Determination of the Endogenous HIV-Restrictive APOBEC3 Repertoire

A DISSERTATION SUBMITTED TO THE FACULTY

OF THE UNIVERSITY OF MINNESOTA BY

Eric W. Refsland

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

Advisor: Reuben S. Harris

August 2014

ACKNOWLEDGEMENTS

I am indebted to the following:

My thesis advisor, Reuben Harris, whose energy and enthusiasm for his research program, is infectious…in a good way.

Every Harris lab member, current and former, for fostering a light-hearted, yet competitive, environment that brings out the best.

My thesis committee members, Eric Hendrickson, Dave Bernlohr, Lou Mansky, and Harry Orr for your patience guidance and career advice.

My parents and brothers whose support and interest has never wavered.

My wife and daughters whose love keeps me grounded in life.

DEDICATION

This work is dedicated to my wife Rhena and my daughters Ada and Zelda for their unwavering support and endless patience.

ABSTRACT

Despite over 3 decades of intense research, an estimated 35.3 million people today are living with HIV-1 (Human Immunodeficiency Virus 1), the etiological agent of AIDS (Acquired Immunodeficiency Syndrome). The good news is that deaths are declining, in large part to combinations of highly active antiretroviral drugs. However, all of the current treatment options do not amount to a cure and a vaccine to prevent future infections is not yet available. With nearly 2 new infections for every 1 patient beginning an antiretroviral treatment regimen, the need is as great as ever to explore every avenue for new therapeutic interventions.

One of these avenues is harnessing proteins endogenous to the very cells targeted for infection by HIV-1. Collectively referred to as host restriction factors, these innate immune proteins share a number of characteristics. First, they decrease virus replication. Second, expression of host restriction factors is often inducible and coupled to the immediate innate immune response to a foreign attack. Third, under the immense pressure of adapting to a pathogenic agent, host restriction factors often exhibit high degrees of rapid evolutionary change. Finally, if a restriction factor is potent enough, the virus will have evolved a counter-restriction mechanism.

My thesis research has focused on the APOBEC3 family of host restriction factors that constitute an important arm of innate immunity. These enzymes, armed with the capacity to trigger cytosine to uracil mutagenesis in single-stranded DNA, function to defend the genome against attacks from foreign DNA elements including the retrovirus HIV-1. A more comprehensive understanding of the entire APOBEC3 family, its cellular functions, and how HIV-1 counteracts these activities with its accessory protein Vif, will

all be essential before the design of therapeutic approaches incorporating endogenous APOBEC3 proteins in the fight against HIV/AIDS can be achieved.

All 7 APOBEC3 proteins have been implicated in restricting HIV-1. The overall objective of my thesis research, detailed in the following chapters, was to define which of the 7 APOBEC3 proteins are capable of restricting the replication of HIV-1 in a T cell and which are potentially the source of the high levels of G-to-A mutagenesis observed in patient-derived HIV-1 sequences. To begin to narrow down which proteins are involved *in vivo*, we solved a long-standing problem in the APOBEC3 field. Quantifying expression levels had been impeded by the degree of homology between APOBEC3s, both at the nucleic acid and protein levels. My colleagues and I developed and characterized specific and efficient expression assays for each APOBEC3 family member. Using these assays we demonstrate that multiple APOBEC3s were expressed in relevant cell types and tissues including the major target of HIV-1, CD4+ T lymphocytes.

To interrogate the contribution of individual APOBEC3s to HIV-1 restriction, we took a genetic approach and performed targeted deletion and knockdown experiments in a cell line that expresses multiple APOBEC3s. This approach allowed for a direct and definitive test of our hypothesis that multiple endogenous APOBEC3s restrict Vif-deficient HIV-1 replication. Based on HIV-1 replication kinetics and the levels of viral genome mutation, we concluded that 4 APOBEC3s are involved in HIV-1 restriction and that any future strategies employing the restrictive APOBEC3s will benefit from liberating all four proteins from Vif counteraction rather than any single one alone.

Finally, to explore the consequences of genetic variation in the *APOBEC3* locus on the HIV-1 restriction capacity of these proteins, we analyzed the expression levels and

activity of the 7 haplotypes of the most genetically diverse *APOBEC3*, *APOBEC3H*. Through a series of primary cell experiments and HIV-1 spreading infections, we found that a subset of *APOBEC3H* haplotypes produce proteins that are stably expressed and capable of restricting naturally occurring HIV-1 variants that haven't evolved or have lost the capacity to neutralize this APOBEC3. The ramifications of this variability in a human restriction factor coinciding with diversity in HIV-1 variants able to counteract it will be an exciting area of future research.

Overall, my research demonstrated that HIV-1 must contend with 4 APOBEC3 proteins to efficiently replicate in T cells. Expression of the APOBEC3s in human cells is widespread and inducible, and these host restriction factors combine to mutagenize the viral genome in the absence of HIV-1 Vif or, in the case of APOBEC3H, with Vif-proficient HIV-1 variants that are unable to mount an effective counter-defense. Consequently, unleashing all of these potent antiviral agents and allowing them to directly attack the virus' genetic code may lead to the first targeted innate immune therapy against HIV-1.

TABLE OF CONTENTS

Acknowledgements ..... **i**

Dedication ..... **ii**

Abstract ..... **iii**

Table of Contents ..... **v**

List of Tables ..... **vii**

List of Figures ..... **ix**

Chapter 1: Introduction: The APOBEC3 Family of Retroelement Restriction Factors **1**

DNA Deaminase Evolution **4**

Biochemical and Structural Insights **8**

Biological Functions **13**

Pathological Consequences **21**

Possible Avenues to APOBEC3-Based Therapeutics **23**

Additional Contributions **26**

Figures **27**

Chapter 2: Quantitative Profiling of the Full *APOBEC3* mRNA Repertoire in Lymphocytes and Tissues: Implications for HIV-1 Restriction **33**

Introduction **36**

Results **38**

Discussion **45**

Materials & Methods **51**

Additional Contributions **56**

Figures **57**

Chapter 3: Endogenous Origins of HIV-1 G-to-A Hypermutation and Restriction in the Nonpermissive T Cell Line CEM2n **71**

Introduction **74**

Results **77**

Discussion **86**

Materials & Methods **89**

Additional Contributions 95
Figures 96
Chapter 4: Stable APOBEC3H Inhibits HIV-1 Replication in Primary T Lymphocytes 116
Introduction 119
Results 123
Discussion 132
Materials & Methods 135
Additional Contributions 141
Figures 142
Chapter 5: Conclusions and Discussion 157
APOBEC3 expression 160
The HIV-Restrictive APOBEC3 Repertoire 162
Bibliography 169

List of Tables

CHAPTER 2

Table 2-1. *APOBEC3* qPCR primers and probes **.....59**

Table 2-S1. *APOBEC3* qPCR primer specificities**.....70**

CHAPTER 3

Table 3-1. Gene targeting statistics in CEM2n **.....96**

Table 3-2. Mutation summary**.....100**

Table 3-S1. Primer sequences **.....107**

CHAPTER 4

Table 4-S1. Selected Donor Genotyping Results **.....148**

List of Figures

CHAPTER 1

Figure 1-1. Evolution and structure-function of the APOBEC deaminases **.....27**

Figure 1-2. The physiological functions of the APOBEC family**.....28**

Figure 1-3. APOBEC expression and model of carcinogenesis **.....30**

Figure 1-4. Potential therapies harnessing APOBEC3 innate immunity**.....31**

CHAPTER 2

Figure 2-1. A panel of quantitative PCR assays to monitor *APOBEC3* mRNA levels **....57**

Figure 2-2. *APOBEC3* expression in human T-cell lines and naïve PBMCs **.....60**

Figure 2-3. *APOBEC3* expression in naïve and stimulated CD4+ lymphocytes**.....62**

Figure 2-4. IFN induces *APOBEC3* expression**.....64**

Figure 2-5. *APOBEC3* expression in human tissues**.....66**

Figure 2-6. APOBEC3 immunoblots**.....67**

Figure 2-S1. Formamide agarose gel analysis **.....68**

Figure 2-S2. *APOBEC3* expression using different priming methods**.....69**

CHAPTER 3

Figure 3-1. CEM2n is a near-diploid, non-permissive T cell line **.....96**

Figure 3-2. Construction and characterization of *A3G*-null CEM2n cells**.....97**

Figure 3-3. Construction and characterization of *A3F*-null CEM2n cells **.....101**

Figure 3-4. Construction and characterization of *A3F*-null/*A3*-knockdown CEM2n cells **.....103**

Figure 3-5. Characterization of independent knockout and knockdown clones**.....105**

Figure 3-S1. CEM is a near-tetraploid T cell line**.....108**

Figure 3-S2. A comparison of *A3* mRNA levels in CEM2n, CEM(4n), and primary CD4+ lymphocytes**.....109**

Figure 3-S3. PCR reactions to detect *A3G* targeting events **.....110**

Figure 3-S4. G-to-A mutation loads for Vif-deficient HIV produced in CEM2n and key derivatives **111**
Figure 3-S5. PCR reactions to detect *A3F* targeting events **113**
Figure 3-S6. Vif-Proficient and Vif-Deficient HIV replication kinetics **114**

CHAPTER 4

Figure 4-1. Endogenous APOBEC3H stability/instability occurs at the protein level **141**
Figure 4-2. APOBEC3H adaptation studies **143**
Figure 4-3. Generation and validation of HIV-1 Vif separation-of-function molecular/viral probes **144**
Figure 4-4. Stable APOBEC3H inhibits HIV-1 replication in primary T lymphocytes and inflicts GA-to-AA hypermutations **146**
Figure 4-5. Correlations between the global distributions of HIV-1 hyper-Vif alleles and human A3H haplotypes **148**
Figure 4-6. Vif separation-of-function substitutions define the likely APOBEC3H interaction surface **149**
Figure 4-S1. Characterization of mouse monoclonal antibodies specific to human APOBEC3H **152**
Figure 4-S2. Stable APOBEC3H inhibits HIV replication in primary T lymphocytes **154**
Figure 4-S3. Stable APOBEC3H alleles inflict GA-to-AA hypermutations in viruses encoding hypo-Vif variants **156**

CHAPTER 1

Introduction - The APOBEC3 Family of Retroelement Restriction Factors

Reprinted with permission from: Eric W. Refsland and Reuben S. Harris (2013) Current Topics in Microbiology and Immunology 371:1-27

Authors' Contributions

EWR and RSH both contributed to the writing, editing, and figure preparation.

FOREWORD

The discovery of the anti-viral activities of the APOBEC3 proteins a little over 10 years ago is regarded as a major breakthrough for the development of endogenous therapeutic approaches to treat HIV-1 infected individuals. Members of this DNA cytosine deaminase family are potent anti-viral proteins because they directly inhibit reverse transcription, both physically and through their hallmark C-to-U enzymatic activity of nascent viral replication intermediates that can result lethal mutagenesis. However, the HIV-1 counterdefense protein, Vif, limits these activities by interacting with APOBEC3 proteins and marking them for cellular degradation. My research details which of the 7 endogenous APOBEC3 proteins are relevant to HIV-1 restriction with an eventual goal of liberating these factors from Vif to control the HIV-1 pandemic. Here, I briefly describe the origins of the family of APOBEC3 innate immune proteins, their beneficial functions in antiviral defenses, deleterious potential, and potential approaches to leverage their potent activities as the next generation of antiviral therapies thereby harnessing an innate immune defense to combat a contemporary foe.

SUMMARY

The ability to regulate and even target mutagenesis is an extremely valuable cellular asset. Enzyme-catalyzed DNA cytosine deamination is a molecular strategy employed by vertebrates to promote antibody diversity and defend against foreign nucleic acids. Ten years ago, a family of cellular enzymes was first described with several proving capable of deaminating DNA and inhibiting HIV-1 replication. Ensuing studies on the apolipoprotein B mRNA-editing enzyme catalytic polypeptide-like 3 (APOBEC3) restriction factors have uncovered a broad-spectrum innate defense network that suppresses the replication of numerous endogenous and exogenous DNA-based parasites. Although many viruses possess equally elaborate counter-defense mechanisms, the APOBEC3 enzymes offer a tantalizing possibility of leveraging innate immunity to fend off viral infection. Here we focus on mechanisms of retroelement restriction by the APOBEC3 family of restriction enzymes and we consider the therapeutic benefits, as well as the possible pathological consequences, of arming cells with active DNA deaminases.

DNA DEAMINASE EVOLUTION

Central to nucleic acid metabolism is the near-ubiquitous process of enzymatic deamination of adenine and cytosine bases, individually or in the context of larger nucleic acid constituents (Conticello et al., 2005; Grosjean, 2009). For instance, in most species the wobble base in several of the tRNA anti-codons is frequently changed by deamination of adenosine to inosine (A-to-I), which can then base pair with cytosine, uracil, or adenosine, thereby increasing the flexibility and decoding capacity of a tRNA anti-codon (Gerber and Keller, 1999). These enzymes belong to the adenosine deaminase acting on tRNA (ADAT) family. Related proteins in most metazoans from nematodes and flies to humans catalyze A-to-I editing of a variety of RNA targets (Kim et al., 1994; Nishikura, 2010). These enzymes are called, appropriately, adenosine deaminases that act on RNA (ADAR). Editing events that occur in the coding region of an mRNA can result in amino acid substitutions in the resulting protein. However, the majority of editing events occur in non-coding regions of mRNA or in non-coding RNAs, and these A-to-I editing events can alter RNA secondary structure, stability, function, and/or capacity to be bound by regulatory RNAs such as siRNAs (Agranat et al., 2008; Levanon et al., 2004; Li et al., 2009; Morse et al., 2002).

Cytosine to uracil (C-to-U) deamination is almost as ancient as A-to-I editing (Conticello et al., 2007b; Grosjean, 2009). The pyrimidine salvage pathways of most organisms use cytidine deaminase (CDA) to produce the essential RNA and DNA building blocks of uridine directly, and thymidine after additional enzymatic steps (Zrenner et al., 2006). However, at some point near the root of the vertebrate tree, polynucleotide cytosine deaminases emerged, with the lamprey CDA being a present day

example (Rogozin et al., 2007). This enzyme is thought to underpin a unique form of adaptive immunity in which the DNA segments that encode arrays of highly diverse leucine rich repeats are assembled into mature variable lymphocyte receptor genes by a recombination-mediated process. It is thought that an ancestor of the present day lamprey enzyme served as the original substrate for expansion of the polynucleotide cytosine deaminase gene family during vertebrate evolution (**Figure 1-1a**) (Conticello, 2008; Rogozin et al., 2007). The result in most vertebrates alive today is a much larger repertoire of polynucleotide C-to-U editing enzymes that execute diverse biological functions from lipid metabolism to adaptive and innate immunity (**Figure 1-1 & 1-2**).

All vertebrate polynucleotide cytosine deaminases belong to the so-called 'APOBEC' family. The defining feature of this family is a conserved His-X-Glu-X_{25-31}-Pro-Cys-X_{2-4}-Cys zinc (Z)-coordinating motif, which is strictly required for deaminase activity (where X can be a variety of amino acids) (Conticello et al., 2005; Harris and Liddament, 2004; LaRue et al., 2009; Wedekind et al., 2003). As described in more detail below, key residues within this motif position zinc at the active site of the enzyme (**Figure 1-1c**). The protein sequences within these motifs enable phylogenetic groupings into three sub-families: APOBEC1, AID, and the APOBEC3s. The APOBEC3s can be further sub-divided into three sub-groups: Z1, Z2, and Z3 (**Figure 1-1a**) (Conticello, 2008; LaRue et al., 2009). Importantly, the number and organization of the A3 Z-domains can vary dramatically from branch to branch throughout the mammalian portion of the vertebrate phylogenetic tree (*e.g.*, human versus mouse loci depicted in **Figure 1-1a**).

Apolipoprotein B mRNA editing catalytic subunit 1 (APOBEC1) has provided the namesake to the larger family. It was discovered as an enzyme that catalyzes the deamination of a specific cytosine within the *APOB* mRNA (**Figure 1-1b & 1-2a**) (Teng et al., 1993). This produces a premature translation stop codon and a smaller secondary gene product. These two APOB proteins (APOB100 and APOB48) differentially regulate the secretion of lipoproteins from the liver (Chan, 1992). Many mammalian APOBEC1 enzymes also possess DNA C-to-U deaminase activity (Harris et al., 2002; Ikeda et al., 2011; Ikeda et al., 2008; Petit et al., 2009). Taken together with the fact that earlier vertebrate lineages, such as the one represented by birds and lizards, lack an *APOB*-like gene, it is probable that DNA editing function preceded involvement in RNA editing (Severi et al., 2011).

The most conserved DNA cytosine deaminase in vertebrates is activation-induced deaminase (AID; gene name *AICDA*) (**Figure 1-1a**). AID has a central role in adaptive immunity by seeding somatic hypermutation, gene conversion, and class switch recombination with its DNA deaminase activity (**Figures 1-1b & 1-2b**) [(Di Noia and Neuberger, 2002; Muramatsu et al., 2000; Muramatsu et al., 1999; Petersen-Mahrt et al., 2002); reviewed by (Conticello, 2008; Di Noia and Neuberger, 2007; Longerich et al., 2006)]. Interestingly, the genes that encode APOBEC1 and AID are positioned adjacent to each other in the genomes of most vertebrates (an inversion has placed the human gene farther away on the same chromosome). This suggests that an ancestral *AID* gene duplicated and diverged to produce *APOBEC1* (**Figure 1-1a**). It is likely that duplication of an ancestral *AID*/*APOBEC1* locus produced the genetic seeds for the mammal-

exclusive *APOBEC3* sub-family (**Figure 1-1a**) (Harris and Liddament, 2004; Jarmuz et al., 2002).

