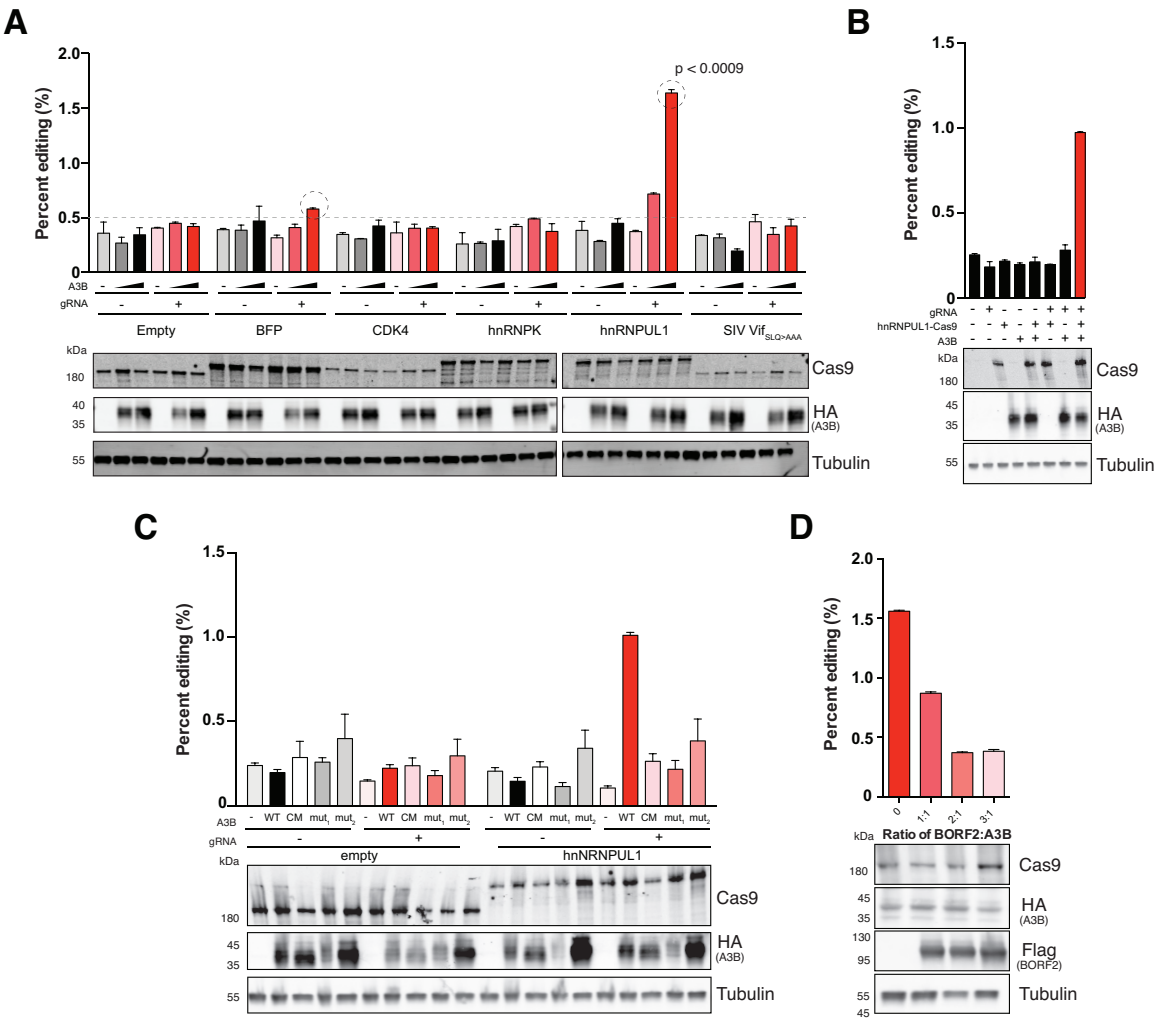


Figure 3.2. Chromosomal DNA editing by MagnEdit

A, Quantification of chromosomal eGFP reporter editing activity of the indicated MagnEdit complexes in 293T cells (n = 3 biologically independent experiments, average ± SD, p<0.0009 by unpaired student t-test for circled reactions). The immunoblots below are from one of these experiments.

B, Chromosomal eGFP editing activity for MagnEdit reactions containing the indicated components (n = 3, average ± SD). The immunoblots below each histogram are from one of the experiments.

C, Chromosomal eGFP editing activity for MagnEdit reactions containing $A3B_{WT}$, $A3B_{CM}$, and A3B localization mutations ($A3B_{V54D}$ and $A3B_{chim22\text{-}32}$, as mut 1 and mut 2 respectively) (n = 3, average ± SD). The immunoblots below each histogram are from one of the experiments.

D, Chromosomal eGFP editing activity for MagnEdit reactions containing increasing ratios of BORF2-F to A3B-HA (n = 3, average ± SD). The immunoblots below each histogram are from one of the experiments.

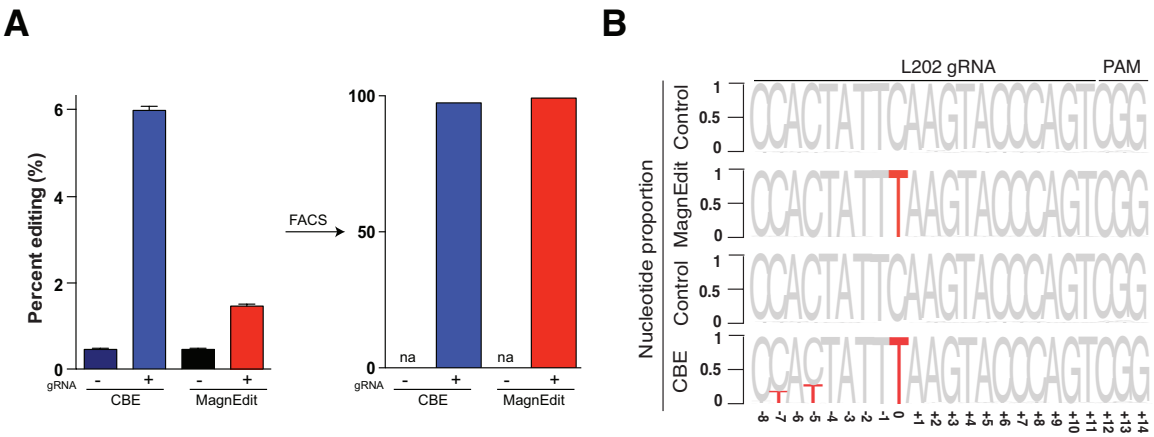


Figure 3.3. Target-adjacent editing by CBE versus MagnEdit

A, Quantification of eGFP-positive 293T cells (Leu202 edited) post-editing and post-enrichment by FACS for the indicated editing reactions (n = 3 technical replicate experiments, average ± SD).

B, Sequence logos summarizing MiSeq data from the same reactions as panel (a). The consensus sequence matches the ssDNA region displaced by gRNA annealing with the target cytosine. Red coloring highlights base substitution mutations that occurred in >5% of the MiSeq reads for each reaction (numbers are nucleobase distances 5' or 3' of the target "C").

Chapter IV.

Conclusions and Discussion

Authors: Jennifer L. McCann[1-4]

Affiliations: [1]Department of Biochemistry, Molecular Biology and Biophysics, [2]Institute for Molecular Virology, [3]Masonic Cancer Center, [4]Center for Genome Engineering, University of Minnesota, Minneapolis, MN 55455, USA.

Conclusion

Chapter II: The DNA Deaminase APOBEC3B Interacts with the Cell Cycle Protein CDK4 and Disrupts CDK4-Mediated Nuclear Import of Cyclin D1 (113)

Since the discovery of APOBEC3B as a major contributor to mutagenesis in breast and other cancer types, the modes by which this enzymatic protein are regulated has proven to be a difficult paradox to address (28,48). Many studies have shown potential axis by which *A3B* can be transcriptionally regulated, while only a handful have reported mechanisms of post-translational regulation, specifically A3B interacting proteins [SIV Vif (111), hnRNPK (110), hnRNPA3 (112), CDK4 (113), BORF2 (19)]. To further complicate the story, only one study, aside from our work, has provided any insight into a functional relevance for a proclaimed A3B interacting protein (110). Cumulatively, these data are incomplete and merely scratch the surface of the dire need for APOBEC research aimed toward addressing mechanisms of A3B inhibition and suppression of cancer progression.

Here, we used a proteomics approach involving affinity purification coupled to mass spectrometry in human 293T cells to identify cellular proteins that interact with A3B. Through this approach we identified a cohort of proteins that specifically bind A3B, including the cell cycle protein CDK4. Using co-immunoprecipitation experiments we validated and mapped this interaction to both the N-terminal domain of A3B and CDK4. Functional studies and immunofluorescence microscopy experiments in multiple cell lines revealed that A3B is not a substrate for CDK4-Cyclin D1 phosphorylation nor is its deaminase activity modulated. Instead, we found that A3B is capable of disrupting the CDK4-dependent nuclear import of Cyclin D1. We concluded that this interaction between A3B and CDK4 proves a novel role for this mutagenic enzyme and likely supports a more effective antiviral response and may simultaneously facilitate cancer mutagenesis.

Chapter III: MagnEdit – Interacting Factors that Recruit DNA Editing Enzymes to Single Target Bases

Base editing is a recently developed gene-editing system that has been successfully applied in many species to induce targeted base substitutions in both DNA and RNA with high precision and efficiency (139,143). This newly developed technology has the ability to revolutionize applications in basic biology (e.g. directed gene evolution and identification of functional amino acids via genetic screening), medicine (e.g. correction of disease-causing mutations) and agriculture (e.g. crop improvement). However, like any new technology, it has several limitations that need to be overcome to make it as efficient and useful as possible. One large hindrance is the production of undesirable target-adjacent and off target mutations that can result in nonsynonymous mutations.

In light of this off target limitation and need for optimized base editing we have devised a method (manuscript in review), deemed “MagnEdit”, to attract an endogenous DNA editor to a target base for efficient and specific editing without gratuitous off-target adjacent damage. In this method, discussed in **Chapter 3**, we demonstrate the feasibility of this method using previously validated A3B-interacting proteins to serve as “magnets” fused to Cas9n and UGI, in order to attract A3B to single-base editing hotspots. We demonstrate that current CBE methodology generates adjacent off-target editing at a frequency of >25%, while MagnEdit catalyzes an on-target efficiency of >95% and no significant off-target events mutations. Overall, our advancements to the genome engineering field demonstrate a modular system that can be adapted for any recruiter (interacting proteins, antibodies, epitopes etc.) and base editor system with current studies producing substantially lower off-target editing than current base editing methodology.

