The evolutionary history and activity of a gain-of-function polymorphism in a human antiviral gene

Cristina Wittkopp

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

Year 2016

Reading Committee:

Michael Emerman, Chair

Denise Galloway

Jaisri Lingappa

Program Authorized to Offer Degree:

Microbiology

University of Washington

Abstract

Primate evolutionary dynamics of APOBEC3 anti-lentiviral activity

Cristina Wittkopp

Chair of the Supervisory Committee:
Professor Michael Emerman
Microbiology

Humans express seven human APOBEC3 proteins, which can inhibit viruses and endogenous retroelements through cytidine deaminase activity. The seven paralogs differ in the potency of their antiviral effects, as well as in their antiviral targets. One APOBEC3, APOBEC3C, is exceptional as it had previously been found to only weakly block viruses and endogenous retroelements compared to other APOBEC3s. However, our positive selection analyses suggested that APOBEC3C has played a role in pathogen defense during primate evolution. Here, I describe a single nucleotide polymorphism in human APOBEC3C, a change from serine to isoleucine at position 188 (I188) that confers potent antiviral activity against HIV-1. The gain-of-function APOBEC3C SNP results in increased enzymatic activity and hypermutation of target sequences when tested *in vitro*, and correlates with increased dimerization of the protein.

Furthermore, I show, using a fully-dimerizing point mutant as well as a synthetic dimer of APOBEC3C, that dimerization correlates with potent antiviral activity. The I188 is widely distributed in human African populations, and is the ancestral primate allele, but is not found in chimpanzees or gorillas. Thus, while other hominids have lost activity of this antiviral gene, it has been maintained, or re-acquired, as a more active antiviral gene in a subset of humans. Taken together, my results suggest that APOBEC3C is in fact involved in protecting hosts from lentiviruses.

TABLE OF CONTENTS

LIST OF FIGURES

ACKNOWLEDGEMENTS

Most importantly, I thank my PhD mentor, Michael Emerman. I am so grateful for the opportunity to train under Michael's guidance. I am also grateful for his patience as a teacher and his willingness to let me think for myself, even though sometimes it meant learning lessons the hard way. Next, I thank the Emerman lab. I am especially grateful for the lab's scientifically rigorous lab meetings and practice talks. Our thorough scientific discussions were a wonderful training opportunity. Thank you to Oliver, your mentorship and encouragement were crucial to helping me navigate the ups and downs of graduate school. Molly, thank you for the laughter that we shared, and for the happiness and energy you bring to the lab. Chelsea, thank you for the wisdom you provided as a senior grad student; thank you for paving the way! Erin, thank for being such a fun bay-mate! Lily, thank you for your research support, and especially for your amazing Western blot abilities. To my friends and fellow Micro grad students, thank you for being there for me and helping me put things in perspective when experiments didn't work. Lastly, I would like to thank my family for their unwavering support. Thank you to my parents, Melinda, Sam, and Craig, for all your encouragement and love. Thank you to my sister, Shannon, for being there for me every step of the way. And lastly, thank you to my son James, for being my inspiration.

Chapter 1. INTRODUCTION

1.1 RETROVIRUSES

Retroviruses are enveloped, positive sense, single stranded RNA viruses(1). The viral RNA is reverse transcribed to make double stranded DNA that is then integrated into the host genome. The integrated DNA, known as the provirus, serves as a template for viral gene expression as well as production of the viral RNA genome, which gets packaged as two copies into budding virions(1). A consequence of the proviral stage is that retroviruses can lie dormant at this step, allowing the virus to evade immune detection and response. Another consequence of the proviral stage is that, if integrated into germline cells, retroviruses can become transmitted vertically. Vertically transmitted retroviruses, known as endogenous retroviruses, can lose the ability to produce virions that bud from the cell. In contrast, exogenous retroviruses must be able to replicated their genomes, but also produce a structure to package their genetic material for dispersal to target cells in a subsequent round of infection. All exogenous retroviruses express the proteins Gag (a structural protein), Env (a membrane-bound protein required for cell entry), and Pol (encoding enzymes needed for replication, such as polymerase and protease).

Retroviruses are subdivided into seven genera; alpha- epsilion retroviruses, lentiviruses, and spumaretroviruses(2). Alpharetroviruses, betaretroviruses and gammaretroviruses are all simple retroviruses, which encode only *gag*, *env*, *pro*, and *pol* genes and each is typified by a different oncovirus; avian sarcoma and leukosis virus, mouse mammary tumor virus, Mason-Pfizer monkey virus, respectively(2). Deltaretroviruses, epsilonretroviruses, lentiviruses and spumaretroviruses are complex retroviruses, in that they encode *gag*, *env*, and *pol*, as well as additional genes that are alternatively spliced. Only two of these five retroviruses genera are

known to include human pathogens; deltaretroviruses include human T cell leukemia viruses, and lentiviruses include the Human Immunodeficiency Virus, HIV.

1.2 HUMAN IMMUNODEFICIENCY VIRUS & THE AIDS EPIDEMIC

HIV was discovered in 1983(3), and was identified as the etiologic agent responsible for a mysterious, devastating epidemic known as Acquired Immunodeficiency Disease, or AIDS(4-7). There are two types of HIV known to be circulating in humans, HIV-1 and HIV-2, but only HIV-1 has been involved in the global AIDS epidemic. HIV-2 causes a much less pathogenic disease(8), however patients can develop AIDS. The AIDS epidemic first became recognized in the early 1980s, when young, otherwise healthy people began succumbing to opportunistic infections. AIDS is still a global epidemic, as there are currently 37 million people infected, and it is estimated that 35 million people have died from the disease since the epidemic began(9). Most of the AIDS cases around the world occur in sub-Saharan Africa, where more than 25 million individuals are currently infected(9).

AIDS is a chronic illness that causes depletion of a certain subset of immune cells, T helper cells or CD4+ T cells, leading to death(10). No cure for AIDS has yet been developed. However, combination antiretroviral therapy, which consists of a cocktail of small molecule drugs that inhibit multiple steps of the HIV life cycle, was approved for use in the mid-1990s. Before combination antiretroviral therapies existed HIV a case fatality rate that approaches 100%(11), but this is no longer fatal for individuals with access to treatment. However, many individuals in developing countries do not have access to treatment. Therefore, efforts to develop a vaccine to protect against HIV transmission continue.

1.3 HIV Life Cycle

HIV-1 has a genome of approximately 10kb and encodes nine genes (*gag*, *pol*, *env, tat, rev*, *nef, vpu, vpr*, and *vif*)(2). Gag, Pol, and Env are conserved across retroviruses and therefore were described in section 1.1. Tat and Rev are regulatory proteins that aid in the efficiency of transcription and in the nuclear export of viral transcripts, respectively(12). Nef, Vpu, Vpr, and Vif are known as accessory proteins and play diverse roles in altering host cell biology to promote infection(13). In order to enter cells, HIV uses CD4 as a receptor and can use two different cellular co-receptors: either CCR5 or CXCR4(14). Most viral strains are CCR5-tropic, however a minority viral strains can use CXCR4 as a co-receptor. CD4 is a glycoprotein expressed on the surface of certain T cells (known as CD4+ T cells), macrophages, monocytes, and dendritic cells. CCR5 is expressed on the surface of memory CD4+ T cells, and these cells are the primary cellular target of HIV. CD4+ T cells are also known as T-helper cells, because this cell type facilitates the activation & recruitment of other immune cell types by the release of cytokines and other signals. The viral protein that facilitates binding to the receptor and coreceptor is Env. Env is a heavily glycosylated, membrane-bound protein produced as a precursor known as gp160, which gets cleaved by a host enzyme to form two subunits, gp120 and gp41(14). On the envelope of a virion, Env is expressed as heterodimers of gp120 and gp41, and these heterodimers come together in trimers. Env binds to the receptor and co-receptor through interactions mediated by the gp120 subunit. Once bound to the receptor/co-receptor pair, Env undergoes a conformational change, exposing the gp41 subunit and allowing its insertion into the host membrane. When gp41 inserts into the host membrane this facilitates fusion of the viral envelope with the host cell membrane. When the envelope fuses with the host cell, this releases the viral capsid into the cytoplasm.

The viral capsid is the inner compartment of the virus and contains the viral genome as well as proteins needed for subsequent stages of the life cycle. Next, uncoating of the capsid and reverse transcription of the viral genome occur, however it is not well understood which order these events occur in. Uncoating of the virus particle must be temporally regulated, as premature uncoating due to capsid instability exposes the virus to host antiviral factors that trigger an immune response(15, 16). At this stage, the homodimers of viral RNA are used as a template for reverse transcription. In order for reverse transcription to occur, the viral enzyme reverse transcriptase, a primer, dNPTs, and the viral genome template are necessary. In the first step of HIV reverse transcription, the primer, which is a host cell lys3 tRNA, binds to complementary sequence near the 5' end of the viral (+) strand RNA, to initiate elongation from the primer binding site to reverse transcribe the unique 5' region of the genome (U5) as well as the 5' repeat region (R) that is also present at the 5' end(2). Next the newly synthesized U5-R (-) DNA segment transfers to the 3' end and the 5' R anneals with the 3' R, based on complementarity. After this step, known as minus strand transfer, elongation continues all the way to the initial primer binding site. An RNA-DNA duplex is formed during this initial stage of synthesis. Reverse transcriptase has RNAseH activity, and this leads to the degradation of the RNA in the RNA-DNA duplex and leaves behind a region of single stranded DNA. However, a purine rich region (PPT) near the initial 3' end of the RNA genome is resistant to RNAseH degradation, so the PPT is left to serve as a primer for elongation of the next DNA strand, the (+) DNA. Elongation from the ppt produces only the unique 3' region (U3), R, and U5, which was transferred during (-) strand production. Next, U3-R-U5 (+) segment is transferred to the other end of the (-) and binds based on complementary to the R and U5 regions. This serves as the primer for the final stage of extension, resulting in double stranded DNA, and this sequence

differs from the RNA genome in that it encodes U5, R, and U3 region at both ends of the molecule. During reverse transcription, the reverse transcriptase enzyme switches templates by jumping from one RNA to the other RNA in the genomic homodimer pair. This template switching facilitates the emergence of recombinant lentiviruses. After (or during) reverse transcription, the new double-stranded DNA is imported into the nucleus as part of a preintegration complex containing viral DNA and other viral proteins. Importantly, the preintegration complex contains the enzyme integrase (IN), which catalyzes the initial steps of integration of the provirus in to the host genome. First, integrase removes two nucleotides from each 3' end of the viral DNA(2). Next, integrase processes these ends and transfers them to form phosphodiester linkages with host DNA, and the host DNA targeted for each link are 5 nucleotides apart from one another. The remaining steps of integration, such as the breakage of the other host strand, and the repair using the viral 5'end are carried out by host DNA repair enzymes.

Once integration is complete, the provirus is the template for viral gene expression (although some minor transcription might also occur from unintegrated templates). The long terminal repeat region of the genome, or LTR, at the 5' end of the genome serves as the promoter for viral transcription. The trans-activation response element is an RNA element encoded by part of the LTR which is bound by viral protein Tat to stimulate transcript elongation by recruitment of the elongation factor P-TEFb to the nascent RNA/polymerase complex(12). Initial transcription is inefficient until Tat proteins are produced, at which time efficiency of transcription is drastically enhanced. HIV transcription leads to the production of completely spliced, incompletely spliced, and unspliced RNA. As for most host mRNAs, the fully spliced viral mRNAs (tat, rev, and nef) are exported out of the nucleus for translation. However, cells

typically degrade incompletely spliced or unspiced RNAs in the nucleus. To export of incompletely spliced viral mRNAs from the nucleus, the HIV protein Rev binds to incompletely spliced transcripts (env, vpr, vpu, and vif) and unspliced transcripts (encoding gag-pol, as well as the viral genome) in a region of the RNA called the RRE (rev-response element), and escorts these mRNAs out of the nucleus via the CRM1 pathway.

After translation of the viral genes occurs, virions begin to assemble. The structural subunits of HIV virions are expressed from the *gag* gene. Gag transcripts are translated from the same RNA sequence that encodes the viral genome. Gag is translated as a polypeptide, which contains four domains; matrix, capsid, and nucleocapsid, as well as two spacer domains. However, due to a ribosomal frame shift that occurs 5% of the time, 1 out of every 20 Gag translation events creates a Gag-Pro-Pol precursor. Gag traffics as oligomers to the plasmid membrane, where the subunits further polymerize. Assembly of Gag particles requires multiple cellular factors(17). To form an immature capsid shell approximately 2,000 Gag polyproteins multimerize, with the matrix facing out and the p6 comprising the inner domain. Interactions between the capsid domain of each subunit are necessary for mutlimerization. Matrix targets the multimers to the plasmid membrane. Nucleocapsid is important for targeting the viral genome to the virion. p6 is not known to be important for assembly, however, it is necessary at a later stage in the life cycle. When Gag multimers assemble to form immature virus particles, several viral components are packaged into the virions- two copies of the viral genome, Reverse transcriptase, Integrase, and Protease. The Env protein is expressed at the plasma membrane as trimers, and is recruited once virions assemble through an interaction with the Matrix domain of Gag. Only approximately 10-15 Env trimers are expressed on the surface of a single HIV-1 virion and strikingly, for most HIV strains expression of only 2-3 trimers on a virion is sufficient for

cellular entry(18). Once the viral particles assemble, a hexagonal lattice of Gag molecules alter the membrane curvature. Next the virions bud from the cell using the endosomal sorting complex required for transport (ESCRT) machinery(19). As HIV buds from the cell, the immature viral particles go through a process known as maturation to form infectious particles. At this step, the viral protein protease begins to cleave Gag polyproteins and Gag-Pro-Pol polyproteins into the separate subunits matrix, capsid, and nucleocapsid. In immature particles, capsid subunits of the Gag polyprotein interact to form a spherical particle. In contrast, protease-cleaved capsid molecules interact to form a conical shape. Protease is also important for cleaving the Protease and Polymerase from the Gag-Pol precursor, as well as for cleaving the different domains of Pol: reverse transcriptase, integrase, and protease. Once virion maturation is complete, the virus is competent to infect a new cell, and the viral lifecycle is then repeated.

1.4 HIV Pathogenesis

HIV is typically transmitted through sexual contact, through blood-to-blood contact such as needle sharing, or through vertical transmission from mother to child during pregnancy, delivery or breastfeeding. During initial infection, HIV infects target cells in the periphery, and these infected cells traffic to regional lymph nodes. Gastrointestinal CD4+ T cells also become infected during early infection and is a major site of HIV replication(20). The only known genetic determinant of HIV resistance is an allele that encodes a deletion in CCR5(21-23). After HIV acquisition, acute infection occurs during the first two to six weeks of infection, corresponding with viremia and with rapid depletion of CD4+ T cells. Acute HIV infection can be asymptomatic, or flu-like symptoms, diarrhea, and/or a rash occur in 40-90% of cases. During this stage, virus becomes disseminated throughout the lymphoid system and sometimes the central nervous system. However, once an adaptive immune response is mounted, this initiates a

recovery stage when the amount of virus present in the blood significantly decreases. Patient studies suggest that a CD8+ T cell response is responsible for the control of HIV replication after the acute stage(24). After acute infection, a steady-state level of viral replication can occur for years and the amount of HIV present in the blood at this time is known as the viral set-point. This constitutes the transition from the acute stage of infection to the chronic stage. Eventually during the chronic stage, the number of CD4+ T cells steadily declines to dangerous and eventually fatal levels when the immune system can no longer fight infections. When a patient's CD4+ T cell count drops to 200 cells per microliter of blood, this is clinically defined as AIDS.

1.5 Primate Lentiviruses

Human immunodeficiency viruses, HIV-1 and HIV-2, originated due to cross-species transmissions of lentiviruses, which infect African primates, known as simian immunodeficiency viruses, or SIVs. HIV-1 is the result of at least four separate cross-species transmissions from the SIVcpz/SIVgor lineage of lentiviruses(25), which infect chimpanzees and gorillas, respectively. These four transmission events gave rise to four HIV-1 groups; group M (responsible for the major pandemic), group N, group O, and group P. The SIVcpz/SIVgor clade that gave rise to all HIV-1 strains derived from a recombination event that likely occurred between two different SIVs (SIVrcm, which infects red-capped mangabeys, and lineage SIVmus/mon/gsn, which infects Cercopithecus monkeys)(26). On the other hand, the closest relative to HIV-2 is a virus that infects sooty mangabeys, SIVsmm, so SIVsmm likely crossed into humans as the origin of HIV-2(27). There is evidence of natural SIV infections in over 40 non-human primate species(28). All of these natural infections occur in primate species in Africa. Phylogenetic analysis of SIVs shows several instances of cross-species transmissions between non-human primate species.

When cross-species of primate lentiviruses occurs, the virus may become more pathogenic in the new host species. For example, SIVmac, a virus that originated from an accidental laboratory cross-species transmission from sooty mangabeys to rhesus macaques, causes severe pathogenesis in macaques (29), but SIVsmm is not pathogenic to sooty mangabeys (30, 31). Furthermore, SIVagm, a virus that naturally infects several species of African green monkeys, has also been shown to be non-pathogenic to its hosts(32). For old world monkeys that harbor natural infections, host adaptation is thought to occur (such as changes in T cell gene expression) that dampens the pathogenesis of the virus(33). In addition to differences across SIV-harboring hosts, there are genetic differences in SIV viruses across virus clades. For instance, in addition to the genes encoded by HIV-1, some SIVs and HIV-2 encode vpx, which is a paralog of vpr. Another genetic difference among HIV/SIVs is that some lineages lack the *vpu* gene. These genetic differences across SIVs likely have significant impacts on host-virus interactions.

1.6 RESTRICTION FACTORS AND VIRAL ANTAGONISTS

Viruses put evolutionary pressure on hosts to evade viral pathogenesis. Therefore, host species have evolved many ways to block viral infection. In addition to the adaptive and innate immune responses, organisms may also express restriction factors, or antiviral intrinsic immunity genes(34). These factors are able to directly detect and respond to infection, when expressed in an infected cell. Many steps of the HIV life cycle can be blocked by restriction factors. However, viruses can adapt to overcome restriction by the host by expressing genes that function to block restriction factors. These viral genes that block restriction factors are known as viral antagonists. If a host species is infected with a virus that encodes an antagonist that neutralizes a restriction factor, this can exert evolutionary pressure on the host to select for restriction factor alleles that

are resistant to the viral antagonist. However, viruses evolve quickly and can adapt to overcome antagonist-resistant restriction factors. Therefore, pressure is placed back on the host to select for new antagonist-resistance alleles. This back-and-forth evolutionary pressure on the host and virus is known as an evolutionary arms race, as the adaptation and counter-adaption occurs in a positive feedback loop, continually feeding the cycle(35). The genetic signature classically associated with host-virus evolutionary arms races is positive selection. Positive selection is marked by an excess of non-synonymous mutations compared to synonymous mutations for a given gene phylogeny. Most genes evolve under negative selection, whereby synonymous mutations occur at a greater rate than non-synonymous mutations, because non-synonymous mutations are often detrimental to gene function. In contrast, for host restriction factors, non-synonymous mutations can be beneficial in order to evade antagonism by a virus. As these restriction factors accrue an abundance of non-synonymous over millions of years, this leads to vast sequence diversity across primate restriction factor genes(36). This can contribute to species-specificity of viruses. Lentiviruses, in particular, infect primates in a species-specific manner, and host restriction factors play an important role in protecting primates from some cross-species transmission events. In the next subsection of this chapter, specific primate restriction factors will be described, along with the viral antagonists or viral defenses against associated with each restriction factor.

Anti-lentiviral TRIM genes

The tripartite motif-containing family or TRIM superfamily includes over 70 known human TRIM proteins(37). TRIM proteins are characterized by a conserved TRIM, or RBCC motif, which includes a RING domain, two B boxes, and a coiled-coil domain. The RING domain can mediate the transfer of ubiquitin or similar molecules. Many TRIM family proteins are induced by interferon and many are involved in the detection of or response to viruses. TRIM proteins that have been reported to restrict HIV are TRIM5, TRIM22, and TRIM34. Of these three TRIMs, TRIM5 is the only one with a well-defined role against HIV. TRIM5 is a restriction factor that can block lentiviruses at the step of uncoating, through a direct interaction with the capsid lattice. In addition to the TRIM motif, TRIM5 also contains a PRYSPRY domain, and this is the region that binds to the capsid. TRIM binding to capsid destabilizes capsid-capsid interactions and also induces capsid degradation. TRIM5-mediated restriction of HIV is dependent on a direct interaction with another host gene, cylophilinA. The human version of TRIM5 cannot block HIV-1, but does block certain strains of murine leukemia virus(38). In contrast, rhesus macaques express a version of TRIM5 that completely blocks HIV infection(39), and in fact this is thought to be a major reason why rhesus monkeys are resistant to laboratory acquisition of HIV. Interestingly, some new world monkeys and certain old world monkey species express a fusion between TRIM5 and cyclophilinA, known as TrimCyp, and these proteins have potent antiviral activity against HIV and other primate lentiviruses(40, 41). Furthermore, it is hypothesized that other TRIMs may play a role in blocking lentiviruses, although more studies are needed to verify this.

Tetherin

In contrast to TRIM5, which specifically binds and inhibits certain viruses based on protein-protein interactions, the restriction factor Tetherin has very broad activity against several classes of viruses. Tetherin blocks budding of virions by tethering the viral membrane to the cell. In addition to HIV, Tetherin also blocks certain other enveloped viruses, including Ebola and Kasposi's sarcoma herpesvirus(42, 43). However, many viruses have evolved mechanisms to evade Tetherin-mediated restriction. HIV-1 is resistant to human Tetherin because the viral gene vpu binds to tetherin and targets it for ubiquitination and degradation. In fact, Tetherin was discovered to be a restriction factor during experiments to uncover why certain cell lines are resistant to laboratory HIV mutants that lack the gene *vpu(44, 45)*. Interestingly, other primate lentiviruses utilize different viral proteins to block Tetherin. Most SIVs encode a protein, Nef, that is able to bind to and mediate degradation of tetherin(46-48).