In humans, the seven APOBEC3 proteins are encoded by a tandemly arranged gene cluster (**Figure 1-1a**) (Jarmuz et al., 2002). These present day genes are the products of continual evolution, in which an ancestral cluster of three Z domains is predicted to have undergone a minimum of eight duplication events over the past 100 million years to produce the locus found in most primates (LaRue et al., 2008; Münk et al., 2012). These domains are either expressed singly or one enzyme may consist of two Z domains (LaRue et al., 2009). In contrast, the ancestral *APOBEC3* locus experienced a deletion in the rodent lineage of one of the ancestral Z-domains leading to the present day two domain locus, which encodes a single protein quite distinct from any of the primate enzymes (**Figure 1-1a**).

One possible explanation for why some mammalian lineages, like primates, have many APOBEC3s, while other lineages, such as rodents have few is that these enzymes have overlapping innate immune functions to protect the host from a variety of parasitic elements (*e.g.*, in HIV-1 restriction, **Figure 1-2c**; mechanism elaborated in section 3, below). Because multiple distinct innate immune mechanisms serve to suppress the spread of such parasitic elements, it is reasonable to postulate that some mammals will be fortified at the *APOBEC3* locus and weaker at other loci, with each mammalian lineage being distinct. For instance, primates encode a single TRIM5α protein, whereas mice have the capacity to encode a total of eight TRIM5α-like proteins [(Sawyer et al., 2007; Tareen et al., 2009); see also Chapter 13 in (Lever et al., 2010)]. It is likely that each species' present day innate immune fortifications were independently shaped by past

pathogenic pressures, which one can only speculate may have been the ancestors of present day viruses and transposable elements.

BIOCHEMICAL AND STRUCTURAL INSIGHTS

Zinc-dependent deaminases, such as the APOBECs, catalyze the conversion of cytosine to uracil in polynucleotide substrates (**Figure 1-1b**). This reaction requires the activation of water by a zinc ion coordinated by the enzyme (**Figure 1-1c**). A glutamic acid in the active site of the enzyme protonates N3, priming the nucleophilic attack on the C4 position of the pyrimidine ring, followed by the removal and subsequent protonation of an amino group (NH_2) that results in the release of ammonia (NH_3) and uracil as products. This conversion can theoretically occur within both RNA and single-stranded DNA substrates. However, apart from APOBEC1, which has both RNA and DNA editing activities, AID and the APOBEC3s have proven specific to DNA substrates *in vitro* and *in vivo*.

The extent of amino acid homology to APOBEC1 originally suggested that the APOBEC3 enzymes might be a family of RNA editing proteins (Jarmuz et al., 2002). Three lines of evidence, however, demonstrated that this view was incorrect and established the APOBEC3 enzymes as single-stranded DNA cytosine deaminases. First, APOBEC3 has a high degree of homology with AID and experiments in *E. coli* demonstrated AID, APOBEC3C, and APOBEC3G are capable of inducing high levels of mutation in an antibiotic resistance gene (Harris et al., 2002; Petersen-Mahrt et al., 2002). This was clearly due to DNA editing because mutation levels rose synergistically in a

bacterial strain deficient for uracil DNA glycosylase (UDG), an enzyme that initiates base excision repair by recognizing and removing uracil exclusively from DNA (Di Noia and Neuberger, 2002; Harris et al., 2002; Lindahl, 2000). Second, unambiguous evidence for DNA versus RNA editing comes from head-to-head biochemical studies using recombinant enzymes. AID and APOBEC3G, have a strong preference for single-stranded DNA substrates, with no detectable RNA editing activity (Bransteitter et al., 2003; Iwatani et al., 2006). Third, a strong preference for single-stranded DNA substrates is also evident in sequencing studies of retroviruses produced in the presence of a given APOBEC3 protein, such as APOBEC3G (**Figure 1-2c**) (Harris et al., 2003; Lecossier et al., 2003; Mangeat et al., 2003; Zhang et al., 2003). In this experimental system, each APOBEC3 protein presumably has a chance to deaminate viral genomic RNA cytosines before the reverse transcription process converts it to a single-stranded cDNA intermediate and then to the double-stranded DNA required for integration. However, the most common APOBEC3-dependent mutations detected in integrated viral DNA that has survived this process are genomic strand G-to-A mutations, entirely attributable to cDNA strand C-to-U deamination events. Genomic strand C-to-T editing events possibly due to RNA editing are rarely detected. Importantly, APOBEC3 DNA editing activity is required to explain previously reported G-to-A mutation biases in HIV-1 substrates *in vivo* (Janini et al., 2001; Vartanian et al., 1994).

The solved structures of bacterial and yeast cytidine and cytosine deaminases were used to inform early functional and structural studies of various APOBEC3 family members (Betts et al., 1994; Ireton et al., 2003; Johansson et al., 2004; Ko et al., 2003; Xie et al., 2004). Each of these bacterial and yeast proteins, in monomeric form, is

globular with a hydrophobic β-stranded core and several surrounding a-helices. The most conserved structural feature is the active site, which is defined by a histidine and two cysteines in the yeast enzyme and three cysteines in the bacterial enzymes (Ireton et al., 2003; Ko et al., 2003; Xiang et al., 1997). In both instances, these residues are positioned similarly by alpha helices and they serve to coordinate a zinc ion in the active site, which, as described above, is essential for the deamination reaction (**Figure 1-1b**). Although these conserved features have been useful for generating models of APOBEC3 structures, they have also been misleading because the oligomeric state of each enzyme is variable. For instance, the *E. coli* CDA is homodimeric and the yeast enzyme is homotetrameric (Betts et al., 1994; Johansson et al., 2002). This has fuelled (likely incorrect) speculation that APOBEC3 family members must also function as oligomers.

Generating high-resolution structures of APOBEC3 family members has proved challenging in large part due to insolubility at higher protein concentrations [*e.g.*, (Iwatani et al., 2006)]. However, several NMR and crystal structures have been achieved for the APOBEC3G catalytic domain (representing Z1-type deaminases) and crystal structures were obtained recently for APOBEC3C and the APOBEC3F catalytic domain (representing Z2-type deaminases) (*e.g.*, **Figure 1-1c**) (Bohn, 2013; Chen et al., 2008; Furukawa et al., 2009; Holden et al., 2008; Kitamura et al., 2012b; Li et al., 2012; Shandilya et al., 2010).

These structures have several conserved features that provide insight into how these enzymes may function. First, these proteins are all globular with a hydrophobic core consisting of five beta strands surrounded by six alpha helices and the hallmark α2-β3-α3 zinc-coordinating motif that defines the larger cytosine deaminase superfamily.

Second, β-strands 3, 4, and 5 are arranged in parallel, similar to the RNA editing enzyme TadA (an ADAT) but different from the antiparallel arrangement found in bacterial and yeast CDAs. This parallel β3-β4-β5 organization may be a key feature that distinguishes polynucleotide from non-polynucleotide deaminases. Third, although many potential oligomeric interfaces have been captured in the crystal lattices, none have proven critical for enzymatic activity and no common themes have emerged (Bohn, 2013; Furukawa et al., 2009; Kitamura et al., 2012b; Shandilya et al., 2010). This is consistent a number of other studies indicating that oligomerization may not be essential for binding and deaminating single-stranded DNA substrates (Nowarski et al., 2008; Opi et al., 2006; Shlyakhtenko et al., 2011; Shlyakhtenko et al., 2012). However, more work on this topic is clearly needed to define the role of oligomerization *in vivo*, because several of the family members, including APOBEC3G, elicit such a property in living cells (Bransteitter et al., 2003; Chiu et al., 2006; Li et al., 2014; Soros et al., 2007). Finally, it is notable that the majority of structural and amino acid differences between APOBEC3 structures are confined to non-catalytic loop regions. Such differences likely relate to substrate targeting and possible co-factor binding, ultimately reflecting physiological function.

A significant remaining question in our understanding of APOBEC3 function is how these enzymes bind single-stranded DNA substrates. A current working model proposes a positively charged brim in the region surrounding the active site consisting of R213, R215, R313, and R320 in APOBEC3G (Chen et al., 2008; Shindo et al., 2012). These residues are predicted to position single-stranded DNA substrates in a manner that allows the target cytosine to enter the active site (Chen et al., 2008). This model also

predicts that in order to access to the catalytic glutamic acid, the target C will be flipped-out with respect to the phosphodiester backbone. A base-flipping mechanism is in good agreement with the structure of the adenosine deaminase TadA complexed with its RNA substrate (Losey et al., 2006).

Finally, the brim-domain model and TadA structures suggest an explanation for the different local single-stranded DNA deamination preferences among APOBEC family members (Chen et al., 2008; Conticello et al., 2007a). Unlike bacterial restriction enzymes with 4, 6, or 8 base palindromic recognition sequences, APOBEC3 family members have a notable preference for the base immediately 5' of the target C (Aguiar et al., 2008; Bishop et al., 2004; Dang et al., 2006; Doehle et al., 2005a; Harari et al., 2009; Harris et al., 2003; Harris et al., 2002; Langlois et al., 2005; Liddament et al., 2004; Mangeat et al., 2003; Stenglein et al., 2010; Wiegand et al., 2004; Yu et al., 2004a; Yu et al., 2004b; Zhang et al., 2003; Zheng et al., 2004). Specifically, AID prefers a 5' purine base (5'-AC or GC), APOBEC3G a 5' cytosine (5'-CC), and all other family members a 5' thymine (5'-TC). Several studies have recently converged on a loop adjacent to the active site, positioned between β4 and α4 secondary structural elements (Chen et al., 2008; Conticello et al., 2007b; Holden et al., 2008; Kohli et al., 2009; Kohli et al., 2010; Rathore et al., 2013). This is most dramatically evidenced by loop grafting experiments, in which this loop in AID can be replaced by the homologous loop from APOBEC3G or APOBEC3F resulting in a complete switch of the preferred base immediately 5' of the target cytosine (Carpenter et al., 2010; Kohli et al., 2009; Kohli et al., 2010; Wang et al., 2010). Moreover, exchanging the same loop (or even a single amino acid) between APOBEC3A and APOBEC3G completely swaps the dinucleotide preference of these

enzymes (Rathore et al., 2013). Despite this progress, the field still anxiously awaits high-resolution structures of enzyme-substrate complexes that will more precisely define the substrate binding mechanism and advance our understanding of how these enzymes function *in vivo*.

BIOLOGICAL FUNCTIONS

APOBEC3 Proteins in HIV Restriction

Pathogens including the retrovirus HIV-1 (hereafter HIV) must both engage and avoid numerous host factors to replicate and cause disease. Genome-wide knockdown and proteomic studies suggest that up to 10% of human proteins either directly or indirectly impact HIV replication (Brass et al., 2008; Jäger et al., 2012a; Jäger et al., 2012b; Konig et al., 2008; Yeung et al., 2009; Zhou et al., 2008). The majority of these proteins are required in some capacity for virus replication (*i.e*., dependency factors). In contrast, a small number of these cellular proteins are dominant proteins that directly suppress virus replication (*i.e*., restriction factors). Restriction factor hallmarks include the capacity to potently inhibit virus replication, signatures of rapid evolution (positive selection), responsiveness to interferon, and neutralization by at least one viral counter-restriction strategy (Harris et al., 2012; Malim and Bieniasz, 2012; Malim and Emerman, 2008b). Here, we focus on the mechanism of HIV restriction by APOBEC3 DNA cytosine deaminases.

Original studies showed that the viral infectivity factor (Vif) protein of HIV is required for virus replication in primary $CD4^+$ lymphocytes and in several common

laboratory T cell lines (*e.g.*, CEM, H9, so-called non-permissive), but it was dispensable in several others (*e.g.*, CEM-SS, SupT1, or permissive lines) (Fisher et al., 1987; Gabuzda et al., 1992; Strebel et al., 1987). This phenotypic difference led to the cloning of *APOBEC3G* as one cDNA sequence expressed differentially (of many) between CEM and CEM-SS (Sheehy et al., 2002). However, APOBEC3G proved remarkable as it could convert a permissive cell line to a non-permissive phenotype (Sheehy et al., 2002). Taken together with the independent and near-simultaneous discoveries of APOBEC3G as a putative RNA editing factor and as a DNA editing enzyme, an editing mechanism of restriction was predicted and shortly after demonstrated (Harris et al., 2003; Harris et al., 2002; Jarmuz et al., 2002; Lecossier et al., 2003; Mangeat et al., 2003; Zhang et al., 2003).

Over five hundred papers have now been published on APOBEC3G and the related APOBEC3 proteins that have culminated in a current Trojan horse-like model for HIV restriction as shown in **Figure 1-2c**. To be effective as an HIV restriction factor, APOBEC3G must first be expressed in the cytoplasmic compartment of an infected virus-producing cell (Mangeat et al., 2003). Second, cytoplasmic APOBEC3G is thought to interact with a Gag-ribonucleoprotein complex, and this interaction is required for APOBEC3G packaging (Alce and Popik, 2004; Bogerd and Cullen, 2008; Burnett and Spearman, 2007; Khan et al., 2005; Luo et al., 2004; Schafer et al., 2004; Svarovskaia et al., 2004). This interaction can be disrupted by RNase treatment and is therefore thought to involve a bridging RNA between APOBEC3G and the nucleocapsid region of Gag (Bogerd and Cullen, 2008; Cen et al., 2004; Iwatani et al., 2007; Svarovskaia et al., 2004). The identity of the bridging RNA is still an active area of investigation; while

several reports indicate a role for the Alu-like 7SL RNA, a role for viral genomic RNA has not been excluded (Bogerd and Cullen, 2008; Khan et al., 2007; Khan et al., 2005; Tian et al., 2007; Wang et al., 2007). Third, by an ill-defined mechanism, packaged APOBEC3G must breach the nucleocapsid core of the viral particle. This is a genetically (but not mechanistically) defined step, as various chimeric constructs have the capacity to be packaged into viral particles, yet do not enter the core or restrict (Haché et al., 2005; Martin et al., 2011; Song et al., 2012).

Once an APOBEC3G-loaded viral core is deposited into a target cell, reverse transcription proceeds and a susceptible single-stranded cDNA intermediate is generated. Here, APOBEC3G deaminates C-to-U (Harris et al., 2003; Lecossier et al., 2003; Mangeat et al., 2003; Yu et al., 2004b; Zhang et al., 2003). The uracilated cDNA is either subjected to degradation, or templates the second-strand synthesis, preceding integration into the target cell's genome (Kaiser and Emerman, 2006; Mbisa et al., 2007; Yang et al., 2007). The uracils in the cDNA strand template the insertion of adenines in the nascent plus strand and thereby immortalize G-to-A mutations that limit subsequent rounds of viral replication (Harris et al., 2003; Lecossier et al., 2003; Mangeat et al., 2003; Zhang et al., 2003).

Human T cell lines that express near-physiological levels of an APOBEC3G catalytic mutant do not suppress the replication of a Vif-deficient virus, indicating that the predominant mechanism of HIV restriction depends upon deaminase activity (Bishop et al., 2004; Browne et al., 2009; Miyagi et al., 2007; Schumacher et al., 2008). However, a component of APOBEC3G's capacity to restrict HIV may be deaminase-independent (Bishop et al., 2006; Bishop et al., 2008; Iwatani et al., 2007; Iwatani et al., 2006; Mbisa

et al., 2007; Newman et al., 2005; Opi et al., 2006). The most convincing of these studies has shown that APOBEC3G is capable of binding viral genomic RNA and sterically hindering the processivity of reverse transcriptase (Bishop et al., 2006; Bishop et al., 2008). Nevertheless, the importance of deaminase-independent mechanisms, at least for APOBEC3G and HIV, is questionable given the aforementioned results at physiologic levels.

Soon after the initial discovery and functional characterization of APOBEC3G, attention turned towards its six, highly related family members. Proviral sequences isolated from HIV-infected individuals exhibit two distinct patterns of G-to-A mutation consistent with APOBEC3-mediated deamination, both 5'-GG-to-AG and 5'-GA-to-AA (Caride et al., 2002; Fitzgibbon et al., 1993; Gandhi et al., 2008; Janini et al., 2001; Kieffer et al., 2005; Land et al., 2008; Pace et al., 2006). As mentioned previously, APOBEC3G has a demonstrated preference for deaminating 5'-CC dinucleotides on the viral minus strand resulting in 5'-GG-to-AG mutations and is therefore the most likely source of this mutational pattern (Harris et al., 2003; Liddament et al., 2004; Mangeat et al., 2003; Yu et al., 2004b; Zhang et al., 2003). However, high levels of proviral 5'-GA-to-AA mutation in patients' proviral sequences suggested the involvement of at least one additional family member. Unfortunately, all six other APOBEC3 family members prefer to deaminate a 5'-TC, therefore excluding the possibility of identifying an additional APOBEC3 source based solely on mutational preference [*e.g.*, (Bishop et al., 2004); reviewed in (Albin and Harris, 2010)].