Discussion

Functional insight into APOBEC3B using interacting proteins

Human APOBEC3B (A3B) is unique among the APOBEC3 family of DNA cytosine deaminase enzymes in that its nuclear localization allows for A3B-mediated mutations to arise in cellular genomic DNA, thereby contributing to the development and progression of cancer. Thus, targeting A3B by inhibiting its activity, localization and/or overall function could result in new opportunities for diagnostic and therapeutic development. Therefore, it is crucial to study the potential interactions of A3B with other cellular proteins to further characterize and understand their contributions toward cancer progression.

Through carefully controlled proteomic affinity purification – mass spectrometry, structural-guided mutagenesis, genetic, and immunofluorescence microscopy approaches, we identified the cell cycle protein, CDK4, as a specific A3B interacting cellular protein. Our findings demonstrate A3B perturbs a key step in the G1/S phase transition of the cell cycle, the nuclear import of Cyclin D1, via its interaction with CDK4. These results advance multiple current models for A3B regulation, the first being a potential role for tight regulation of A3B throughout the phases of the cell cycle. While this may seem obvious, irregular cell cycle phenotypes such as cell cycle arrest, cell cycle delay, and aborted cell division (multi-nucleation) have been reported in the literature for induced/upregulated A3B but still remain poorly understood (43,180-182). However, these studies agree that when *A3B* is overexpressed it elicits a cell cycle response, and for cells to tolerate the expression of this mutagenic enzyme, a balance needs to occur within the regulatory mechanisms in place to avoid catastrophic damage to the host genome. As our findings demonstrate, the A3B-CDK4 interaction perturbs a crucial step in the G1/S phase transition, playing into the phenotypes mentioned above. As the G1/S phase transition involves numerous points of exposed ssDNA (DNA replication, transcription, and DNA

repair), this may result in increased substrate availability and could lead to increased genomic DNA deamination, non-clustered and clustered mutation, and overall genomic instability. One could imagine that other imbalances such as defects in the DNA damage response pathway could also result in accumulation of genomic mutations. For example, a recent study showed p53 inactivation renders cells more permissive for A3B mutagenesis (181). This model also explains why somatic *TP53* mutations were identified as a significant global correlate with *A3B* overexpression in cancer (28). Collectively, these data synergize to demonstrate a post-translational mechanistic insights into the intricate balance between cell cycle and A3B.

These proteomic studies also synergize with an additive model for transcriptional regulation of *A3B*. Playing into the cell-cycle arrest studies mentioned above, *A3B* has been implicated as a cell cycle regulated gene as its expression is highly correlated with both proliferative features as well as the *E2F* family of cell cycle transcription factors (73,82). Therefore, it is plausible to link these two models with a potential negative-feedback loop. As one could image, in a normal cell *A3B* becomes grossly overexpressed by an unknown mechanism (ex. viral infection of HPV), it binds CDK4 and elicits a G1/S phase stalling (21,113). This would suppress its own expression and act as a fail-safe and/or cell-cycle checkpoint mechanism for the cell, thereby avoiding transformation into a pre-cancerous cell or spread of viral infection. As is known, a hallmark of cancer is the disruption of the controlled/normal cell cycle progression; thereby, if *A3B* induction occurs within a cell harboring one or more driving alterations (ex. inactivated *RB*) within their compartments, this may provide a survival edge and allow for escape from A3B-mediated cell cycle arrest (214). This would ultimately lead to continued expression of *A3B* and genomic DNA damage accumulation. While these models are enticing, many questions still remain in regard to A3B regulation, its association with the cell cycle, and its

contribution to A3B-mediated mutagenesis in cancer. For example, a complete understanding of how *A3B* becomes grossly overexpressed in cancer has remained elusive. Utilizing CRISPR-Cas9 knock-in methods and CRISPR library screens, an endogenous epitope tag (such as GFP) could be engineered on A3B followed by a CRISPR library screen to reveal novel players in *A3B* regulation. Additionally, little is known about A3B target substrates within a cell; ChIP-seq of A3B along with whole genome sequencing could provide insights on A3B genomic occupancy as well as editing. However, the proteomic insight into A3B interacting proteins has advanced the understanding of both known A3B phenotypes, as well as documenting a novel role for this enzyme.

Achieving true single-base editing

Base editing is a powerful and rapidly advancing field. The applications of this technology are vast and highly impactful, but we and others are always expanding with new editors that target different bases, have altered editing windows, increased efficiencies and target more accurately. A bottleneck in this process has been the need for covalent fusion of the editor to the Cas9 complex, which traps the tethered enzyme locally and links both on-target and off-target mutational events. The non-covalent method, MagnEdit, described here can be used to decouple the fates of on-target and target-adjacent editing events and thereby increase the likelihood of achieving precise single base editing. Although we demonstrate A3B and hnRNPUL1 here as a proof-of-concept with successful C-to-T editing of both episomal and chromosomal substrates with negligible target-adjacent affects, the MagnEdit system has implications for other editing technologies and other recruiters such as antibodies, small molecules, and epitope tags. The system is transferable to many cell types and presents both promise in base editing evolution as well as high throughput screening applications for potential A3B inhibitors

using both cDNA and small molecule libraries in addition to A3B functional mutation screens. While the future applications of this system will require further optimization, the Harris lab has generated preliminary data supporting the use of this system for these operations.

In addition to reporting a novel system to “attract” an endogenous DNA deaminase, we also reveal another A3B interacting protein from our AP-MS studies, hnRNPUL1. When challenged against published A3B interactors hnRNPUL1 was the only one capable of recruiting A3B to the chromosomal target site. Although not directly tested here, this could be due to increased affinity and/or binding conformation. Nevertheless, the single “hit-and-run” kinetics of the non-covalent MagnEdit system, along with the newly identified A3B interacting protein, is a fundamentally different approach to single base editing and provides a new unexplored route for base editing improvements.

Advancements to the base editing field thus far, including the work described here, focus on improving the specificity, efficiency and delivery of editing complexes. A significant trade-off has been seen between specificity and efficiency; when specificity increases, efficiency decreases, and vice-versa. So, here comes the big question: will truly pure single nucleotide editing be achieved? Perhaps some targets with accessible chromatin status, optimal base editing windows (i.e. placement of the target cytosine relative to the PAM), and a low number of neighboring cytosines will be successful. However, many therapeutic targets do not fall into this small category, but despite these limitations I believe that MagnEdit is another step forward in achieving this overarching goal.

Closing remarks

Historically, the study of protein-protein interactions has proven to be integral in regard to understanding oncoproteins, tumor suppressors, and many other key biological functions in the context of both normal, healthy cells as well as virally infected and/or cancer cells. Since A3B has been shown to play a large role in tumor survival and evolution, understanding modes of A3B regulation has now become an ever-growing area of cancer therapeutic research. The primary goal of my thesis has been to better understand how the cancer-associated mutagenic enzyme, APOBEC3B (A3B), binds and/or is regulated by interacting cellular proteins. Here I have shown through an affinity purification mass spectrometry approach a specific interaction between A3B and the cell cycle protein CDK4 (**Chapter 2**). These studies revealed that A3B is capable of disrupting the CDK4-dependent nuclear import of Cyclin D1, thereby providing novel insights into an alternative function of A3B. Additionally, these proteomic studies on A3B interacting proteins, have led me to develop a novel technology that merges the current CRISPR/Cas9 cytosine base editing (CBE) technology with our new understanding of the A3B proteome. This system, termed "MagnEdit" relies on these interactors to serve as "magnets" fused to Cas9n to hone A3B to single-base editing hotspots (**Chapter 3**). This is the first fundamentally different approach to single base editing through the use of non-covalent interactions to "attract" editors to single target sites. This method largely advances single base editing technologies due to its novelty and highly specific on-target editing frequency with implications for other editing technologies and alternative recruiters such as antibodies, small molecules and epitope tags. Together, these studies combine to provide novel insights into the complex interplay between A3B and cellular proteins thereby proving both an alternative function for this mutagenetic enzyme as well as

CRISPR/Cas9 base editing advancements through the repurposing of these known interactions as added tools in the ever-expanding genome engineering “toolkit”.

References

1. Zhao, Y., Aguilar, A., Bernard, D., and Wang, S. (2015) Small-molecule inhibitors of the MDM2-p53 protein-protein interaction (MDM2 Inhibitors) in clinical trials for cancer treatment. *J Med Chem* **58**, 1038-1052
2. Münk, C., Willemsen, A., and Bravo, I. G. (2012) An ancient history of gene duplications, fusions and losses in the evolution of APOBEC3 mutators in mammals. *BMC Evol Biol* **12**, 71
3. Harris, R. S., and Liddament, M. T. (2004) Retroviral restriction by APOBEC proteins. *Nat Rev Immunol* **4**, 868-877
4. Conticello, S. G., Thomas, C. J., Petersen-Mahrt, S. K., and Neuberger, M. S. (2005) Evolution of the AID/APOBEC family of polynucleotide (deoxy)cytidine deaminases. *Mol Biol Evol* **22**, 367-377
5. LaRue, R. S., Andrésdóttir, V., Blanchard, Y., Conticello, S. G., Derse, D., Emerman, M., Greene, W. C., Jonsson, S. R., Landau, N. R., Lochelt, M., Malik, H. S., Malim, M. H., Munk, C., O'Brien, S. J., Pathak, V. K., Strebel, K., Wain-Hobson, S., Yu, X. F., Yuhki, N., and Harris, R. S. (2009) Guidelines for naming nonprimate APOBEC3 genes and proteins. *J Virol* **83**, 494-497
6. Conticello, S. G. (2008) The AID/APOBEC family of nucleic acid mutators. *Genome Biology* **9**, 10
7. Hultquist, J. F., Lengyel, J. A., Refsland, E. W., LaRue, R. S., Lackey, L., Brown, W. L., and Harris, R. S. (2011) Human and rhesus APOBEC3D, APOBEC3F, APOBEC3G, and APOBEC3H demonstrate a conserved capacity to restrict Vif-deficient HIV-1. *Journal of Virology* **85**, 11220-11234

8. Jarmuz, A., Chester, A., Bayliss, J., Gisbourne, J., Dunham, I., Scott, J., and Navaratnam, N. (2002) An anthropoid-specific locus of orphan C to U RNA-editing enzymes on chromosome 22. *Genomics* **79**, 285-296