The APOBEC3 gene family

APOBEC3 proteins are part of the AID/APOBEC gene family, which are all cytidine deaminases that can either mutate mRNA or ssDNA. To summarize what is known about the other AID/APOBEC family members, first, Activation-Induced Cytidine Deaminase (AID) mutates DNA in B cells to promote antibody diversity. APOBEC1 is expressed only in the small intestines (for humans) and is an mRNA editing cytidine deaminase, which introduces an early stop codon in the mRNA of ApolipoproteinB, a gene involved in lipid metabolism(49, 50). APOBEC2 is not well-studied, but has been shown to be important for muscle development in both mice and zebrafish(51, 52). Humans encode seven APOBEC3 genes in tandem on chromosome 22, APOBEC3A APOBEC3B APOBEC3C APOBEC3D APOBEC3F APOBEC3G, and APOBEC3H. Each human APOBEC3 is composed of either a single cytidine

deaminase domain, or two tandem cytidine deaminase domains (Figure 1A). APOBEC3 structure is composed of a core of five beta sheets, surrounded by six alpha helices(53).

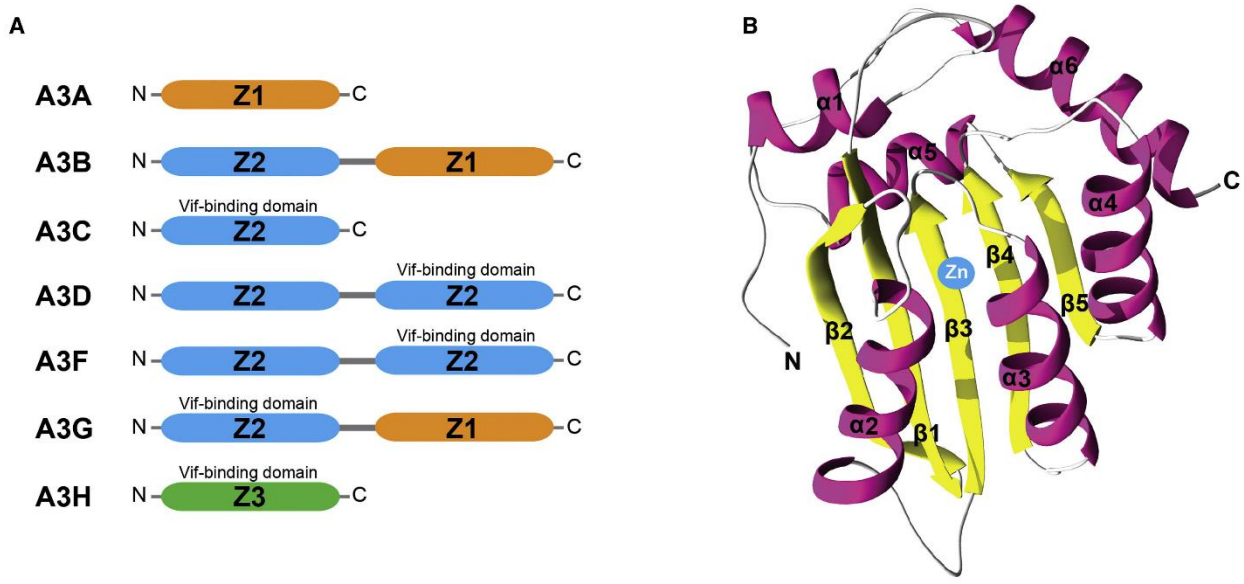


Figure 1. A. The seven human APOBEC3s, either composed of a single cytidine deaminase domain or two cytidine deaminase domains. B. The structure of APOBEC3 cytidine deaminase domains. Figure adapted from Aydin, 2014 (53).

APOBEC3s are mammal-specific, and although it however there have been many expansions/contractions in the number of APOBEC3s expressed across mammals (54). For instance, rodents only encode one APOBEC3, horses encode six and cats encode four. APOBEC3s all have a conserved (H/C)-(A/V)-E-($X_{24–30}$)-(P-C-$X_{2–4}$-C) sequence that coordinates a zinc molecule and is required for deaminase activity. In the APOBEC3-catalyzed deamination reaction, the conserved histines coordinate zinc, which stabilizes the formation of a nucleophilic hydroxyl molecule from water through a reaction with the conserved glutamic acid. The hydroxyl performs a nucleophilic attack on the targeted cytidine, in a reaction that leads to the loss of the cytidine 4-amino group. This deamination event creates an uracil base.

Antiviral activity of the seven human APOBEC3s

Here, what is known about the biological function of each of the seven human APOBEC3 paralogs will be described.

APOBEC3G

APOBEC3G was the first APOBEC3 protein to be characterized and is one of the most well-studied restriction factors. Several cell lines had been shown to be resistant to infection with HIV lacking the accessory protein Vif(55). APOBEC3G was identified as the host protein that renders certain cell lines resistant to *vif*-deleted viruses(56). However, Vif expression relieves the APOBEC3G-mediated block to infection. Cytidine deaminase activity is likely required for APOBEC3G restriction of viral replication(57, 58), although this is somewhat controversial(59, 60). Subsequently, other APOBEC3 paralogs were described to have anti-lentiviral activity. Expression of APOBEC3 genes in virus-producing cells can lead to the production of non-infectious viral particles. APOBEC3s can be packaged in into viral particles through interactions with RNA and nucleocapsid (61, 62), and taken to a target cell where they mutate the viral DNA during reverse transcription. Oligomerization of APOBEC3G has been shown to be important for antiviral deaminase activity. Since APOBEC3s only mutate single stranded DNA, the C to U mutations occur on the (-) strand of the viral DNA, leading to G to A mutations on the (+) strand. This can lead to premature stop codons or other deleterious mutations in the viral genome. This mode of viral restriction is dependent on cytidine deaminase activity; however, deaminase-independent restriction of lentiviruses has reported by certain APOBEC3 proteins(59, 63). For example, several studies report that APOBEC3 can inhibit reverse transcription, independent of

cytidine deaminase activity(59, 63, 64). For the APOBEC3 proteins that restrict lentiviruses, APOBEC3G has the most potent restriction activity(65). HIV-1 (-vif) produced in the presence of APOBEC3G becomes extensively hypermutated in the target cell and productive infection is drastically inhibited. Strikingly, only approximately 10 APOBEC3G proteins need to be packaged into a single virion to restrict its replication in a target cell(58). One reason APOBEC3G may be more restrictive than other anti-lentiviral APOBEC3s is the specific deamination motif that it targets. APOBEC3 deamination of cytidines is dependent on the nucleotide context surrounding the cytidine. APOBEC3G preferentially mutates cytidines in a 5'-CT-3' context, whereas the other antilentiviral APOBEC3s (APOBEC3D, APOBEC3F, and APOBEC3H) mutate target cytidines in a 5'-CC-3' context. Therefore, APOBEC3G mutations cause GG to GA changes in the viral genome, leading to the addition of stop codons. In contrast, the other anti-lentiviral APOBEC3s cause GA to AA mutations, which are not as likely to induce stop codons but may disrupt viral coding sequences by inducing missense mutations.

APOBEC3F

APOBEC3F is a double-domain APOBEC3 with activity against HIV in the absence of Vif. There is a general consensus in the field that APOBEC3F is a potent inhibitor of lentiviruses, although it does not pose as strong of a block to viral replication as APOBEC3G. However, APOBEC3F is the APOBEC3 that has been most implicated in deaminase-independent restriction of HIV(57, 63). Furthermore, APOBEC3F has been reported to potently block LINE-1 retrotransposition through deaminase-independent activity(66).

APOBEC3H

APOBEC3H is a single domain APOBEC3, and is the most evolutionarily divergent of the seven human APOBEC3s. Although most humans express a variant of this protein that does not have anti-lentiviral activity, there are several haplotypes that do inhibit HIV(67). APOBEC3H will be described in more detail in section 1.8.

APOBEC3D

APOBEC3D a double-domain APOBEC3 with a demonstrated capacity to restrict lentiviruses(68), however APOBEC3D does not hypermutate HIV-1 (-vif) as extensively as APOBEC3G, APOBEC3F, and APOBEC3H. Although humans express an APOBEC3D with relatively weak anti-HIV activity, chimpanzees encode a version with more potent antilentiviral activity(69). Despite human APOBEC3D's relatively weak restriction activity against HIV, APOBEC3D is under positive selection in primates, suggesting that APOBEC3D is an antiviral restriction factor.

APOBEC3A

Although several APOBEC3s potently restricts lentivirus replication, not all human APOBEC3s have activity against HIV-1. For instance, APOBEC3A does not package into HIV particles and inhibit replication in a target cell(68). However, APOBEC3A does play a role in cytidine deamination of HIV in monocytes(70). APOBEC3A also potently inhibits certain LTR retrotransposons, as well as adeno-associated virus(71).

APOBEC3B

APOBEC3B likely does not play a role in restricting lentiviruses(68). However, APOBEC3B has been shown to block LTR retrotransposons and certain DNA viruses such hepatitis B virus(72, 73). Additionally, APOBEC3B has been implicated in certain cancers, as APOBEC3B is over-expressed in many cancer cell lines as well as tumor biopsies and there is evidence of extensive APOBEC3-induced mutagenesis in many of these same cancer samples in which APOBEC3B is over-expressed(74, 75). Interestingly, APOBEC3B is the only APOBEC3 that is predominantly localized in the nucleus(76).

APOBEC3C

APOBEC3C is another APOBEC3 that does not have potent activity against lentiviruses(68). In one report APOBEC3C is shown to potently block SIVagm and SIVmac(77). However, in a study that compares the activity of the seven human APOBEC3s, APOBEC3C is shown to be a weak inhibitor of HIV-1 and SIVmac, as compared to other APOBEC3s(68). Whereas most APOBEC3 proteins have the potential to oligomerize, APOBEC3C is monomeric in solution and in cells. APOBEC3C does not potently restrict endogenous retroelements(78). There are reports that APOBEC3C inhibits human papilloma virus(79, 80), however in one of these studies it is shown that APOBEC3C-mediated restriction of human papillomavirus is weaker than the restriction induced by APOBEC3A(80).

Summary of antilentiviral activity of APOBEC3s

To summarize the differences in anti-lentiviral activity across human APOBEC3s: APOBEC3G, APOBEC3F, and APOBEC3H have the capacity to potently block HIV-1

replication. APOBEC3D also blocks HIV-1, but to a lesser extent. In contrast, APOBEC3A, APOBEC3B, and APOBEC3C do not have a strong inhibitory effect on HIV in vitro. However, APOBEC3A expressed monocytes can block incoming virus. Also, APOBEC3B transfection in HEK293 cells can inhibit HIV, although expression in this cell line does not embody a physiologically relevant system.

		Restriction	
		T cells	HEK293
APOBEC3	A3A	-	-
	A3B	-	++
	A3C	-	-
	A3D	+	+
	A3F	++	++
	A3G	++	++
	A3H	++	++

Figure 2. Comparison of APOBEC3-mediated restriction of HIV-1 (-vif) in a T cell line (SUPT11) or in HEK293 cells. Adapted from reference Hultquist, 2011(68).

In addition to differences in antiviral activity function across APOBEC3s, the seven human paralogs also differ in both their oligomerization potential, as well as their intracellular localization. First, certain APOBEC3s form dimers and higher-ordered oligomers in cells and in vitro (APOBEC3B, APOBEC3D, APOBEC3F, APOBEC3G, and APOBEC3H), whereas

APOBEC3A and APOBEC3C are both found to be monomeric. In fact, this difference in oligomeric potential correlates with the ability to restrict lentiviruses. One possible reason for this is because oligomerization affects DNA binding, so oligomerization may be important in scanning for cytidine deamination events. Additionally, there are differences across the APOBEC3s in their intracellular localization. APOBEC3B is predominantly found in the nucleus, APOBEC3A and APOBEC3C are found cell-wide, and APOBEC3D, APOBEC3F, APOBEC3G, and APOBEC3H are found in the cytoplasm. These localization differences also roughly correlate with antilentiviral activity, as the antilentiviral paralogs are found predominantly in the cytoplasm. Interestingly, nuclear localization correlates with cancer-associations, as APOBEC3B expression is highly associated with certain cancers and APOBEC3A has been implicated in genomic mutations in cell culture studies.

1.7 Vif antagonism of APOBEC3s

Vif hijacks the host ubiquitin ligase machinery and induces K48-linked ubiquitination of APOBEC3s to target APOBEC3s for proteosomal degradation. Vif binds directly to APOBEC3 proteins, and also binds to the Cul5/EloB/EloC ubiquitin ligase complex. APOBEC3C, APOBEC3D, APOBEC3F, APOBECG, and APOBEC3H are all susceptible to Vif-mediated degradation, whereas APOBEC3A and APOBEC3B are not targeted by Vif. Vif targets a conserved domain present on APOBEC3C, APOBEC3D, and APOBEC3F. This surface includes the glutamic acid on alpha helix 3 that is required for Vif binding, as well as a histidine five nucleotides away, and several residues between alpha helix-2 and beta fold 3. Vif does not target these same residues on APOBEC3G but instead interacts with a different domain. The residues required for Vif binding to APOBEC3G were identified because HIV-1 Vif is able to mediate degradation of human APOBEC3G, but cannot induce degradation of the APOBEC3G expressed

by African green monkeys. In contrast, SIVagm Vif can mediate degradation of the APOBEC3G expressed by African green monkeys, but not human APOBEC3G. This led to the identification of APOBEC3G residues 128-130 as the major determinants of the specificity of Vif binding, as swapping this domain on the two APOBEC3 orthologs switches the Vif binding phenotype. Further studies show that adaptive evolution has occurred at APOBEC3G positions 128 and 130 across diverse old world monkey species, and the sequence variation at these sites has led to variation in the sequence motif targeted by diverse SIV Vifs(81). In fact, for one group of old world monkeys, Colobinae, adaptation in a region distant from the 128/130 Vif-binding motif confers resistance to most SIV Vifs (81). The significant variation across primates in the susceptibility of APOBEC3Gs to lentiviral Vifs indicates that primate adaptation to lentiviruses has been occurring for millions of years. Lastly, a third Vif-binding domain exists on APOBEC3H. APOBEC3H shares the same contact sites as APOBEC3D/APOBEC3F for half of the interface (on alpha helix-3), but uses a distinct set of residues for the other half (Figure 3).

Interestingly, there is evidence that despite the expression of Vif during infection, APOBEC3s still have a significant effect on HIV sequences within infected patients. Across different studies, 9-43% of virus sequences derived from patients contain APOBEC3-induced mutations(82-85). APOBEC3 activity against HIV in patients could have protective benefits, as APOBEC3-induced hypermutation correlates with slower disease progression. In contrast, another potential outcome of the viral hypermutation induced by APOBEC3s is that it may contribute to viral diversity that allows for adaptive evolution of the virus. For instance, APOBEC3-induced mutations may allow HIV to more quickly evolve resistance to antiretrovirals. In fact, one study correlated APOBEC3G-induced hypermutation with anti-

retroviral therapy failure in patients infected with a particular subtype of HIV-1(84). However, a recent study concluded that APOBEC3-induced mutations contribute to viral sequence diversity at a much lower level than the random errors introduced during reverse transcription(86).

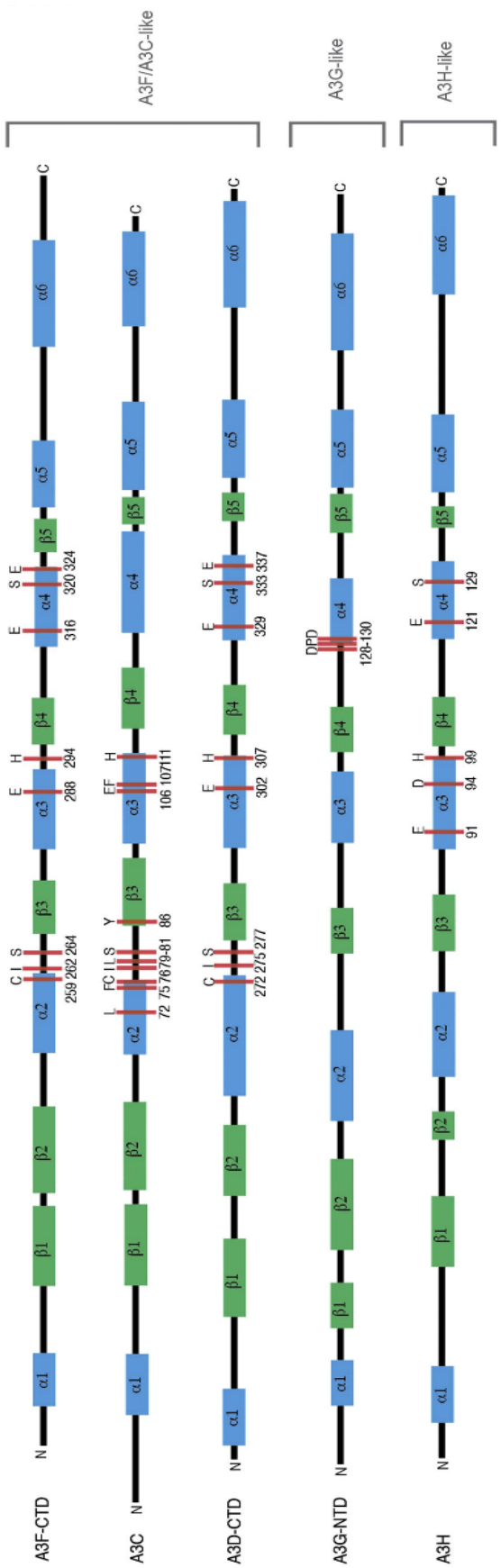

A3F/A3C-like
A3G-like
A3H-like
A3F-CTD
A3C
A3D-CTD
A3G-NTD
A3H
N
C
α1
β1
β2
α2
β3
α3
β4
α4
β5
α5
α6
E
H
S
I
C
Y
F
L
D
DPD
316
320 324
288
294
259 262 264
72 75 76 79-81
86
106 107 111
302
307
272 275 277
329
333 337
126-130
91
94
99
121
129

Figure 3. The Vif binding domains of human APOBEC3 proteins. Three different Vif binding domains exist for the five human APOBEC3 proteins that are antagonized by Vif. First there is the A3C/A3F-like domain present on A3C, A3F, and A3D. This domain lies on the loop between helix-2 and beta fold-3 and on alpha helix-3. Secondly, there is a distinct binding domain on APOBEC3G at the very beginning of alpha helix-4. Thirdly, there is a distinct binding domain present on APOBEC3H, which is similar to the residues targeted on A3D and A3F, except the loop between alpha helix-2 and beta fold-3 does not appear to be involved. From Aydin, 2014(53).

1.8 HUMAN POLYMORPHISM IN APOBEC3 GENES

There are two APOBEC3 human polymorphisms that have been well-characterized in their impacts on disease. First, some humans carry a large deletion in the APOBEC3 locus that includes the last exon of APOBEC3A and the first seven exons of APOBEC3B(87). This deletion leads to the expression of an APOBEC3A/APOBEC3B fusion protein. This allele is present at a frequency of 93% in Oceania, 37% in East Asia, and 58% in American Natives. The deletion is associated with increased risk of breast cancer, and ovarian cancer(88, 89). APOBEC3B does not block lentivirus replication, but instead potently inhibits certain LTR-retrotransposons, such as LINE-1(78, 90). Furthermore, the deletion allele is associated with increased risk of hepatitis B infection, as well as infection with several other pathogens(91).

Only one of the anti-lentiviral APOBEC3s has a significant level of polymorphism among humans: APOBEC3H. Strikingly, there are seven distinct haplotypes expressed globally (global frequencies ranging from 6-86%). The haplotype with the highest global frequency, encodes a protein that is unstable, meaning it has a short-half life and quickly becomes degraded in the cellular environment(67). Three other haplotypes also encode unstable proteins. In contrast, the remaining three haplotypes (haplotypes II, V, VII), encode stable proteins. These stable APOBEC3Hs potently restrict HIV in the absence of Vif(67, 92). Intriguingly, stable APOBEC3H alleles have been associated with resistance to HIV infections in two separate studies, one in a Japanese cohort and one in an Indian cohort(93, 94). These findings were somewhat surprising given that HIV Vif is thought to successfully antagonize APOBEC3H. However, further studies have shown that Vifs sequenced from different patients show varied activity against APOBEC3H. In fact, Vifs sequenced from patients encoding stable APOBEC3H

are more likely to be competent to antagonize APOBEC3H(95, 96). In contrast, for individuals carrying unstable APOBEC3H alleles, Vifs sequenced from these patients are more likely to be unable to antagonize APOBEC3H, and the Vif loss-of-function is specific to APOBEC3H (antagonism of other APOBEC3s remains unaffected). The differences in APOBEC3H-specific activity in Vifs from patients with different haplotypes may help explain the increased HIV-resistance for individuals encoding stable APOBEC3H haplotypes. For instance, if someone carrying an unstable APOBEC3H is infected with HIV, this may correlate with the presence of a Vif that is defective against APOBEC3H. If this virus is transmitted to an individual carrying a stable APOBEC3H allele, then that individual may be protected because the Vif cannot overcome their APOBEC3H. These findings indicate that even though HIV encodes Vif as a defense against APOBEC3s, APOBEC3 polymorphism can impact HIV pathogenesis.

Another APOBEC3 with a fairly common human polymorphism is APOBEC3C. Approximately 10% of the population in African encodes an APOBEC3C variant with a serine to isoleucine change at position 188. Interestingly, the isoleucine that is present in the variant is a conserved residue at the homologous position in the other six human APOBEC3s. Although APOBEC3C has not been implicated as an antilentiviral restriction factor, we hypothesized that the variant encoding isoleucine at position 188 may confer antilentiviral activity.