In cell-based model systems with forced cDNA expression, all seven APOBEC3s have reported activity [*e.g.*, (Bishop et al., 2004; Dang et al., 2008; Dang et al., 2006;

Doehle et al., 2005a; Goila-Gaur et al., 2007; Liddament et al., 2004; OhAinle et al., 2008; Rose et al., 2005; Wiegand et al., 2004; Yu et al., 2004a; Zheng et al., 2004); reviewed in (Albin and Harris, 2010)]. However, cellular expression patterns, ability to encapsidate into viral particles, and restriction in relevant T cell models indicate not all the APOBEC3s function *in vivo* as HIV restriction factors. For example, in primary $CD4^+$ T lymphocytes, a major target of HIV *in vivo*, six APOBEC3s are expressed at appreciable levels, but *APOBEC3A* mRNA is undetectable (Koning et al., 2009; Refsland et al., 2010). In addition, only four: APOBEC3D, APOBEC3F, APOBEC3G, and APOBEC3H (but not APOBEC3A, APOBEC3B, and APOBEC3C), package into viral-like particles and inhibit viral replication when stably expressed in human T cell lines (Hultquist et al., 2011). Finally, endogenous APOBEC3D and APOBEC3F combine to inflict the 5'-GA-to-AA mutation pattern observed in the non-permissive T cell line CEM2n (Refsland et al., 2012). Of note, CEM2n does not express appreciable levels of *APOBEC3H* mRNA; therefore the endogenous contribution of this protein could not be evaluated. Seven variants of APOBEC3H have been reported and cell-based studies indicate only a subset of these haplotypes are stable at the protein level and capable of HIV restriction (Dang et al., 2008; Harari et al., 2009; OhAinle et al., 2008; Wang et al., 2011). The significance of this natural variation in human populations and the relevance of it to HIV restriction require further investigation. Taken together, while all of the APOBEC3s can be compelled to deaminate a single-stranded viral DNA *in vitro* or in heterologous cell lines by gross overexpression, only four APOBEC3s — APOBEC3D, APOBEC3F, APOBEC3G, and APOBEC3H — appear to have the capacity to function as HIV restriction factors in T lymphocytes.

All successful viruses have evolved sophisticated immune suppression and/or evasion mechanisms to neutralize the cellular defense systems they must face. One of the best-studied APOBEC3 counter-restriction mechanisms is orchestrated by the HIV viral infectivity factor, Vif. This protein is small (23 kDa), highly basic, and strictly required for pathogenesis *in vivo* as well as for virus replication *ex vivo* in monocytes, macrophages, primary $CD4^+$ T lymphocytes, and non-permissive T cell lines (Fisher et al., 1987; Gabuzda et al., 1992; Kan et al., 1986; Lee et al., 1986; Sodroski et al., 1986; Strebel et al., 1987; von Schwedler et al., 1993). Vif recruits a multiprotein E3 ubiquitin ligase complex consisting of CUL5/NEDD8, ELOB, ELOC, RBX2 and CBFb to mediate proteasomal degradation of the cellular APOBEC3 restriction factors (**Figure 1-2c**) (Conticello et al., 2003; Jäger et al., 2012b; Marin et al., 2003; Mehle et al., 2004; Sheehy et al., 2003; Yu et al., 2003). In addition to this primary role, it has been suggested that Vif can relieve APOBEC3G-mediated restriction by alternative mechanisms such as directly blocking deaminase activity, preventing encapsidation, sequestering the protein in catalytically inactive conformations, or impeding APOBEC3G translation (Britan-Rosich et al., 2011; Goila-Gaur et al., 2008; Kao et al., 2004; Mariani et al., 2003b; Opi et al., 2007; Santa-Marta et al., 2005; Stopak et al., 2003). The extent to which these mechanisms function to counteract APOBEC3 during a productive infection warrants further investigation but is certain to be secondary to the proteasomal degradation mechanism.

APOBEC3 proteins in general innate immune defense

Exogenous retroviruses are rare in humans, consistent with the idea that the multifaceted APOBEC3 defense provides robust protection against this type of pathogen. HIV is one exception to this rule. It is successful, at least in part, due to Vif and degradation of the cellular APOBEC3s. However, evidence for a long history of positive selection acting on the human *APOBEC3* locus suggests this family has been defending the genomic integrity of its host's cells long before HIV was transmitted into the human population (Sanville et al., 2010; Sawyer et al., 2004; Zhang and Webb, 2004). Given the significant presence of endogenous retroelements in the genomes of mammals (nearly 50%), suppression of these elements may represent the sub-family's true *raison d'être* and the primary source of the positive selection that has shaped the complex present day locus and the diverse functions of these innate immunity factors.

Endogenous retroelements, including those containing long terminal repeats (LTR) like endogenous retroviruses, as well as non-LTR elements like long interspersed nuclear elements (LINEs) and short interspersed nuclear elements (SINEs), may have provided the evolutionary pressure necessary for the maintained expansion of the *APOBEC3* locus in primates. Additionally, the differences in retrotransposition frequency between rodents and primates could be attributable to possessing an arsenal of seven APOBEC3 proteins versus only a single one (Maksakova et al., 2006; Stenglein and Harris, 2006). In support of this hypothesis, human APOBEC3s have demonstrated activity on LTR retrotransposons from mice and yeast (Bogerd et al., 2006a; Chen et al., 2006; Dutko et al., 2005; Esnault et al., 2005; Esnault et al., 2006; Jern and Coffin, 2008; Lee and Bieniasz, 2007; Lee et al., 2008; Schumacher et al., 2005). Additionally, some

endogenous retroviruses exhibit the characteristic scars from APOBEC3F and APOBEC3G activity in their genomes (Anwar et al., 2013). Non-LTR elements, including LINE1 and Alu, are also restricted by human APOBEC3s but, in contrast to the HIV restriction mechanism described above, this mechanism appears entirely deamination-independent (Bogerd et al., 2006b; Carmi et al., 2011; Chiu et al., 2006; Muckenfuss et al., 2006; Stenglein and Harris, 2006).

The generation of a single-stranded DNA intermediate in the life cycle of a parasitic element may render it susceptible to APOBEC3 restriction. In addition to HIV, other viruses including simian immunodeficiency virus, murine leukemia virus, foamy virus, porcine endogenous retrovirus, human T-cell leukemia virus, and hepatitis B virus have all been reported to be susceptible to APOBEC3-mediated editing of their genomes (Abudu et al., 2006; Delebecque et al., 2006; Doehle et al., 2005b; Harris et al., 2003; Jonsson et al., 2007; Kobayashi et al., 2004; Lochelt et al., 2005; Mangeat et al., 2003; Mariani et al., 2003b; Russell et al., 2005; Suspène et al., 2005a; Turelli et al., 2004; Yu et al., 2004a). Whether or not the APOBEC3-mediated restriction of all of these viral pathogens is part of a natural innate immune response or a by-product of the particular model systems used awaits further investigation.

Endogenous retroelements and their mobility are believed to have played a central role early in shaping the human genome during speciation (Carmi et al., 2011; Kazazian, 2004). However, this process has its associated costs. Ultimately, cells have devised strategies to defend and preserve genomic integrity by curbing the movement of these genetic elements. The evolution of the APOBEC3 family has likely played a prominent

role in this defense and in diversifying the retroelements to make them more useful for the host species [*e.g.*, (Carmi et al., 2011)].

PATHOLOGICAL CONSEQUENCES OF DNA DEAMINATION

Although DNA cytosine deamination has obvious benefits for individual cells and the organism as a whole, this process may have considerable pathological consequences, most notably cancer. For instance, overexpression of APOBEC1 in the liver of mice has been shown to cause hepatocellular carcinoma and liver dysplasia (Yamanaka et al., 1995). Likewise, AID transgenesis also causes cancer, and it has been implicated in initiating the chromosomal translocations responsible for some lymphocyte neoplasias including the hallmark c-myc/IgH rearrangement in Burkitt's lymphoma [(Okazaki et al., 2003; Ramiro et al., 2004; Unniraman et al., 2004); reviewed by (Perez-Duran et al., 2007)]. AID may also promote resistance to the chemotherapeutic drug imatinib (Klemm et al., 2009). However, the overall impact of APOBEC1 and AID on human cancer is questionable because their expression is largely limited to tissues associated with their biological functions, *APOBEC1* in the enterocytes of the small intestine and *AID* in B cells (**Figure 1-3a**) (Muramatsu et al., 2000; Muramatsu et al., 1999; Powell et al., 1987; Teng et al., 1993).

In contrast, most *APOBEC3s* have much broader expression ranges that span most tissues in the human body (**Figure 1-3a**) (Burns et al., 2013; Koning et al., 2009; Refsland et al., 2010). Broad expression profiles, potent DNA deaminase activity, and C-to-T transition biases in tumor genome sequences strongly suggested that one or more of

the APOBEC3 proteins may be a source of mutation in different cancers (Burns et al., 2013; Harris et al., 2002; Nik-Zainal et al., 2012; Roberts et al., 2012). APOBEC3B became a leading candidate as it uniquely and constitutively localizes to the nucleus by inheriting a nuclear import mechanism from AID (Bogerd et al., 2006b; Bonvin et al., 2006; Hultquist et al., 2011; Kinomoto et al., 2007; Lackey et al., 2012; Lackey et al., 2013; Muckenfuss et al., 2006; Pak et al., 2011; Stenglein and Harris, 2006; Stenglein et al., 2008). Recently, APOBEC3B was found overexpressed in some laboratory breast cancer cell lines, but not in available control cell lines (**Figure 1-3a**; *e.g.*, compare HCC1569, MCF-MB-453, and MCF-MB-463 to telomerase immortalized human mammary epithelial cells, hTERT HMECS) (Burns et al., 2013). APOBEC3B up-regulation was shown to be responsible for elevated genomic uracil levels and increased mutation rates in breast cancer cell lines (Burns et al., 2013). APOBEC3B up-regulation was similarly robust in the majority of human breast tumors, in contrast to barely detectable levels in normal breast tissue (Burns et al., 2013). Remarkably, APOBEC3B overexpression correlated with a doubling in the tumor genomic mutation loads, and the majority of C-to-T transition mutations occurred within the preferred motif of recombinant APOBEC3B (Burns et al., 2013).

Overall, a model is emerging for how APOBEC3B provides genetic fuel for tumorigenesis, which, coupled with selection, may help explain many hallmarks of cancer such as increased DNA damage, elevated proliferation, decreased apoptosis, and massive heterogeneity (**Fig. 3b**). In particular, *APOBEC3B* up-regulation correlates with inactivation of the tumor suppressor gene *TP53*, which strongly suggests that it may be an early tumor-initiating event (Burns et al., 2013). Obviously, the potential benefits to

encoding APOBEC3B must outweigh potential costs of carcinogenesis. An attractive explanation for this apparent conundrum may be that its innate immune function is important early in life and for the health of the species, for instance in germ cells or early development (Bogerd et al., 2006b; Wissing et al., 2011), whereas the toll of cancer is not imposed in most instances until after the reproductive years. In any event, much more work is now justified on APOBEC3B and its role in breast and, potentially, other human cancers.

POSSIBLE AVENUES TO APOBEC3-BASED THERAPEUTICS

Therapy by hypermutation

If left unimpeded by Vif, APOBEC3 proteins such as APOBEC3G can convert up to 10% of viral cDNA cytosines into uracils in a single-round of virus replication [*e.g.*, (Harris et al., 2003; Yu et al., 2004b)]. The resulting massive levels of G-to-A mutations effectively ruin the genetic potential of the retrovirus in a process called lethal mutagenesis (Haché et al., 2006; Loeb et al., 1999). Moreover, the preferred context of APOBEC3G deamination events often results in the conversion of the tryptophan codon TGG into a premature stop codon TAG, which is more detrimental to the virus than a simple amino acid change. Overall, physiological levels of APOBEC3 proteins largely, if not fully, suppress the replication of Vif-deficient HIV.

This remarkable potency therefore raises the prospect of developing drugs to leverage the APOBEC3 restriction mechanism against HIV (**Figure 1-4a**). Direct inhibition of Vif is certainly one strategy, but such an approach is destined to be

susceptible to problems imposed by extensive natural HIV variation and the rapid evolution of drug resistance. A more appealing alternative may be to develop a drug toward one of the more genetically stable cellular proteins recruited to the APOBEC3 degradation complex. In particular, Vif requires at least four heterologous protein-protein interactions to successfully counteract the APOBEC3 proteins. The first strategy, of course, is targeting the direct interaction with the APOBEC3 proteins themselves, which may occur through conserved structural motifs (Albin et al., 2010b; Kitamura et al., 2012a). Second is targeting the recently discovered interaction with CBFβ, which is essential for Vif stability and function (Jäger et al., 2012b; Zhang et al., 2012). Third and fourth are targeting the distinct Vif interaction motifs in ELOC (the SLQ motif) or CUL5, which are also essential for activity of the APOBEC3 ubiquitin ligase complex (Marin et al., 2003; Yu et al., 2003). Finally, it may be possible to target an upstream component of the proteasomal pathway involved in APOBEC3 degradation [*e.g.*, (Kim et al., 2013)].

However, despite these clear opportunities for drug development, progress has been relatively underwhelming. Proof of concept experiments have been achieved through cell-based screens for preservation of APOBEC3G-GFP fluorescence in the presence of HIV Vif (Nathans et al., 2008). The lead molecule from these studies, RN18, however, has only been subject to modest additional development (Ali et al., 2012). An additional concern holding back development of RN18 is the major challenge of identifying the molecular target, which is essential for structural studies and rational improvements through medicinal chemistry.

Independent lead compounds have also been identified based on predicted fit into the Vif-binding pocket in ELOC, which would effectively outcompete the SLQ motif of

Vif (Huang et al., 2013a; Huang et al., 2013b). However, the best of these indolizine type compounds has a modest IC50 value of 11 μM. This will likely need improvements in potency and solubility before efficacy can be achieved in live cell studies. In any event, much more work in this area is needed, including larger scale cellular screens, biochemical screens, and computational screens, followed by a comprehensive sets of secondary and tertiary screens to narrow in on the most effective drug candidates.

Therapy by hypomutation

Vif efficiently counteracts the HIV-relevant APOBEC3 repertoire (Conticello et al., 2003; Hultquist et al., 2011; Kao et al., 2003; Marin et al., 2003; Sheehy et al., 2003; Stopak et al., 2003). Nonetheless, evidence of APOBEC3 activity is found in the proviral sequences derived from infected patients [*e.g.*, (Janini et al., 2001)]. This raises the possibilities that either Vif cannot completely neutralize all of the cytoplasmic APOBEC3 before encapsidation and/or HIV regulates its high mutation rate, in part, through controlled degradation of the APOBEC3 proteins with Vif acting as a molecular rheostat optimizing the levels of cytosine deamination necessary for immune evasion and potentially drug resistance (Haché et al., 2006; Harris, 2008).

A counterintuitive but potentially more effective strategy to decrease the pathogenesis of HIV may be to inhibit the enzymatic activity of the APOBEC3s, thereby eliminating a potential source of variation for the virus (**Figure 1-4b**) (Harris, 2008; Li et al., 2012; Olson et al., 2013). Toward this end, high-throughput screens have resulted in proof-of-concept experiments that APOBEC3G activity can be blocked with small molecules (Li et al., 2012; Olson et al., 2013). However, much additional work is still

necessary to improve the solubility, potency, and bioavailability of APOBEC3 inhibitors, and efficacy will need to be demonstrated in cell-based studies before critical animal experiments can be done. Nevertheless, a hypomutation strategy is attractive because decreasing the HIV mutation rate has the potential to limit the diversity of the viral population and render it susceptible to normal immune clearance mechanisms (**Figure 1-4b**). Indeed, the adaptive T and B cell-mediated immune responses manage to keep the virus in-check initially in most patients, and it is tempting to speculate that even a slight tip in favor of host immunity may enable complete virus clearance (*i.e.*, analogous to most other viral infections). Given the fact that existing anti-retroviral drugs are numerous and largely effective, a big challenge facing the field is the development of a curative therapy. It remains possible that promoting either virus hypermutation or hypomutation could be part of such a cure.

ADDITIONAL CONTRIBUTIONS

We apologize to colleagues whose work could not be cited due to space limitations. Support was provided by grants from the National Institutes of Health R01-AI064046 and P01-GM091743 to RSH and F31-DA033186 to EWR.