9. Rogozin, I. B., Basu, M. K., Jordan, I. K., Pavlov, Y. I., and 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

10. Teng, B., Burant, C. F., and Davidson, N. O. (1993) Molecular cloning of an apolipoprotein B messenger RNA editing protein. *Science* **260**, 1816-1819

11. Mehta, A., Kinter, M. T., Sherman, N. E., and Driscoll, D. M. (2000) Molecular cloning of apobec-1 complementation factor, a novel RNA-binding protein involved in the editing of apolipoprotein B mRNA. *Mol Cell Biol* **20**, 1846-1854

12. Davidson, N. O., Anant, S., and MacGinnitie, A. J. (1995) Apolipoprotein B messenger RNA editing: insights into the molecular regulation of post-transcriptional cytidine deamination. *Curr Opin Lipidol* **6**, 70-74

13. Muramatsu, M., Sankaranand, V. S., Anant, S., Sugai, M., Kinoshita, K., Davidson, N. O., and Honjo, T. (1999) Specific expression of activation-induced cytidine deaminase (AID), a novel member of the RNA-editing deaminase family in germinal center B cells. *J Biol Chem* **274**, 18470-18476

14. Pham, P., Bransteitter, R., Petruska, J., and Goodman, M. F. (2003) Processive AID-catalysed cytosine deamination on single-stranded DNA simulates somatic hypermutation. *Nature* **424**, 103-107

15. Schrader, C. E., Guikema, J. E., Wu, X., and Stavnezer, J. (2009) The roles of APE1, APE2, DNA polymerase beta and mismatch repair in creating S region

DNA breaks during antibody class switch. *Philos Trans R Soc Lond B Biol Sci* **364**, 645-652

16. Harris, R. S., Petersen-Mahrt, S. K., and Neuberger, M. S. (2002) RNA editing enzyme APOBEC1 and some of its homologs can act as DNA mutators. *Mol Cell* **10**, 1247-1253

17. Lada, A. G., Krick, C. F., Kozmin, S. G., Mayorov, V. I., Karpova, T. S., Rogozin, I. B., and Pavlov, Y. I. (2011) Mutator effects and mutation signatures of editing deaminases produced in bacteria and yeast. *Biochemistry (Mosc)* **76**, 131-146

18. Liao, W., Hong, S. H., Chan, B. H., Rudolph, F. B., Clark, S. C., and 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

19. Cheng, A. Z., Yockteng-Melgar, J., Jarvis, M. C., Malik-Soni, N., Borozan, I., Carpenter, M. A., McCann, J. L., Ebrahimi, D., Shaban, N. M., Marcon, E., Greenblatt, J., Brown, W. L., Frappier, L., and Harris, R. S. (2019) Epstein-Barr virus BORF2 inhibits cellular APOBEC3B to preserve viral genome integrity. *Nature Microbiology* **4**, 78-88

20. Verhalen, B., Starrett, G. J., Harris, R. S., and Jiang, M. (2016) Functional upregulation of the DNA cytosine deaminase APOBEC3B by polyomaviruses. *Journal of Virology* **90**, 6379-6386

21. Vieira, V. C., Leonard, B., White, E. A., Starrett, G. J., Temiz, N. A., Lorenz, L. D., Lee, D., Soares, M. A., Lambert, P. F., Howley, P. M., and Harris, R. S. (2014) Human Papillomavirus E6 Triggers Upregulation of the Antiviral and Cancer Genomic DNA Deaminase APOBEC3B. *mBio* **5**, e02234-02214

22. Warren, C. J., Westrich, J. A., Doorslaer, K. V., and Pyeon, D. (2017) Roles of APOBEC3A and APOBEC3B in human papillomavirus infection and disease progression. *Viruses* **9**, 233

23. Ooms, M., Krikoni, A., Kress, A. K., Simon, V., and Munk, C. (2012) APOBEC3A, APOBEC3B, and APOBEC3H haplotype 2 restrict human T-lymphotropic virus type 1. *Journal of Virology* **86**, 6097-6108

24. Stavrou, S., and Ross, S. R. (2015) APOBEC3 proteins in viral immunity. *Journal of Immunology* **195**, 4565-4570

25. Stenglein, M. D., Burns, M. B., Li, M., Lengyel, J., and Harris, R. S. (2010) APOBEC3 proteins mediate the clearance of foreign DNA from human cells. *Nature Structural and Molecular Biology* **17**, 222-229

26. Refsland, E. W., and Harris, R. S. (2013) The APOBEC3 family of retroelement restriction factors. in *Intrinsic Immunity* (Cullen, B. R. ed.), Springer-Verlag Berlin, Berlin. pp 1-27

27. Refsland, E. W., Stenglein, M. D., Shindo, K., Albin, J. S., Brown, W. L., and Harris, R. S. (2010) Quantitative profiling of the full APOBEC3 mRNA repertoire in lymphocytes and tissues: implications for HIV-1 restriction. *Nucleic Acids Research* **38**, 4274-4284

28. Burns, M. B., Temiz, N. A., and Harris, R. S. (2013) Evidence for APOBEC3B mutagenesis in multiple human cancers. *Nature Genetics* **45**, 977-983

29. Schumann, G. G. (2007) APOBEC3 proteins: major players in intracellular defence against LINE-1-mediated retrotransposition. *Biochem Soc Trans* **35**, 637-642

30. Horn, A. V., Klawitter, S., Held, U., Berger, A., Vasudevan, A. A., Bock, A., Hofmann, H., Hanschmann, K. M., Trösemeier, J. H., Flory, E., Jabulowsky, R.

A., Han, J. S., Löwer, J., Löwer, R., Münk, C., and Schumann, G. G. (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

31. Muckenfuss, H., Hamdorf, M., Held, U., Perkovic, M., Lower, J., Cichutek, K., Flory, E., Schumann, G. G., and Munk, C. (2006) APOBEC3 proteins inhibit human LINE-1 retrotransposition. *J Biol Chem* **281**, 22161-22172
32. Chiu, Y. L., Witkowska, H. E., Hall, S. C., Santiago, M., Soros, V. B., Esnault, C., Heidmann, T., and Greene, W. C. (2006) High-molecular-mass APOBEC3G complexes restrict Alu retrotransposition. *Proc Natl Acad Sci U S A* **103**, 15588-15593
33. Esnault, C., Heidmann, O., Delebecque, F., Dewannieux, M., Ribet, D., Hance, A. J., Heidmann, T., and Schwartz, O. (2005) APOBEC3G cytidine deaminase inhibits retrotransposition of endogenous retroviruses. *Nature* **433**, 430-433
34. Wissing, S., Montano, M., Garcia-Perez, J. L., Moran, J. V., and Greene, W. C. (2011) Endogenous APOBEC3B restricts LINE-1 retrotransposition in transformed cells and human embryonic stem cells. *J Biol Chem* **286**, 36427-36437
35. Dutko, J. A., Schäfer, A., Kenny, A. E., Cullen, B. R., and Curcio, M. J. (2005) Inhibition of a yeast LTR retrotransposon by human APOBEC3 cytidine deaminases. *Curr Biol* **15**, 661-666
36. Kinomoto, M., Kanno, T., Shimura, M., Ishizaka, Y., Kojima, A., Kurata, T., Sata, T., and Tokunaga, K. (2007) All APOBEC3 family proteins differentially inhibit LINE-1 retrotransposition. *Nucleic Acids Research* **35**, 2955-2964

37. Stenglein, M. D., and Harris, R. S. (2006) APOBEC3B and APOBEC3F inhibit L1 retrotransposition by a DNA deamination-independent mechanism. *J Biol Chem* **281**, 16837-16841

38. Bogerd, H. P., Wiegand, H. L., Doehle, B. P., Lueders, K. K., and Cullen, B. R. (2006) APOBEC3A and APOBEC3B are potent inhibitors of LTR-retrotransposon function in human cells. *Nucleic Acids Res* **34**, 89-95

39. Bulliard, Y., Narvaiza, I., Bertero, A., Peddi, S., Röhrig, U. F., Ortiz, M., Zoete, V., Castro-Díaz, N., Turelli, P., Telenti, A., Michielin, O., Weitzman, M. D., and Trono, D. (2011) Structure-function analyses point to a polynucleotide-accommodating groove essential for APOBEC3A restriction activities. *J Virol* **85**, 1765-1776

40. Carpenter, M. A., Li, M., Rathore, A., Lackey, L., Law, E. K., Land, A. M., Leonard, B., Shandilya, S. M., Bohn, M. F., Schiffer, C. A., Brown, W. L., and Harris, R. S. (2012) Methylcytosine and normal cytosine deamination by the foreign DNA restriction enzyme APOBEC3A. *J Biol Chem* **287**, 34801-34808

41. Koning, F. A., Newman, E. N., Kim, E. Y., Kunstman, K. J., Wolinsky, S. M., and Malim, M. H. (2009) Defining APOBEC3 expression patterns in human tissues and hematopoietic cell subsets. *J Virol* **83**, 9474-9485

42. Lackey, L., Demorest, Z. L., Land, A. M., Hultquist, J. F., Brown, W. L., and Harris, R. S. (2012) APOBEC3B and AID have similar nuclear import mechanisms. *Journal of Molecular Biology* **419**, 301-314

43. Lackey, L., Law, E. K., Brown, W. L., and Harris, R. S. (2013) Subcellular localization of the APOBEC3 proteins during mitosis and implications for genomic DNA deamination. *Cell Cycle* **12**, 762-772

44. Pak, V., Heidecker, G., Pathak, V. K., and Derse, D. (2011) The role of amino-terminal sequences in cellular localization and antiviral activity of APOBEC3B. *Journal of Virology* **85**, 8538-8547

45. Stenglein, M. D., Matsuo, H., and Harris, R. S. (2008) Two regions within the amino-terminal half of APOBEC3G cooperate to determine cytoplasmic localization. *Journal of Virology* **82**, 9591-9599

46. Salamango, D. J., McCann, J. L., Demir, O., Brown, W. L., Amaro, R. E., and Harris, R. S. (2018) APOBEC3B Nuclear localization requires two distinct n-terminal domain surfaces. *Journal of Molecular Biology* **430**, 2695-2708

47. Salamango, D. J., Becker, J. T., McCann, J. L., Cheng, A. Z., Demir, Ö., Amaro, R. E., Brown, W. L., Shaban, N. M., and Harris, R. S. (2018) APOBEC3H subcellular localization determinants define zipcode for targeting HIV-1 for restriction. *Molecular and Cellular Biology* **38**, e00356-00318