Chapter 2. MATERIALS & METHODS

APOBEC3C sequences

APOBEC3C was amplified by RT-PCR from total RNA extracted from chimpanzee, gorilla, orangutan, white-cheeked gibbon, siamang, baboon, sooty mangabey, and red-capped mangabey cells (either fibroblast or lymphoid) obtained from Corriell Repository (GM03448, AG05251, AG05252, PR01038, PR00598, PR00976, G077, PR00485) as well as from the vervet monkey cell line Vero, the tantalus monkey cell line CV-1, and the sabeus cell line V038 provided by the Nonhuman Primate Research Resource (NPRR). Primers were designed to amplify from the 3' and 5' UTRs of APOBEC3C mRNA transcripts (5'UTR: CTAAGAGGCTGAACATGAATC'3, 3'UTR: 5'GGCTAGAGGAGACAGACCATGA'3). The APOBEC3C amplicons were cloned into pGEM vectors, and then sequenced. The S188-S188 forced dimer was designed to mimic the linker between the two domains of the double-domain APOBEC3F. The N-terminal subunit consists of APOBEC3C residues 1-189 (the last residue is removed), followed by the residues RNP, which serve as a linker. The C-terminal APOBEC3C begins at the second start codon, M12. The dimer S188-S188 APOBEC3C was constructed by overlap extension PCR. Two separate PCRs were performed for the N terminal and C terminal APOBEC3C subunits (1^{st} domain, For: TTCAGGATCCATGAATCCAGAGATC, 1^{st} domain, Rev: GCCTCCATTGGGTCCCGGAGACTCTCCCGTAGCCTTCTTT, 2nd domain, For: TCCAGGATCCATGAATCCACAGATC, 2^{nd} Rev: GCCCTCTAGATTAGGCGTAGTCAGG), and these amplicons were annealed in a third PCR reaction using the 1^{st} domain For and the 2^{nd} domain Rev primers.

Sequence Analysis

APOBEC3C genes were aligned using Geneious software. To test for positive selection, maximum likelihood tests were performed using the PAML statistical software suite(97). The *APOBEC3C* genes were subjected to tests that allowed for positive selection (M8 model), or disallowed positive selection (M8a model). The analyses were performed with the F3X4 codon model, and multiple starting omega values were used, ranging between 0.5 and 1.4. Specific residues with signatures of positive selection with a posterior probability of 99% or greater were identified by Bayes Empirical Bayes analysis. Ancestral APOBEC3C sequences were reconstructed by the likelihood/Empirical Bayes approach using the codeml program in PAML. Brach analysis to identify particular primate branches with signatures of positive selection in APOBEC3C were performed in two ways. Overall dN/dS values were calculated with PAML, using the free ratio model. Additionally, a branch-site test to identify statistically significant signatures of episodic selection was performed using the Branch-site REL method in the HyPhy software suite (98) .

APOBEC3, provirus, and LINE-1 plasmids

APOBEC3Cs were cloned into the BamHI and XhoI sites of pCDNA3.1 by PCR addition of restriction sites (BamHI and XhoI) to the N and C termini of APOBEC3C. The human APOBEC3C plasmid we previously obtained from the AIDS Repository contained the SNP rs11551111, which is not common (no reported frequency according to dbSNP). Therefore, we used site-directed mutagenesis to change the asparagine at position 31 to aspartic acid (For: GCCAACGATCGGGACGAAACTTGGC, Rev: GCCAAGTTTCGTCCCGATCGTTGGC). A hemagglutinin tag was inserted into the XhoI and XbaI sites of pCDNA3.1, at the C-terminus of

each APOBEC3C sequence. APOBEC3G and APOBEC3A were also in a pCDNA3.1 backbone, with a kozak sequence, as well as a hemagglutinin tag at the N-terminus. HIVΔ*env*,Δ*vif,* HIVΔ*vif* + HIV-1 *vif*, HIVΔ*vif* +HIV-2 *vif* have been described elsewhere(99). SIVagm Δ*env*, Δ*vif* was kindly provided by Nathaniel Landau.

Infectivity Assays

Single round HIV-1 and SIVagm infectivity assays were performed as previously described (100). 293Ts were plated at a density of 0.05 X 10^5 cells per well of a 24-well plate. The next day, the cells were transfected with 0.3µg provirus encoding luciferase as a marker gene 0.1µg pL-VSV-G, and 0.3µg pCDNA3.1.APOBEC3.HA or empty pCDNA3.1 plasmid. For the dose response infectivity assay, either 0.1 µg, 0.2µg, or 0.3µg APOBEC3 plasmid was used. For experiments involving Vif expression, 0.2µg of APOBEC3 was use. Forty-eight hours after transfection, virions were harvested. For SIVagm infectivity assays, SupT1 cells were infected with 10µl of each virus and treated with 20µg/ml DEAE/dextran. For HIV infectivity assays, ELISA was performed to quantify p24, and virus equivalent to 2ng p24 was used for infections. For all infectivity assays, 5 X 10^4 were infected in a 96-well dish. Seventy-two hours later, infected cells were lysed in luciferase lysis reagent (Brightglo, Promega) and luciferase expression was measured on a luminometer (LUMISTAR Omega, BMG). Infectivity of each virus was compared by setting infectivity of the "No APOBEC3" control to 100%.

LINE-1 Assays

To assay for restriction of LINE-1 retrotransposition 293T cells were transfected with 200ng LINE-1 plasmids pYX016 and pYX015(101), along with 100ng of APOBEC3C S188 or

I188, APOBEC3C 10ng APOBEC3A, or empty pCDNA3.1 plasmid. The next day, the cells were treated with 2.5 uμ/ul puromycin to select for transformants. Three days later, expression of renilla and firefly luciferase were assayed using a luminometer. The LINE-1 plasmids encode firefly luciferase disrupted by a splice site, so expression only occurs after retrotransposition, whereas renilla luciferase expression is not dependent upon retrotransposition. Percent retrotransposition is reported by setting retrotransposition (firefly luciferase values divided by renilla luciferase values) in the absence of APOBEC3 to 100%.

Intracellular Protein Expression

Intracellular expression of the APOBEC3 proteins during virion production was evaluated by lysis of the virion-producing 293T cells with Radio Immunoprecipitation Assay buffer, with protease inhibitor (50mM Tris, 150mM sodium chloride, 0.1% SDS, 0.5% sodium deoxycholate, 1% NP-40, protease inhibitor cocktail cOmplete by Roche). Lysates were resolved on an SDS-PAGE gel in MES buffer, and transferred to a PVDF membrane for Western blot analysis, using and anti-HA (BioLegend) antibody and anti-tubulin (Sigma-Aldrich) antibody. Endogenous levels of APOBEC3C were measured by Western blotting with antibody purchased from Fisher (product # PA5- 27629). HRP-conjugated secondary antibodies (Santa Cruz) were used to detect primary antibodies.

Recombinant protein expression and purification

Recombinant baculovirus production for APOBEC3C S188 was carried out in the pACG2T transfer vector (BD Biosciences), as described previously (102). Recombinant baculovirus production for APOBEC3C I188 was carried out in the pFastbac1-GST-APOBEC3C

vector according to the Bac-to-Bac expression system (Life Technologies) and as described previously (103). Recombinant virus was then used to infect *Sf9* cells. Cells were harvested 72 hours after infection and, clarified cell lysates were incubated with glutathione-sepharose 4B resin (GE Healthcare) at 4°C and subjected to a series of salt washes, as described previously(104). The APOBEC3C S188, APOBEC3C I188 enzymes were eluted from the glutathione-sepharose resin (GE Healthcare) with the GST tag, as previously described (104). The samples were then treated with thrombin (GE Healthcare) for 6 hr at 21°C to cleave the GST tag.

Size exclusion chromatography

The oligomerization states of the APOBEC3C enzymes were determined by loading 10 μg of purified enzyme on a 10 mL Superdex 200 (GE Healthcare) size exclusion column. The column was prepared by pouring the resin bed in a column with 16-cm height and 0.5-cm diameter. The running buffer contained 50 mM Tris pH 8.0, 200 mM NaCl and 1 mM DTT. The Bio-Rad standard set was used to generate a standard curve from which molecular masses and oligomerization states of the enzymes were determined.

***In vitro* deamination assay**

All ssDNA substrates were obtained from Tri-Link Biotechnologies as previously published (105). Reactions were carried out under single-hit conditions (*i.e.* <15% substrate usage) to ensure that a single enzyme carried out the deaminations on the ssDNA(106). A ssDNA substrate containing two 5′-TTC motifs (100 nm) was incubated with 350 nM of APOBEC3C I188 or 700 nM of APOBEC3C S188 for 5 to 30 min at 37 °C in RT buffer (50 mM

Tris, pH 7.5, 40 mM KCl, 10 mM $MgCl_2$, and 1 mM DTT). The reaction time was varied on each ssDNA according to the specific activity of the enzymes to ensure <15% substrate usage. Reactions were started by the addition of the ssDNA substrate. APOBEC3C-catalyzed deaminations were detected by treating the ssDNA with uracil DNA glycosylase (New England Biolabs) and heating under alkaline conditions before resolving the fluorescein-labeled ssDNA on 10 or 20% (v/v) denaturing polyacrylamide gels, depending on the sizes of the ssDNA fragments. Gel photos were obtained using a Typhoon Trio multipurpose scanner (GE Healthcare), and integrated gel band intensities were analyzed using ImageQuant (GE Healthcare). The specific activity was calculated from single-hit condition reactions by determining the picomoles of substrate used per minute for a microgram of enzyme.

In vitro reverse transcription assay

Mutagenesis of ssDNA by A3 enzymes during reverse transcription of an RNA template was assessed using an *in vitro* assay, which models reverse transcription of an RNA template and second-strand synthesis. The method is described in detail in Feng and Chelico 2011 (107). This system uses an *in vitro* synthesized RNA, which contains a polypurine tract (PPT), a protease gene (prot) of HIV, and a lacZα reporter for blue/white screening. The RNA is reverse transcribed to (−)DNA by reverse transcriptase (RT) by annealing a DNA primer and after the RNaseH domain of RT removes the RNA, the PPT enables second-strand (+)DNA synthesis by acting as a primer. A 368-nt RNA template (50 nM) is annealed to a DNA primer (24-nt) and incubated with 1.5 μM of nucleocapsid (NC), 1.2 μM of reverse transcriptase (RT) and 500 μM of dNTPs in RT buffer in the presence or absence of 350 nM of each APOBEC3C enzyme. The RNA template contained an HIV-1 PPT, nucleotides (nt 2282–2401) from the HIV-1 clone

93th253.3 (accession number U51189), and *lacZα*. The resulting dsDNA that is synthesized from this *in vitro* system was PCR amplified using Pfu C_x Turbo Hotstart (Agilent Technologies) that can use uracils as a template with high fidelity. These amplicons were then cloned into a pET-Blue vector backbone that allows for blue-white screening of the synthesized *lacZα.* At least twenty-five mutated clones for each condition were tested.

SNP analysis

1000 Genomes Project data was mined for the presence of SNPs at position 188 of APOBEC3C (SNP ID rs112120857). To further elucidate the frequency of the APOBEC3C I188 SNP across Africa, we analyzed the genomes reported by Schuster et al.(108) and Lachance et al.(109) for the presence of the I188 allele. To assay for the presence of SNP at position 188 in other hominoids, we mined the Great Ape Genome Project (110) (accession number SRP018689) sequences in the NCBI short read archive.

Chapter 3. A SINGLE NUCLEOTIDE POLYMORPHISM IN HUMAN APOBEC3C ENHANCES RESTRICTION OF LENTIVIRUSES

3.1 INTRODUCTION

The *APOBEC3* locus encodes seven cytidine deaminase proteins that inhibit endogenous retroelements, lentiviruses such as HIV-1, and other viruses (111). The *APOBEC3* locus arose through duplication events on chromosome 22(112) of cytidine deaminase domains, resulting in single domain *APOBEC3*s (*APOBEC3A*, *APOBEC3C*, and *APOBEC3H*) and double-domain *APOBEC3* genes (*APOBEC3B*, *APOBEC3D*, *APOBEC3F*, and *APOBEC3G*). In order for APOBEC3 proteins to restrict lentiviruses such as HIV-1, they are packaged into virions, brought to a target cell, and deaminate cytidines on ssDNA during reverse transcription, resulting in cytidine to uracil mutations in the viral genome. APOBEC3 proteins exert selective pressure on primate lentiviruses, which have evolved to encode a protein, Vif, which targets APOBEC3 proteins for proteasomal degradation.

Vif-mediated antagonism leads to non-synonymous mutations in *APOBEC3* that allow for escape from Vif but maintenance of antiviral activity(34). Lentiviruses, in turn, select for Vif alleles that target these APOBEC3 variants, leading to further adaptive evolution of *APOBEC3* genes through non-synonymous mutations. As such, enrichment of the rate of nonsynonymous mutations (dN) compared to the rate of synonymous mutations (dS) is a common signature of antiviral genes (34). This enrichment, or dN/dS>1, is referred to as positive selection. *APOBEC3* genes involved in blocking viral replication are expected to exhibit signatures of positive

selection. Specifically, APOBEC3s involved in lentiviral restriction should have signatures of positive selection at the Vif:APOBEC3 interface(81).

There is considerable variation in the antiviral activity of each of the seven human APOBEC3 paralogs. APOBEC3G potently inhibits *vif*-deleted-HIV-1 (Δ*vif*) (65) . Human APOBEC3D, APOBEC3F, and APOBEC3H also inhibit HIV-1 (Δ*vif*), but to a lesser extent than APOBEC3G (65, 68, 113, 114). In contrast, APOBEC3A and APOBEC3B do not potently block HIV infection of T cells (65, 68, 77, 113), which are the primary target of HIV (although a target-cell effect has been reported in monocytes for APOBEC3A) (115) . Instead, APOBEC3A and APOBEC3B drastically inhibit replication of endogenous retroelements and some DNA viruses (71, 72, 79, 80, 90, 116). In studies that compare the ability of the seven human APOBEC3s to restrict lentiviruses and endogenous retroelements, the only APOBEC3 that has weak activity against both lentiviruses and endogenous retroelements is APOBEC3C (65, 68, 76, 80, 116-121). For another *APOBEC3* gene, *APOBEC3H*, the most common human variant does not block HIV infection although other haplotypes exist that potently restrict lentivirus replication(67). In fact, one haplotype of APOBEC3H restricts HIV-1(Δ*vif*) as potently as APOBEC3G(67) and has been shown to impact clinical outcomes in HIV-1+ patients(93-95). Thus, we considered the possibility that while the common human haplotype of *APOBEC3C* encodes a protein with little antiviral activity, other variants of *APOBEC3C* may in fact encode more potent anti-lentiviral proteins. Compellingly, the Vif protein of HIV-1 targets human APOBEC3C for proteosomal degradation (122). Moreover, APOBEC3C mRNA is highly expressed in the major HIV-1 target cells, activated T cells(123). Thus, the high expression of APOBEC3C in HIV target cells and the antagonism of APOBEC3C by HIV-1 Vif are consistent

with the hypothesis that APOBEC3C may have an overlooked role in combating lentivirus infection.

In this study, we found that *APOBEC3C* has evolved under positive selection in primates in a manner that suggests that APOBEC3C has played a role in blocking primate lentiviruses. This provided motivation to determine if there are naturally occurring variants of APOBEC3C that potently block lentivirus replication. In humans, only one APOBEC3C coding variant is present at a frequency above 1% and this is a serine to isoleucine change at position 188, here called APOBEC3C I188 (124). Here, we show that the polymorphism at APOBEC3C I188 is present at about 10% frequency in diverse populations throughout Africa, and thus did not recently arise in a particular subpopulation of humans, but is an ancient allele that has likely been circulating in humans for much of human history. Moreover, we show that the APOBEC3C I188 single nucleotide polymorphism (SNP) has about 10-fold more potent anti-lentiviral activity than the common human APOBEC3C variant and has greater *in vitro* cytidine deaminase specific activity. We show that the APOBEC3C I188 allele is likely the ancestral state since all sequenced Old World monkeys and some great apes carry isoleucine at position 188. However, gorillas, chimpanzees and most humans carry the S188, the apparent loss of function allele. Taken together, our results suggest that APOBEC3C is involved in protecting hosts from lentiviruses, and we speculate that some humans may be afforded some level of additional protection from lentiviruses by a more active antiviral version of this protein.

3.2 RESULTS

APOBEC3C has evolved under positive selection in primates, suggesting an ancient role in protection from pathogens

In studies that compare the antiviral activity of the seven APOBEC3 paralogs, APOBEC3C consistently has poorer restriction activity than the other paralogs (65, 68, 118, 120, 121). However, we reasoned that if APOBEC3C is in fact a bona-fide restriction factor then we would expect that the gene has an evolutionary signature of positive selection(34) . We performed positive selection analyses of twenty-two *APOBEC3C* sequences derived from eighteen primate species with sequences representing diverse clades of catarrhines, a subdivision of primates including old world monkeys and apes (Figure 4A). Among these, multiple sequences were obtained from African green monkeys, because we chose to include three subspecies (vervet, tantalus, and sabeus). The sequences were aligned and tests for positive selection were conducted using maximum likelihood ratio tests comparing M8 (a model that allows positive selection across the gene) to M8a (a model that disallows positive selection). Our results indicate *APOBEC3C* shows a gene-wide signature of positive selection ($p<0.0008$) (Figure 4B).

We next analyzed individual lineages to determine which branches of the *APOBEC3C* tree have signatures of positive selection. Branch analysis identified two branches with statistically significant signatures of positive selection, both in Old World monkeys (Figure 1A), and while most were not statistically significant, 21 out of 39 branches had a dN/dS >1. (Figure 1A). Furthermore, we performed M8 vs M8a analysis of the hominoid and Old World monkey clades of the tree separately, and found that the Old World monkey clade has a statistically

significant signature of positive selection ($p<0.05$) (Figure 4B). We did not see a statistically significant signature of positive selection in the hominoid-only branch ($p=0.15$), although this could be due to a smaller sample size ($n= 7$).

For antiviral genes, sites under positive selection often correlate with sites of interaction with a viral antagonist (36). APOBEC3C is antagonized by the lentiviral protein Vif and the interface of Vif binding has been extensively mapped (125). If APOBEC3C is in fact an anti-lentiviral gene, the Vif binding interface may be evolving under positive selection. Therefore, we performed a site-analysis to determine which amino acids are under positive selection across the tree. Our analysis indicated seven sites under positive selection (posterior probability > 99%) (Figure 4C). Next, we mapped the positively selected sites onto the structure of human APOBEC3C and compared these to the Vif interface of APOBEC3C. Of the seven positively selected sites, two of these, residues 106 and 77, are located within the two helices that are targeted by Vif (Figure 4C). Strikingly, residue 106 has been identified as the most important for Vif binding and this interaction has been documented in two separate studies (122, 125). Thus, *APOBEC3C* has evolved under selection, gene-wide, as well as at the Vif-binding interface. These results suggest that although the common human APOBEC3C variant does not potently block lentivirus replication, primate *APOBEC3C* may have evolved as an anti-lentiviral protein.

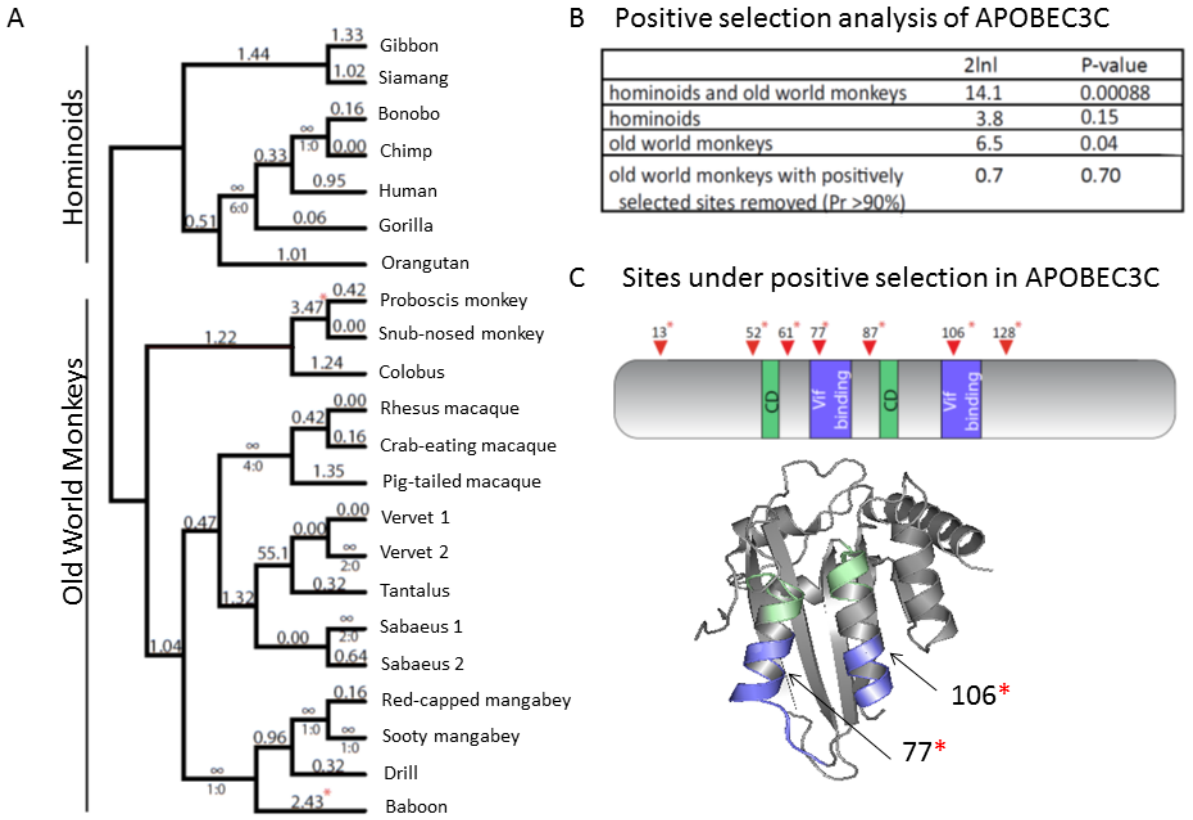


Figure 4: APOBEC3C is rapidly evolving in primates. (A) Twenty-two primate APOBEC3C coding sequences were obtained by PCR or from the NCBI sequence database. A phylogeny of APOBEC3C, indicating the branch analysis results of the positive selection tests. The ratio of rate of nonsynonymous changes (dS) and the rate of synonymous changes (dS) that occurred along each branch are shown above each branch. For dN/dS values = ∞, the total number of non-synonymous changes (N) and synonymous changes (S) are shown (N:S) below the branch. (B) Maximum likelihood tests for positive selection, with 2lnl values indicating twice the log difference between the model that allows for positive selection (M8) and the model that does not allow for positive selection (M8a), as well as a P-value to indicate whether the M8 model better fits the data than the M8a model. (C) Sites under positive selection in APOBEC3C are shown in a cartoon diagram, comparing these sites to the Vif binding domain and the cytidine deaminase enzymatic domain (CD). Red triangles depict sites with a posterior probability >0.99 (red triangle). The structure of APOBEC3C(125) is represented, with the Vif binding domain(125) shown in blue. Two of the seven positively selected sites (PP > 0.99) overlap with this domain, are shown with arrows. The cytidine deaminase domain is shown in green.