FIGURES

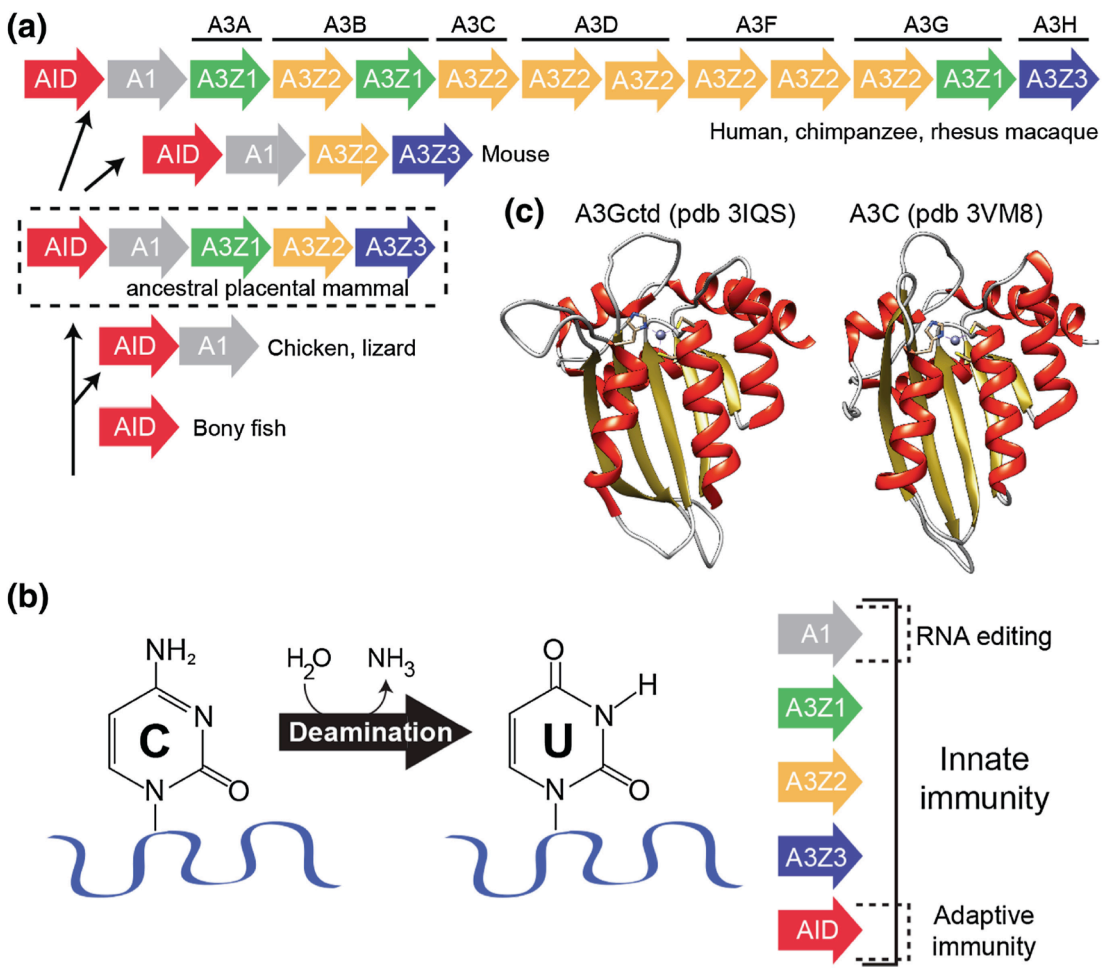


Fig. 1-1 Evolution and structure-function of APOBEC cytosine deaminases. **(a)** Expansion of the modern primate *APOBEC3* locus encoding seven *APOBEC3* genes with eleven zinc-coordinating (Z) domains [reprinted with permission from (Lackey et al., 2012)]. **(b)** Deamination of cytosine to uracil plays a central role in innate immunity. **(c)** Three-dimensional structures of the APOBEC3G C-terminal domain and APOBEC3C. A zinc ion (purple) is shown coordinated in both proteins by one histidine and two cysteine residues in the α2-β3-α3 core.

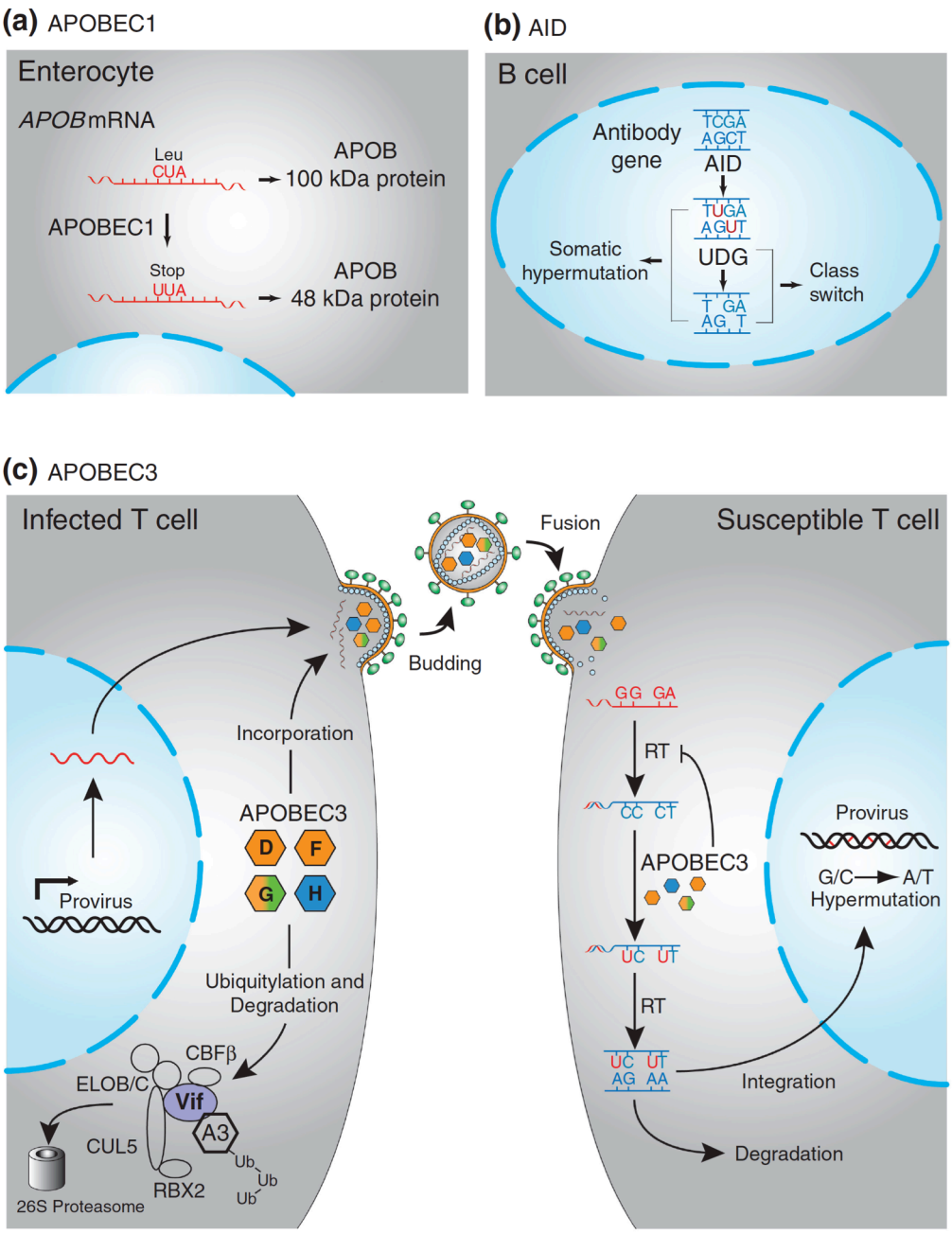

(a) APOBEC1
Enterocyte
APOB mRNA
Leu
CUA
APOB
100 kDa protein
APOBEC1
Stop
UUA
APOB
48 kDa protein
(b) AID
B cell
Antibody
gene
TCGA
AGCT
AID
TUGA
AGUT
UDG
Somatic
hypermutation
Class
switch
T GA
AG T
(c) APOBEC3
Infected T cell
Fusion
Susceptible T cell
Budding
Incorporation
APOBEC3
D
F
G
H
Provirus
Ubiquitylation and
Degradation
CBFβ
ELOB/C
Vif
A3
CUL5
RBX2
Ub
Ub
Ub
26S Proteasome
GG GA
RT
CC CT
APOBEC3
UC UT
RT
UC UT
AG AA
Integration
Degradation
Provirus
G/C → A/T
Hypermutation

Fig. 1-2 The physiological functions of the APOBEC family. **(a)** RNA editing by APOBEC1 generates a truncated APOB protein. **(b)** AID activity underpins the processes of somatic hypermutation and class switch recombination in B-cell germinal centers. **(c)** Four APOBEC3 proteins restrict HIV through cytosine deamination in the absence of Vif [reprinted with permission from (Hultquist et al., 2011)].

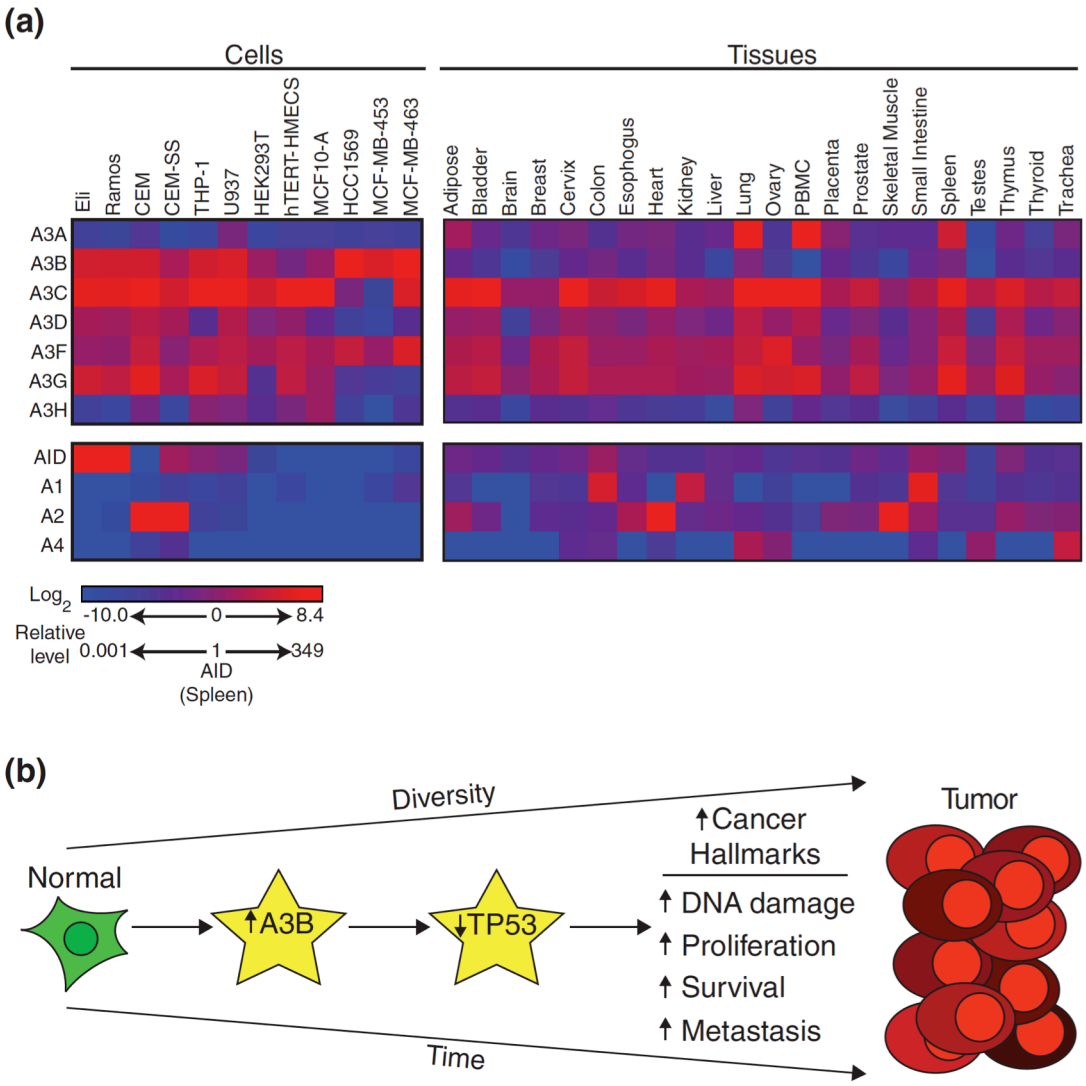
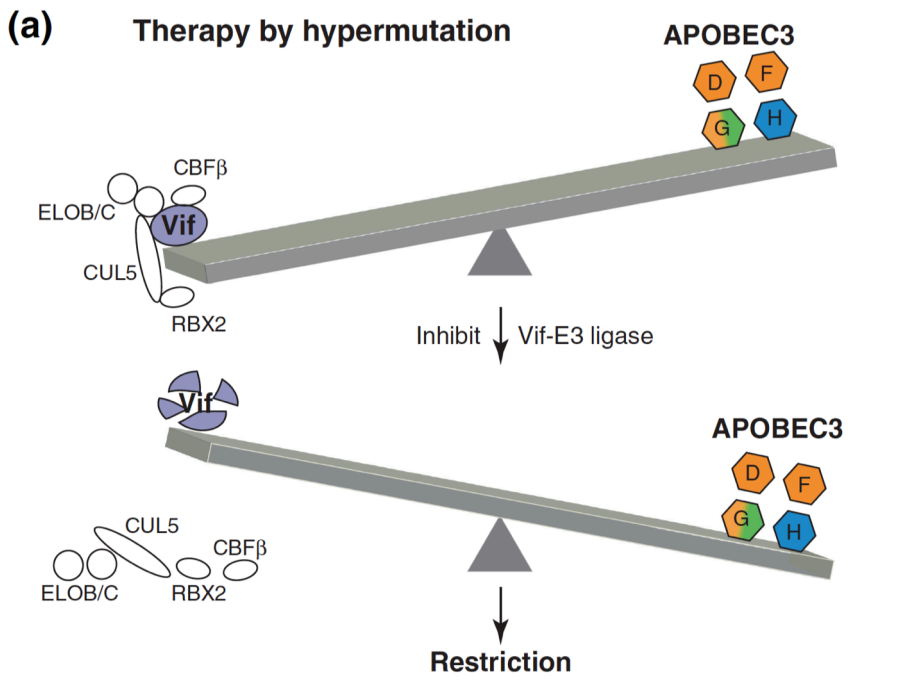


Fig. 1-3 *APOBEC* expression and model of carcinogenesis. **(a)** *APOBEC* family mRNA expression in the indicated cell lines and normal human tissues. Data are relative to levels of *AID* in the spleen [reprinted with permission from (Burns et al., 2013)]. **(b)** Model depicting the effect of APOBEC3B overexpression over time.


(a)
Therapy by hypermutation
APOBEC3
D
F
G
H
CBFβ
ELOB/C
Vif
CUL5
RBX2
Inhibit
Vif-E3 ligase
Vif
APOBEC3
D
F
G
H
CUL5
CBFβ
ELOB/C
RBX2
Restriction

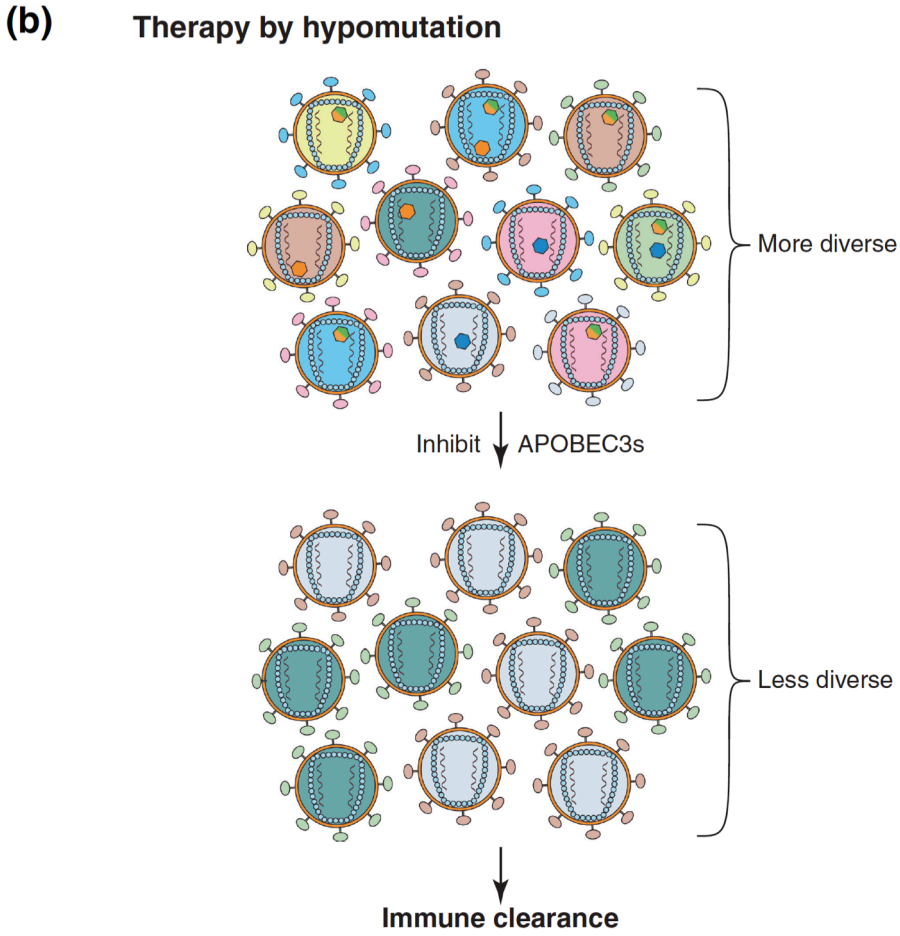

(b)
Therapy by hypomutation
More diverse
Inhibit
APOBEC3s
Less diverse
Immune clearance

Fig. 1-4 Potential therapies harnessing APOBEC3 innate immunity. **(a)** Inhibition of the Vif-E3 ligase complex could result in virus restriction. **(b)** Inhibiting the deaminase activity of APOBEC3 proteins could deprive HIV of a source of variation and allow immune clearance.

CHAPTER 2

Quantitative Profiling of the Full *APOBEC3* mRNA Repertoire in Lymphocytes and Tissues: Implications for HIV-1 Restriction

Reprinted with permission from: Eric W. Refsland[1], Mark D. Stenglein[1] Keisuke Shindo, John S. Albin, William L. Brown and Reuben S. Harris. (2010) Nucleic Acids Research 38:4274-84.

[1]Equal primary contributions

Authors' Contributions

EWR and MDS performed all of the experiments, wrote and revised the manuscript, and designed all of the figures except as follows. KS, JSA, and WLB performed the APOBEC3 immunobloting of the cell lines in Figure 6A. RSH aided in the design of the experiments and in manuscript preparation/revision.

FORWARD

While there is substantial evidence that multiple APOBEC3 proteins can inhibit HIV replication, methods for detection and quantification of the expression level of each individual APOBEC3 family member is challenging due to the high degree of identity between members. This homology often confounds interpretations of data attributing deaminase function of a specific APOBEC3 family member in a particular cell type or tissue. Before the experiments described here, very few reagents existed commercially or otherwise to detect the expression of the APOBEC3 family members. Here I described my colleagues' and my efforts to establish the most reliable methods of APOBEC3 detection. Since their publication in 2010, these methods have advanced the field of HIV restriction factors as well as opened up new directions of research in our own laboratory (see Chapter 5). This is supported by the greater than 100 citations this publication has received as of this writing. Definitively identifying the individual APOBEC3s expressed in a cell type or tissue is an important step towards the ultimate goal of mobilizing these anti-viral proteins as new options in the treatment of HIV infection.