48. Burns, M. B., Lackey, L., Carpenter, M. A., Rathore, A., Land, A. M., Leonard, B., Refsland, E. W., Kotandeniya, D., Tretyakova, N., Nikas, J. B., Yee, D., Temiz, N. A., Donohue, D. E., McDougle, R. M., Brown, W. L., Law, E. K., and Harris, R. S. (2013) APOBEC3B is an enzymatic source of mutation in breast cancer. *Nature* **494**, 366-370

49. Roberts, S. A., Lawrence, M. S., Klimczak, L. J., Grimm, S. A., Fargo, D., Stojanov, P., Kiezun, A., Kryukov, G. V., Carter, S. L., Saksena, G., Harris, S., Shah, R. R., Resnick, M. A., Getz, G., and Gordenin, D. A. (2013) An APOBEC cytidine deaminase mutagenesis pattern is widespread in human cancers. *Nature Genetics* **45**, 970-976

50. Lefebvre, C., Bachelot, T., Filleron, T., Pedrero, M., Campone, M., Soria, J. C., Massard, C., Levy, C., Arnedos, M., Lacroix-Triki, M., Garrabey, J., Boursin, Y.,

Deloger, M., Fu, Y., Commo, F., Scott, V., Lacroix, L., Dieci, M. V., Kamal, M., Dieras, V., Goncalves, A., Ferrerro, J. M., Romieu, G., Vanlemmens, L., Mouret Reynier, M. A., Thery, J. C., Le Du, F., Guiu, S., Dalenc, F., Clapisson, G., Bonnefoi, H., Jimenez, M., Le Tourneau, C., and Andre, F. (2016) Mutational profile of metastatic breast cancers: a retrospective analysis. *PLoS Medicine* **13**, e1002201

51. Alexandrov, L. B., Nik-Zainal, S., Wedge, D. C., Aparicio, S. A., Behjati, S., Biankin, A. V., Bignell, G. R., Bolli, N., Borg, A., Borresen-Dale, A. L., Boyault, S., Burkhardt, B., Butler, A. P., Caldas, C., Davies, H. R., Desmedt, C., Eils, R., Eyfjord, J. E., Foekens, J. A., Greaves, M., Hosoda, F., Hutter, B., Ilicic, T., Imbeaud, S., Imielinsk, M., Jager, N., Jones, D. T., Jones, D., Knappskog, S., Kool, M., Lakhani, S. R., Lopez-Otin, C., Martin, S., Munshi, N. C., Nakamura, H., Northcott, P. A., Pajic, M., Papaemmanuil, E., Paradiso, A., Pearson, J. V., Puente, X. S., Raine, K., Ramakrishna, M., Richardson, A. L., Richter, J., Rosenstiel, P., Schlesner, M., Schumacher, T. N., Span, P. N., Teague, J. W., Totoki, Y., Tutt, A. N., Valdes-Mas, R., van Buuren, M. M., van 't Veer, L., Vincent-Salomon, A., Waddell, N., Yates, L. R., Zucman-Rossi, J., Futreal, P. A., McDermott, U., Lichter, P., Meyerson, M., Grimmond, S. M., Siebert, R., Campo, E., Shibata, T., Pfister, S. M., Campbell, P. J., and Stratton, M. R. (2013) Signatures of mutational processes in human cancer. *Nature* **500**, 415-421

52. Nik-Zainal, S., Alexandrov, L. B., Wedge, D. C., Van Loo, P., Greenman, C. D., Raine, K., Jones, D., Hinton, J., Marshall, J., Stebbings, L. A., Menzies, A., Martin, S., Leung, K., Chen, L., Leroy, C., Ramakrishna, M., Rance, R., Lau, K. W., Mudie, L. J., Varela, I., McBride, D. J., Bignell, G. R., Cooke, S. L., Shlien, A., Gamble, J., Whitmore, I., Maddison, M., Tarpey, P. S., Davies, H. R.,

Papaemmanuil, E., Stephens, P. J., McLaren, S., Butler, A. P., Teague, J. W., Jonsson, G., Garber, J. E., Silver, D., Miron, P., Fatima, A., Boyault, S., Langerod, A., Tutt, A., Martens, J. W., Aparicio, S. A., Borg, A., Salomon, A. V., Thomas, G., Borresen-Dale, A. L., Richardson, A. L., Neuberger, M. S., Futreal, P. A., Campbell, P. J., and Stratton, M. R. (2012) Mutational processes molding the genomes of 21 breast cancers. *Cell* **149**, 979-993

53. Swanton, C., McGranahan, N., Starrett, G. J., and Harris, R. S. (2015) APOBEC enzymes: mutagenic fuel for cancer evolution and heterogeneity. *Cancer Discovery* **5**, 704-712

54. Roberts, S. A., and Gordenin, D. A. (2014) Hypermutation in human cancer genomes: footprints and mechanisms. *Nature Reviews Cancer* **14**, 786-800

55. Helleday, T., Eshtad, S., and Nik-Zainal, S. (2014) Mechanisms underlying mutational signatures in human cancers. *Nature Reviews Genetics* **15**, 585-598

56. Venkatesan, S., Rosenthal, R., Kanu, N., McGranahan, N., Bartek, J., Quezada, S. A., Hare, J., Harris, R. S., and Swanton, C. (2018) Perspective: APOBEC mutagenesis in drug resistance and immune escape in HIV and cancer evolution. *Annals Oncology* **29**, 563-572

57. Ullah, I., Karthik, G. M., Alkodsi, A., Kjallquist, U., Stalhammar, G., Lovrot, J., Martinez, N. F., Lagergren, J., Hautaniemi, S., Hartman, J., and Bergh, J. (2018) Evolutionary history of metastatic breast cancer reveals minimal seeding from axillary lymph nodes. *Journal of Clinical Investigation* **128**, 1355-1370

58. Bertucci, F., Ng, C. K. Y., Patsouris, A., Droin, N., Piscuoglio, S., Carbuccia, N., Soria, J. C., Dien, A. T., Adnani, Y., Kamal, M., Garnier, S., Meurice, G., Jimenez, M., Dogan, S., Verret, B., Chaffanet, M., Bachelot, T., Campone, M., Lefeuvre, C., Bonnefoi, H., Dalenc, F., Jacquet, A., De Filippo, M. R., Babbar, N.,

Birnbaum, D., Filleron, T., Le Tourneau, C., and André, F. (2019) Genomic characterization of metastatic breast cancers. *Nature* **569**, 560-564

59. Green, A. M., Landry, S., Budagyan, K., Avgousti, D. C., Shalhout, S., Bhagwat, A. S., and Weitzman, M. D. (2016) APOBEC3A damages the cellular genome during DNA replication. *Cell Cycle* **15**, 998-1008

60. Starrett, G. J., Luengas, E. M., McCann, J. L., Ebrahimi, D., Temiz, N. A., Love, R. P., Feng, Y., Adolph, M. B., Chelico, L., Law, E. K., Carpenter, M. A., and Harris, R. S. (2016) The DNA cytosine deaminase APOBEC3H haplotype I likely contributes to breast and lung cancer mutagenesis. *Nat Commun* **7**, 12918

61. Chen, J., and MacCarthy, T. (2017) The preferred nucleotide contexts of the AID/APOBEC cytidine deaminases have differential effects when mutating retrotransposon and virus sequences compared to host genes. *PLoS Comput Biol* **13**, e1005471

62. Nabel, C. S., Lee, J. W., Wang, L. C., and Kohli, R. M. (2013) Nucleic acid determinants for selective deamination of DNA over RNA by activation-induced deaminase. *Proc Natl Acad Sci U S A* **110**, 14225-14230

63. Kohli, R. M., Maul, R. W., Guminski, A. F., McClure, R. L., Gajula, K. S., Saribasak, H., McMahon, M. A., Siliciano, R. F., Gearhart, P. J., and Stivers, J. T. (2010) Local sequence targeting in the AID/APOBEC family differentially impacts retroviral restriction and antibody diversification. *J Biol Chem* **285**, 40956-40964

64. Seplyarskiy, V. B., Andrianova, M. A., and Bazykin, G. A. (2017) APOBEC3A/B-induced mutagenesis is responsible for 20% of heritable mutations in the TpCpW context. *Genome Res* **27**, 175-184

65. Prasad, R., Poltoratsky, V., Hou, E. W., and Wilson, S. H. (2016) Rev1 is a base excision repair enzyme with 5'-deoxyribose phosphate lyase activity. *Nucleic Acids Res* **44**, 10824-10833

66. Chan, K., Resnick, M. A., and Gordenin, D. A. (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

67. Kim, N., Mudrak, S. V., and 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

68. Doseth, B., Ekre, C., Slupphaug, G., Krokan, H. E., and Kavli, B. (2012) Strikingly different properties of uracil-DNA glycosylases UNG2 and SMUG1 may explain divergent roles in processing of genomic uracil. *DNA Repair (Amst)* **11**, 587-593

69. Pulukuri, S. M., Knost, J. A., Estes, N., and Rao, J. S. (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

70. Law, E. K., Sieuwerts, A. M., LaPara, K., Leonard, B., Starrett, G. J., Molan, A. M., Temiz, N. A., Vogel, R. I., Meijer-van Gelder, M. E., Sweep, F. C., Span, P. N., Foekens, J. A., Martens, J. W., Yee, D., and Harris, R. S. (2016) The DNA cytosine deaminase APOBEC3B promotes tamoxifen resistance in ER-positive breast cancer. *Science Advances* **2**, e1601737

71. Sieuwerts, A. M., Lyng, M. B., Meijer-van Gelder, M. E., de Weerd, V., Sweep, F. C., Foekens, J. A., Span, P. N., Martens, J. W., and Ditzel, H. J. (2014) Evaluation of the ability of adjuvant tamoxifen-benefit gene signatures to predict

outcome of hormone-naive estrogen receptor-positive breast cancer patients treated with tamoxifen in the advanced setting. *Molecular Oncology* **8**, 1679-1689

72. Tsuboi, M., Yamane, A., Horiguchi, J., Yokobori, T., Kawabata-Iwakawa, R., Yoshiyama, S., Rokudai, S., Odawara, H., Tokiniwa, H., Oyama, T., Takeyoshi, I., and Nishiyama, M. (2016) APOBEC3B high expression status is associated with aggressive phenotype in Japanese breast cancers. *Breast Cancer* **23**, 780-788

73. Cescon, D. W., Haibe-Kains, B., and Mak, T. W. (2015) APOBEC3B expression in breast cancer reflects cellular proliferation, while a deletion polymorphism is associated with immune activation. *PNAS* **112**, 2841-2846