Human APOBEC3C SNP I188 increases antiviral activity

Because the positive selection analyses suggested an ancient or ongoing role of APOBEC3C in lentiviral restriction (Figure 4), we re-evaluated human polymorphisms in *APOBEC3C* for potential variants with increased activity. There is only one SNP in *APOBEC3C* above 1% frequency globally, and this is a serine to isoleucine change at position 188 (124). To evaluate the potential significance of this SNP, we aligned this region of *APOBEC3C* to other human *APOBEC3* genes. Strikingly, we found that in contrast to *APOBEC3C*, the other ten *APOBEC3* deaminase domains all encode a conserved isoleucine at the position homologous to APOBEC3C 188 (Figure 5). Thus, the human I188 polymorphism in APOBEC3C actually encodes an amino acid that is highly conserved at this position across human APOBEC3s, while the more common APOBEC3C in the human population has a different amino acid at position 188.

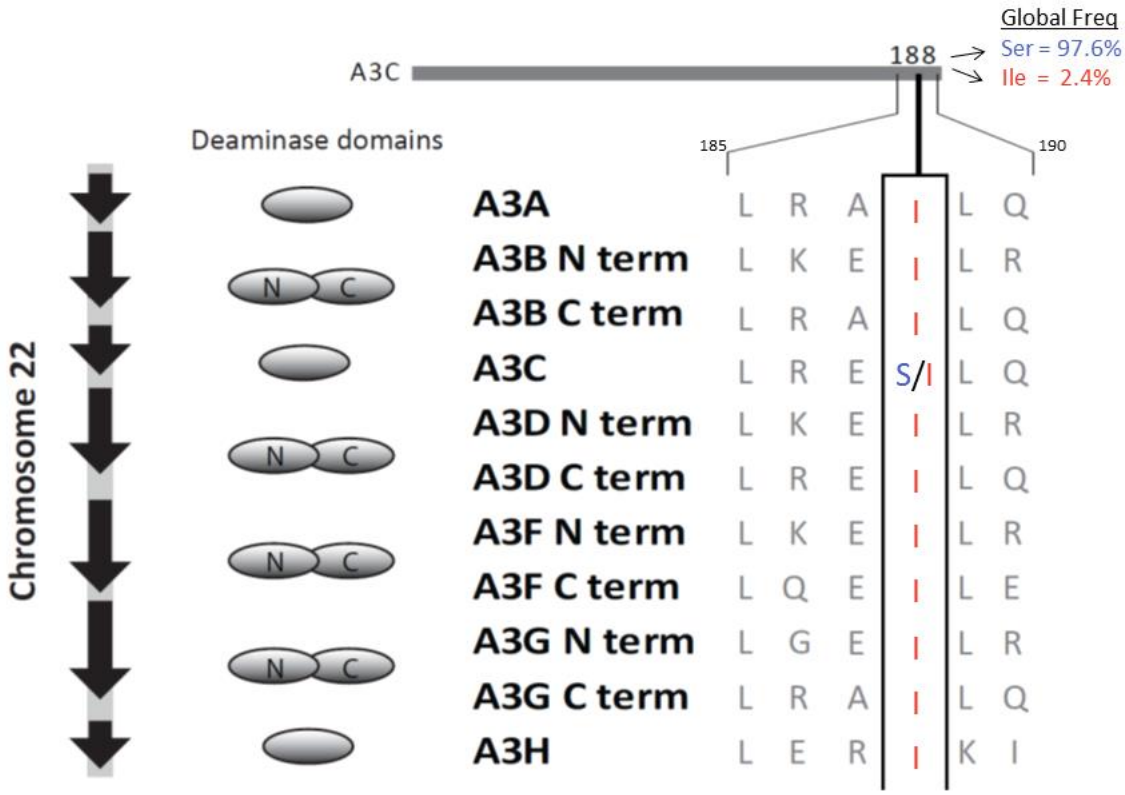


Figure 5: I188 is a SNP in APOBEC3C and is the conserved residue in the other six human APOBEC3 paralogs (A) Alignment of the seven APOBEC3 proteins, using both deaminase domains (N and C terminal) of the double-domain APOBEC3 proteins. The residue homologous to A POBEC3C 188 in the other ten deaminase domains is conserved as an isoleucine, whereas APOBEC3C is the only domain with a serine at that position. However, human APOBEC3C is polymorphic at that position, with an isoleucine at an allele frequency of 2.4% globally(124).

Since conserved sequences are often important for function and comparative studies indicate that human APOBEC3C (S188) has weak antiviral/anti-retroelement activity compared to the other human APOBEC3s, we posited that the serine change may contribute to the weak restriction activity of the common variant of APOBEC3C. Therefore, we directly compared APOBEC3C S188 and APOBEC3C I188 for their ability to restrict HIV-1. We transfected the two APOBEC3C variants, S188 and I188, along with VSV-G and an *env*- *vif*- deleted luciferase-expressing HIV-1 provirus (Δ*env*, Δ*vif*). Normalized amounts of virus were subsequently used to infect SupT1 cells and infectivity of the viruses was compared by measuring virus-encoded

luciferase. Viral infectivity in the presence of no APOBEC3 is set to 100%. APOBEC3G was used as a positive control because it potently inhibits HIV-1 (Δ*vif*). We found that APOBEC3C I188 restricts infectivity of HIV-1(Δ*vif*) to a level approximately ten-fold greater than the common APOBEC3C, S188, (approx. 30% infectivity versus 3%, respectively) (Figure 6A) even though both proteins are expressed at similar levels. Furthermore, infectivity assays were conducted as a dose-response in the presence of decreasing concentrations of APOBEC3, and the 188 isoleucine variant restricts HIV-1(Δ*vif*) more potently for all conditions (Figure 6B) at similar protein expression levels.

To determine if the APOBEC3C I188 variant has increased potency against another lentivirus, we evaluated its activity against SIVagm, which is a simian immunodeficiency virus that infects African green monkeys. As shown by others, the S188 variant of APOBEC3C restricted infectivity of SIVagm to a greater extent than HIV-1 (77). However, the SIVagm restriction caused by APOBEC3C I188 was ten-fold greater than the restriction caused by APOBEC3C S188(10 % versus 1% infectivity, respectively, $p<0.05$) (Fig.6C).

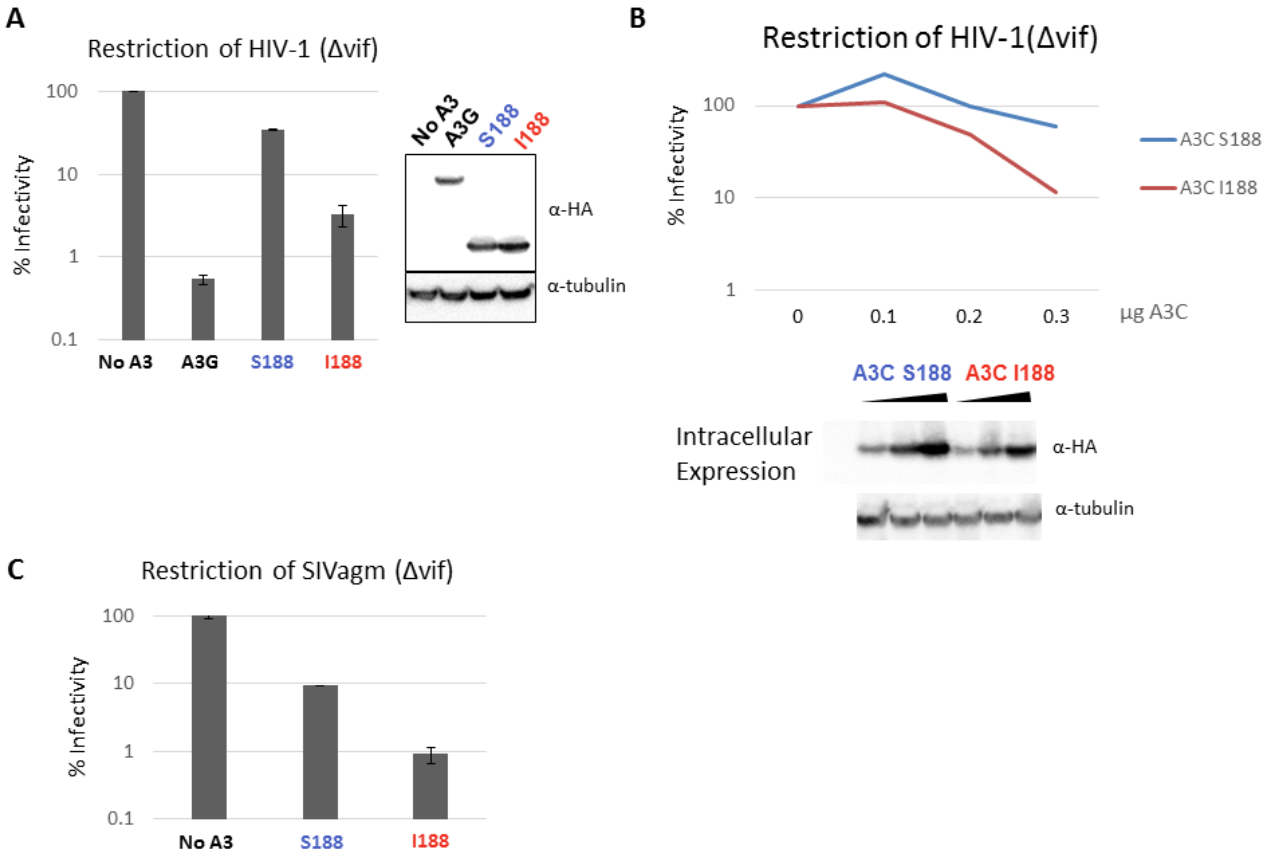


Figure 6: APOBEC3C SNP Isoleucine 188 confers increased antiviral activity. (A) Infectivity of HIV-1 Δvif in the absence of APOBEC3 (No A3), APOBEC3G (A3G), APOBEC3C S188 (A3C S188) and APOBEC3C I188 (A3C I188). 0.3µg of HA-tagged APOBEC3 was expressed in virus-pro ducing cells, and viruses were collected and use for infection. Infectivity in the absence of APOBEC3 is set to 100%. Error bars indicate the standard deviation of triplicate transfections and infections, and this experiment was repeated four times with similar results. Intracellular expression of APOBEC3 was measured by Western Blot using an anti-HA antibody. A section of the blot was probed with an anti-tubulin antibody as a loading control. (B) Dose-response analysis showing restriction of HIVΔ*vif* in the presence of two-fold dilutions of transfected APOBEC3C S188, or APOBEC3C plasmids I188 along with Western blot analysis of APOBEC3C S188, and APOBEC3C I188 protein expression during virus production. This experiment was performed three times, and a representative result is shown. (C) Infectivity of Simian Immunodeficiency virus SIVagmΔ*vif*, in the presence of APOBEC3C S188, APOBEC3C I188. Infectivity is set to 100% for infection with No APOBEC3 present. Error bars indicate the standard deviation for four transfections and infections for each condition, and this experiment was repeated twice.

Some APOBEC3s also restrict endogenous retroelements, such as LINE-1s (71, 90). However, the APOBEC3C I188 variant does not confer increased restriction of LINE-1 as we have previously published (124) and have repeated for this study (Figure 7). Therefore, the human polymorphism in APOBEC3C at position 188 enhances restriction of at least two primate lentiviruses. Thus, we conclude that a SNP in human APOBEC3C has increased anti-lentiviral activity relative to the APOBEC3C encoded by most humans.

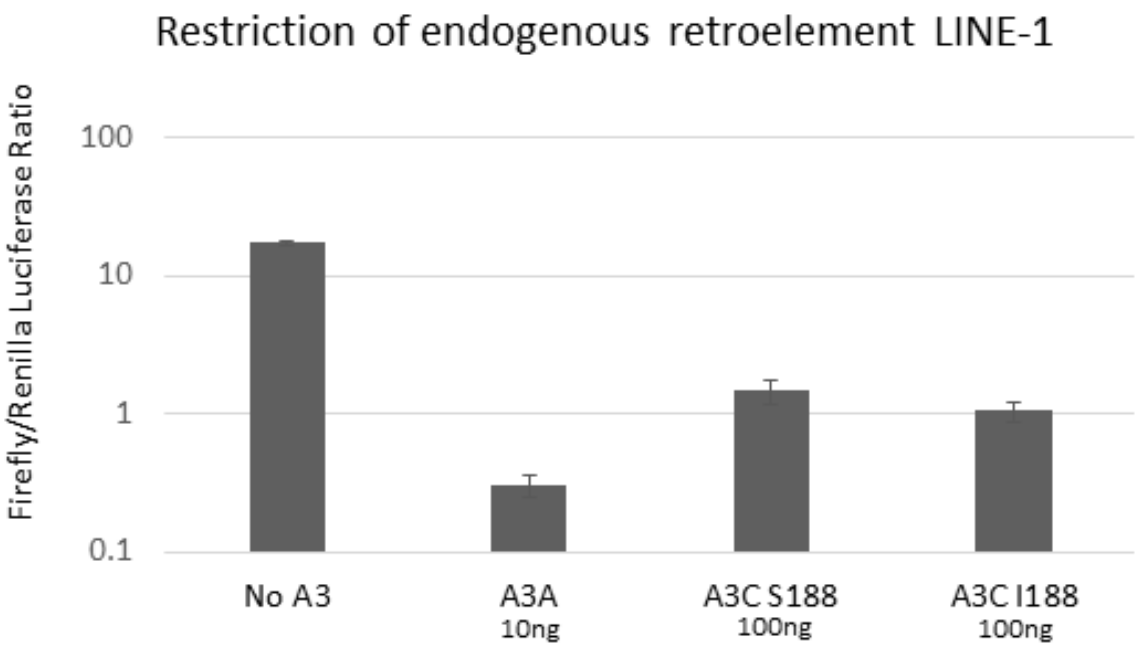


Figure 7. Restriction of LINE-1 by APOBEC3C S188 and I188.

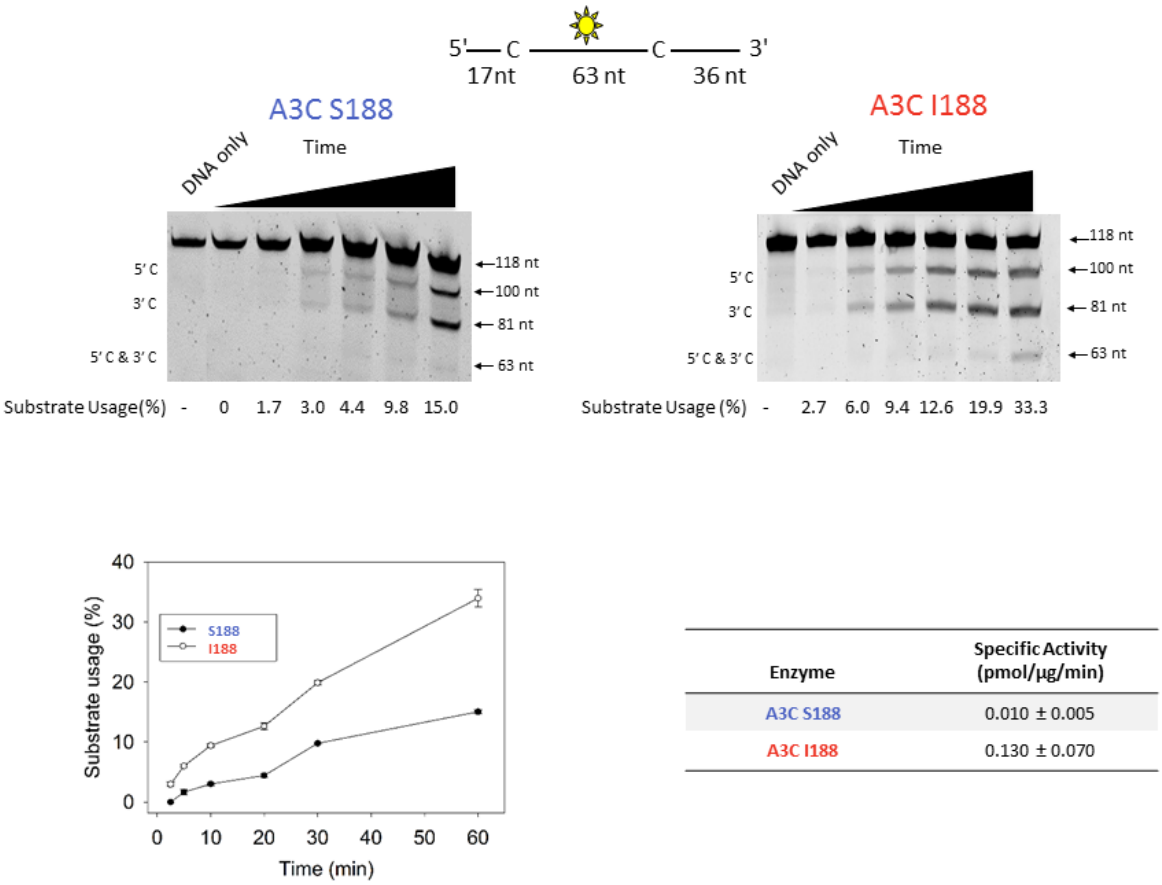
Isoleucine at position 188 of APOBEC3C enhances enzymatic activity *in vitro*

We wished to investigate whether or not the more potent antiviral activity of APOBEC3C I188 compared to APOBEC3C S188 could be explained by differences in their inherent enzymatic activity. Thus, each protein was produced by expression in a recombinant baculovirus

system, purified as described in the Materials and Methods, and tested for its ability to cause cytidine deamination. We examined APOBEC3C S188 and I188 activity using a ssDNA substrate containing two deamination target motifs (Figure 8, top sketch). 5' TTC deamination motifs were used because APOBEC3C preferentially targets this motif (121). Reactions were carried out as a time-course over 60 minutes and next the substrates were incubated with uracil DNA glycosylase, which modifies uracil-containing DNA and makes it sensitive to cleavage at high pH. Cytidine to uracil mutations leading to DNA cleavage were detected based on a fluorescein label placed between the two deamination motifs. Substrate usage was calculated from integrated gel band intensity of cleaved product at either deamination motif relative to the uncleaved substrate. We found that at all time points substrate usage of APOBEC3 I188 was higher than S188, and by 60 minutes I188 had led to twice as many cleavage events as S188 (Figue 8, top). The specific activity of APOBEC3C was determined by calculating the picomoles of substrate used (or deamination events) per microgram of enzyme per minute on a 118 nt ssDNA. The specific activity values were calculated using initial reaction times where the substrate usage was in the linear range (Figure 8, bottom left). We found that APOBEC3C S188 had a specific activity approximately 10-fold lower than I188 (0.010 pmol/μg/min vs 0.130 pmol/μg/min) (Figure 4, bottom right). Therefore, the I188 APOBEC3C more rapidly deaminated cytosines *in vitro* than S188.

Since APOBEC3C I188 has greater cytidine deaminase activity *in vitro* than APOBEC3C S188 (Figure 8), we predicted that it would also have a higher mutational frequency than the APOBEC3C S188. To test this prediction, we used a model *in vitro* system that reconstitutes reverse transcription of RNA to DNA, and observed the ability of APOBEC3 enzymes to induce mutagenesis. The template includes the gene *lacZα*, and blue/white screening was performed to

identify mutated reverse transcription products. White colonies, representing templates that were mutated, were then sequencing and the number of mutations induced by each APOBEC3 were quantified. We found that addition of APOBEC3C I188 induced two-fold higher clonal mutation frequency compared to APOBEC3C S188 (data not shown, 0.33×10^{-2} mutations/bp versus 0.15×10^{-2} mutations/bp, respectively). For reactions containing APOBEC3C S188, 100% of clones had zero to one G→A mutation. In contrast, the presence of APOBEC3C I188 caused a noticeable shift in the number of G→A mutations with 32% of clones having more than one mutation and up to four to five mutations in some individual clones (data not shown). Overall, isoleucine at position 188 increased the APOBEC3C-induced mutagenesis of ssDNA *in vitro*.