SUMMARY

The human APOBEC3 proteins are DNA cytidine deaminases that impede the replication of many different transposons and viruses. The genes that encode APOBEC3A, APOBEC3B, APOBEC3C, APOBEC3D, APOBEC3F, APOBEC3G, and APOBEC3H were generated through relatively recent recombination events. The resulting high degree of inter-relatedness has complicated the development of specific quantitative PCR assays for these genes despite considerable interest in understanding their expression profiles. Here, we describe a set of quantitative PCR assays that specifically measures the mRNA levels of each *APOBEC3* gene. The specificity and sensitivity of each assay was validated using a full matrix of *APOBEC3* cDNA templates. The assays were used to quantify the *APOBEC3* repertoire in multiple human T-cell lines, bulk leukocytes and leukocyte subsets, and 20 different human tissues. The data demonstrate that multiple *APOBEC3* genes are expressed constitutively in most types of cells and tissues, and that distinct *APOBEC3* genes are induced upon T cell activation and interferon treatment. These data help define the *APOBEC3* repertoire relevant to HIV-1 restriction in T cells, and they suggest a general model in which multiple APOBEC3 proteins function together to provide a constitutive barrier to foreign genetic elements, which can be fortified by transcriptional induction.

INTRODUCTION

The APOBEC3 (A3; apolipoprotein B mRNA-editing enzyme, catalytic polypeptide-like 3) proteins are Zn^{2+}-dependent DNA cytidine deaminases, which are capable of inhibiting the replication of an incredible range of mobile genetic elements [reviewed by (Chiu and Greene, 2008; Goila-Gaur and Strebel, 2008; Hultquist and Harris, 2009; Malim and Emerman, 2008a)]. In humans, representative A3 substrates include exogenous retroviruses such as HIV-1 and HTLV, endogenous retroviruses such as HERV, endogenous retrotransposons such as L1 and Alu, and DNA viruses such as HBV and HPV [*e.g.*, (Bogerd et al., 2006b; Chiu et al., 2006; Derse et al., 2007; Lee et al., 2008; Mangeat et al., 2003; Muckenfuss et al., 2006; Sheehy et al., 2002; Stenglein et al., 2010; Suspène et al., 2005a; Vartanian et al., 2008; Zhang et al., 2003)]. Despite the relevance of many of these parasitic elements to human health, the A3 proteins that function *in vivo* against each particular element have yet to be defined. For instance, A3F and A3G have been shown to strongly restrict HIV-1 *in vitro* and very likely contribute to HIV-1 restriction *in vivo*, but strong arguments have also implicated several other human A3 proteins [reviewed by (Albin and Harris, 2010; Chiu and Greene, 2008; Goila-Gaur and Strebel, 2008; Hultquist and Harris, 2009; Malim and Emerman, 2008a)].

The seven *A3* genes are positioned in tandem on human chromosome 22: *A3A*, *A3B*, *A3C*, *A3D* (formerly *A3DE*), *A3F*, *A3G*, and *A3H* (Jarmuz et al., 2002). A defining feature of each *A3* gene is that it encodes a protein with one or two conserved zinc (Z)-coordinating deaminase domains. Each Z domain belongs to one of three distinct phylogenetic groups: Z1 (A3A and the C-terminal halves of A3B and A3G), Z2 (A3C, both halves of A3D and A3F, and the N-terminal halves of A3B and A3G), and Z3

(A3H) (LaRue et al., 2008). Based on the relatedness of these Z domains, the human *A3* repertoire appears to be the result of a minimum of eight unequal crossing-over recombination events, which mostly occurred during the radiation of primates (LaRue et al., 2008). The net result is that the human *A3* mRNAs share considerable identity, ranging from 30% to nearly 100% (**Figure 2-1A and Table 2-S1**). For example, nucleotides 1-473 of the *A3F* mRNA are 98.7% identical to *A3G* nucleotides 50-522. Similarly, *A3A* nucleotides 259-937 are 96.6% identical to the corresponding region of *A3B*, nucleotides 692-1370. The *A3* genes are also under positive selection and accordingly are highly polymorphic (Kidd et al., 2007; OhAinle et al., 2008; Sawyer et al., 2004). These inter-domain identities and polymorphisms present considerable challenges to *A3* expression profiling and quantification.

Here we report specific quantitative PCR (qPCR) assays for each of the seven human *A3* cDNA sequences. We use these assays to profile the expressed *A3* repertoire in common T cell lines, primary CD4+ T lymphocytes, and twenty distinct human tissues. We also quantify the effects of T cell stimulation and interferon (IFN) induction on *A3* expression. These expression data indicate that several A3 proteins, in addition to A3F and A3G, are expressed in CD4+ cells and are therefore positioned to contribute to HIV-1 restriction. More generally, nearly every cell type and tissue expresses multiple *A3*s, consistent with a model in which parasitic elements must evolve ways to cope with a constitutive set of restriction factors that can be further fortified by transcriptional induction.

RESULTS

APOBEC3 qPCR assay designs and optimization experiments

To determine where and when each *A3* mRNA is expressed we developed a panel of specific qPCR assays. For each assay, several computational and manual procedures were used to develop specific and efficient qPCR primers. Each primer was selected to span two exons (**Figure 2-1A and Table 2-1**). The primers and probes were screened to avoid known single nucleotide polymorphisms and splice variations (data not shown). Assay specificity was tested by pitting all seven qPCR probe/primer sets against each of the seven *A3* cDNA templates in a 49-reaction qPCR matrix. The amplification results demonstrated that each assay was specific to the intended target *A3* (**Figure 2-1B**).

The next most important parameter to test and optimize was reaction efficiency. Serial dilutions of linearized *A3* control cDNA templates were used in qPCR reactions to measure efficiencies. After as many as four primer/probe redesigns for some *A3*s, high efficiencies ranging from 1.97 – 1.99 were achieved for each qPCR assay **(Figure 2-1C**). These nearly ideal efficiencies indicated that each assay would provide a robust and quantitative measure of the level of expression of each *A3* gene. In other words, these data indicated that quantitative comparisons of each *A3* relative to the others are possible.

Assay standardization and normalization are crucial parameters to ensure quantitative data set comparisons. We therefore also designed and tested qPCR assays for three common housekeeping genes, *TBP*, *RPL13A*, and *HPRT* (**Materials and Methods**). The mRNA levels of each of these genes were consistent between cell types and tissues. To facilitate data normalization and comparisons, only one of these genes (*TBP*) was used as an internal standard.

Finally, each of the aforementioned *A3* qPCR assays was evaluated on two independent platforms to ensure that all data could be reproduced and generalized. The Roche Lightcycler 480 platform, which uses fluorescent probes, was used to develop the assays and for all data reported here. A similar set of reaction specificities was observed using different instrumentation and detection methods (iCycler and SybrGreen, respectively; BioRad, data not shown). These data indicated that our *A3* qPCR assays could be applied to a range of experimental questions.

APOBEC3 expression profiles in permissive and non-permissive T cell lines

Three groups independently discovered A3G in 2002 (Harris et al., 2002; Jarmuz et al., 2002; Sheehy et al., 2002). One of these studies was based on the observation that there are two distinct phenotypes in human T cell lines with respect to Vif-deficient HIV-1 infection (Gabuzda et al., 1992; Sheehy et al., 2002; Strebel et al., 1987; von Schwedler et al., 1993). Cell lines such as CEM and H9 are resistant to infection by Vif-deficient viruses and, accordingly, are termed non-permissive. On the other hand, cell lines such as CEM-SS (a clonal derivative of CEM) and SupT1 support HIV-1 replication in the presence or absence of Vif and are thus deemed permissive. A3G was originally identified as the dominant restriction factor whose presence accounted for the non-permissive phenotype of CEM cells in comparison to its permissive daughter line, CEM-SS. Since then, overexpression of several A3 family members has been shown to render CEM-SS non-permissive to Vif-deficient HIV replication [*e.g.*, (Haché et al., 2008; Sheehy et al., 2002); reviewed by (Albin and Harris, 2010; Chiu and Greene, 2008;

Goila-Gaur and Strebel, 2008; Hultquist and Harris, 2009; Malim and Emerman, 2008a)]. Despite the elegance of this permissive and non-permissive dichotomy, the full nature of the Vif-deficient HIV-1 replication defect on non-permissive cells has not been elucidated. In other words, which A3s are required for HIV-1 restriction in non-permissive cells: A3G or some combination of A3G and these other A3 proteins?

To shed additional light on this important question, we profiled the expressed *A3* repertoire in the two related cell lines, non-permissive CEM and permissive CEM-SS. These cell lines were used to clone *A3G* by subtractive hybridization (Sheehy et al., 2002) and, as expected, *A3G* mRNA levels were 12-fold higher in the non-permissive line CEM (**Figure 2-2A)**. However, levels of *A3B*, *A3C*, *A3D*, and *A3F* were also significantly higher than those in CEM-SS, by 19-, 2-, 3-, and 28-fold, respectively. These observations suggest that other A3s, in addition to A3G, may contribute to Vif-deficient HIV-1 restriction in CEM. However, the presence of significant *A3C* mRNA levels in CEM-SS is concordant with it alone not being sufficient to restrict the replication of Vif-deficient HIV-1 (Bishop et al., 2004; Yu et al., 2004a).

Further consistent with the possibility that multiple A3s contribute to the non-permissive phenotype, the non-permissive T cell line H9 was also found to express multiple *A3* mRNAs at levels well beyond those in permissive cells (**Figure 2-2B**). *A3C*, *A3D*, *A3F*, *A3G*, and *A3H* were 2-, 8-, 16-, 19-, and 1000-fold above those in CEM-SS cells, respectively. In contrast to CEM cells, *A3B* was virtually absent, and *A3H* was well expressed (considered further below). Notably, the expression levels of all seven *A3* genes were extremely low and approaching detection limits in the permissive line SupT11. Thus far, SupT11 is the only T cell line that we have found that is practically

devoid of *A3* gene expression. The molecular explanation for this is not known, but it is not due to deletion of the locus because specific qPCR products are still detected.

APOBEC3 expression profiles in primary leukocytes

Next we sought to define the *A3* repertoire in fresh, unstimulated peripheral blood mononuclear cells (PBMCs) (**Figure 2-2B**). Multiple *A3*s were expressed in PBMCs with the overall pattern resembling that of CEM and H9 non-permissive lines (significant levels of *A3C*, *A3D*, *A3F*, and *A3G*). However, some major differences were detected. First and most strikingly, PBMCs were found to express high *A3A* levels, which were approximately 5000-fold greater than CEM. These observations were consistent with recent reports indicating that *A3A* expression is specific to the CD14+ lineage, which includes macrophages and monocytes (Chen et al., 2006; Koning et al., 2009; Peng et al., 2007; Stenglein et al., 2010). Second, PBMCs expressed virtually no *A3B* mRNA, which is approximately 100-fold less than CEM. Third, PBMCs expressed 4-fold more A3G than CEM, and 40-fold more than CEM-SS. Finally, PBMCs were found to express high levels of A3H, similar to those observed in H9, approximately 13-fold more than CEM, and nearly 1000-fold more than CEM-SS or SupT11.

The expressed *APOBEC3* repertoire in naïve and stimulated CD4+ T lymphocytes

CD4+ T lymphocytes are a major target of HIV-1. To define the expressed *A3* repertoire in this important T cell subset, we used negative selection to isolate CD4+ lymphocytes from fresh PBMCs (**Figure 2-3A**). These cells were stimulated with IL-2 and PHA and shown to proliferate by CFSE staining and flow cytometry (**Figure 2-3B**).

T-cell stimulation caused the induction of every *A3* with the exception of *A3A* (**Figure 2-3C**). Mitogen activation had been shown previously to induce A3G expression in CD4+ T cells (Koning et al., 2009; Stopak et al., 2007). The increase in *A3H* expression levels is particularly remarkable, rising 22-fold over naïve levels.

Several *APOBEC3*s are interferon-responsive

In response to viral infection, cytokines such as type I interferons (IFN) are induced, and these in turn activate the expression of hundreds of genes. Several *A3*s have been shown to be IFN-responsive, but only two reports have considered the entire *A3* repertoire and procedures have varied considerably [(Argyris et al., 2007; Bonvin et al., 2006; Chen et al., 2006; Koning et al., 2009; Peng et al., 2007; Stenglein et al., 2010; Tanaka et al., 2006; Vetter et al., 2009; Wang et al., 2008); see **Discussion**]. To extend this work, fresh PBMCs were treated with leukocyte IFN and the resulting *A3* mRNA levels were quantified by qPCR (**Figure 2-4A**). In bulk PBMCs, *A3A* was clearly induced, in agreement with prior reports (Koning et al., 2009; Peng et al., 2006; Stenglein et al., 2010; Taylor et al., 2004). The other *A3*s were also found to be upregulated but to lesser extents.

We next examined the effects of IL-2, PHA, and IFN on *A3* expression in naïve and stimulated CD4+ T cells (**Figure 2-4B**). As shown above, the T cell mitogens IL-2 and PHA induced the mRNA levels of all of the *A3*s except *A3A* (compare **Figure 2-3C** and **Figure 2-4B**). In contrast, IFN treatment did little to alter these expression profiles. Thus, the *A3*s in CD4+ T lymphocytes are not IFN inducible, whereas the *A3*s in at least one other PBMC cell type are IFN-responsive. Based on recent reports, CD14+

phagocytic cells such as monocytes and macrophages are likely to be the only leukocyte subset in which *A3* expression is IFN-inducible [(Koning et al., 2009; Peng et al., 2006; Stenglein et al., 2010; Taylor et al., 2004) and data not shown].

APOBEC3 expression profile in human tissues

An important question with respect to *A3* expression is whether or not it is confined to immune cell compartments. To help address this issue, we quantified the expressed *A3* repertoire of twenty normal human tissues. Two cell lines, CEM and SupT11, and fresh PBMCs were assayed in parallel to facilitate comparisons with our other data sets. Several interesting observations emerged from these studies (**Figure 2-5**).

First, consistent with central roles in innate immunity, *A3* expression levels were high in lymphoid organs such as the thymus and spleen. The thymus, which is rich in T lymphocytes, has an expressed *A3* repertoire similar to that of primary CD4+ T cells: low *A3A* and *A3B*, high *A3C*, *A3D*, *A3F*, *A3G*, and *A3H* (compare **Figure 2-5** and **Figure 2-3C**). The spleen is rich in both B and T lymphocytes, but it is also a major reservoir for undifferentiated monocytes (Swirski et al., 2009). This helped reconcile observations showing that the spleen has relatively high levels of all seven *A3*s, consistent with previous observations indicating that *A3A* is restricted to CD14+ cells including monocytes (Koning et al., 2009; Peng et al., 2006; Stenglein et al., 2010).

Second, many *A3*s were clearly expressed outside of the blood compartment or common immune tissues. For example, five of seven *A3*s showed peak expression levels in lung tissue: *A3A*, *A3B*, *A3C*, *A3D*, and *A3H*. The most remarkable was *A3A*, which was detected at levels 70-fold higher in the lung than the median value from all 20 tissues.

This observation is consistent with large numbers of CD14+ macrophages residing in lung alveoli. A bronchoalveolar lavage cell preparation from a non-smoking individual yields 1 x 10^7 cells, of which 90-95% are macrophages (Daniele et al., 1977). In addition, *A3A* was expressed in adipose tissue at levels 5-fold higher than the median value. This level is compatible with the finding that adipose from an obese individual is 40% macrophages by weight (Weisberg et al., 2003). Also of note is *A3* expression in the ovary. Levels of *A3C*, *A3F*, and *A3G* were 3-, 5-, and nearly 3-fold higher, respectively than median values from the other tissues. This may indicate a requirement for viral and transposable element restriction in female germ cells [*e.g.*, (Stenglein and Harris, 2006)].

Finally, some tissues expressed virtually no *A3*s. The brain and the testes, for instance, had low *A3* levels comparable to those of SupT11 (compare **Figure 2-5** and **Figure 2-2B**). These are largely immune privileged organs that are physically separated from other tissues and from potential pathogen infections by the blood-brain barrier and the blood-tubular barrier, respectively. The apparent *A3* expression deficiency in these tissues may also be due in part to the absence of appropriate gene expression activators such as interferons, cytokines, and/or growth factors.

Strong correlations between *APOBEC3* mRNA and protein expression levels

To ask whether the observed *APOBEC3* mRNA levels correlate with protein levels, we separated cell extracts by SDS-PAGE and probed the resulting blots with antibodies raised against A3F and A3G (**Figure 2-6**). The A3F polyclonal antibody is specific to the best of our knowledge [(Haché et al., 2008; Holmes et al., 2007) and unpublished observations]. The A3G polyclonal antibody is semi-specific as it reacts with

both A3A and A3G, but these proteins are distinguishable by size (Koning et al., 2009; Newman et al., 2005; Stenglein et al., 2010).