74. Walker, B. A., Wardell, C. P., Murison, A., Boyle, E. M., Begum, D. B., Dahir, N. M., Proszek, P. Z., Melchor, L., Pawlyn, C., Kaiser, M. F., Johnson, D. C., Qiang, Y. W., Jones, J. R., Cairns, D. A., Gregory, W. M., Owen, R. G., Cook, G., Drayson, M. T., Jackson, G. H., Davies, F. E., and Morgan, G. J. (2015) APOBEC family mutational signatures are associated with poor prognosis translocations in multiple myeloma. *Nature Communications* **6**, 6997

75. Nik-Zainal, S., Wedge, D. C., Alexandrov, L. B., Petljak, M., Butler, A. P., Bolli, N., Davies, H. R., Knappskog, S., Martin, S., Papaemmanuil, E., Ramakrishna, M., Shlien, A., Simonic, I., Xue, Y. L., Tyler-Smith, C., Campbell, P. J., and Stratton, M. R. (2014) Association of a germline copy number polymorphism of APOBEC3A and APOBEC3B with burden of putative APOBEC-dependent mutations in breast cancer. *Nature Genetics* **46**, 487-491

76. Tokunaga, E., Yamashita, N., Tanaka, K., Inoue, Y., Akiyoshi, S., Saeki, H., Oki, E., Kitao, H., and Maehara, Y. (2016) Expression of APOBEC3B mRNA in primary breast cancer of Japanese women. *PLoS One* **11**, e0168090

77. Fanourakis, G., Tosios, K., Papanikolaou, N., Chatzistamou, I., Xydous, M., Tseleni-Balafouta, S., Sklavounou, A., Voutsinas, G. E., and Vastardis, H. (2016) Evidence for APOBEC3B mRNA and protein expression in oral squamous cell carcinomas. *Experimental and Molecular Pathology* **101**, 314-319

78. Middlebrooks, C. D., Banday, A. R., Matsuda, K., Udquim, K. I., Onabajo, O. O., Paquin, A., Figueroa, J. D., Zhu, B., Koutros, S., Kubo, M., Shuin, T., Freedman, N. D., Kogevinas, M., Malats, N., Chanock, S. J., Garcia-Closas, M., Silverman, D. T., Rothman, N., and Prokunina-Olsson, L. (2016) Association of germline variants in the APOBEC3 region with cancer risk and enrichment with APOBEC-signature mutations in tumors. *Nature Genetics* **48**, 1330-1338

79. Chan, K., Roberts, S. A., Klimczak, L. J., Sterling, J. F., Saini, N., Malc, E. P., Kim, J., Kwiatkowski, D. J., Fargo, D. C., Mieczkowski, P. A., Getz, G., and Gordenin, D. A. (2015) An APOBEC3A hypermutation signature is distinguishable from the signature of background mutagenesis by APOBEC3B in human cancers. *Nat Genet* **47**, 1067-1072

80. Hoopes, J. I., Cortez, L. M., Mertz, T. M., Malc, E. P., Mieczkowski, P. A., and Roberts, S. A. (2016) APOBEC3A and APOBEC3B preferentially deaminate the lagging strand template during DNA replication. *Cell Reports* **14**, 1273-1282

81. Bransteitter, R., Prochnow, C., and Chen, X. J. S. (2009) The current structural and functional understanding of APOBEC deaminases. *Cellular and Molecular Life Sciences* **66**, 3137-3147

82. Starrett, G. J., Serebrenik, A. A., Roelofs, P. A., McCann, J. L., Verhalen, B., Jarvis, M. C., Stewart, T. A., Law, E. K., Krupp, A., Jiang, M., Martens, J. W. M., Cahir-McFarland, E., Span, P. N., and Harris, R. S. (2019) Polyomavirus T

antigen induces *APOBEC3B* expression using an LXCXE-dependent and TP53-independent mechanism. *MBio* **10**

83. Bhagwat, A. S., Hao, W., Townes, J. P., Lee, H., Tang, H., and Foster, P. L. (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**, 2176-2181

84. Kanu, N., Cerone, M. A., Goh, G., Zalmas, L. P., Bartkova, J., Dietzen, M., McGranahan, N., Rogers, R., Law, E. K., Gromova, I., Kschischo, M., Walton, M. I., Rossanese, O. W., Bartek, J., Harris, R. S., Venkatesan, S., and Swanton, C. (2016) DNA replication stress mediates APOBEC3 family mutagenesis in breast cancer. *Genome Biol* **17**, 185

85. Henderson, S., Chakravarthy, A., Su, X., Boshoff, C., and Fenton, T. R. (2014) APOBEC-mediated cytosine deamination links PIK3CA helical domain mutations to human papillomavirus-driven tumor development. *Cell Rep* **7**, 1833-1841

86. McGranahan, N., Favero, F., de Bruin, E. C., Birkbak, N. J., Szallasi, Z., and Swanton, C. (2015) Clonal status of actionable driver events and the timing of mutational processes in cancer evolution. *Sci Transl Med* **7**, 283ra254

87. Yan, S., He, F., Gao, B., Wu, H., Li, M., Huang, L., Liang, J., Wu, Q., and Li, Y. (2016) Increased APOBEC3B predicts worse outcomes in lung cancer: a comprehensive retrospective study. *J Cancer* **7**, 618-625

88. Xu, L., Chang, Y., An, H., Zhu, Y., Yang, Y., and Xu, J. (2015) High APOBEC3B expression is a predictor of recurrence in patients with low-risk clear cell renal cell carcinoma. *Urol Oncol* **33**, 340 e341-348

89. Sieuwerts, A. M., Willis, S., Burns, M. B., Look, M. P., Meijer-Van Gelder, M. E., Schlicker, A., Heideman, M. R., Jacobs, H., Wessels, L., Leyland-Jones, B.,

Gray, K. P., Foekens, J. A., Harris, R. S., and Martens, J. W. (2014) Elevated APOBEC3B correlates with poor outcomes for estrogen-receptor-positive breast cancers. *Hormones and Cancer* **5**, 405-413

90. Zhang, J., Wei, W., Jin, H. C., Ying, R. C., Zhu, A. K., and Zhang, F. J. (2015) The roles of APOBEC3B in gastric cancer. *Int J Clin Exp Pathol* **8**, 5089-5096

91. Jin, Z., Han, Y. X., and Han, X. R. (2014) The role of APOBEC3B in chondrosarcoma. *Oncol Rep* **32**, 1867-1872

92. Luo, X., Huang, Y., Chen, Y., Tu, Z., Hu, J., Tavis, J. E., Huang, A., and Hu, Y. (2016) Association of hepatitis B virus covalently closed circular DNA and human APOBEC3B in hepatitis B virus-Related hepatocellular carcinoma. *PLoS One* **11**, e0157708

93. Wu, P. F., Chen, Y. S., Kuo, T. Y., Lin, H. H., Liu, C. W., and Chang, L. C. (2015) APOBEC3B: a potential factor suppressing growth of human hepatocellular carcinoma cells. *Anticancer Res* **35**, 1521-1527

94. Gwak, M., Choi, Y. J., Yoo, N. J., and Lee, S. (2014) Expression of DNA cytosine deaminase APOBEC3 proteins, a potential source for producing mutations, in gastric, colorectal and prostate cancers. *Tumori* **100**, 112e-117e

95. Lamy, P., Nordentoft, I., Birkenkamp-Demtröder, K., Thomsen, M. B., Villesen, P., Vang, S., Hedegaard, J., Borre, M., Jensen, J. B., Høyer, S., Pedersen, J. S., Ørntoft, T. F., and Dyrskjøt, L. (2016) Paired exome analysis reveals clonal evolution and potential therapeutic targets in urothelial carcinoma. *Cancer Res* **76**, 5894-5906

96. Sasaki, H., Suzuki, A., Tatematsu, T., Shitara, M., Hikosaka, Y., Okuda, K., Moriyama, S., Yano, M., and Fujii, Y. (2014) gene overexpression in non-small-cell lung cancer. *Biomed Rep* **2**, 392-395

97. Kosumi, K., Baba, Y., Ishimoto, T., Harada, K., Nakamura, K., Ohuchi, M., Kiyozumi, Y., Izumi, D., Tokunaga, R., Taki, K., Higashi, T., Miyata, T., Shigaki, H., Kurashige, J., Hiyoshi, Y., Iwatsuki, M., Iwagami, S., Sakamoto, Y., Miyamoto, Y., Yoshida, N., Oki, E., Watanabe, M., and Baba, H. (2016) APOBEC3B is an enzymatic source of molecular alterations in esophageal squamous cell carcinoma. *Med Oncol* **33**, 26

98. Leonard, B., Hart, S. N., Burns, M. B., Carpenter, M. A., Temiz, N. A., Rathore, A., Vogel, R. I., Nikas, J. B., Law, E. K., Brown, W. L., Li, Y., Zhang, Y., Maurer, M. J., Oberg, A. L., Cunningham, J. M., Shridhar, V., Bell, D. A., April, C., Bentley, D., Bibikova, M., Cheetham, R. K., Fan, J. B., Grocock, R., Humphray, S., Kingsbury, Z., Peden, J., Chien, J., Swisher, E. M., Hartmann, L. C., Kalli, K. R., Goode, E. L., Sicotte, H., Kaufmann, S. H., and Harris, R. S. (2013) APOBEC3B upregulation and genomic mutation patterns in serous ovarian carcinoma. *Cancer Res* **73**, 7222-7231

99. Shi, K., Carpenter, M. A., Kurahashi, K., Harris, R. S., and Aihara, H. (2015) Crystal structure of the DNA deaminase APOBEC3B catalytic domain. *J Biol Chem* **290**, 28120-28130

100. Shi, K., Demir, O., Carpenter, M. A., Wagner, J., Kurahashi, K., Harris, R. S., Amaro, R. E., and Aihara, H. (2017) Conformational switch regulates the DNA cytosine deaminase activity of human APOBEC3B. *Sci Rep* **7**, 17415

101. Shi, K., Carpenter, M. A., Banerjee, S., Shaban, N. M., Kurahashi, K., Salamango, D. J., McCann, J. L., Starrett, G. J., Duffy, J. V., Demir, O., Amaro, R. E., Harki, D. A., Harris, R. S., and Aihara, H. (2017) Structural basis for targeted DNA cytosine deamination and mutagenesis by APOBEC3A and APOBEC3B. *Nature Structural and Molecular Biology* **24**, 131-139