Enzyme	Specific Activity (pmol/µg/min)
A3C S188	0.010 ± 0.005
A3C I188	0.130 ± 0.070

Figure 8: In vitro characterization of APOBEC3C S188 and I188. (Top) The specific activity of APOBEC3C S188 and I188 was determined by incubating the enzyme with a 118 nt ssDNA substrate with an internal fluorescein label (yellow star). Single deaminations of the 5'C and 3'C are detected as the appearance of fluorescently labeled 100 nt and 81 nt fragments, respectively; double deamination of both C residues on the same molecule results in a 63 nt labeled fragment. Substrate usage is below gels. (Bottom) The substrate usage during a 60 min time course was plotted from three independent experiments and the specific activity was calculated.

APOBEC3C is expressed in activated PBMCs and both variants are targeted by Vif

In the absence of direct clinical or cohort data, we next sought to further evaluate the relevance of APOBEC3C to HIV infection. Although previous studies had found that APOBEC3C mRNA is expressed in primary T cells, the protein levels had not been investigated. We found that APOBEC3C protein is expressed in PHA-activated primary PBMC (Figure 9A). We also found that the protein is expressed in various monocyte and lymphocyte cell lines (Figure 9B).

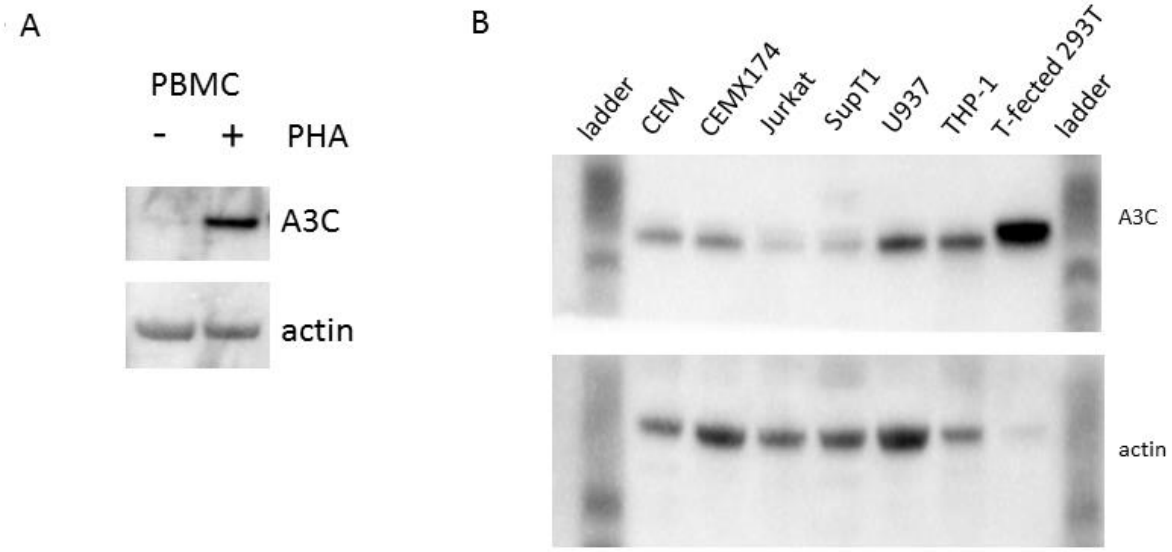


Figure 9. APOBEC3C is expressed in activated PBMCs and is targeted by Vif. (A)Western blot analysis of endogenous expression of APOBEC3C (A3C) in lymphocyte and monocyte cell lines, as well as PBMCs (unactivated, and 48 hours after activation with PHA).

We then reasoned that if APOBEC3C is indeed a restriction factor relevant to HIV, then one would expect it to be antagonized by the viral Vif protein. To test this, we produced HIV-1 (either lacking *vif*, or expressing either HIV-1 or HIV-2 *vif*) in the presence of APOBEC3C. When we express APOBEC3C I188 during HIV production, the infectivity of the virus is reduced by more ten-fold. However, in the presence of APOBEC3 S188 or I188, both HIV-1 Vif and HIV-2 Vif restored viral infectivity (Figure 10). Thus, APOBEC3C I188 is effectively antagonized by HIV-1 and HIV-2 Vif which suggests that even the more active form of APOBEC3C can still be targeted by both human lentiviral pathogens. In conclusion, both variants of APOBBEC3C are targeted by Vif and APOBEC3C protein is expressed in HIV target cells. These results are consistent with the hypothesis that APOBEC3C is relevant to HIV infection.

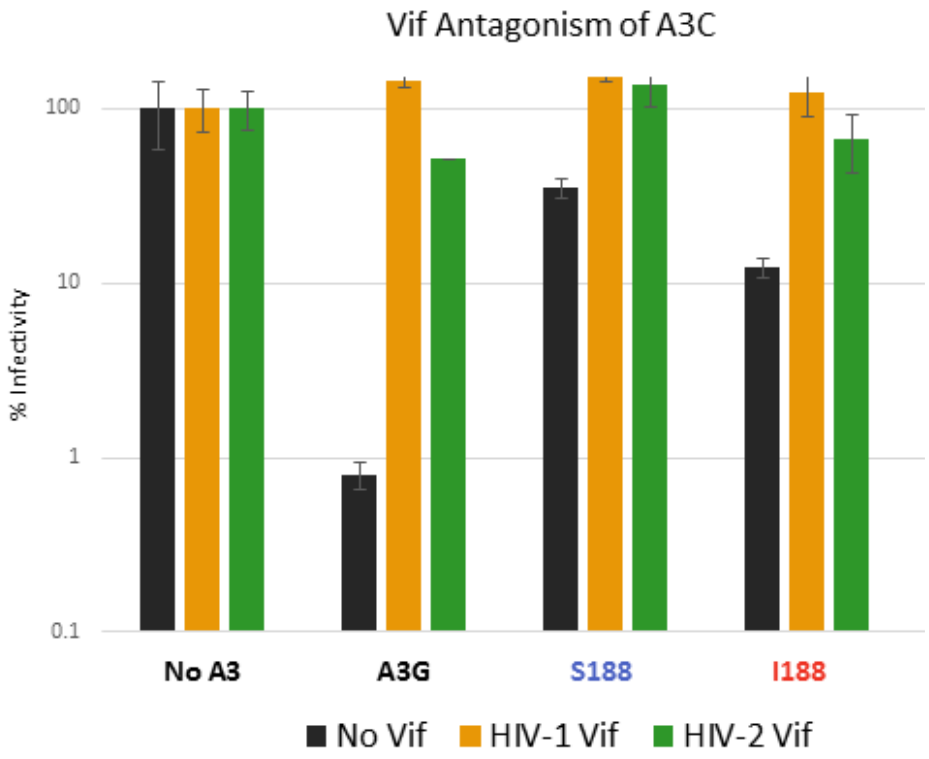


Figure 10: APOBEC3C is expressed in activated PBMCs and is targeted by Vif. Restric tion of HIV-1 Δ*vif* by APOBEC3C S188 (A3C S188), APOBEC3C I188 (A3C I188), and full or partial recovery of infectivity by the presence of HIV-1 (LAI strain) and HIV-2 (ROD strains) *vif*. 0.4ug of APOBEC3 plasmid was used for each condition. Infectivity of each virus is set to 100% for infection with No APOBEC3 (No A3) present. Error bars represent standard deviation of triplicate transfections and infection, and this experiment was repeated twice.

APOBEC3C I188 is an ancient human polymorphism that is not found in other hominids

The APOBEC3C I188 variant is present at frequency of 2.4% in the 1000 Genomes Project (124). Therefore, a relatively small proportion of humans carry a variant of APOBEC3C that is more enzymatically active against lentiviruses. To determine which allele is ancestral at position 188, we constructed a phylogeny of primate *APOBEC3C* sequences. All old world monkeys (N=15) analyzed encode an isoleucine at position 188 (Fig. 11A). Moreover, orangutans, siamangs, and gibbons also encode isoleucine, but the serine change at amino acid 188 occurred in the lineage leading to gorillas, chimpanzees, and humans (Figure 7A). Thus, isoleucine at position 188 is likely the ancestral state, and changed during the evolution of hominids. There are two possible explanations for the existence of the I188 in humans: 1) a reversion back to isoleucine may have occurred in a subpopulation or 2) a polymorphism has been maintained at this site for millions of years, since humans split from their ancestor with gorillas and chimpanzees.

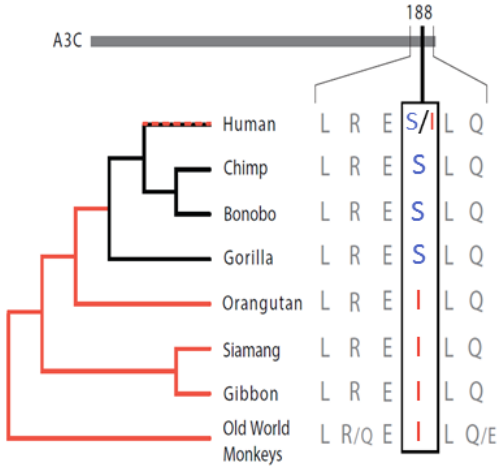
If a serine to isoleucine reversion mutation occurred in recent human evolution, we would expect it to be present only in a limited subset of humans. The frequency of the allele in the 1000 Genomes Project data is 8.9% in populations of African descent, less than 1% frequency in the Americas, and not present in Asia and Europe (124) (Figure 11). Humans are dramatically more genetically diverse in Africa than on any other continent, so the presence of the allele almost exclusively in Africa does not provide strong evidence that the isoleucine reverted only in African populations. Therefore, we sought to determine if the APOBEC3C I188 allele is distributed across divergent populations in Africa, or if it is present in only a particular subpopulation. The APOBEC3C I188 allele is present in all six African subpopulations analyzed by the 1000 Genomes project, with a frequency ranging between 5.6% and 13% (Figure

11B). However, many of the sub-populations included in the 1000 Genomes Project live in regions affected by the Bantu Expansion, a migration event when Bantu-speaking tribes swept across the continent approximately 3,000 years ago(126, 127). To determine if the isoleucine allele is present in more diverse African genomes, we determined the *APOBEC3C* sequence from individuals from four hunter-gatherer groups (Hadza, Sandawe, Mbuti, and Khoe-San)(108, 109). We found that one of the four Khoe-San individuals was heterozygous for the I188 allele, and two out of five Sandawe individuals were heterozygous for the I188 allele (Figure 11B). In conclusion, I188 seems to be a widely distributed SNP in African populations suggesting that the more active allele is very ancient, and may have even been circulating in humans since the birth of the species. Presence of the I188 in the ancient human relative *Homo neanderthalensis* would have provided evidence that the allele has been present in the Homo lineage for at least 600,000 years but we failed to find the I188 SNP in the published Neanderthal genomes.

To determine if other hominoids also possess variation at position 188 we probed the *APOBEC3C* sequences from the Great Ape Genome project(128), and found that none of the great apes included in the study (n=79) were polymorphic at position 188 (Figure 11C). Ten orangutans were included in the study, and all encoded isoleucine at position 188. In contrast, all gorillas (n= 31), and chimpanzees and bonobos (n=38), encoded serine at position 188. Humans, gorillas, and chimpanzees diverged from their most recent common ancestor approximately 10 to 20 million years ago(129, 130), and in this ancestral lineage the more active isoleucine allele was lost. However, since some humans express the I188 allele, it is possible S188 never rose to fixation and I188 was maintained as a minor allele for a long period of the evolutionary history of hominoids. Alternatively, it is possible that serine became fixed in the ancestor to gorillas, chimpanzees and humans, but more recently the serine reverted to isoleucine in a subpopulation

of humans. Nonetheless, we find that the APOBEC3C I188 is relatively ancient to humans, but is not present to an appreciable extent in out-of-Africa human populations, nor have we found it in other hominids.

A



B

Allele frequency of A3C S188I SNP in Modern Humans & Ancient Relatives

	% frequency
Humans- Global[1]	**2.44** (5008)
African[1]	**8.85** (1322)
African Carribean[1]	13.02 (192)
African Ancestry (Southern U.S.)[1]	4.92 (122)
Esan, Nigeria[1]	5.56 (198)
Luhya, Webuye	10.60 (198)
Mandinka, The Gambia[1]	7.96 (226)
Mende, Sierra Leone[1]	7.06 (170)
Yoruba, Kenya[1]	11.11 (216)
Hadza, Tanzania[3]	0.00 (10)
Khoisan & Bantu, Namibia & S. Africa[2]	10.00 (10)
Mbuti, DRC[3]	0.00 (10)
Sandawe, Tanzania[3]	20.00 (10)
American[1]	**0.72** (694)
Columbia[1]	0.53 (188)
Mexican Ancestry, California[1]	0.78 (128)
Peru[1]	0.59 (170)
Puerto Rico[1]	0.96 (208)
East Asian[1]	**0.00** (1008)
South Asian[1]	**0.00** (978)
European[1]	**0.00** (1006)
Homo neanderthalensis	**0.00 (1)**

n= allele count

Data sources: [1]= 1K Genomes Project, [2]= Schuster et al, [3]= Lachance et al

C

Allele frequency of A3C I188 in Great Apes

Genus	% Freq Ile (n)
Pan (chimpanzee & bonobo)	0.00 (76)
Gorilla	0.00 (62)
Pongo	100.00 (20)

n = allele count , Data Source: Great Ape Genome Project

Figure 11: Isoleucine 188 changes to serine in some hominoids but was maintained or reverted back to isoleucine for some human populations. (A) A phylogram of cattarhines, along with an alignment of the C-terminus of APOBEC3C. All old world monkey sequences encoded isoleucine at 188 (15 sequences total). (B) Allele frequency of Ile SNP in global human populations, as well as *Homo neanderthalensis*. (C) Allele frequency of Ile 188 in chimpanzees, bonobos, gorillas, and orangutans (n=79). Sequences were derived from the Great Ape Genome Project(128).

3.3 DISCUSSION

APOBEC3C stands out among the seven human APOBEC3 paralogs as it has no known function. We observed that the six APOBEC3s with known functions possess a conserved isoleucine at the residue homologous to APOBEC3C position 188, whereas APOBEC3C encodes a serine at this position. However, human APOBEC3C is in fact polymorphic at this site, and some humans encode an isoleucine, the residue that correlates with APOBEC3 antiviral/anti-retroelement function. This led us to hypothesize that APOBEC3C may have an as yet overlooked role as a restriction factor, and that the I188 variant may have enhanced antiviral activity compared to the more common variant, S188. *APOBEC3C* has evolved under positive selection in primates and within the interface of binding by the viral protein Vif, suggesting that this gene may have played a role in restriction of lentiviruses over primate evolution. Furthermore, we found that *APOBEC3C* I188 encodes a protein with increased antiviral activity, increased enzymatic activity, and the ability to dimerize in solution. Consistent with this conclusion, an artificial forced dimer of APOBEC3C S188 has vastly increased antiviral activity. We find that the isoleucine at position 188 was lost during hominid evolution but was either reacquired by some humans since humans split with our most recent common ancestor with chimpanzees, or alternatively, has never been lost as an allele and has been maintained as a polymorphism through several million years of hominoid evolution.

Positive selection of APOBEC3C in primates suggests an ancient role in antiviral defense

Previous studies have shown that APOBEC3C binds to HIV-1 Vif and that E106 is important for Vif binding (122, 125). We found that this residue within the Vif binding interface is evolving under positive selection, and another residue in the Vif-binding region, 77, is also

under positive selection. Additionally, it is possible that Vifs from other lentiviruses target APOBEC3C at different motifs, driving the positive selection in other regions of the protein. For example, APOBEC3C is under positive selection at residues 128 and 130. While these residues are not in the known APOBEC3C:HIV-1 Vif binding interface, the homologous residues of APOBEC3G are involved in HIV-1 Vif binding(110, 131). Therefore, it is possible that other Vif proteins from other lentiviruses target APOBEC3C at positions 128 and 130, or that ancient lentiviruses have targeted these residues in the past. In summary, rapid evolution of APOBEC3C at the known APOBEC3C:Vif binding interface suggests that APOBEC3C has evolved to block lentiviruses in primates.

Mechanism of increased activity of APOBEC3C I188 relative to APOBEC3C S188

Our results indicate that the difference in the anti-HIV activity of the APOBEC3C variants S188 and I188 lies in the enzymatic efficiency of the two APOBEC3C proteins. We found that I188 more rapidly deaminates ssDNA *in vitro*. Furthermore, in an *in vitro* RT model system, the presence of APOBEC3C cause a higher mutation frequency than APOBEC3C S188.

Ongoing work suggests that the APOBEC3C I188 is more processive (able to consecutively deaminate cytidine motifs without dissociating from substrate), than APOBEC3C S188 (Adolph, Wittkopp et al, Cytidine deaminase efficiency of the hominid viral restriction factor APOBEC3C is mediated by dimerization, submitted). Taken together, these results bolster the model that the APOBEC3C I188 protein has greater antiviral activity than the more common APOBEC3C protein due to better enzymatic activity.

Population genetics of human APOBEC3C suggests that the I188 polymorphism is ancient

The isoleucine at position 188 of APOBEC3C is present at approximately 10% frequency across diverse African populations, but almost absent from all other global populations. All human populations outside of Africa are thought to have descended from one or a few migration events out of Africa(132). As such, humans from non-African populations may lack the APOBEC3C I188 allele because it was excluded in a population bottleneck during the migrations. Or, the allele may have been lost in non-African populations due to drift or a lack of selective pressure, and this possibility is consistent with the fact that no primate lentiviruses are known to originate outside of Africa. Alternatively, it is possible that loss of the allele was selected for non-African populations. Expression of another APOBEC3, APOBEC3B, has been associated with increased risk of cancer (75, 133). Therefore, the antiviral function of APOBEC3s may come at an evolutionary trade-off. In fact, this may have driven the maintenance of the less enzymatically active S188 allele for millions of years in humans and ancient human ancestors.

Our phylogenetic analysis shows that APOBEC3C I188 is ancestral in primates, but changed to serine in the clade of apes including gorillas, chimpanzee, and humans. The fact that humans have a polymorphism that corresponds with the ancestral residue could be due to a reversion back to the amino acid present in other primates, but not in gorillas nor chimpanzees. If a reversion occurred it must have happened long ago in human history, since the allele is present in such deeply divergent populations across Africa. However, the allele was likely lost due to a bottleneck in the out-of-Africa populations because it is almost completely missing from non-African populations. Alternatively, it is possible that the isoleucine allele has continued in the human lineage through incomplete lineage sorting (the maintenance of a polymorphism after the

divergence of species), since before humans split with their most recent common ancestor with gorillas more than 10 million years ago. Notably, the isoleucine codon, ATT, at position 188 is the same in the human SNP as in all other primates with an Ile at this position in APOBEC3C. While we did not find support for incomplete lineage sorting since we did not find any other hominids that were polymorphic at position 188, the limited number of great ape sequences were included does not allow us to completely rule out this second possibility. Nonetheless, given the increased antiviral activity of APOBEC3C I188 and its fixation in primates other than hominids argues that the gain (or maintenance) of this allele in humans have been driven by a function for protection against pathogens.

Potential impact on human health

We discovered that an APOBEC3C single nucleotide polymorphism (SNP) that is common in Africa enhances anti-lentiviral activity. This polymorphism may impact human susceptibility to cross-species transmissions of lentiviruses because Vifs from other lentiviruses may not antagonize human APOBEC3C HIV-1 and HIV-2 Vif are able to antagonize both variants of APOBEC3C so the I188 SNP may not block HIV transmission. However, the fact that APOBEC3C is antagonized by Vif is consistent with the possibililty that APOBEC3C is an important barrier that must be countered by the virus during natural infections. Alternatively, it is possible that APOBEC3C antagonism by Vif is an unintended consequence due to Vif binding to another APOBEC3, APOBEC3F, as APOBEC3C has a Vif binding pocket that is nearly identical to the Vif binding pocket of APOBEC3F(122, 125, 134). Despite the ability of Vif to antagonize APOBEC3C, it is possible that APOBEC3C I188 influences HIV susceptibility. In infected individuals possessing the whole APOBEC3 repertoire, Vif has to adapt to counteract

multiple antiviral proteins and this may constrain Vif and weaken its activity. In fact, viral genomes sequenced from HIV-1-infected patient cells are extensively mutated by APOBEC3s despite the presence of Vif(135, 136) and the extent of APOBEC3-induced mutagenesis negatively correlates with disease progression rate(137). As such, it is possible that I188 may provide some level of protection from HIV transmission or pathogenesis.

Chapter 4.

4.1 INTRODUCTION

The human APOBEC3 family of single-stranded (ss) DNA cytidine deaminases has seven members that act as host restriction factors against retroelements (including HIV) and certain DNA viruses (138). In order for APOBEC3s to restrict HIV, they are encapsidated into budding virions and then mutate viral DNA by deaminating cytosines (producing uracils) on the (-) strand DNA synthesized by reverse transcriptase (139). When the (-) DNA is copied to form the (+) DNA the uracils template the addition of adenine, resulting in G→A mutations that reduce the infectivity of HIV (140-142).