In strong agreement with our qPCR data and many prior studies, A3F and A3G protein levels are considerably lower in CEM-SS than they are in the non-permissive lines CEM [*e.g.*, **Figure 2-6A**; (Hache and Harris, 2009; Haché et al., 2008)]. The low but still measurable levels of A3G in the CEM-SS population also enabled quantification by densitometry, revealing a 5- to 10-fold difference at the protein level between these two lines. This value closely resembles the 10-fold difference in *A3G* mRNA levels between CEM and CEM-SS (**Figure 2**). In further agreement with our qPCR data, A3G protein was too low to detect by immunoblotting in SupT11 and approximately 1.6-fold higher in H9 than CEM. As for A3F, nearly equivalent protein levels were observed in H9 and CEM, and it was undetectable in SupT11 and CEM-SS (**Figure 2-2 and Figure 2-6A**). Finally, we compared the protein expression of A3F and A3G in primary CD4+ T lymphocytes under naïve and stimulated conditions. Consistent with our qPCR data, A3F was induced slightly and A3G levels increased approximately 2-fold upon T cell stimulation (**Figure 2-3C and Figure 2-6B**). Overall, we conclude that strong correlations are evident between the qPCR results and the immunoblot data.

DISCUSSION

Despite remarkable progress identifying A3 restriction substrates such as HIV-1, transposons, and foreign DNA, the overarching biological functions of the A3 repertoire are still being elucidated [(Esnault et al., 2005; Sheehy et al., 2002; Stenglein et al.,

2010); reviewed by (Chiu and Greene, 2008; Goila-Gaur and Strebel, 2008; Hultquist and Harris, 2009; Malim and Emerman, 2008a)]. What events drove the rapid expansion of the *A3* locus to encode seven A3 proteins, and a total of eleven zinc-coordinating domains? Does each A3 protein serve some unique, divergent function or do they serve the same function only in different places and/or at different times? How much redundancy exists? How much specificity? How much of this is governed by transcriptional programs? To begin to address the latter question, in particular, here we report qPCR assays that enable the specific and quantitative detection of the entire seven gene human *A3* repertoire. We have improved on previous commercial methods by reporting the identity, specificity and efficiency of each *A3* primer/probe set, by demonstrating the assay's functionality and usefulness across detection platforms and, most importantly, by quantifying *A3* expression in multiple human tissues and cell types. During the course of completing our experiments, an independent study on *A3* expression in hematopoietic cells and tissues was published by Koning *et al.* (Koning et al., 2009). These two studies are largely complementary, but some differences merit discussion.

The first point is technical. We developed our own qPCR assays and, as such, were able to ensure target specificity and near identical reaction efficiencies (**Figure 2-1**). We also report the nucleotide sequences of the primers and probes so the assays are fully compatible with many different qPCR methods and instruments (**Table 2-1**). In contrast, Koning *et al.* used proprietary primers and probes from Applied Biosystems, which are optimized for specific reagents and instrumentation and less readily adaptable to other platforms. Knowledge of the primer binding sites is also crucial for detecting and ultimately ascertaining the function of alternatively spliced *A3* variants.

Second, our study uniquely focuses on the question of which A3s are expressed in such a manner that they may contribute to HIV-1 restriction *in vivo*. The observation that the non-permissive T cell lines CEM and H9 express significantly higher levels of *A3G* and *A3F* is not surprising (**Figure 2-2**). However, what is notable is that six *A3*s (*i.e.*, all but *A3A*) are expressed at higher levels in these non-permissive cell lines. The constitutive *A3* repertoire in H9 appears remarkably similar to the induced repertoire in stimulated, primary CD4+ T lymphocytes (**Figures 2-2B and 2-3C**). *A3H* levels, in particular, increase over 20-fold upon T cell stimulation with IL-2 and PHA. Together with prior cell culture overexpression studies indicating that A3H can restrict HIV-1 in a Vif susceptible manner (Dang et al., 2008; Harari et al., 2009; Li et al., 2010; OhAinle et al., 2008; OhAinle et al., 2006; Tan et al., 2009), our expression data strengthen the case that A3H may contribute to HIV-1 restriction *in vivo*. In contrast, a role for A3A in HIV-1 restriction in CD4+ T cells is unlikely, because it is not expressed in this cell type and it does not restrict Vif-deficient HIV-1 in overexpression experiments [this study and (Bishop et al., 2004; Chen et al., 2006; Koning et al., 2009; Stenglein et al., 2010)]. Unambiguously testing the involvement of other A3s in HIV-1 restriction, in addition to A3F and A3G, will require much further work.

Third, based on an extensive survey of *A3* expression in 20 tissues, we conclude that the *A3* mRNAs are expressed broadly and not confined to cells of the immune compartment (although there is clearly some bias to immune cell types; **Figure 2-5**). For example, adipose, colon, cervix, bladder, and the heart all express at least two *A3*s at levels 2-fold above the median among all tissues. In partial contrast, Koning *et al.* concluded that, apart from germ cell tissues, *A3* expression is confined to the

hematopoietic compartment and that their detection in many tissues is simply due to infiltrating leukocytes (Koning et al., 2009). This is likely the case for some tissues such as lung and adipose where significant numbers of *A3A*-expressing macrophages are found. Although the true breadth of *A3* expression in human tissues (and a different developmental stages) will not be fully appreciated until suitably specific immunohistochemistry antibodies are developed, at least two additional lines of evidence support the likelihood that *A3*s are expressed broadly. First, as noted by Koning *et al* and as is also evident in our data sets, the *A3G* mRNA levels typically exceed those of *A3F* in PBMCs and in specific leukocyte types [**Figures 2-3C, 2-4 and 2-5** and (Koning et al., 2009)]. This ratio is inverted in some tissues such as the cervix, colon and ovary, strongly suggesting that immune cell infiltration is not the whole story [**Figure 2-5** and (Koning et al., 2009)]. Second, many clonally derived cancer cell lines of non-immune cell origins have been shown to express multiple human A3s, such as colorectal hepatocarcinoma, adenocarcinoma, melanoma, and lung carcinoma lines [(Bonvin et al., 2006; Harris et al., 2002; Henry et al., 2009; Jarmuz et al., 2002; Jost et al., 2007; Kock and Blum, 2008; Liddament et al., 2004) and www.oncomine.org]. It is possible that *A3* expression switches on at some stage during oncogenesis, but given the abundance of *A3*s in cancer cell lines it is likely that part of this reflects the normal gene expression program that existed prior to immortalization. Thus, the broad and apparently constitutive *A3* expression profile of many human tissues is consistent with a general role for A3 proteins in innate immunity. Substrates may not only include endogenous and exogenous retroelements, but also DNA viruses and even naked, foreign double-stranded DNA [*e.g.*,

(Narvaiza et al., 2009; Stenglein et al., 2010; Suspène et al., 2005a; Vartanian et al., 2008)].

Fourth, our data and those of Koning *et al.* are somewhat discordant on the relative difference between *A3G* and *A3F* expression levels. They concluded that the A3G mRNA levels are 10-fold higher than those of A3F (Koning et al., 2009). In contrast, we did not observe such a large expression bias. In some tissues such as lung we observed a modest bias in favor of *A3G*. Conversely, in other tissues such as the colon and ovary, we observed a slight bias toward *A3F*. However, in general, we observed that *A3F* and *A3G* levels are fairly similar and that these mRNAs are invariably expressed together. Our data are consistent with prior *A3F* and *A3G* multi-tissue northern blot results and the fact that the promoter regions of *A3F* and *A3G* are 96% identical over a 5kb region including exons 1 and 2 (Liddament et al., 2004). We favor a model in which A3F and A3G are coordinately expressed [proposed originally by (Liddament et al., 2004)]. A greater understanding of the promoters and their associated transcription factors will help provide future tests of this model. It is possible that differences between the results reported here and the literature are due to protocol differences such as random primed versus oligo-dT mediated cDNA synthesis, the latter being much less efficient and susceptible to repetitive elements in 3' UTRs, and/or to imperfect PCR reaction efficiencies (*e.g.*, **Figure 2-S2**).

On a strong complementary note, a major point detailed here and in the study by Koning and coworkers is the inducibility of *A3* mRNA expression. Several prior studies have examined the IFN-responsiveness of the *A3*s and particularly *A3G* and the conclusions have been variable (Argyris et al., 2007; Bonvin et al., 2006; Chen et al.,

2006; Koning et al., 2009; Peng et al., 2006; Stenglein et al., 2010; Tanaka et al., 2006; Wang et al., 2008). However, in agreement with Koning and coworkers, our data indicate that none of the *A3*s are induced by IFN in primary CD4+ T lymphocytes (**Figure 2-4B**). Small increases in *A3* expression, such as for *A3A*, can be attributed readily to a few contaminating CD14+ cells in the T lymphocyte preparations. Rather, the strong IFN responsiveness of the *A3*s, highlighted by *A3A*, may be a property of other blood cell lineages such as CD14+ monocytes and macrophages (Chen et al., 2006; Koning et al., 2009; Stenglein et al., 2010). We also found that T cell activation with IL-2 and PHA causes the induction of 6/7 *A3*s [*i.e*., all but A3A; **Figure 2-4B** and (Koning et al., 2009; Stopak et al., 2007)]. These studies combine to indicate that cells have multiple mechanisms to regulate *A3* expression. Thus, in addition to constitutive expression of most of the *A3* repertoire in many cell types and tissues, transcriptional regulation is likely to have a major role in ultimately determining the efficacy and potency of A3-dependent defenses against viral and non-viral challenges.

Overall, turning back to the question of HIV-1 restriction, our studies are consistent with A3F, A3G, and potentially several other A3s being important in CD4+ T lymphocytes. A3H is the most likely additional restriction factor because (i) it is expressed in the non-permissive T cell line H9 (**Figure 2-2**), (ii) it is expressed in naïve CD4+ cells (**Figure 2-3**), (iii) it is induced 20-fold by T cell activation (**Figure 2-3**), and (iv) it is capable of HIV-1 restriction in a Vif-susceptible manner (Dang et al., 2008; Harari et al., 2009; Li et al., 2010; OhAinle et al., 2008; OhAinle et al., 2006; Tan et al., 2009). Although the literature overwhelmingly favors a role for A3G in HIV-1

restriction, with A3F a distant second, A3H and several of the other A3s warrant substantive additional investigation.

MATERIALS AND METHODS

Cell lines

Human T cell lines were cultured in RPMI (Thermo-Fisher) supplemented with 10% fetal bovine serum (Denville). SupT1 and H9 were obtained from the AIDS Research and Reference Reagent Program. SupT11 is a sub-clone of SupT1 obtained by limiting dilution. CEM and CEM-SS were described (Haché et al., 2008).

Enrichment and culture of primary cells

Peripheral blood mononuclear cells (PBMCs) were isolated from blood (Memorial Blood Center, St. Paul, MN) by Ficoll-Paque gradient centrifugation (GE Healthcare). The resulting buffy coat was subjected to negative selection to enrich for naïve primary CD4+ T cells (Miltenyi Biotech). Primary cells were cultured in RPMI supplemented with 10% FBS and stimulated to proliferate with 50 U ml^{-1} recombinant interleukin-2 (IL-2; Sigma), and 10 μg ml^{-1} phytohemagglutinin (PHA; Fisher Healthcare). To confirm proliferation, 10^6 cells were stained with 10 μM carboxyfluorescein diacetate, succinimidyl ester (CFSE) according to the manufacturer (Invitrogen) and analyzed by flow cytometry. PBMCs and CD4+ T cells were treated with 2 ng ml^{-1} universal type I interferon (R&D Systems).

RNA preparation

PBS-washed cells were prepared for RNA isolation by lysis in ice-cold buffer RLT (Roche) supplemented with 10 µl ml^{-1} 2-mercaptoethanol (Sigma). RNA was isolated using the RNeasy kit (Qiagen). Two optional components of the kit, QiaShredder and on-column DNase digestion, were used to homogenize lysates and remove genomic DNA. RNA concentrations were determined spectroscopically and RNA quality was assessed by gel analysis (**Figure 2-S1**). Total RNA from 20 distinct tissues, each from a minimum of 3 donors, was purchased from Ambion. The control qPCRs for housekeeping genes (below) were similarly efficient for all RNA preparations, indicating equally high qualities and integrities.

cDNA synthesis

1 µg of total RNA was used to synthesize cDNA using avian myeloblastosis virus reverse transcriptase (AMV RT; Roche) and random hexameric primers. We found that random hexamer mRNA priming was preferable to oligo-dT priming, which resulted in inefficient synthesis of some *A3* cDNAs, particularly *A3D* and *A3F*, most likely because of their long, repeat-rich 3' untranslated regions (UTRs) (**Figure 2-1A** and **Figure 2-S2**). cDNA synthesis reactions were performed by mixing 3 µl of 333 µM random hexamer primer (50 µM final concentration) with 1 µg of total RNA diluted in 10 µl of RNase-free water. This mixture was heated for 10 minutes at 65˚C to remove RNA secondary structure and then cooled briefly on ice. 7 µl of master mix containing 0.5 µl Protector RNase inhibitor (Roche), 0.4 µl AMV RT, 4 µl 5x AMV RT reaction buffer, and 2.1 µl dNTPs (Roche, 10 mM each; ~1 mM each final concentration) were then added to these

reactions, which were subsequently incubated at 25˚C for 10 min, 42˚C for 1 hour, then 95˚C for 5 min. Reactions were then diluted by addition of 80 µl RNase-free H_2O, and used in qPCR reactions.

Quantitative PCR

cDNAs levels were quantified by PCR using a Roche Lightcycler 480 instrument according to the manufacturer's protocols. Reactions were performed in 96 well plates with each containing 7.5 µl 2x probe master mix (Roche), 1.25 µl H_2O, 1.05 µl primers (5 µm each), 0.2 µl UPL probe (Roche), and 5 µl cDNA (prepared as above). Reactions were incubated at 95˚C for 10 min, then 40 cycles of 95˚C for 10 sec, 58˚C for 15 sec, then 72˚C for 2 sec. **Table 2-1** lists the full sequences of all primers and probes used in this study; the design rationale and validation experiments are described below. The PCR cycle at which amplification was detectable above a background threshold (threshold cycle, or Ct) was calculated with the maximum second derivative method using the Lightcycler 480 software (Roche, version 1.5.0). cDNA was synthesized and qPCR performed in triplicate for each sample, and the mean values and standard deviations for each triplicate are reported.

Primer Design

Primer pairs were designed to avoid inter-*A3* identity and only amplify the intended *A3* target (**Figure 2-1A, Figure 2-S2, Table 2-1, and Table 2-S1**). Primer pair specificity was verified by manually inspecting alignments of primers and *A3* cDNA sequences, by using BLAST (version 2.2.20; http://blast.ncbi.nlm.nih.gov/) and Primer-BLAST

software (http://www.ncbi.nlm.nih.gov/tools/primer-blast/; default parameters except E set to 1) and, experimentally, by attempting to amplify 10^4 copies of every *A3* control template with each set of *A3* qPCR primers (**Figure 2-1B**). Primers were additionally designed to have similar melting temperatures (**Table 2-1**) and were confirmed to have similar reaction efficiencies (**Figure 2-1C** and see below). Furthermore, primers were designed such that they would amplify all described variants of each *A3* mRNA (*i.e.*, splice variants and SNPs; data not shown). Design was assisted by the Roche ProbeFinder software (version 2.43; http://qpcr.probefinder.com/roche3.html) and by Primer3 software (version 0.4.0; (http://primer3.sourceforge.net/).

Data Analysis

Expression level normalization: *A3* expression data were normalized to the expression of TATA-box binding protein (*TBP*; see below). Essentially identical normalized results were obtained using *ribosomal protein 13A* (*RPL13A*), or *hypoxanthine-guanine phosphoribosyltransferase* (*HPRT*) (data not shown). None of these controls showed altered expression levels upon treatment of cells with IL-2, PMA and/or IFN.

Determination of reaction efficiencies: For all assays, the primer efficiency was determined using serial dilutions of control templates (*e.g.*, **Figure 2-1C**). The control templates consisted of portions of the 5' UTRs and coding regions of the *A3* cDNAs amplified by PCR and cloned into pCR2.1-TOPO (Invitrogen). The same templates were used to validate assay specificity (**Figure 2-1B**). These plasmids were linearized by digestion with XmnI or BglII, phenol-chloroform extracted, ethanol precipitated, and the DNA concentration was determined spectroscopically. Ten-fold serial dilutions of these

plasmids were prepared with 10^1-10^8 plasmid molecules per 5 µl. These dilutions were used as templates in qPCR reactions. Ct values were plotted against the $\log_{10}$ number of template molecules in each reaction (**Figure 2-1C**), and the slope of this line determines the reaction efficiency according to the equation: efficiency = $10^{(-1/slope)}$ (Pfaffl, 2001). Reaction efficiencies were also calculated from the actual amplification curves from several experiments (Peirson et al., 2003). Reaction efficiencies calculated by these two methods agreed closely (data not shown).

The reaction efficiencies (E) determined in this manner were then used to calculate relative expression levels between two samples (S1 and S2), normalized to a reference gene (ref, *TBP*), according to published methods (Pfaffl, 2001):

$$\Delta Ct_{(A3)} = Ct_{A3,S1} - Ct_{A3,S2}$$

$$\Delta Ct_{(ref)} = Ct_{ref,S1} - Ct_{ref,S2}$$

$$\text{Relative Expression} = (E_{A3})^{-\Delta Ct(A3)} / (E_{ref})^{-\Delta Ct(ref)}$$

Immunoblotting

Cells were cultured and/or isolated as described above, washed in PBS and lysed in 0.2% NP40 buffer with 1x EDTA-free protease inhibitor mixture (Roche). Proteins were subjected to ice-cold acetone precipitation prior to quantification (Bio-Rad). An equal amount of protein was fractionated on a 10% SDS-polyacrylamide gel and transferred to a PVDF membrane (Millipore). Membranes were blotted with antibodies specific to A3F, A3G, or tubulin (Covance). The A3F and A3G polyclonal antibodies were obtained from the AIDS Research and Reference Reagent Program courtesy of M. Malim (Kings College London) and J. Lingappa (University of Washington), respectively (Holmes et

al., 2007; Newman et al., 2005). Band intensities were quantified using Image J software (NIH).