102. Xiao, X., Yang, H., Arutiunian, V., Fang, Y., Besse, G., Morimoto, C., Zirkle, B., and Chen, X. S. (2017) Structural determinants of APOBEC3B non-catalytic domain for molecular assembly and catalytic regulation. *Nucleic Acids Research* **45**, 7494-7506

103. Chou, W. C., Chen, W. T., Hsiung, C. N., Hu, L. Y., Yu, J. C., Hsu, H. M., and Shen, C. Y. (2017) B-Myb induces APOBEC3B expression leading to somatic mutation in multiple cancers. *Sci Rep* **7**, 44089

104. Mori, S., Takeuchi, T., Ishii, Y., and Kukimoto, I. (2015) Identification of APOBEC3B promoter elements responsible for activation by human papillomavirus type 16 E6. *Biochem Biophys Res Commun* **460**, 555-560

105. Mori, S., Takeuchi, T., Ishii, Y., Yugawa, T., Kiyono, T., Nishina, H., and Kukimoto, I. (2017) Human Papillomavirus 16 E6 upregulates APOBEC3B via the TEAD transcription factor. *J Virol* **91**

106. Leonard, B., McCann, J. L., Starrett, G. J., Kosyakovsky, L., Luengas, E. M., Molan, A. M., Burns, M. B., McDougle, R. M., Parker, P. J., Brown, W. L., and Harris, R. S. (2015) The PKC/NF-kappaB signaling pathway induces APOBEC3B expression in multiple human cancers. *Cancer Res* **75**, 4538-4547

107. Periyasamy, M., Singh, A. K., Gemma, C., Kranjec, C., Farzan, R., Leach, D. A., Navaratnam, N., Palinkas, H. L., Vertessy, B. G., Fenton, T. R., Doorbar, J., Fuller-Pace, F., Meek, D. W., Coombes, R. C., Buluwela, L., and Ali, S. (2017) p53 controls expression of the DNA deaminase APOBEC3B to limit its potential mutagenic activity in cancer cells. *Nucleic Acids Res* **45**, 11056-11069

108. Matsumoto, T., Shirakawa, K., Yokoyama, M., Fukuda, H., Sarca, A. D., Koyabu, S., Yamazaki, H., Kazuma, Y., Matsui, H., Maruyama, W., Nagata, K., Tanabe, F., Kobayashi, M., Shindo, K., Morishita, R., Sato, H., and Takaori-Kondo, A.

(2019) Protein kinase A inhibits tumor mutator APOBEC3B through phosphorylation. *Sci Rep* **9**, 8307

109. Molan, A. M., Hanson, H. M., Chweya, C. M., Anderson, B. D., Starrett, G. J., Richards, C. M., and Harris, R. S. (2017) APOBEC3B lysine residues are dispensable for DNA cytosine deamination, HIV-1 restriction, and nuclear localization. *Virology* **511**, 74-81

110. Zhang, W., Zhang, X., Tian, C., Wang, T., Sarkis, P. T., Fang, Y., Zheng, S., Yu, X. F., and Xu, R. (2008) Cytidine deaminase APOBEC3B interacts with heterogeneous nuclear ribonucleoprotein K and suppresses hepatitis B virus expression. *Cell Microbiol* **10**, 112-121

111. Land, A. M., Wang, J., Law, E. K., Aberle, R., Kirmaier, A., Krupp, A., Johnson, W. E., and Harris, R. S. (2015) Degradation of the cancer genomic DNA deaminase APOBEC3B by SIV Vif. *Oncotarget* **6**, 39969-39979

112. Mishra, N., Reddy, K. S., Timilsina, U., Gaur, D., and Gaur, R. (2018) Human APOBEC3B interacts with the heterogenous nuclear ribonucleoprotein A3 in cancer cells. *J Cell Biochem* **119**, 6695-6703

113. McCann, J. L., Klein, M. M., Leland, E. M., Law, E. K., Brown, W. L., Salamango, D. J., and Harris, R. S. (2019) The DNA deaminase APOBEC3B interacts with the cell cycle protein CDK4 and disrupts CDK4-mediated nuclear import of Cyclin D1. *J Biol Chem*

114. Komor, A. C., Kim, Y. B., Packer, M. S., Zuris, J. A., and Liu, D. R. (2016) Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* **533**, 420-424

115. Nishida, K., Arazoe, T., Yachie, N., Banno, S., Kakimoto, M., Tabata, M., Mochizuki, M., Miyabe, A., Araki, M., Hara, K. Y., Shimatani, Z., and Kondo, A.

(2016) Targeted nucleotide editing using hybrid prokaryotic and vertebrate adaptive immune systems. *Science* **353**, pii: aaf8729

116. Nishimasu, H., Ran, F. A., Hsu, P. D., Konermann, S., Shehata, S. I., Dohmae, N., Ishitani, R., Zhang, F., and Nureki, O. (2014) Crystal structure of Cas9 in complex with guide RNA and target DNA. *Cell* **156**, 935-949

117. Gaudelli, N. M., Komor, A. C., Rees, H. A., Packer, M. S., Badran, A. H., Bryson, D. I., and Liu, D. R. (2017) Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage. *Nature* **551**, 464-471

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

119. Tsai, S. Q., Wyvekens, N., Khayter, C., Foden, J. A., Thapar, V., Reyon, D., Goodwin, M. J., Aryee, M. J., and Joung, J. K. (2014) Dimeric CRISPR RNA-guided FokI nucleases for highly specific genome editing. *Nat Biotechnol* **32**, 569-576

120. Chaudhuri, J., Tian, M., Khuong, C., Chua, K., Pinaud, E., and Alt, F. W. (2003) Transcription-targeted DNA deamination by the AID antibody diversification enzyme. *Nature* **422**, 726-730

121. Larson, E. D., and Maizels, N. (2004) Transcription-coupled mutagenesis by the DNA deaminase AID. *Genome Biology* **5**, 211

122. Ramiro, A. R., Stavropoulos, P., Jankovic, M., and Nussenzweig, M. C. (2003) Transcription enhances AID-mediated cytidine deamination by exposing single-stranded DNA on the nontemplate strand. *Nature Immunology* **4**, 452-456

123. Pavri, R. (2017) R Loops in the regulation of antibody gene diversification. *Genes (Basel)* **8**, pii: E154

124. Liu, M., and Schatz, D. G. (2009) Balancing AID and DNA repair during somatic hypermutation. *Trends Immunol* **30**, 173-181

125. Peters, A., and Storb, U. (1996) Somatic hypermutation of immunoglobulin genes is linked to transcription initiation. *Immunity* **4**, 57-65

126. Adolph, M. B., Love, R. P., Feng, Y., and Chelico, L. (2017) Enzyme cycling contributes to efficient induction of genome mutagenesis by the cytidine deaminase APOBEC3B. *Nucleic Acids Res* **45**, 11925-11940

127. Tang, W., and Liu, D. R. (2018) Rewritable multi-event analog recording in bacterial and mammalian cells. *Science* **360**, pii: eaap8992

128. Zafra, M. P., Schatoff, E. M., Katti, A., Foronda, M., Breinig, M., Schweitzer, A. Y., Simon, A., Han, T., Goswami, S., Montgomery, E., Thibado, J., Kastenhuber, E. R., Sánchez-Rivera, F. J., Shi, J., Vakoc, C. R., Lowe, S. W., Tschaharganeh, D. F., and Dow, L. E. (2018) Optimized base editors enable efficient editing in cells, organoids and mice. *Nat Biotechnol* **36**, 888-893

129. Kim, D., Lim, K., Kim, S. T., Yoon, S. H., Kim, K., Ryu, S. M., and Kim, J. S. (2017) Genome-wide target specificities of CRISPR RNA-guided programmable deaminases. *Nat Biotechnol* **35**, 475-480

130. Gehrke, J. M., Cervantes, O., Clement, M. K., Wu, Y., Zeng, J., Bauer, D. E., Pinello, L., and Joung, J. K. (2018) An APOBEC3A-Cas9 base editor with minimized bystander and off-target activities. *Nat Biotechnol* **36**, 977-982

131. Zuo, E., Sun, Y., Wei, W., Yuan, T., Ying, W., Sun, H., Yuan, L., Steinmetz, L. M., Li, Y., and Yang, H. (2019) Cytosine base editor generates substantial off-target single-nucleotide variants in mouse embryos. *Science* **364**, 289-292

132. Jin, S., Zong, Y., Gao, Q., Zhu, Z., Wang, Y., Qin, P., Liang, C., Wang, D., Qiu, J. L., Zhang, F., and Gao, C. (2019) Cytosine, but not adenine, base editors induce genome-wide off-target mutations in rice. *Science* **364**, 292-295

133. St.Martin, A., Salamango, D., Serebrenik, A., Shaban, N., Brown, W. L., Donati, F., Munagala, U., Conticello, S. G., and Harris, R. S. (2018) A fluorescent reporter for quantification and enrichment of DNA editing by APOBEC-Cas9 or cleavage by Cas9 in living cells. *Nucleic Acids Res* **46**, e84

134. St.Martin, A. R., Salamango, D. J., Serebrenik, A. A., Shaban, N. M., Brown, W. L., and Harris, R. S. (2019) A panel of eGFP reporters for single base editing by APOBEC-Cas9 editosome complexes. *Scientific Reports* **9**, 497

135. Zhou, C., Sun, Y., Yan, R., Liu, Y., Zuo, E., Gu, C., Han, L., Wei, Y., Hu, X., Zeng, R., Li, Y., Zhou, H., Guo, F., and Yang, H. (2019) Off-target RNA mutation induced by DNA base editing and its elimination by mutagenesis. *Nature* **571**, 275 – 278

136. Coelho, M. A., Li, S., Pane, L. S., Firth, M., Ciotta, G., Wrigley, J. D., Cuomo, M. E., Maresca, M., and Taylor, B. J. M. (2018) BE-FLARE: a fluorescent reporter of base editing activity reveals editing characteristics of APOBEC3A and APOBEC3B. *BMC Biol* **16**, 150

137. Kim, Y. B., Komor, A. C., Levy, J. M., Packer, M. S., Zhao, K. T., and Liu, D. R. (2017) Increasing the genome-targeting scope and precision of base editing with engineered Cas9-cytidine deaminase fusions. *Nat Biotechnol* **35**, 371-376

138. Zischewski, J., Fischer, R., and Bortesi, L. (2017) Detection of on-target and off-target mutations generated by CRISPR/Cas9 and other sequence-specific nucleases. *Biotechnol Adv* **35**, 95-104

139. Rees, H. A., and Liu, D. R. (2018) Base editing: precision chemistry on the genome and transcriptome of living cells. *Nat Rev Genet* **19**, 770-788