The seven human APOBEC3s derived from ancient duplications of deaminase domains. Certain APOBEC3 proteins consist of a single deaminase domain (APOBEC3A, APOBEC3C, and APOBEC3H), whereas the other APOBEC3s consist of two deaminase domains (APOBEC3B, APOBEC3D, APOBEC3F, and APOBEC3G). Another difference among APOBEC3 proteins is that they vary in their potency and specificity of viral/retroelement restriction. Human APOBEC3D, APOBEC3F, and APOBEC3H also inhibit HIV-1 (Δ*vif*), but to a lesser extent than APOBEC3G (65, 68, 113, 114). In contrast, APOBEC3A and APOBEC3B do not potently block HIV infection of T cells (65, 68, 77, 113), which are the primary target of

HIV (although a target-cell effect has been reported in monocytes for APOBEC3A) (115) . Instead, APOBEC3A and APOBEC3B drastically inhibit replication of endogenous retroelements and some DNA viruses (71, 72, 79, 80, 90, 116). In studies that compare the ability of the seven human APOBEC3s to restrict lentiviruses and endogenous retroelements, the only APOBEC3 that has weak activity against both lentiviruses and endogenous retroelements is APOBEC3C (65, 68, 76, 80, 116-121). However, a naturally occurring APOBEC3C variant encoding isoleucine at position 188 has enhanced enhances anti-HIV-1 activity (see Chapter 3). Furthermore, this I188 variant is a more potent cytidine deaminase enzyme in vitro. Dimerization/oligomerization has been correlated with the ability of APOBEC3s to restrict lentiviruses (143). Therefore, we sought to determine if the I188 variant is a more potent anti-HIV restriction factor due to enhanced dimerization.

We found that while the S188 APOBEC3C is monomeric in solution, the I188 variant exists as monomers and dimers in solution. Furthermore, we found that a forced dimer, or double-domain APOBEC3C, potently enhances HIV-1 restriction. Therefore, we propose a model where the dimerization of I188 enhances its cytidine deaminase activity, and in turn, causes the variant to more potently restrict HIV-1(Δ*vif*). The I188 variant only partially dimerizes in solution. Therefore, we next sought to identify APOBEC3C point mutants that would enhance dimerization in the background of I188 in order to determine if increased dimerization correlates with antiviral activity. Interestingly, chimpanzees express an APOBEC3Cs that more potently restrict HIV-1 (Δ*vif*) than human APOBEC3C, despite encoding serine at position 188. We found that lysine at position 115 of chimpanzee APOBEC3C is sufficient to increase antiviral activity and dimerization. Next, we constructed a mutant of APOBEC3C with isoleucine at 188 and lysine at position 115, to determine if these

residues when expressed together, can lead to full dimerization. We found that indeed, the double mutant of human APOBEC3C (K115/I188) fully dimerizes in solution. This double mutant human APOBEC3C also potently inhibits HIV-1 (Δ*vif*) replication in cell culture, to the same level as the APOBEC3C double-domain forced dimer. These results support our model that dimerization is important for APOBEC3C antiviral activity.

4.2 RESULTS

Dimerization correlates with enhanced antiviral activity of human APOBEC3C

Previous studies have reported that the S188 variant of APOBEC3C is a monomeric protein, both in solution(125) and in cells(143). Indeed, by size exclusion chromatography we also found that baculovirus-produced APOBEC3C S188 (the common variant) is monomeric (Figure 12, apparent molecular weight 17 kDa). However, the baculovirus-produced APOBEC3C I188 was in equilibrium between monomer and dimer forms (Figure 12, apparent molecular weight 21 kDa and 42 kDa, respectively). Therefore, dimerization of APOBEC3C correlates with antiviral activity. The observation that the isoleucine residue at position 188 was able to shift the oligomeric profile of APOBEC3C suggests that residue 188 is important for dimerization.

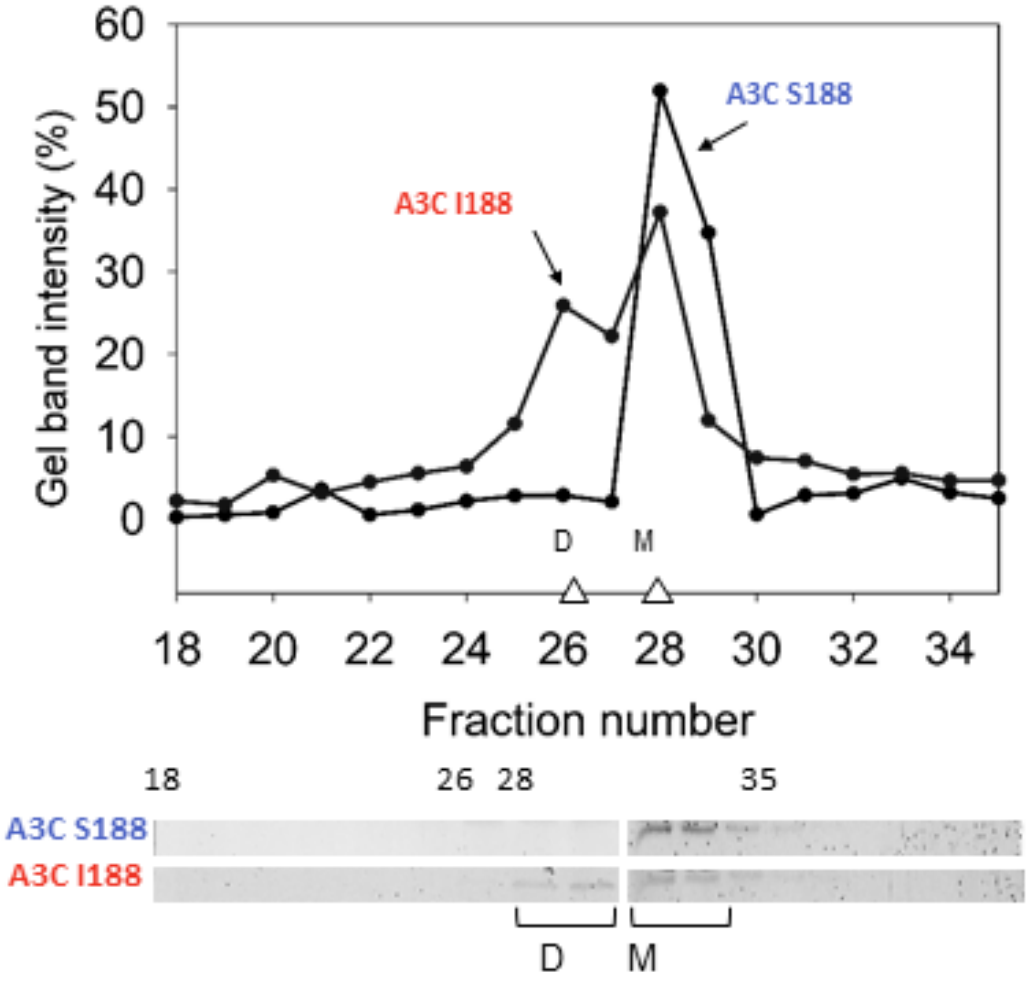


Figure 12. APOBEC3C dimerization. (A) Size exclusion chromatography profiles of the APOBEC3C S188 and I188 from a 10mL G200 Superdex column were used to calculate the oligomerization states of the enzymes. When APOBEC3C S188 and I188 were loaded onto the column, APOBEC3C S188 eluted at an apparent molecular weight 17 kD (consistent with a monomeric form), whereas APOBEC3C I188 eluted in two peaks, with apparent molecular weights of 42 kD and 21 kD (consistent with the presence of both monomers and dimers). Chromatograms were made using the integrated gel band intensities of each protein fraction after resolution by SDS-PAGE.

Interestingly, dimerization has been previously been correlated with improved APOBEC3 catalytic activity because it enables efficient scanning of ssDNA to find cytosines targets for deamination (107). Therefore, we considered the possibility that dimerization is important for efficient antiviral activity of APOBEC3C. In order to test this hypothesis, we first sought to identify other residues that contribute to dimerization. The I188 variant of APOBEC3C has enhanced antiviral activity, and this correlates with its ability to dimerize in solution. However, the I188 variant only partially dimerizes. Therefore, we sought to construct a mutant version of APOBEC3C I188 that fully dimerizes in order to test whether a fully dimerizing point mutant has enhanced antiviral activity, as compared to the naturally occurring I188 variant.

We found that despite encoding serine at position 188 (present in the monomeric form of human APOBEC3C), chimpanzee APOBEC3C more potently restricts HIV-1 (Δ*vif*) than the common human APOBEC3C variant (Figure 13A). Therefore, we hypothesized that chimpanzee APOBEC3C may better dimerize in solution. To test this, our collaborators produced human and chimpanzee APOBEC3C in a baculovirus expression system and purified the proteins using size exclusion chromatography. Our collaborators found that while S188 human APOBEC3C is a monomer in solution, chimpanzee APOBEC3C partially dimerizes in a pattern similar to the human APOBEC3C variant, I188 (Figure 13B). Next, we sought to determine which residue or residues of chimpanzee APOBEC3C are responsible for conferring the ability to dimerize. Human and chimpanzee APOBEC3Cs differ at four residue positions (Figure 13C). We found that the lysine at position 115 of chimpanzee APOBEC3C is sufficient to confer dimerization, because human APOBEC3C S188/K115 point mutant partially dimerizes, in contrast to human APOBEC3C S188/N115. Interestingly, in the background of I188, K115 confers human

APOBEC3C the ability to fully dimerize (Figure 13D). Therefore, we were able to identify a mutant of human APOBEC3C that fully dimerizes in solution.

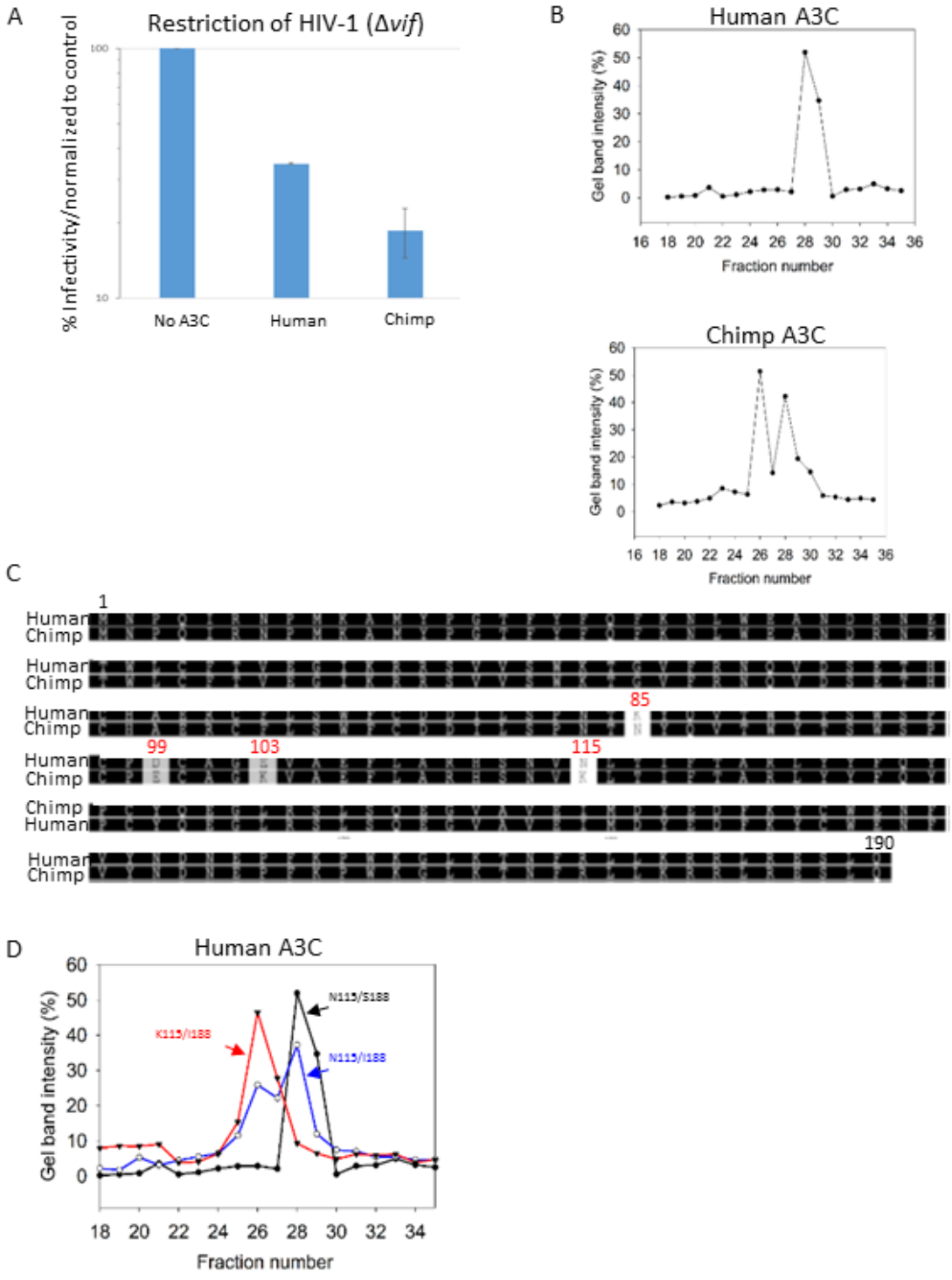


Figure 13: Lysine 115 confers dimerization. (A) Infectivity of HIV-1 Δvif in the absence of APOBEC3C (No A3C), and in the presence of human APOBEC3C or chimpanzee APOBEC3C. 0.2µg of HA-tagged APOBEC3 was expressed in virus-producing cells, and viruses were collected and use for infection. Infectivity in the absence of APOBEC3 is set to 100%. Error bars indicate standard deviation of triplicate transfections and infections. (B) Size exclusion chromotography of human and chimpanzee APOBEC3Cs. (C). Alignment of human and chimp APOBEC3Cs. (D) Size exclusion chromatography of purified human APOBEC3C (both the S188 and I188 variants, marked N115/S188 and N115/I188, respectively), as well as the mutant APOBEC3C K115/I188.

To further test our hypothesis that dimerization is important for APOBEC3 antiviral activity, we constructed an artificial dimer that consists of two tandem S188 APOBEC3Cs (Figure 14) and tested the anti-lentiviral activity of this protein. We used the linker that naturally exists between the N- and C-terminal domains of the two double-domain APOBEC3s, APOBEC3D and APOBEC3F, which are the APOBEC3 proteins with the highest sequence identity shared with APOBEC3C. This linker consists of amino acids Arg-Asn-Pro followed by the second APOBEC3 domain starting at Met12 (labeled Met12' here—see schematic at top of Figure 14). Western blot analysis shows that this artificial double domain APOBEC3C is expressed in cells and runs at approximately the same molecular weight as the natural double domain APOBEC3 protein, APOBEC3G (Figure 8).

Next, we examined the antiviral activity of the synthetic dimer APOBEC3C (with S188 in both domains, called S188-S188) compared to APOBEC3C S188 and APOBEC3C I188. Again, APOBEC3G was used as a positive control. As shown previously, APOBEC3C I188 restricted 5-10 fold better than APOBEC3C S188 (Figure 14: 30% infectivity compared to 8 % infectivity). Strikingly, APOBEC3C S188-S188 dimer restricted infection as efficiently as APOBEC3G (Figure 14: approximately 1% infectivity for both conditions). Importantly, the fact that the APOBEC3C S188-S188 synthetic dimer restricts infection far greater than two-fold more than the APOBEC3C S188 monomer (Figure 14: 30% infectivity relative to 1% infectivity), suggests that the increased antiviral activity is not simply the result of having twice as many active sites. Thus, these results indicate that forced dimerization is sufficient to induce anti-HIV activity of APOBEC3C, and suggest that the increased activity of APOBEC3C I188 is due to its increased ability to spontaneously form dimers.

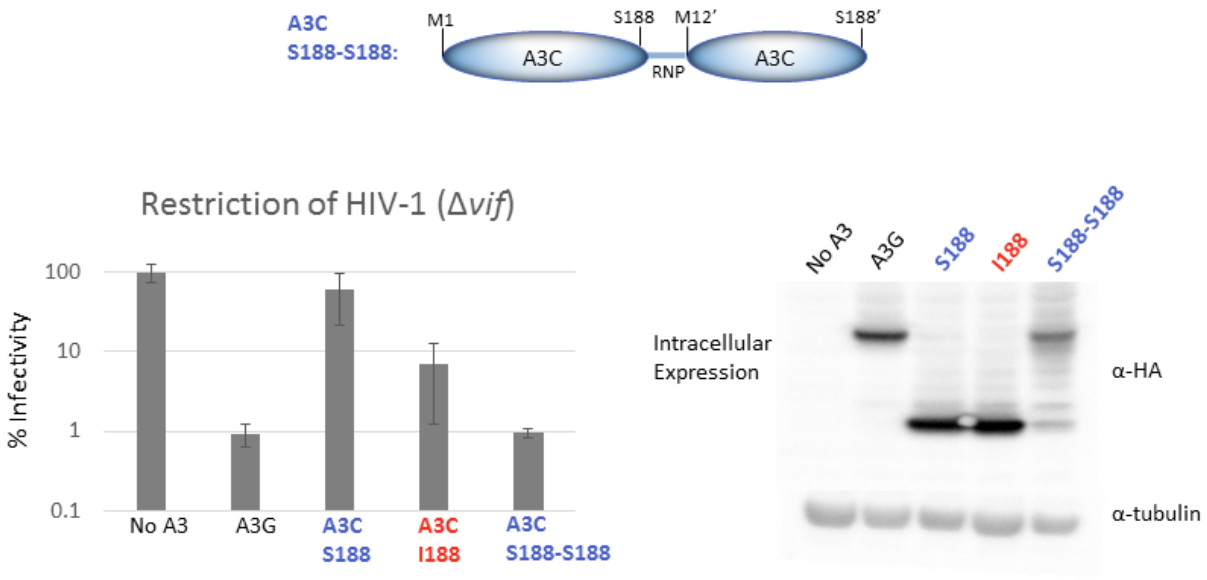


Figure 14: The impact of APOBEC3C dimerization on HIV restriction. (top) Cartoon schematic showing the sequence of the double-domain APOBEC3C. (bottom left) Restriction of HIV-1 Δvif by APOBEC3G(A3G), APOBEC3C S188 (A3C S188), APOBEC3C I188 (A3C I188) and the double-domain APOBEC3C (S188-S188). 0.3μg APOBEC3 was used in this assay. Error bars represent standard deviation of four independent transfections and infections, and this experiment was repeated twice. (bottom right). A Western blot for expression of the APOBEC3 proteins is shown, and is representative of three experiments.

We identified a mutant of human APOBEC3C that fully dimerizes in solution. To further test our model that dimerization is important for APOBEC3C antiviral activity, we tested the ability of the I188/K115 mutant to restrict HIV-1 (Δ*vif*). We found that the double mutant (I188/K115) potently restricts HIV-1(Δ*vif*), and this restriction is significantly more potent than that of the human APOEC3C variant I188 (12% versus 3% in this assay) (Figure 15).

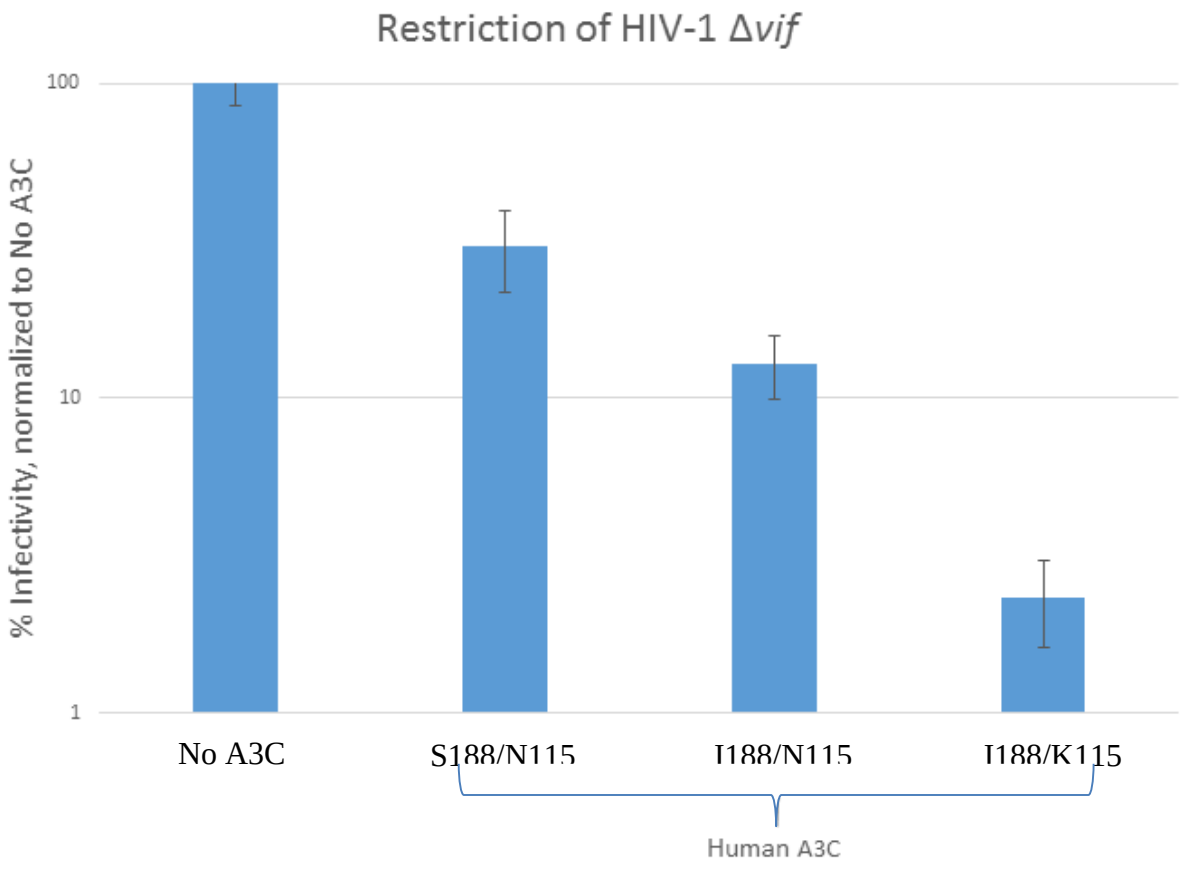


Figure 15: Infectivity of HIV-1 (Δ*vif*) in the absence of APOBEC3C (No A3C), APOBEC3C S188, APOBEC3C I188, and APOBEC3C I188/N115. 0.2µg of HA-tagged APOBEC3 was expressed in virus-producing cells, and viruses were collected and use for infection. Infectivity in the absence of APOBEC3 is set to 100%. Error bars indicate the standard deviation of triplicate transfections and infections.