ADDITIONAL CONTRIBUTIONS

This work was supported by grants from the National Institutes of Health, GM090437 and AI064046, and by the Bill and Melinda Gates Foundation. M.D.S. was supported in part by a 3M Graduate Fellowship and a Cancer Biology Training Grant (CA009138). We thank J. Hultquist for comments on the manuscript, M. Malim for a provocative discussion at Cold Spring Harbor, the AIDS Research and Reference Reagent Program for cell lines and antibodies, and D. Bernlohr for the use of a BioRad iCycler.

FIGURES

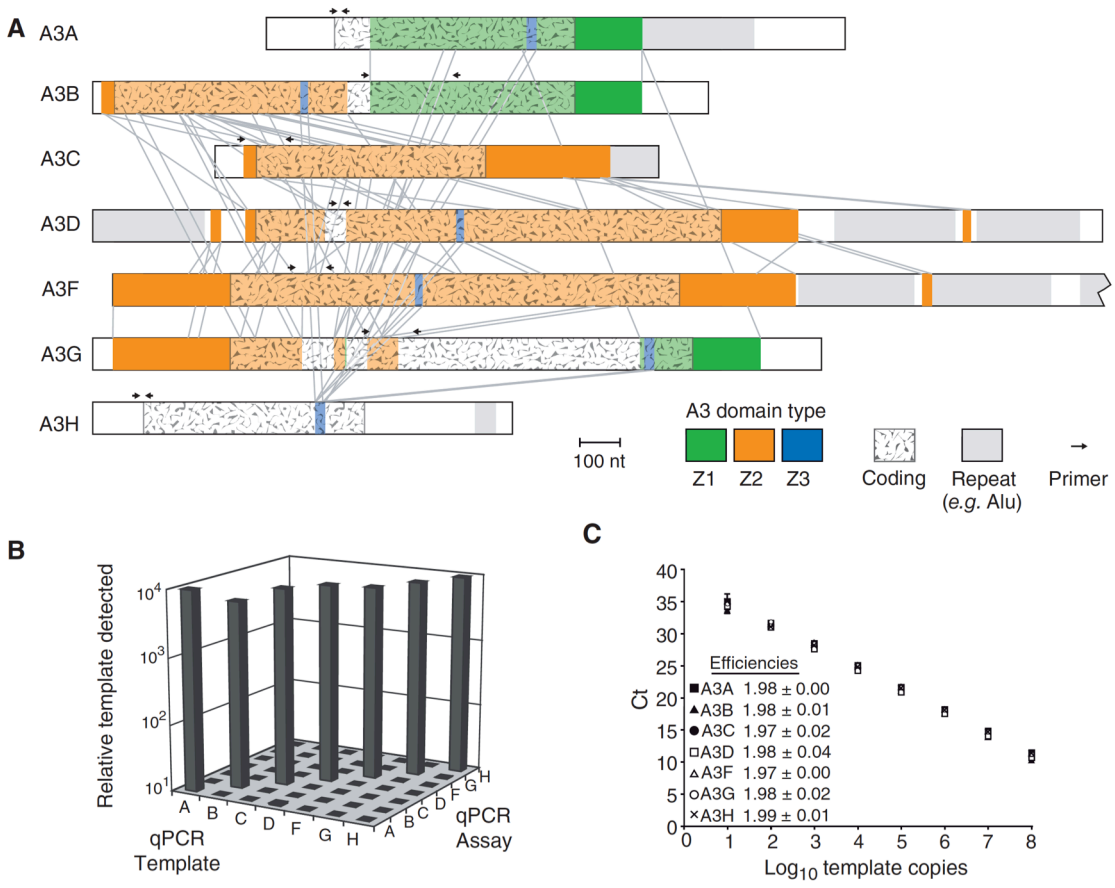

A
A3A
A3B
A3C
A3D
A3F
A3G
A3H
100 nt
A3 domain type
Z1
Z2
Z3
Coding
Repeat
(e.g. Alu)
Primer
B
Relative template detected
10^4
10^3
10^2
10^1
A B C D F G H
qPCR
Template
A B C D F G H
qPCR
Assay
C
Ct
40
35
30
25
20
15
10
5
0
Efficiencies
A3A 1.98 ± 0.00
A3B 1.98 ± 0.01
A3C 1.97 ± 0.02
A3D 1.98 ± 0.04
A3F 1.97 ± 0.00
A3G 1.98 ± 0.02
A3H 1.99 ± 0.01
0 1 2 3 4 5 6 7 8
Log10 template copies

Figure 2-1. A panel of quantitative PCR assays to monitor *APOBEC3* mRNA levels. (A) Overview of *A3* mRNA features. Each *A3* mRNA is depicted to scale except 2100 nucleotides of the *A3F* 3' UTR are not shown. Inter-*A3* regions of 90% or greater identity over ≥18 nucleotides are highlighted in color.

(B) A histogram depicting the results of running each *A3* qPCR assay against each of the seven *A3* cDNA control templates.

(C) A graph showing the *A3* qPCR assay amplification ranges and efficiencies. The mean and standard deviation of two independent experiments each consisting of three replica reactions is shown (in most instances the error is smaller than the symbol).

Gene symbol	mRNA NCBI accession	5′ Primer Name	Seq (5′–3′)	3′ Primer name	Seq (5′–3′)	Probe name	Seq[a]
APOBEC3s							
APOBEC3A	NM_145699	RSH2742	gagaagggacaagcacatgg	RSH2743	tggatccatcaagtgtctgg	UPL26	ctgggctg
APOBEC3B	NM_004900	RSH3220	gacctttggtccttcgac	RSH3221	gcacagccccaggagaag	UPL1	cctggagc
APOBEC3C	NM_014508	RSH3085	agcgcttcagaaaagagtgg	RSH3086	aagtttcgttccgatcgttg	UPL155	ttgccttc
APOBEC3D	NM_152426	RSH2749	acccaaacgtcagtcgaatc	RSH2750	cacatttctgcgtggttctc	UPL51	ggcaggag
APOBEC3F	NM_145298	RSH2751	ccgtttggacgcaaagat	RSH2752	ccaggtgatctggaaacactt	UPL27	gctgcctg
APOBEC3G	NM_021822	RSH2753	ccgaggacccgaaggttac	RSH2754	tccaacagtgctgaaattcg	UPL79	ccaggagg
APOBEC3H	NM_181773	RSH2757	agctgtggccagaagcac	RSH2758	cggaatgtttcggctgtt	UPL21	tggctctg
Reference genes							
TBP	NM_003194	RSH3231	cccatgactcccatgacc	RSH3232	tttacaaccaagattcactgtgg	UPL51	ggcaggag
RPL13A	NM_012423	RSH3227	ctggaccgtctcaaggtgtt	RSH3228	gccccagataggcaaacttt	UPL74	ctgctgcc
HPRT1	NM_000194	RSH2959	tgaccttgatttattttgcatacc	RSH2960	cgagcaagacgttcagtcct	UPL73	gctgagga

[a]It is not known whether the UPL probes correspond to the coding or template DNA strands of their target sequences (Roche proprietary information).

Table 1. *APOBEC3* qPCR primers and probes.

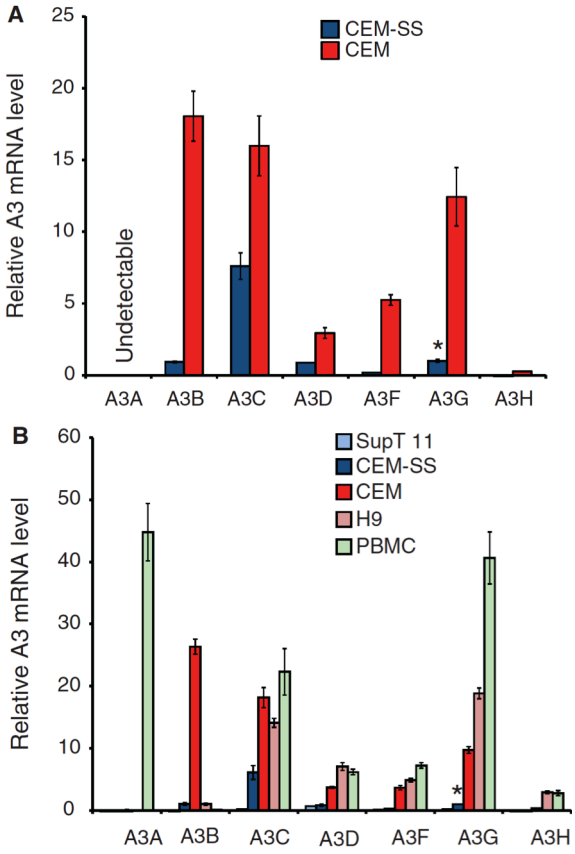

A
Relative A3 mRNA level
25
20
15
10
5
0
CEM-SS
CEM
Undetectable
*
A3A
A3B
A3C
A3D
A3F
A3G
A3H
B
Relative A3 mRNA level
60
50
40
30
20
10
0
SupT 11
CEM-SS
CEM
H9
PBMC
*
A3A
A3B
A3C
A3D
A3F
A3G
A3H

Figure 2-2. *APOBEC3* expression in human T cell lines and naïve PBMCs. (A) *A3* expression in the permissive T cell line CEM-SS and the non-permissive line CEM. Mean values and standard deviations of three independent qPCR reactions are shown for each condition. Expression is normalized to the reference gene *TBP* and the level of CEM-SS *A3G* is set to 1 to facilitate comparison (denoted by the asterisk).

(B) *A3* expression in the permissive T cell lines SupT11 and CEM-SS in comparison to non-permissive lines CEM and H9. The expressed *A3* repertoire in PBMCs is shown for comparison (data from an independent experiment in which CEM and CEM-SS yielded results similar to those shown here). The experimental parameters are identical to those used in panel A.

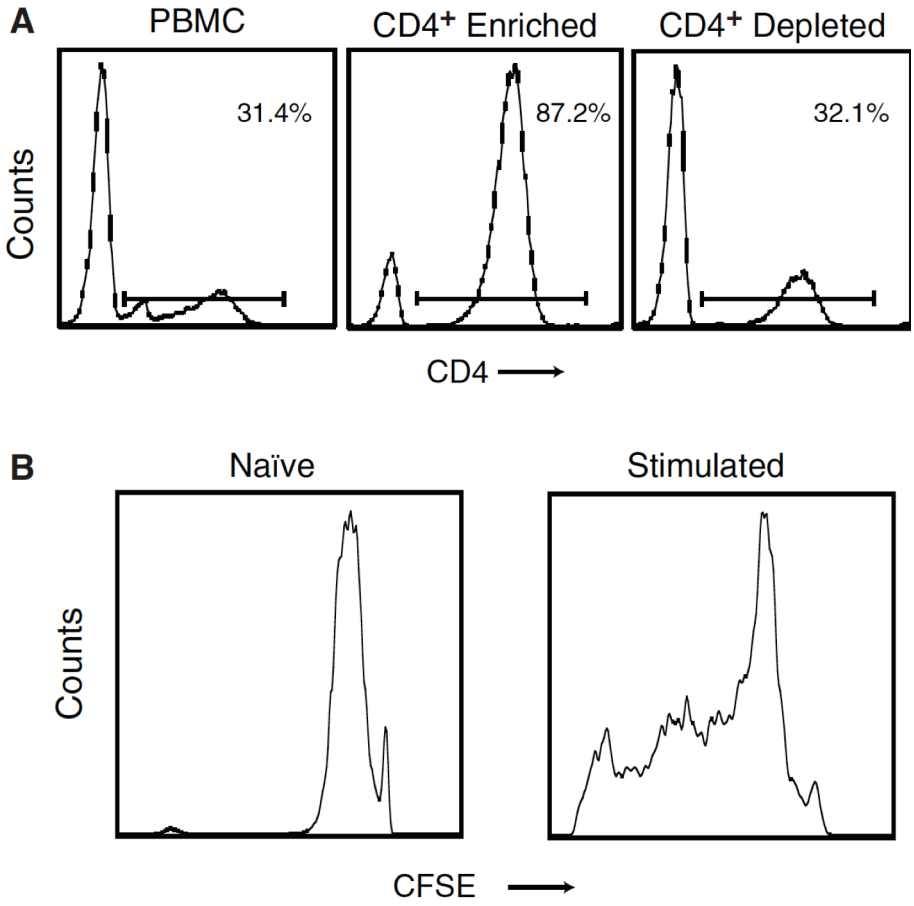

A
PBMC
CD4+ Enriched
CD4+ Depleted
31.4%
87.2%
32.1%
Counts
CD4
B
Naïve
Stimulated
Counts
CFSE
C
CD4+ Naïve
CD4+ Stimulated
CEM
Relative A3 mRNA level
0
10
20
30
40
50
60
70
80
A3A
A3B
A3C
A3D
A3F
A3G
A3H
*

Figure 2-3. *APOBEC3* expression in naïve and stimulated CD4+ lymphocytes. (A) Flow cytometry histograms depicting the results of CD4+ lymphocyte purification by negative selection.

(B) Flow cytometry histograms of CFSE-labeled cells 4 days after mock or IL-2/PHA treatment, naïve and stimulated, respectively.

(C) *A3* expression in naïve and 3 day stimulated CD4+ lymphocytes. Data from CEM are shown for comparison. Expression is normalized to *TBP* and the level of *A3H* in CEM is set to 1 (denoted by the asterisk). Mean values and standard deviations of three independent qPCR reactions are shown for each condition.

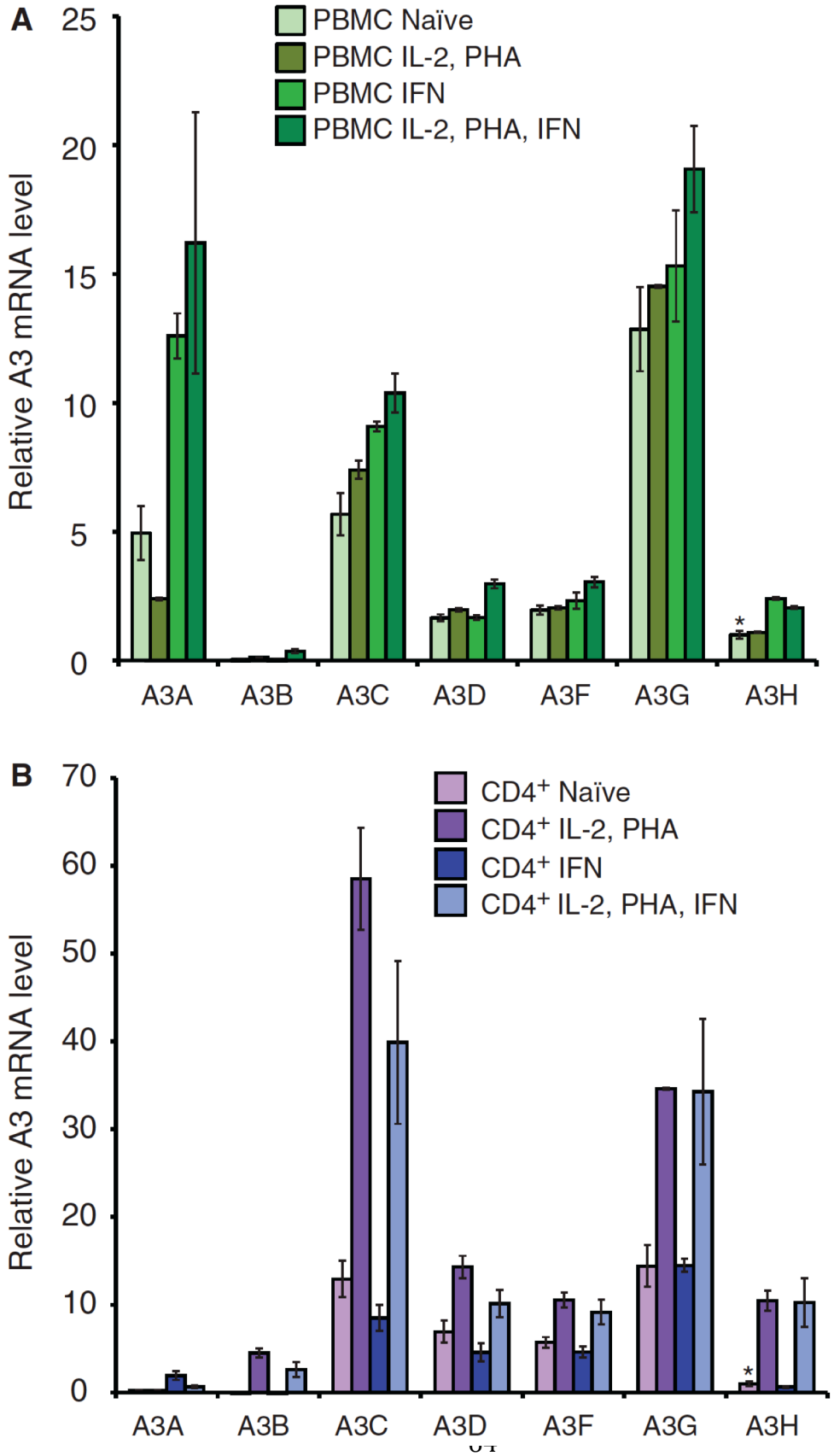

A
Relative A3 mRNA level
25
20
15
10
5
0
PBMC Naïve
PBMC IL-2, PHA
PBMC IFN
PBMC IL-2, PHA, IFN
*
A3A
A3B
A3C
A3D
A3F
A3G
A3H
B
Relative A3 mRNA level
70
60
50
40
30
20
10
0
CD4+ Naïve
CD4+ IL-2, PHA
CD4+ IFN
CD4+ IL-2, PHA, IFN
*
A3A
A3B
A3C
A3D
A3F
A3G
A3H

Figure 2-4. IFN induces *APOBEC3* expression in PBMCs but not CD4+ T cells. The relative *A3* mRNA levels in (A) PBMCs and (B) CD4+ T lymphocytes treated for 48 hours with IL-2/PHA and/or IFN. Expression is normalized to *TBP* and the naïve *A3H* level is set to 1 (denoted by the asterisk). Mean values and standard deviations of three independent qPCR reactions are shown for each condition.