140. Liu, Z., Lu, Z., Yang, G., Huang, S., Li, G., Feng, S., Liu, Y., Li, J., Yu, W., Zhang, Y., Chen, J., Sun, Q., and Huang, X. (2018) Efficient generation of mouse models of human diseases via ABE- and BE-mediated base editing. *Nat Commun* **9**, 2338

141. Ryu, S. M., Koo, T., Kim, K., Lim, K., Baek, G., Kim, S. T., Kim, H. S., Kim, D. E., Lee, H., Chung, E., and Kim, J. S. (2018) Adenine base editing in mouse embryos and an adult mouse model of Duchenne muscular dystrophy. *Nat Biotechnol* **36**, 536-539

142. Burns, M. B., Leonard, B., and Harris, R. S. (2015) APOBEC3B: pathological consequences of an innate immune DNA mutator. *Biomed J* **38**, 102-110

143. Schatoff, E. M., Zafra, M. P., and Dow, L. E. (2019) Base editing the mammalian genome. *Methods* **S1046-2023**, 30318-30319

144. Harris, R. S. (2015) Molecular mechanism and clinical impact of APOBEC3B-catalyzed mutagenesis in breast cancer. *Breast Cancer Research* **17**, 8 (10 pages)

145. Kuong, K. J., and Loeb, L. A. (2013) APOBEC3B mutagenesis in cancer. *Nature Genetics* **45**, 964-965

146. Wagner, J. R., Demir, Ö., Carpenter, M. A., Aihara, H., Harki, D. A., Harris, R. S., and Amaro, R. E. (2019) Determinants of oligonucleotide selectivity of APOBEC3B. *J Chem Inf Model* **59**, 2264-2273

147. Jeffrey, P. D., Russo, A. A., Polyak, K., Gibbs, E., Hurwitz, J., Massague, J., and Pavletich, N. P. (1995) Mechanism of CDK activation revealed by the structure of a cyclinA-CDK2 complex. *Nature* **376**, 313-320

148. Jeffrey, P. D., Tong, L., and Pavletich, N. P. (2000) Structural basis of inhibition of CDK-cyclin complexes by INK4 inhibitors. *Genes and Development* **14**, 3115-3125

149. Malumbres, M., and Barbacid, M. (2009) Cell cycle, CDKs and cancer: a changing paradigm. *Nature Reviews Cancer* **9**, 153-166

150. Malumbres, M. (2014) Cyclin-dependent kinases. *Genome Biology* **15**, 10

151. Murray, A. W. (2004) Recycling the cell cycle: Cyclins revisited. *Cell* **116**, 221-234

152. Day, P. J., Cleasby, A., Tickle, I. J., O'Reilly, M., Coyle, J. E., Holding, F. P., McMenamin, R. L., Yon, J., Chopra, R., Lengauer, C., and Jhoti, H. (2009) Crystal structure of human CDK4 in complex with a D-type cyclin. *PNAS* **106**, 4166-4170

153. Chen, H. B., Xu, X. W., Wang, G. P., Zhang, B. Y., Wang, G., Xin, G. W., Liu, J. J., Jiang, Q., Zhang, H. Y., and Zhang, C. M. (2017) CDK4 protein is degraded by anaphase-promoting complex/cyclosome in mitosis and reaccumulates in early G(1) phase to initiate a new cell cycle in HeLa cells. *Journal of Biological Chemistry* **292**, 10131-10141

154. Nevins, J. R. (2001) The Rb/E2F pathway and cancer. *Human Molecular Genetics* **10**, 699-703

155. Dyson, N. (1998) The regulation of E2F by pRB-family proteins. *Genes & Development* **12**, 2245-2262

156. Massague, J. (2004) G1 cell-cycle control and cancer. *Nature* **432**, 298-306

157. Baker, S. J., and Reddy, E. P. (2012) CDK4: A key player in the cell cycle, development, and cancer. *Genes & Cancer* **3**, 658-669

158. Barretina, J., Taylor, B. S., Banerji, S., Ramos, A. H., Lagos-Quintana, M., DeCarolis, P. L., Shah, K., Socci, N. D., Weir, B. A., Ho, A., Chiang, D. Y., Reva, B., Mermel, C. H., Getz, G., Antipin, Y., Beroukhim, R., Major, J. E., Hatton, C., Nicoletti, R., Hanna, M., Sharpe, T., Fennell, T. J., Cibulskis, K., Onofrio, R. C., Saito, T., Shukla, N., Lau, C., Nelander, S., Silver, S. J., Sougnez, C., Viale, A., Winckler, W., Maki, R. G., Garraway, L. A., Lash, A., Greulich, H., Root, D. E., Sellers, W. R., Schwartz, G. K., Antonescu, C. R., Lander, E. S., Varmus, H. E., Ladanyi, M., Sander, C., Meyerson, M., and Singer, S. (2010) Subtype-specific genomic alterations define new targets for soft-tissue sarcoma therapy. *Nature Genetics* **42**, 715-721

159. Italiano, A., Bianchini, L., Gjernes, E., Keslair, F., Ranchere-Vince, D., Dumollard, J. M., Haudebourg, J., Leroux, A., Mainguene, C., Terrier, P., Chibon, F., Coindre, J. M., and Pedeutour, F. (2009) Clinical and biological significance of CDK4 amplification in well-differentiated and dedifferentiated liposarcomas. *Clinical Cancer Research* **15**, 5696-5703

160. Sheppard, K. E., and McArthur, G. A. (2013) The cell-cycle regulator CDK4: an emerging therapeutic target in melanoma. *Clinical Cancer Research* **19**, 5320-5328

161. Feng, Y. Q., Baig, T. T., Love, R. P., and Chelico, L. (2014) Suppression of APOBEC3-mediated restriction of HIV-1 by Vif. *Frontiers in Microbiology* **5**, 23

162. Lee, C. R., Park, Y. H., Min, H., Kim, Y. R., and Seok, Y. J. (2019) Determination of protein phosphorylation by polyacrylamide gel electrophoresis. *Journal of Microbiology* **57**, 93-100

163. Wegener, A. D., and Jones, L. R. (1984) Phosphorylation-induced mobility shift in phospholamban in sodium dodecyl sulfate-polyacrylamide gels. Evidence for a

protein structure consisting of multiple identical phosphorylatable subunits. *Journal of Biological Chemistry* **259**, 1834-1841

164. Xue, Y., Ren, J., Gao, X., Jin, C., Wen, L., and Yao, X. (2008) GPS 2.0, a tool to predict kinase-specific phosphorylation sites in hierarchy. *Molecular and Cellular Proteomics* **7**, 1598-1608

165. Obenauer, J. C., Cantley, L. C., and Yaffe, M. B. (2003) Scansite 2.0: Proteome-wide prediction of cell signaling interactions using short sequence motifs. *Nucleic Acids Research* **31**, 3635-3641

166. Stevenson-Lindert, L. M., Fowler, P., and Lew, J. (2003) Substrate specificity of CDK2-cyclin A. What is optimal? *Journal of Biological Chemistry* **278**, 50956-50960

167. Nigg, E. A. (1991) The substrates of the cdc2 kinase. *Seminars in Cell Biology* **2**, 261-270

168. Moreno, S., and Nurse, P. (1990) Substrates for p34cdc2: in vivo veritas? *Cell* **61**, 549-551

169. Grafstrom, R. H., Pan, W., and Hoess, R. H. (1999) Defining the substrate specificity of cdk4 kinase-cyclin D1 complex. *Carcinogenesis* **20**, 193-198

170. Pan, W., Sun, T., Hoess, R., and Grafstrom, R. (1998) Defining the minimal portion of the retinoblastoma protein that serves as an efficient substrate for Cdk4 kinase/cyclin D1 complex. *Carcinogenesis* **19**, 765-769

171. Kouno, T., Silvas, T. V., Hilbert, B. J., Shandilya, S. M. D., Bohn, M. F., Kelch, B. A., Royer, W. E., Somasundaran, M., Kurt Yilmaz, N., Matsuo, H., and Schiffer, C. A. (2017) Crystal structure of APOBEC3A bound to single-stranded DNA reveals structural basis for cytidine deamination and specificity. *Nature Communications* **8**, 15024

172. Land, A. M., Law, E. K., Carpenter, M. A., Lackey, L., Brown, W. L., and Harris, R. S. (2013) Endogenous APOBEC3A DNA cytosine deaminase is cytoplasmic and nongenotoxic. *Journal of Biological Chemistry* **288**, 17253-17260

173. Ayaz, P., Andres, D., Kwiatkowski, D. A., Kolbe, C. C., Lienau, P., Siemeister, G., Lücking, U., and Stegmann, C. M. (2016) Conformational adaption may explain the slow dissociation kinetics of Roniciclib (BAY 1000394), a type I CDK inhibitor with kinetic selectivity for CDK2 and CDK9. *ACS Chemical Biology* **11**, 1710-1719

174. Peng, G., Lei, K. J., Jin, W., Greenwell-Wild, T., and Wahl, S. M. (2006) Induction of APOBEC3 family proteins, a defensive maneuver underlying interferon-induced anti-HIV-1 activity. *Journal of Experimental Medicine* **203**, 41-46

175. Bonvin, M., Achermann, F., Greeve, I., Stroka, D., Keogh, A., Inderbitzin, D., Candinas, D., Sommer, P., Wain-Hobson, S., Vartanian, J. P., and Greeve, J. (2006) Interferon-inducible expression of APOBEC3 editing enzymes in human hepatocytes and inhibition of hepatitis B virus replication. *Hepatology* **43**, 1364-1374

176. Refsland, E. W., Hultquist, J. F., Luengas, E. M., Ikeda, T., Shaban, N. M., Law, E. K., Brown, W. L., Reilly, C., Emerman, M., and Harris, R. S. (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 Genetics* **10**, e1004761

177. Paladino, P., Marcon, E., Greenblatt, J., and Frappier, L. (2014) Identification of herpesvirus proteins that contribute to G1/S arrest. *J Virol* **88**, 4480-4492

178. Desimmie, B. A., Delviks-Frankenberrry, K. A., Burdick, R. C., Qi, D., Izumi, T., and Pathak, V. K. (2014) Multiple APOBEC3 restriction factors for HIV-1 and one Vif to rule them all. *J Mol Biol* **426**, 1220-1245