4.3 DISCUSSION

In Chapter 3 we identified that a naturally occurring APOBEC3C variant (I188) has enhanced anti-lentiviral activity, and this can be attributed to its cytidine deaminase activity (see Chapter 3). In this chapter, we show that the I188 variant forms dimers in vitro, in contrast to the S188 variant. A previous study correlated multimerization of APOBEC3s with the capacity to restrict lentiviruses, and our finding that the monomeric variant (S188) was less antivirally active than the dimer-forming, more active variant (I188), is consistent with this conclusion. Therefore, our model is that dimerization, which is conferred by isoleucine at position 188, improves APOBEC3C antiviral activity. This model is further supported by the fact that a synthetic dimer formed by linking two tandem S188 APOBEC3Cs drastically enhances antiviral activity. In fact, activity is improved even in comparison to I188, the more active variant. I188 only partially dimerizes, and compared to S188 and S188-S188 has an intermediate ability to restrict HIV. Furthermore, through an evolution guided approach we identified another residue that can contribute to dimer formation, lysine at position 115. Chimpanzees encode lysine at position 115, and this residue confers the ability to partially dimerize. Furthermore, a mutant of human APOBEC3C expressing lysine at position 155 and isoleucine at position 188 is able to fully dimerize, and this mutant more potently restricts HIV infection than the naturally occurring I188 variant (which encodes asparagine at position 115). This result bolsters the finding that dimerization is important for APOBEC3 restriction activity.

Isoleucine at position 188 of APOBEC3C confers the ability to dimerize. This residue is located on α-helix 6 of the protein, which is located at the C-terminus of the protein. In the study

reporting the crystal structure of APOBEC3C, the protein was purified in predominantly a monomeric form, but a very small fraction eluted as dimers and oligomers. Therefore, in order to predict naturally occurring dimer interfaces, the authors performed a PISA analysis, which calculates the surface area of interactions between crystal monomers. The PISA analysis indicated that the crystal packing interaction with the largest surface area (865Å) includes α-helix 6, but residue 188 is not directly involved in this interface and is not surface exposed. Therefore, if this crystal structure interaction is indicative of a naturally occurring interface, I188 may indirectly stabilize homodimerization.

One possible reason dimerization contributes to APOBEC3C activity, is that it could improve the protein's ability to scan DNA substrates for cytidine deamination motifs. In fact, I188 lies within α-helix 6, which has been implicated as important for DNA scanning of another APOBEC3, APOBEC3G (105). Furthermore, homodimers of APOBEC3C would have increased avidity for deamination substrates, and this could enhance the enzymatic activity of the protein.

Taken together, these results bolster the model that the APOBEC3C I188 protein has greater antiviral activity than the more common APOBEC3C variant due to its dimerization potential.

Chapter 5. PERSPECTIVES & FUTURE DIRECTIONS

In this dissertation I showed that a single nucleotide variant of APOBEC3C, I188, more potently inhibits HIV-1 (Δ*vif*) replication in cell culture than the more common variant, S188. Furthermore, based on our collaborator's in vitro experiments as well as our infectivity assays, we show that the I188 variant is more active against HIV-1 due to enhanced enzymatic activity and dimerization. Dimerization appears to be important for APOBEC3C antiviral activity, because an artificial dimer of APOBEC3C, as well as a fully-dimerizing point mutant of APOBEC3C (I188/K155), both restrict HIV potently, and to a greater extent than the partially-dimerizing I188 variant. Furthermore, I showed that the more active (I188) allele was lost during hominoid evolution in the most recent common ancestor to gorillas, chimpanzees, and humans. However, the more active allele was either re-gained during human evolution, or has been maintained through incomplete lineage sorting ever since the I→S change occurred, approximately ten million years ago. The position homologous to APOBEC3C 188 in all other ten cytidine deaminase domains of human APOBEC3Cs encodes a conserved isoleucine, and the common APOBEC3C variant, S188, is the only human APOBEC3 that encodes a different residue at this position. The conservation of isoleucine at this position in all other APOBEC3s suggests that the isoleucine is important for APOBEC3 function.

In our experiments to compare the antiviral activity of S188 and I188 APOBEC3C, we exogenously expressed APOBEC3C (S188 and I188) in virus producing cells, and then virions were used for single cycle infectivity assays. However, we have not demonstrated whether these two variants have differential activity against HIV when expressed under more physiological conditions (such as endogenously expressed APOBEC3C or HIV spreading infection). It would also be important to determine whether the I188 has a protective effect in HIV-infected patients

or if the I188 allele protects against HIV acquisition. Lastly, identification of the dimerization interface of I188 would strengthen our model that dimerization is important for APOBEC3 antilentiviral activity. Together, addressing these three outstanding questions would contribute to our understanding of the role of I188 APOBEC3C in blocking lentivirus replication.

Further characterization of the I188 variant under more physiologically relevant conditions

One outstanding question is whether or not APOBEC3C expressed at endogenous levels impacts lentivirus replication. Stable expression of S188 APOBEC3C does not inhibit HIV replication in SUPT11 cells, a cell line that does not express endogenous APOBEC3s. To test whether the I188 variant has an impact on HIV replication at endogenous levels this experiment could be repeated, including the I188 variant as well as the S188 variant. First, SUPT11 cells will be transduced with APOBEC3C lentivirus constructs, and clones will be selected that have APOBEC3C expression levels that are comparable to activated primary T cells. Next, infectivity assays will be performed to determine if APOBEC3C endogenous expression potently inhibits HIV (-vif). This would be important to show because infectivity assays using transiently over-expressed APOBEC3C do not necessarily recapitulate the amount of APOBEC3C that would be present in human cells in vivo during exposure to HIV. This experiment would be important in order to understand whether APOBEC3C plays a role during natural HIV infections, in vivo.

Does the I188 variant protect against HIV acquisition or pathogenesis?

I have demonstrated in a cell-culture system that exogenously expressed I188 APOBEC3C more potently inhibits HIV in single-round infectivity experiments. However, we have yet to determine whether this difference in activity between the two APOBEC3Cs impacts

HIV infection of patients. This is a reasonable hypothesis because variation in another APOBEC3, APOBEC3H, has been shown to impact the course of HIV infection in patients(93). If the I188 allele is protective against HIV, it is possible that HIV-infected individuals who carry the I188 variant could attain a lower set point viral load than individuals who carry the S188 variant. To test this possibility, we would utilize samples derived from a cohort of individuals in an area that is high-risk for HIV acquisition and is also a region where the I188 allele is fairly common. One cohort that would fulfill these criteria is the Mombasa cohort of Mombasa, Kenya. This is a large cohort of female sex-workers in Mombasa, Kenya who were HIV (-) at the time of enrollment. Women in the cohort were followed for HIV seroconversion, and then were followed to determine their set point viral load, as well as their CD4+ T cell counts. We would expect the I188 to be present at an allele frequency of approximately 10% in Kenya, based on what is reported in the 1000 Genomes Project. Therefore, to test our hypothesis, we would genotype samples obtained from HIV (+) members of the cohort to determine whether each individual is homozygous for S188, heterozygote (S188/I188), or homozygous for I188. We would next obtain the data for the set point viral load of each individual, and determine if a presence of the I188 allele correlates with lower set points. In these experiments, it would be important to calculate the statistical power of the study, because if we did not see a correlation, it may be because the study was not well powered enough to see a difference. However, if we did see that I188 correlates with a lower set point viral load, this would suggest that the I188 is protective against HIV pathogenesis. However, this result would need to be verified in multiple cohorts.

An alternative hypothesis is that instead of the APOBEC3C I188 variant affecting viral load, it may decrease the likelihood of HIV transmission. During genotyping of samples from the Mombasa cohort for example, we may find that a strikingly low number of the HIV(+)

individuals carry the I188 allele (as opposed to the approximately 10% we would expect). If this result were obtained, we would then genotype samples from a comparable number of HIV (-) members of the cohort. If the I188 allele is significantly underrepresented in HIV(+) members, this would suggest that APOBEC3C I188 may protect against HIV acquisition. Another possible way to test this would be to use samples from a discordant couple cohort. Serodiscordant couples cohorts include an HIV (+) member, and their couple who is HIV (-), and the HIV (-) member is followed to determine if they seroconvert. If the hypothesis that the I188 APOBEC3C variant is correct, then I188-carrying individuals might seroconvert at a lower rate than S188-carrying individuals.

Understanding the role APOBEC3C I188 plays in HIV-1 infection of patients may uncover an important determinant of HIV susceptibility/resistance.

Characterization of the I188 dimerization interface

Antilentiviral activity of APOBEC3s has previously been correlated with the ability to oligomerize. Purified S188 APOBEC3C is monomeric in solution, whereas the I188 variant is present as both monomers and dimers in solution. Therefore, isoleucine at 188 promotes dimerization of APOBEC3C. Interestingly, position 188 of APOBEC3C is not surface-exposed, so it does not likely play a direct role in dimerization. However, position 188 is present on alpha-helix 6 of APOBEC3C and there is some evidence that this helix could be involved in a dimerization interface. There is a crystal structure of S188 APOBEC3C solved, and the authors who published the structure reported that upon purification of APOBEC3C, a very small fraction eluted as dimers during size exclusion chromatography. During protein crystallization, the contacts formed between crystal subunits can indicate physiologically relevant dimer interfaces.

Therefore, the authors analyzed the interfaces of the crystal subunits to identify dimer interfaces that may naturally occur. Protein-protein interactions often make contacts a region that is 1,100-1,700 angstroms (squared). Therefore, subunit contacts in a crystal structure that are greater than 1,000 angstroms (squared), could be indicative of a naturally-occurring dimerization interface. The analysis of the APOBEC3C crystal subunit interactions revealed that the largest interface between subunits was 800 angstroms squared. This is not as large as most naturally-occurring dimerization interfaces. However, S188 APOBEC3C is not efficient at forming dimers. Therefore, it is possible that this identified interface maybe become larger when isoleucine is introduced at position 188.

Crystallization of the I188 APOBEC3C may help to identify the naturally-occurring dimerization interface. However, I188 only partially forms dimers in solution. Luckily, our collaborators discovered that mutation at another position of APOBEC3C (N115K), in the background of I188, causes complete dimerization. Therefore, to identify the dimerization interface of APOBEC3C, we would work express the fully-dimerizing N115/I188 mutant APOBEC3C in E. coli, purify this protein, and conduct crystallography to solve the structure. Identification of the dimerization domain would strengthen the model that dimerization is an important determinant of APOBEC3C antiviral activity, and also add to our understanding of general APOBEC3 biology.

Previously, the functional role of APOBEC3C was undetermined. Here, we present evidence that it may play a role against lentiviruses, such as HIV-1. Future work will help determine whether variation in APOBEC3C impact HIV in patients. Furthermore, these studies more broadly contribute to the understanding of the mechanism of APOBEC3 antiviral activity.

BIBLIOGRAPHY

1. Coffin JM, Cold Spring Harbor Laboratory. Press. Retroviruses. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press,; 2002. Available from: Publisher description http://www.loc.gov/catdir/enhancements/fy1208/2004565154-d.html.
2. Knipe DM, Howley PM. Fields virology. 6th ed. Philadelphia, PA: Wolters Kluwer/Lippincott Williams & Wilkins Health; 2013. 2 volumes p.
3. Barre-Sinoussi F, Chermann JC, Rey F, Nugeyre MT, Chamaret S, Gruest J, et al. Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). Science. 1983;220(4599):868-71.
4. Gallo RC, Salahuddin SZ, Popovic M, Shearer GM, Kaplan M, Haynes BF, et al. Frequent detection and isolation of cytopathic retroviruses (HTLV-III) from patients with AIDS and at risk for AIDS. Science. 1984;224(4648):500-3.
5. Popovic M, Sarngadharan MG, Read E, Gallo RC. Detection, isolation, and continuous production of cytopathic retroviruses (HTLV-III) from patients with AIDS and pre-AIDS. Science. 1984;224(4648):497-500.
6. Sarngadharan MG, Popovic M, Bruch L, Schupbach J, Gallo RC. Antibodies reactive with human T-lymphotropic retroviruses (HTLV-III) in the serum of patients with AIDS. Science. 1984;224(4648):506-8.
7. Schupbach J, Popovic M, Gilden RV, Gonda MA, Sarngadharan MG, Gallo RC. Serological analysis of a subgroup of human T-lymphotropic retroviruses (HTLV-III) associated with AIDS. Science. 1984;224(4648):503-5.
8. Nyamweya S, Hegedus A, Jaye A, Rowland-Jones S, Flanagan KL, Macallan DC. Comparing HIV-1 and HIV-2 infection: Lessons for viral immunopathogenesis. Rev Med Virol. 2013;23(4):221-40.
9. UNAIDS. AIDS by the Numbers. Geneva, 2015.
10. Lackner AA, Lederman MM, Rodriguez B. HIV pathogenesis: the host. Cold Spring Harb Perspect Med. 2012;2(9):a007005.
11. Organization WH. World Health Report. 2003.
12. Karn J, Stoltzfus CM. Transcriptional and posttranscriptional regulation of HIV-1 gene expression. Cold Spring Harb Perspect Med. 2012;2(2):a006916.
13. Collins DR, Collins KL. HIV-1 accessory proteins adapt cellular adaptors to facilitate immune evasion. PLoS Pathog. 2014;10(1):e1003851.
14. Wilen CB, Tilton JC, Doms RW. HIV: cell binding and entry. Cold Spring Harb Perspect Med. 2012;2(8).
15. Rasaiyaah J, Tan CP, Fletcher AJ, Price AJ, Blondeau C, Hilditch L, et al. HIV-1 evades innate immune recognition through specific cofactor recruitment. Nature. 2013;503(7476):402-5.
16. Lahaye X, Satoh T, Gentili M, Cerboni S, Conrad C, Hurbain I, et al. The capsids of HIV-1 and HIV-2 determine immune detection of the viral cDNA by the innate sensor cGAS in dendritic cells. Immunity. 2013;39(6):1132-42.
17. Lingappa JR, Reed JC, Tanaka M, Chutiraka K, Robinson BA. How HIV-1 Gag assembles in cells: Putting together pieces of the puzzle. Virus Res. 2014;193:89-107.
18. Brandenberg OF, Magnus C, Rusert P, Regoes RR, Trkola A. Different infectivity of HIV-1 strains is linked to number of envelope trimers required for entry. PLoS Pathog. 2015;11(1):e1004595.
19. Garrus JE, von Schwedler UK, Pornillos OW, Morham SG, Zavitz KH, Wang HE, et al. Tsg101 and the vacuolar protein sorting pathway are essential for HIV-1 budding. Cell. 2001;107(1):55-65.
20. Brenchley JM, Douek DC. HIV infection and the gastrointestinal immune system. Mucosal Immunol. 2008;1(1):23-30.
21. Dean M, Carrington M, Winkler C, Huttley GA, Smith MW, Allikmets R, et al. Genetic restriction of HIV-1 infection and progression to AIDS by a deletion allele of the CKR5 structural gene. Hemophilia Growth and Development Study, Multicenter AIDS Cohort Study, Multicenter Hemophilia Cohort Study, San Francisco City Cohort, ALIVE Study. Science. 1996;273(5283):1856-62.
22. Liu R, Paxton WA, Choe S, Ceradini D, Martin SR, Horuk R, et al. Homozygous defect in HIV-1 coreceptor accounts for resistance of some multiply-exposed individuals to HIV-1 infection. Cell. 1996;86(3):367-77.
23. Samson M, Libert F, Doranz BJ, Rucker J, Liesnard C, Farber CM, et al. Resistance to HIV-1 infection in caucasian individuals bearing mutant alleles of the CCR-5 chemokine receptor gene. Nature. 1996;382(6593):722-5.
24. Emu B, Sinclair E, Hatano H, Ferre A, Shacklett B, Martin JN, et al. HLA class I-restricted T-cell responses may contribute to the control of human immunodeficiency virus infection, but such responses are not always necessary for long-term virus control. J Virol. 2008;82(11):5398-407.
25. Sharp PM, Hahn BH. Origins of HIV and the AIDS pandemic. Cold Spring Harb Perspect Med. 2011;1(1):a006841.
26. Bailes E, Gao F, Bibollet-Ruche F, Courgnaud V, Peeters M, Marx PA, et al. Hybrid origin of SIV in chimpanzees. Science. 2003;300(5626):1713.
27. Hirsch VM, Olmsted RA, Murphey-Corb M, Purcell RH, Johnson PR. An African primate lentivirus (SIVsm) closely related to HIV-2. Nature. 1989;339(6223):389-92.
28. VandeWoude S, Apetrei C. Going wild: lessons from naturally occurring T-lymphotropic lentiviruses. Clin Microbiol Rev. 2006;19(4):728-62.
29. Simon MA, Brodie SJ, Sasseville VG, Chalifoux LV, Desrosiers RC, Ringler DJ. Immunopathogenesis of SIVmac. Virus Res. 1994;32(2):227-51.
30. Silvestri G, Sodora DL, Koup RA, Paiardini M, O'Neil SP, McClure HM, et al. Nonpathogenic SIV infection of sooty mangabeys is characterized by limited bystander immunopathology despite chronic high-level viremia. Immunity. 2003;18(3):441-52.
31. Fultz PN, Gordon TP, Anderson DC, McClure HM. Prevalence of natural infection with simian immunodeficiency virus and simian T-cell leukemia virus type I in a breeding colony of sooty mangabey monkeys. AIDS. 1990;4(7):619-25.
32. Kraus G, Werner A, Baier M, Binniger D, Ferdinand FJ, Norley S, et al. Isolation of human immunodeficiency virus-related simian immunodeficiency viruses from African green monkeys. Proc Natl Acad Sci U S A. 1989;86(8):2892-6.
33. Liovat AS, Jacquelin B, Ploquin MJ, Barre-Sinoussi F, Muller-Trutwin MC. African non human primates infected by SIV - why don't they get sick? Lessons from studies on the early phase of non-pathogenic SIV infection. Curr HIV Res. 2009;7(1):39-50.
34. Duggal NK, Emerman M. Evolutionary conflicts between viruses and restriction factors shape immunity. Nature reviews Immunology. 2012;12(10):687-95.
35. Daugherty MD, Malik HS. Rules of engagement: molecular insights from host-virus arms races. Annu Rev Genet. 2012;46:677-700.