179. Izumi, T., Io, K., Matsui, M., Shirakawa, K., Shinohara, M., Nagai, Y., Kawahara, M., Kobayashi, M., Kondoh, H., Misawa, N., Koyanagi, Y., Uchiyama, T., and Takaori-Kondo, A. (2010) HIV-1 viral infectivity factor interacts with TP53 to induce G2 cell cycle arrest and positively regulate viral replication. *Proc Natl Acad Sci U S A* **107**, 20798-20803

180. Ma, W., Ho, D. W., Sze, K. M., Tsui, Y. M., Chan, L. K., Lee, J. M., and Ng, I. O. (2019) APOBEC3B promotes hepatocarcinogenesis and metastasis through novel deaminase-independent activity. *Molecular Carcinogenesis* **58**, 643-653

181. Nikkila, J., Kumar, R., Campbell, J., Brandsma, I., Pemberton, H. N., Wallberg, F., Nagy, K., Scheer, I., Vertessy, B. G., Serebrenik, A. A., Monni, V., Harris, R. S., Pettitt, S. J., Ashworth, A., and Lord, C. J. (2017) Elevated APOBEC3B expression drives a kataegic-like mutation signature and replication stress-related therapeutic vulnerabilities in p53-defective cells. *British Journal of Cancer* **117**, 113-123

182. Ohba, K., Ichiyama, K., Yajima, M., Gemma, N., Nikaido, M., Wu, Q., Chong, P., Mori, S., Yamamoto, R., Wong, J. E., and 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

183. Parker, G. A., Crook, T., Bain, M., Sara, E. A., Farrell, P. J., and Allday, M. J. (1996) Epstein-Barr virus nuclear antigen (EBNA)3C is an immortalizing oncoprotein with similar properties to adenovirus E1A and papillomavirus E7. *Oncogene* **13**, 2541-2549

184. Yim, E. K., and Park, J. S. (2005) The role of HPV E6 and E7 oncoproteins in HPV-associated cervical carcinogenesis. *Cancer Research and Treatment* **37**, 319-324

185. Martin, L. G., Demers, G. W., and Galloway, D. A. (1998) Disruption of the G1/S transition in human papillomavirus type 16 E7-expressing human cells is associated with altered regulation of cyclin E. *Journal of Virology* **72**, 975-985

186. Nascimento, R., Dias, J. D., and Parkhouse, R. M. (2009) The conserved UL24 family of human alpha, beta and gamma herpesviruses induces cell cycle arrest and inactivation of the cyclinB/cdc2 complex. *Archives of Virology* **154**, 1143-1149

187. Davy, C. E., Jackson, D. J., Raj, K., Peh, W. L., Southern, S. A., Das, P., Sorathia, R., Laskey, P., Middleton, K., Nakahara, T., Wang, Q., Masterson, P. J., Lambert, P. F., Cuthill, S., Millar, J. B., and Doorbar, J. (2005) Human papillomavirus type 16 E1 E4-induced G2 arrest is associated with cytoplasmic retention of active Cdk1/cyclin B1 complexes. *Journal of Virology* **79**, 3998-4011

188. Andersen, J. L., Le Rouzic, E., and Planelles, V. (2008) HIV-1 Vpr: mechanisms of G2 arrest and apoptosis. *Experimental and Molecular Pathology* **85**, 2-10

189. Alt, J. R., Cleveland, J. L., Hannink, M., and Diehl, J. A. (2000) Phosphorylation-dependent regulation of cyclin D1 nuclear export and cyclin D1-dependent cellular transformation. *Genes and Development* **14**, 3102-3114

190. Alt, J. R., Gladden, A. B., and Diehl, J. A. (2002) p21(Cip1) promotes Cyclin D1 nuclear accumulation via direct inhibition of nuclear export. *Journal of Biological Chemistry* **277**, 8517-8523

191. Dhar, K. K., Branigan, K., Parkes, J., Howells, R. E., Hand, P., Musgrove, C., Strange, R. C., Fryer, A. A., Redman, C. W., and Hoban, P. R. (1999) Expression

and subcellular localization of Cyclin D1 protein in epithelial ovarian tumour cells. *British Journal of Cancer* **81**, 1174-1181

192. Musgrove, E. A., Caldon, C. E., Barraclough, J., Stone, A., and Sutherland, R. L. (2011) Cyclin D as a therapeutic target in cancer. *Nature Reviews Cancer* **11**, 558-572

193. Diehl, J. A., and Sherr, C. J. (1997) A dominant-negative cyclin D1 mutant prevents nuclear import of cyclin-dependent kinase 4 (CDK4) and its phosphorylation by CDK-activating kinase. *Molecular and Cellular Biology* **17**, 7362-7374

194. Holland, T. A., Elder, J., McCloud, J. M., Hall, C., Deakin, M., Fryer, A. A., Elder, J. B., and Hoban, P. R. (2001) Subcellular localisation of cyclin D1 protein in colorectal tumours is associated with p21(WAF1/CIP1) expression and correlates with patient survival. *International Journal of Cancer* **95**, 302-306

195. Carpenter, M. A., Law, E. K., Serebrenik, A., Brown, W. L., and Harris, R. S. (2019) A lentivirus-based system for Cas9/gRNA expression and subsequent removal by Cre-mediated recombination. *Methods* **156**, 79-84

196. Brown, W., Law, E. , Carpenter, M. ,Argyris, P. ,Levin-Klein, R. ,Ranum, A. ,Molan, A. ,Forster, C. ,Anderson, B. ,Lackey, L., Harris, R. (2019) A rabbit monoclonal antibody against the antiviral and cancer genomic DNA mutating enzyme APOBEC3B. mAbs

197. McNamara, R. P., McCann, J. L., Gudipaty, S. A., and D'Orso, I. (2013) Transcription factors mediate the enzymatic disassembly of promoter-bound 7SK snRNP to locally recruit P-TEFb for transcription elongation. *Cell Reports* **5**, 1256-1268

198. Hess, G. T., Tycko, J., Yao, D., and Bassik, M. C. (2017) Methods and applications of CRISPR-mediated base editing in eukaryotic genomes. *Mol Cell* **68**, 26-43

199. Grünewald, J., Zhou, R., Garcia, S. P., Iyer, S., Lareau, C. A., Aryee, M. J., and Joung, J. K. (2019) Transcriptome-wide off-target RNA editing induced by CRISPR-guided DNA base editors. *Nature* **569**, 433-437

200. Koblan, L. W., Doman, J. L., Wilson, C., Levy, J. M., Tay, T., Newby, G. A., Maianti, J. P., Raguram, A., and Liu, D. R. (2018) Improving cytidine and adenine base editors by expression optimization and ancestral reconstruction. *Nat Biotechnol* **36**, 843-846

201. Zong, Y., Song, Q., Li, C., Jin, S., Zhang, D., Wang, Y., Qiu, J. L., and Gao, C. (2018) Efficient C-to-T base editing in plants using a fusion of nCas9 and human APOBEC3A. *Nat Biotechnol* **36**, 950-953

202. Wang, X., Li, J., Wang, Y., Yang, B., Wei, J., Wu, J., Wang, R., Huang, X., Chen, J., and Yang, L. (2018) Efficient base editing in methylated regions with a human APOBEC3A-Cas9 fusion. *Nat Biotechnol* **36**, 946-949

203. Hu, J. H., Miller, S. M., Geurts, M. H., Tang, W., Chen, L., Sun, N., Zeina, C. M., Gao, X., Rees, H. A., Lin, Z., and Liu, D. R. (2018) Evolved Cas9 variants with broad PAM compatibility and high DNA specificity. *Nature* **556**, 57-63

204. Thuronyi, B. W., Koblan, L. W., Levy, J. M., Yeh, W. H., Zheng, C., Newby, G. A., Wilson, C., Bhaumik, M., Shubina-Oleinik, O., Holt, J. R., and Liu, D. R. (2019) Continuous evolution of base editors with expanded target compatibility and improved activity. *Nat Biotechnol*

205. Komor, A. C., Zhao, K. T., Packer, M. S., Gaudelli, N. M., Waterbury, A. L., Koblan, L. W., Kim, Y. B., Badran, A. H., and Liu, D. R. (2017) Improved base

excision repair inhibition and bacteriophage Mu Gam protein yields C:G-to-T:A base editors with higher efficiency and product purity. *Sci Adv* **3**, eaao4774

206. Kleinstiver, B. P., Sousa, A. A., Walton, R. T., Tak, Y. E., Hsu, J. Y., Clement, K., Welch, M. M., Horng, J. E., Malagon-Lopez, J., Scarfò, I., Maus, M. V., Pinello, L., Aryee, M. J., and Joung, J. K. (2019) Engineered CRISPR-Cas12a variants with increased activities and improved targeting ranges for gene, epigenetic and base editing. *Nat Biotechnol* **37**, 276-282

207. Li, X., Wang, Y., Liu, Y., Yang, B., Wang, X., Wei, J., Lu, Z., Zhang, Y., Wu, J., Huang, X., Yang, L., and Chen, J. (2018) Base editing with a Cpf1-cytidine deaminase fusion. *Nat Biotechnol* **36**, 324-327

208. Bennett, R. P., Diner, E., Sowden, M. P., Lees, J. A., Wedekind, J. E., and Smith, H. C. (2006) APOBEC-1 and AID are nucleo-cytoplasmic trafficking proteins but APOBEC3G cannot traffic. *Biochem Biophys Res Commun* **350**, 214-219

209. Patenaude, A. M., Orthwein, A., Hu, Y., Campo, V. A., Kavli, B., Buschiazzo, A., and Di Noia, J. M. (2009) Active nuclear import and cytoplasmic retention of activation-induced deaminase. *Nat Struct Mol Biol* **16**, 517-527

210. Wang, J., Shaban, N. M., Land, A. M., Brown, W. L., and Harris, R. S. (2018) Simian immunodeficiency virus Vif and human APOBEC3B interactions resemble those between HIV-1 Vif and human APOBEC3G. *Journal of Virology* **92**, e00447-00418.

211. Magoč, T., and Salzberg, S. L. (2011) FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* **27**, 2957-2963

212. Alam, K. K., Chang, J. L., and Burke, D. H. (2015) FASTAptamer: A bioinformatic toolkit for high-throughput sequence analysis of combinatorial selections. *Molecular Therapy Nucleic Acids* **4**, e230

213. Martincorena, I., Raine, K. M., Gerstung, M., Dawson, K. J., Haase, K., Van Loo, P., Davies, H., Stratton, M. R., and Campbell, P. J. (2018) Universal Patterns of Selection in Cancer and Somatic Tissues. *Cell* **173**, 1823

214. Hanahan, D., and Weinberg, R. A. (2011) Hallmarks of cancer: the next generation. *Cell* **144**, 646-674