36. Compton AA, Malik HS, Emerman M. Host gene evolution traces the evolutionary history of ancient primate lentiviruses. Philosophical transactions of the Royal Society of London Series B, Biological sciences. 2013;368(1626):20120496.
37. Rajsbaum R, Garcia-Sastre A, Versteeg GA. TRIMmunity: the roles of the TRIM E3-ubiquitin ligase family in innate antiviral immunity. J Mol Biol. 2014;426(6):1265-84.
38. Perron MJ, Stremlau M, Song B, Ulm W, Mulligan RC, Sodroski J. TRIM5alpha mediates the postentry block to N-tropic murine leukemia viruses in human cells. Proc Natl Acad Sci U S A. 2004;101(32):11827-32.
39. Stremlau M, Owens CM, Perron MJ, Kiessling M, Autissier P, Sodroski J. The cytoplasmic body component TRIM5alpha restricts HIV-1 infection in Old World monkeys. Nature. 2004;427(6977):848-53.
40. Sayah DM, Sokolskaja E, Berthoux L, Luban J. Cyclophilin A retrotransposition into TRIM5 explains owl monkey resistance to HIV-1. Nature. 2004;430(6999):569-73.
41. Wilson SJ, Webb BL, Ylinen LM, Verschoor E, Heeney JL, Towers GJ. Independent evolution of an antiviral TRIMCyp in rhesus macaques. Proc Natl Acad Sci U S A. 2008;105(9):3557-62.
42. Kaletsky RL, Francica JR, Agrawal-Gamse C, Bates P. Tetherin-mediated restriction of filovirus budding is antagonized by the Ebola glycoprotein. Proc Natl Acad Sci U S A. 2009;106(8):2886-91.
43. Pardieu C, Vigan R, Wilson SJ, Calvi A, Zang T, Bieniasz P, et al. The RING-CH ligase K5 antagonizes restriction of KSHV and HIV-1 particle release by mediating ubiquitin-dependent endosomal degradation of tetherin. PLoS Pathog. 2010;6(4):e1000843.
44. Neil SJ, Zang T, Bieniasz PD. Tetherin inhibits retrovirus release and is antagonized by HIV-1 Vpu. Nature. 2008;451(7177):425-30.
45. Van Damme N, Goff D, Katsura C, Jorgenson RL, Mitchell R, Johnson MC, et al. The interferon-induced protein BST-2 restricts HIV-1 release and is downregulated from the cell surface by the viral Vpu protein. Cell Host Microbe. 2008;3(4):245-52.
46. Jia B, Serra-Moreno R, Neidermyer W, Rahmberg A, Mackey J, Fofana IB, et al. Species-specific activity of SIV Nef and HIV-1 Vpu in overcoming restriction by tetherin/BST2. PLoS Pathog. 2009;5(5):e1000429.
47. Zhang F, Wilson SJ, Landford WC, Virgen B, Gregory D, Johnson MC, et al. Nef proteins from simian immunodeficiency viruses are tetherin antagonists. Cell Host Microbe. 2009;6(1):54-67.
48. Lim ES, Malik HS, Emerman M. Ancient adaptive evolution of tetherin shaped the functions of Vpu and Nef in human immunodeficiency virus and primate lentiviruses. J Virol. 2010;84(14):7124-34.
49. Teng B, Burant CF, Davidson NO. Molecular cloning of an apolipoprotein B messenger RNA editing protein. Science. 1993;260(5115):1816-9.
50. Navaratnam N, Shah R, Patel D, Fay V, Scott J. Apolipoprotein B mRNA editing is associated with UV crosslinking of proteins to the editing site. Proc Natl Acad Sci U S A. 1993;90(1):222-6.
51. Etard C, Roostalu U, Strahle U. Lack of Apobec2-related proteins causes a dystrophic muscle phenotype in zebrafish embryos. J Cell Biol. 2010;189(3):527-39.
52. Sato Y, Probst HC, Tatsumi R, Ikeuchi Y, Neuberger MS, Rada C. Deficiency in APOBEC2 leads to a shift in muscle fiber type, diminished body mass, and myopathy. J Biol Chem. 2010;285(10):7111-8.
53. Aydin H, Taylor MW, Lee JE. Structure-guided analysis of the human APOBEC3-HIV restrictome. Structure. 2014;22(5):668-84.
54. Munk C, Willemsen A, Bravo IG. An ancient history of gene duplications, fusions and losses in the evolution of APOBEC3 mutators in mammals. BMC Evol Biol. 2012;12:71.
55. Gabuzda DH, Lawrence K, Langhoff E, Terwilliger E, Dorfman T, Haseltine WA, et al. Role of vif in replication of human immunodeficiency virus type 1 in CD4+ T lymphocytes. J Virol. 1992;66(11):6489-95.
56. Sheehy AM, Gaddis NC, Choi JD, Malim MH. Isolation of a human gene that inhibits HIV-1 infection and is suppressed by the viral Vif protein. Nature. 2002;418(6898):646-50.
57. Kobayashi T, Koizumi Y, Takeuchi JS, Misawa N, Kimura Y, Morita S, et al. Quantification of deaminase activity-dependent and -independent restriction of HIV-1 replication mediated by APOBEC3F and APOBEC3G through experimental-mathematical investigation. J Virol. 2014;88(10):5881-7.
58. Browne EP, Allers C, Landau NR. Restriction of HIV-1 by APOBEC3G is cytidine deaminase-dependent. Virology. 2009;387(2):313-21.
59. Bishop KN, Holmes RK, Malim MH. Antiviral potency of APOBEC proteins does not correlate with cytidine deamination. J Virol. 2006;80(17):8450-8.
60. Newman EN, Holmes RK, Craig HM, Klein KC, Lingappa JR, Malim MH, et al. Antiviral function of APOBEC3G can be dissociated from cytidine deaminase activity. Curr Biol. 2005;15(2):166-70.
61. Alce TM, Popik W. APOBEC3G is incorporated into virus-like particles by a direct interaction with HIV-1 Gag nucleocapsid protein. J Biol Chem. 2004;279(33):34083-6.
62. Schafer A, Bogerd HP, Cullen BR. Specific packaging of APOBEC3G into HIV-1 virions is mediated by the nucleocapsid domain of the gag polyprotein precursor. Virology. 2004;328(2):163-8.
63. Holmes RK, Koning FA, Bishop KN, Malim MH. APOBEC3F can inhibit the accumulation of HIV-1 reverse transcription products in the absence of hypermutation. Comparisons with APOBEC3G. J Biol Chem. 2007;282(4):2587-95.
64. Iwatani Y, Chan DS, Wang F, Maynard KS, Sugiura W, Gronenborn AM, et al. Deaminase-independent inhibition of HIV-1 reverse transcription by APOBEC3G. Nucleic Acids Res. 2007;35(21):7096-108.
65. Chaipan C, Smith JL, Hu WS, Pathak VK. APOBEC3G restricts HIV-1 to a greater extent than APOBEC3F and APOBEC3DE in human primary CD4+ T cells and macrophages. J Virol. 2013;87(1):444-53.
66. Stenglein MD, Harris RS. APOBEC3B and APOBEC3F inhibit L1 retrotransposition by a DNA deamination-independent mechanism. J Biol Chem. 2006;281(25):16837-41.
67. OhAinle M, Kerns JA, Li MM, Malik HS, Emerman M. Antiretroelement activity of APOBEC3H was lost twice in recent human evolution. Cell host & microbe. 2008;4(3):249-59.
68. Hultquist JF, Lengyel JA, Refsland EW, LaRue RS, Lackey L, Brown WL, et al. Human and rhesus APOBEC3D, APOBEC3F, APOBEC3G, and APOBEC3H demonstrate a conserved capacity to restrict Vif-deficient HIV-1. J Virol. 2011;85(21):11220-34.
69. Duggal NK, Malik HS, Emerman M. The breadth of antiviral activity of Apobec3DE in chimpanzees has been driven by positive selection. Journal of virology. 2011;85(21):11361-71.
70. Thielen BK, McNevin JP, McElrath MJ, Hunt BV, Klein KC, Lingappa JR. Innate immune signaling induces high levels of TC-specific deaminase activity in primary monocyte-derived cells through expression of APOBEC3A isoforms. J Biol Chem. 2010;285(36):27753-66.

71. Chen H, Lilley CE, Yu Q, Lee DV, Chou J, Narvaiza I, et al. APOBEC3A is a potent inhibitor of adeno-associated virus and retrotransposons. Current biology : CB. 2006;16(5):480-5.
72. Zhang W, Zhang X, Tian C, Wang T, Sarkis PT, Fang Y, et al. Cytidine deaminase APOBEC3B interacts with heterogeneous nuclear ribonucleoprotein K and suppresses hepatitis B virus expression. Cell Microbiol. 2008;10(1):112-21.
73. Bogerd HP, Wiegand HL, Doehle BP, Lueders KK, Cullen BR. APOBEC3A and APOBEC3B are potent inhibitors of LTR-retrotransposon function in human cells. Nucleic Acids Res. 2006;34(1):89-95.
74. Onguru O, Yalcin S, Rosemblit C, Zhang PJ, Kilic S, Gunduz U. APOBEC3B expression in drug resistant MCF-7 breast cancer cell lines. Biomed Pharmacother. 2016;79:87-92.
75. Burns MB, Lackey L, Carpenter MA, Rathore A, Land AM, Leonard B, et al. APOBEC3B is an enzymatic source of mutation in breast cancer. Nature. 2013;494(7437):366-70.
76. Kinomoto M, Kanno T, Shimura M, Ishizaka Y, Kojima A, Kurata T, et al. All APOBEC3 family proteins differentially inhibit LINE-1 retrotransposition. Nucleic acids research. 2007;35(9):2955-64.
77. Yu Q, Chen D, Konig R, Mariani R, Unutmaz D, Landau NR. APOBEC3B and APOBEC3C are potent inhibitors of simian immunodeficiency virus replication. The Journal of biological chemistry. 2004;279(51):53379-86.
78. Bogerd HP, Wiegand HL, Hulme AE, Garcia-Perez JL, O'Shea KS, Moran JV, et al. Cellular inhibitors of long interspersed element 1 and Alu retrotransposition. Proc Natl Acad Sci U S A. 2006;103(23):8780-5.
79. Ahasan MM, Wakae K, Wang Z, Kitamura K, Liu G, Koura M, et al. APOBEC3A and 3C decrease human papillomavirus 16 pseudovirion infectivity. Biochem Biophys Res Commun. 2015;457(3):295-9.
80. Warren CJ, Xu T, Guo K, Griffin LM, Westrich JA, Lee D, et al. APOBEC3A functions as a restriction factor of human papillomavirus. Journal of virology. 2015;89(1):688-702.
81. Compton AA, Emerman M. Convergence and divergence in the evolution of the APOBEC3G-Vif interaction reveal ancient origins of simian immunodeficiency viruses. PLoS pathogens. 2013;9(1):e1003135.
82. Janini M, Rogers M, Birx DR, McCutchan FE. Human immunodeficiency virus type 1 DNA sequences genetically damaged by hypermutation are often abundant in patient peripheral blood mononuclear cells and may be generated during near-simultaneous infection and activation of CD4(+) T cells. J Virol. 2001;75(17):7973-86.
83. Gandhi SK, Siliciano JD, Bailey JR, Siliciano RF, Blankson JN. Role of APOBEC3G/F-mediated hypermutation in the control of human immunodeficiency virus type 1 in elite suppressors. J Virol. 2008;82(6):3125-30.
84. Neogi U, Shet A, Sahoo PN, Bontell I, Ekstrand ML, Banerjea AC, et al. Human APOBEC3G-mediated hypermutation is associated with antiretroviral therapy failure in HIV-1 subtype C-infected individuals. J Int AIDS Soc. 2013;16:18472.
85. Kieffer TL, Kwon P, Nettles RE, Han Y, Ray SC, Siliciano RF. G-->A hypermutation in protease and reverse transcriptase regions of human immunodeficiency virus type 1 residing in resting CD4+ T cells in vivo. J Virol. 2005;79(3):1975-80.
86. Delviks-Frankenberry KA, Nikolaitchik OA, Burdick RC, Gorelick RJ, Keele BF, Hu WS, et al. Minimal Contribution of APOBEC3-Induced G-to-A Hypermutation to HIV-1 Recombination and Genetic Variation. PLoS Pathog. 2016;12(5):e1005646.
87. Kidd JM, Newman TL, Tuzun E, Kaul R, Eichler EE. Population stratification of a common APOBEC gene deletion polymorphism. PLoS Genet. 2007;3(4):e63.
88. Wen WX, Soo JS, Kwan PY, Hong E, Khang TF, Mariapun S, et al. Germline APOBEC3B deletion is associated with breast cancer risk in an Asian multi-ethnic cohort and with immune cell presentation. Breast Cancer Res. 2016;18(1):56.
89. Qi G, Xiong H, Zhou C. APOBEC3 deletion polymorphism is associated with epithelial ovarian cancer risk among Chinese women. Tumour Biol. 2014;35(6):5723-6.
90. Muckenfuss H, Hamdorf M, Held U, Perkovic M, Lower J, Cichutek K, et al. APOBEC3 proteins inhibit human LINE-1 retrotransposition. The Journal of biological chemistry. 2006;281(31):22161-72.
91. Prasetyo AA, Sariyatun R, Reviono, Sari Y, Hudiyono, Haryati S, et al. The APOBEC3B deletion polymorphism is associated with prevalence of hepatitis B virus, hepatitis C virus, Torque Teno virus, and Toxoplasma gondii co-infection among HIV-infected individuals. J Clin Virol. 2015;70:67-71.
92. Wang X, Abudu A, Son S, Dang Y, Venta PJ, Zheng YH. Analysis of human APOBEC3H haplotypes and anti-human immunodeficiency virus type 1 activity. J Virol. 2011;85(7):3142-52.
93. Naruse TK, Sakurai D, Ohtani H, Sharma G, Sharma SK, Vajpayee M, et al. APOBEC3H polymorphisms and susceptibility to HIV-1 infection in an Indian population. J Hum Genet. 2016;61(3):263-5.
94. Sakurai D, Iwatani Y, Ohtani H, Naruse TK, Terunuma H, Sugiura W, et al. APOBEC3H polymorphisms associated with the susceptibility to HIV-1 infection and AIDS progression in Japanese. Immunogenetics. 2015;67(4):253-7.
95. Ooms M, Brayton B, Letko M, Maio SM, Pilcher CD, Hecht FM, et al. HIV-1 Vif adaptation to human APOBEC3H haplotypes. Cell host & microbe. 2013;14(4):411-21.
96. Refsland EW, Hultquist JF, Luengas EM, Ikeda T, Shaban NM, Law EK, et al. 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. 2014;10(11):e1004761.
97. Yang Z. PAML 4: phylogenetic analysis by maximum likelihood. Molecular biology and evolution. 2007;24(8):1586-91.
98. Kosakovsky Pond SL, Murrell B, Fourment M, Frost SD, Delport W, Scheffler K. A random effects branch-site model for detecting episodic diversifying selection. Molecular biology and evolution. 2011;28(11):3033-43.
99. Li MM, Wu LI, Emerman M. The range of human APOBEC3H sensitivity to lentiviral Vif proteins. Journal of virology. 2010;84(1):88-95.
100. Yamashita M, Emerman M. Capsid is a dominant determinant of retrovirus infectivity in nondividing cells. Journal of virology. 2004;78(11):5670-8.
101. Xie Y, Rosser JM, Thompson TL, Boeke JD, An W. Characterization of L1 retrotransposition with high-throughput dual-luciferase assays. Nucleic acids research. 2011;39(3):e16.
102. Chelico L, Pham P, Calabrese P, Goodman MF. APOBEC3G DNA deaminase acts processively 3' --> 5' on single-stranded DNA. Nature structural & molecular biology. 2006;13(5):392-9.
103. Feng Y, Love RP, Ara A, Baig TT, Adolph MB, Chelico L. Natural Polymorphisms and Oligomerization of Human APOBEC3H Contribute to Single-stranded DNA Scanning Ability. The Journal of biological chemistry. 2015;290(45):27188-203.
104. Chelico L, Sacho EJ, Erie DA, Goodman MF. A model for oligomeric regulation of APOBEC3G cytosine deaminase-dependent restriction of HIV. The Journal of biological chemistry. 2008;283(20):13780-91.

105. Ara A, Love RP, Chelico L. Different mutagenic potential of HIV-1 restriction factors APOBEC3G and APOBEC3F is determined by distinct single-stranded DNA scanning mechanisms. PLoS pathogens. 2014;10(3):e1004024.
106. Creighton S, Bloom LB, Goodman MF. Gel fidelity assay measuring nucleotide misinsertion, exonucleolytic proofreading, and lesion bypass efficiencies. Methods Enzymol. 1995;262:232-56.
107. Feng Y, Chelico L. Intensity of deoxycytidine deamination of HIV-1 proviral DNA by the retroviral restriction factor APOBEC3G is mediated by the noncatalytic domain. The Journal of biological chemistry. 2011;286(13):11415-26.
108. Schuster SC, Miller W, Ratan A, Tomsho LP, Giardine B, Kasson LR, et al. Complete Khoisan and Bantu genomes from southern Africa. Nature. 2010;463(7283):943-7.
109. Lachance J, Vernot B, Elbers CC, Ferwerda B, Froment A, Bodo JM, et al. Evolutionary history and adaptation from high-coverage whole-genome sequences of diverse African hunter-gatherers. Cell. 2012;150(3):457-69.
110. Mangeat B, Turelli P, Liao S, Trono D. A single amino acid determinant governs the species-specific sensitivity of APOBEC3G to Vif action. The Journal of biological chemistry. 2004;279(15):14481-3.
111. Harris RS, Dudley JP. APOBECs and virus restriction. Virology. 2015;479-480:131-45.
112. Jarmuz A, Chester A, Bayliss J, Gisbourne J, Dunham I, Scott J, et al. An anthropoid-specific locus of orphan C to U RNA-editing enzymes on chromosome 22. Genomics. 2002;79(3):285-96.
113. Zheng YH, Irwin D, Kurosu T, Tokunaga K, Sata T, Peterlin BM. Human APOBEC3F is another host factor that blocks human immunodeficiency virus type 1 replication. Journal of virology. 2004;78(11):6073-6.
114. Dang Y, Wang X, Esselman WJ, Zheng YH. Identification of APOBEC3DE as another antiretroviral factor from the human APOBEC family. Journal of virology. 2006;80(21):10522-33.
115. Sharma S, Patnaik SK, Taggart RT, Kannisto ED, Enriquez SM, Gollnick P, et al. APOBEC3A cytidine deaminase induces RNA editing in monocytes and macrophages. Nat Commun. 2015;6:6881.
116. Wissing S, Montano M, Garcia-Perez JL, Moran JV, Greene WC. Endogenous APOBEC3B restricts LINE-1 retrotransposition in transformed cells and human embryonic stem cells. The Journal of biological chemistry. 2011;286(42):36427-37.
117. Niewiadomska AM, Tian C, Tan L, Wang T, Sarkis PT, Yu XF. Differential inhibition of long interspersed element 1 by APOBEC3 does not correlate with high-molecular-mass-complex formation or P-body association. Journal of virology. 2007;81(17):9577-83.
118. Bishop KN, Holmes RK, Sheehy AM, Davidson NO, Cho SJ, Malim MH. Cytidine deamination of retroviral DNA by diverse APOBEC proteins. Current biology : CB. 2004;14(15):1392-6.
119. Wiegand HL, Doehle BP, Bogerd HP, Cullen BR. A second human antiretroviral factor, APOBEC3F, is suppressed by the HIV-1 and HIV-2 Vif proteins. EMBO J. 2004;23(12):2451-8.
120. Rose KM, Marin M, Kozak SL, Kabat D. Regulated production and anti-HIV type 1 activities of cytidine deaminases APOBEC3B, 3F, and 3G. AIDS research and human retroviruses. 2005;21(7):611-9.
121. Langlois MA, Beale RC, Conticello SG, Neuberger MS. Mutational comparison of the single-domained APOBEC3C and double-domained APOBEC3F/G anti-retroviral cytidine deaminases provides insight into their DNA target site specificities. Nucleic acids research. 2005;33(6):1913-23.
122. Smith JL, Pathak VK. Identification of specific determinants of human APOBEC3F, APOBEC3C, and APOBEC3DE and African green monkey APOBEC3F that interact with HIV-1 Vif. Journal of virology. 2010;84(24):12599-608.
123. Refsland EW, Stenglein MD, Shindo K, Albin JS, Brown WL, Harris RS. Quantitative profiling of the full APOBEC3 mRNA repertoire in lymphocytes and tissues: implications for HIV-1 restriction. Nucleic acids research. 2010;38(13):4274-84.
124. Duggal NK, Fu W, Akey JM, Emerman M. Identification and antiviral activity of common polymorphisms in the APOBEC3 locus in human populations. Virology. 2013;443(2):329-37.
125. Kitamura S, Ode H, Nakashima M, Imahashi M, Naganawa Y, Kurosawa T, et al. The APOBEC3C crystal structure and the interface for HIV-1 Vif binding. Nature structural & molecular biology. 2012;19(10):1005-10.
126. Berniell-Lee G, Calafell F, Bosch E, Heyer E, Sica L, Mouguiama-Daouda P, et al. Genetic and demographic implications of the Bantu expansion: insights from human paternal lineages. Molecular biology and evolution. 2009;26(7):1581-9.
127. Li S, Schlebusch C, Jakobsson M. Genetic variation reveals large-scale population expansion and migration during the expansion of Bantu-speaking peoples. Proc Biol Sci. 2014;281(1793).
128. Prado-Martinez J, Sudmant PH, Kidd JM, Li H, Kelley JL, Lorente-Galdos B, et al. Great ape genetic diversity and population history. Nature. 2013;499(7459):471-5.
129. Scally A, Dutheil JY, Hillier LW, Jordan GE, Goodhead I, Herrero J, et al. Insights into hominid evolution from the gorilla genome sequence. Nature. 2012;483(7388):169-75.
130. Langergraber KE, Prufer K, Rowney C, Boesch C, Crockford C, Fawcett K, et al. Generation times in wild chimpanzees and gorillas suggest earlier divergence times in great ape and human evolution. Proceedings of the National Academy of Sciences of the United States of America. 2012;109(39):15716-21.
131. Huthoff H, Malim MH. Identification of amino acid residues in APOBEC3G required for regulation by human immunodeficiency virus type 1 Vif and Virion encapsidation. Journal of virology. 2007;81(8):3807-15.
132. Satta Y, Takahata N. Out of Africa with regional interbreeding? Modern human origins. Bioessays. 2002;24(10):871-5.
133. Burns MB, Temiz NA, Harris RS. Evidence for APOBEC3B mutagenesis in multiple human cancers. Nat Genet. 2013;45(9):977-83.
134. Albin JS, LaRue RS, Weaver JA, Brown WL, Shindo K, Harjes E, et al. A single amino acid in human APOBEC3F alters susceptibility to HIV-1 Vif. The Journal of biological chemistry. 2010;285(52):40785-92.
135. Cuevas JM, Geller R, Garijo R, Lopez-Aldeguer J, Sanjuan R. Extremely High Mutation Rate of HIV-1 In Vivo. PLoS Biol. 2015;13(9):e1002251.
136. Ho YC, Shan L, Hosmane NN, Wang J, Laskey SB, Rosenbloom DI, et al. Replication-competent noninduced proviruses in the latent reservoir increase barrier to HIV-1 cure. Cell. 2013;155(3):540-51.
137. Kourteva Y, De Pasquale M, Allos T, McMunn C, D'Aquila RT. APOBEC3G expression and hypermutation are inversely associated with human immunodeficiency virus type 1 (HIV-1) burden in vivo. Virology. 2012;430(1):1-9.
138. Feng Y, Baig TT, Love RP, Chelico L. Suppression of APOBEC3-mediated restriction of HIV-1 by Vif. Frontiers in microbiology. 2014;5:450.
139. Yu Q, Konig R, Pillai S, Chiles K, Kearney M, Palmer S, et al. Single-strand specificity of APOBEC3G accounts for minus-strand deamination of the HIV genome. Nature structural & molecular biology. 2004;11(5):435-42.

140. Mangeat B, Turelli P, Caron G, Friedli M, Perrin L, Trono D. Broad antiretroviral defence by human APOBEC3G through lethal editing of nascent reverse transcripts. Nature. 2003;424(6944):99-103.
141. Zhang H, Yang B, Pomerantz RJ, Zhang C, Arunachalam SC, Gao L. The cytidine deaminase CEM15 induces hypermutation in newly synthesized HIV-1 DNA. Nature. 2003;424(6944):94-8.
142. Harris RS, Bishop KN, Sheehy AM, Craig HM, Petersen-Mahrt SK, Watt IN, et al. DNA deamination mediates innate immunity to retroviral infection. Cell. 2003;113(6):803-9.
143. Li J, Chen Y, Li M, Carpenter MA, McDougle RM, Luengas EM, et al. APOBEC3 multimerization correlates with HIV-1 packaging and restriction activity in living cells. Journal of molecular biology. 2014;426(6):1296-307.