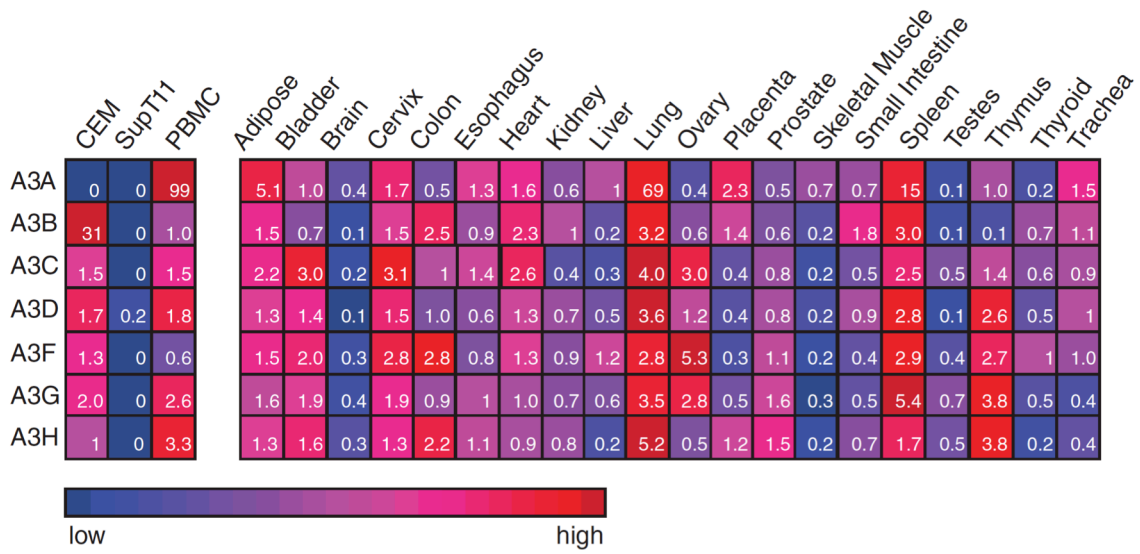


Figure 2-5. *APOBEC3* expression in human tissues. A summary of qPCR data showing the relative *A3* mRNA levels in the indicated cells and tissues. The color scheme provides qualitative information, as the full range of blue (low expression) to red (high expression) color is used for each row of data. The inset numbers represent the relative levels of each *A3* mRNA across the panel with the median value in each row set to 1. Since the assay efficiencies are almost identical (Figure 1C), these quantitative data can be used to compare any of the values within the table. Three replicas were done for each condition and the average values were used to construct the table (the errors were less than 10% and are not shown for simplicity).

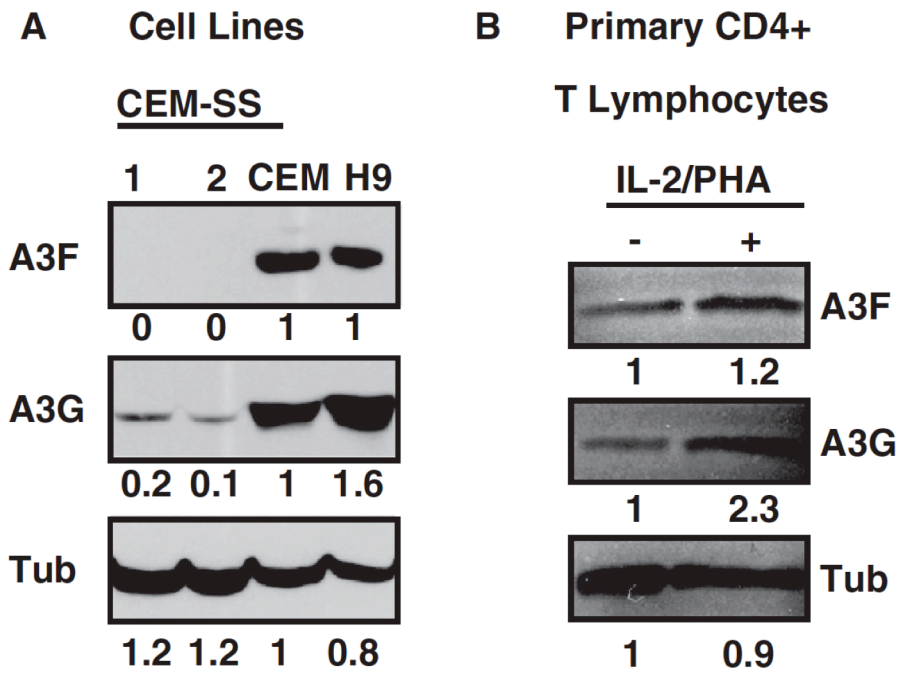


Figure 2-6. APOBEC3 immunoblots. (A) T cell line or (B) primary CD4+ lymphocyte (IL-2/PHA or mock treated) protein extracts were examined by immunoblotting with antibodies specific to A3F or A3G. The blots were stripped and re-probed with anti-tubulin to control for protein loading.

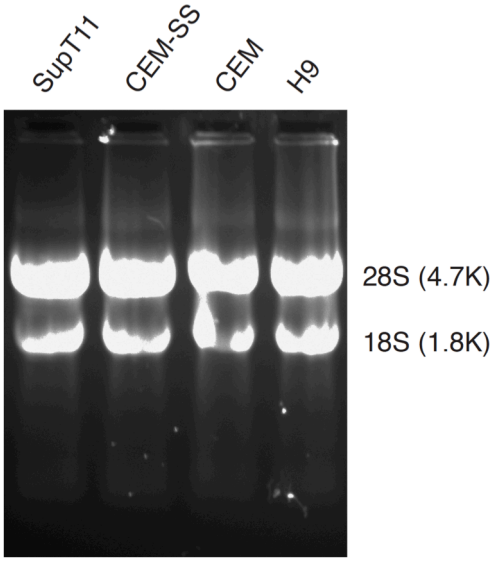


Figure 2-S1. Formamide agarose gel analysis of representative total RNA preparations from the indicated cell lines (10 μg/lane). The large and small ribosomal RNAs are indicated.

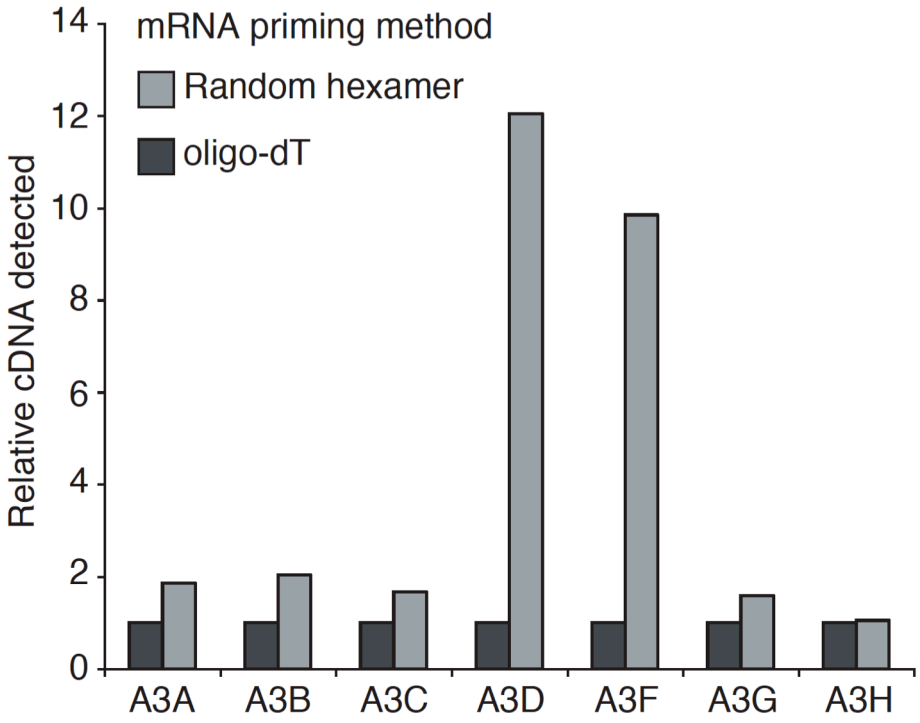


Figure 2-S2. A comparison of qPCR results using cDNA templates primed by poly-dT or random hexamers. See the main text and methods for details.

Gene	cDNA RefSeq[a]	Primer Name	Tm (˚C)[b]	Fraction identical to indicated cDNA sequences[c]						
				A3A	***A3B***	***A3C***	***A3D***	***A3F***	***A3G***	***A3H***
A3A	NM_145699	RSH2742	60.7	20/20	14/20(1) & 10/20	11/20				
		RSH2743	59.5	20/20	11/20				19/20 (1)	
A3B	NM_004900	RSH3220	59.5		19/19					
		RSH3221	62.0	18/18	18/18					13/18
A3C	NM_014508	RSH3085	59.8			20/20			11/20	
		RSH3086	60.1			20/20				
A3D	NM_152426	RSH2749	60.0				20/20			
		RSH2750	59.3		16/20	12/20 (1)	20/20	16/20		
A3F	NM_145298	RSH2751	59.6					18/18	15/18	
		RSH2752	59.0		21/21	13/21 (1)	15/21	21/21		
A3G	NM_021822	RSH2753	61.8						19/19	
		RSH2754	59.8	14/20 (1)					20/20	
A3H	NM_181773	RSH2757	60.1						12/18 (1)	18/18
		RSH2758	59.6							18/18

(a) NCBI accession number of reference cDNA sequence (RefSeq) for each A3.
(b) Primer melting temperature as calculated by primer3 (version 0.4.0; http://primer3.sourceforge.net/), using default parameters.
(c) Shown are all regions of identity between primer and A3 cDNA sequences greater than or equal to 10 nucleotides, as determined by blasting (http://blast.ncbi.nlm.nih.gov/) the primer sequences against a database of the A3 cDNA reference sequences (RefSeqs). In cases where the region of identity contained a mismatch, the number of mismatches is indicated in parenthesis. (BLAST queries run using blastall program (version 2.2.20; http://blast.ncbi.nlm.nih.gov/Blast.cgi), using parameters: -p blastn -q 1 -W 7).

Table 2-S1. *APOBEC3* qPCR primer specificities

CHAPTER 3

Endogenous Origins of HIV-1 G-to-A Hypermutation and Restriction in the Nonpermissive T Cell Line CEM2n

Reprinted with permission from: Eric W. Refsland, Judd F. Hultquist, and Reuben S. Harris. (2012) PLoS Pathogens 8:e1002800.

Authors' Contributions

EWR designed and performed all of the experiments, wrote and revised the manuscript, and designed all of the figures except as follows. JFH aided in the HIV infection experiments and revised the manuscript. RSH aided in the design of the experiments and in manuscript preparation/revision.

FOREWORD

HIV uses its accessory protein Vif to shield itself from lethal mutagenesis by cellular APOBEC3 proteins. The APOBEC3-Vif interaction is therefore a promising target for drug development, and a clear definition of all proteins involved on both sides of this critical host-pathogen conflict is crucial. However, apart from APOBEC3G, conflicting over-expression studies have been published on the six other human APOBEC3s, some supporting and some refuting each protein's involvement in HIV restriction. Here, we use a combination of gene targeting and RNA interference to define the endogenous sources of Vif-deficient HIV restriction and G-to-A hypermutation in a nonpermissive T cell line. As anticipated, APOBEC3G contributed partially to restriction and proved to be the source of hypermutations at 5'GG dinucleotides. However, a major and unexpected role for APOBEC3D became evident in the *APOBEC3F*-null background, with both proteins combining to suppress viral infectivity and form hypermutations at 5'GA dinucleotides. We conclude that endogenous levels of APOBEC3D, APOBEC3F, and APOBEC3G act redundantly to restrict Vif-deficient HIV and produce the two hallmark dinucleotide hypermutation patterns observed in patient-derived viral sequences. Therapeutic strategies that unleash this full 'swarm' of restrictive APOBEC3 proteins are likely to be more effective than those that focus on enabling a single enzyme.

SUMMARY

The DNA deaminase APOBEC3G converts cytosines to uracils in retroviral cDNA, which are immortalized as genomic strand G-to-A hypermutations by reverse transcription. A single round of APOBEC3G-dependent mutagenesis can be catastrophic, but evidence suggests that sublethal levels contribute to viral genetic diversity and the associated problems of drug resistance and immune escape. APOBEC3G exhibits an intrinsic preference for the second cytosine in a 5'CC dinucleotide motif leading to 5'GG-to-AG mutations. However, an additional hypermutation signature is commonly observed in proviral sequences from HIV-1 infected patients, 5'GA-to-AA, and it has been attributed controversially to one or more of the six other APOBEC3 deaminases. An unambiguous resolution of this problem has been difficult to achieve, in part due to dominant effects of protein over-expression. Here, we employ gene targeting to dissect the endogenous APOBEC3 contribution to Vif-deficient HIV-1 restriction and hypermutation in a nonpermissive T cell line CEM2n. We report that *APOBEC3G*-null cells, as predicted from previous studies, lose the capacity to inflict 5'GG-to-AG mutations. In contrast, *APOBEC3F*-null cells produced viruses with near-normal mutational patterns. Systematic knockdown of other *APOBEC3* genes in an *APOBEC3F*-null background revealed a significant contribution from APOBEC3D in promoting 5'GA-to-AA hypermutations. Furthermore, Vif-deficient HIV-1 restriction was strong in parental CEM2n and *APOBEC3D*-knockdown cells, partially alleviated in *APOBEC3G*- or *APOBEC3F*-null cells, further alleviated in *APOBEC3F*-null/*APOBEC3D*-knockdown cells, and alleviated to the greatest extent in *APOBEC3F*-null/*APOBEC3G*-knockdown cells revealing clear redundancy in the HIV-1 restriction mechanism. We conclude that

endogenous levels of APOBEC3D, APOBEC3F, and APOBEC3G combine to restrict Vif-deficient HIV-1 and cause the hallmark dinucleotide hypermutation patterns in CEM2n. Primary T lymphocytes express a similar set of *APOBEC3* genes suggesting that the same repertoire may be important *in vivo*.

INTRODUCTION

Human cells can express up to seven APOBEC3 (A3) proteins: A3A, A3B, A3C, A3D, A3F, A3G, and A3H (Koning et al., 2009; Refsland et al., 2010). A3G is the archetypal restriction factor, capable of restricting Vif-deficient HIV-1 (hereafter HIV) by packaging into viral cores and then suppressing reverse transcription and deaminating viral cDNA cytosines to uracils (C-to-U) (reviewed by (Albin and Harris, 2010; Malim and Emerman, 2008a)). The hallmark of A3G activity is viral genomic plus-strand G-to-A mutations within 5'GG-to-AG dinucleotide motifs, which reflects its minus-strand 5'CC-to-CU preference (Harris et al., 2003; Yu et al., 2004a). However, HIV single-cycle and spreading infection experiments have yet to provide an overall consensus to explain the additional 5'GA-to-AA dinucleotide bias that is also commonly found in patient-derived viral sequences (Caride et al., 2002; Fitzgibbon et al., 1993; Gandhi et al., 2008; Janini et al., 2001; Wood et al., 2009). In fact, over-expression studies have implicated all six of the other A3 proteins in generating this mutation pattern, with multiple studies for and against each enzyme (reviewed by (Albin and Harris, 2010)). For example, despite several early studies strongly implicating A3F as a major source of the 5'GA-to-AA mutations ((Bishop et al., 2004; Liddament et al., 2004; Wiegand et al., 2004; Zheng et

al., 2004) and nearly twenty more thereafter), two recent papers have questioned its relevance to HIV restriction and hypermutation (Miyagi et al., 2010; Mulder et al., 2010). The data are even murkier and more conflicting for the other five human APOBEC3 proteins (reviewed by (Albin and Harris, 2010); summarized in Discussion).

Three major problems have made it difficult to address which endogenous A3 proteins cause HIV restriction and hypermutation. First, most prior studies have relied on transient or stable over-expression of a single A3 coupled to assays for viral infectivity and hypermutation. Although powerful for answering some questions, over-expression of a dominant and active DNA deaminase may overwhelm regulatory mechanisms and adversely affect the cell, result in nonspecific packaging into the virus, and create sequence artifacts by gratuitous deamination of non-productive viral replication intermediates. Most over-expression approaches are complicated further by being done in HEK293 (kidney) or HeLa (cervical) cells, which may not recapitulate as many aspects of restriction and/or viral replication as CD4+ T cell lines. Even best attempts to stably express physiological levels of APOBEC3 proteins in permissive CD4+ T cell lines are imperfect, because each A3 is expressed without other family members (*i.e.*, out of normal endogenous context) and the expression of each protein is driven by heterologous promoters with foreign 5' and 3' untranslated sequences that are unlikely to be responsive to cellular signaling pathways triggered by viral infection and/or interferon signaling (Albin et al., 2010a; Haché et al., 2008; Hultquist et al., 2011; Mariani et al., 2003b; Miyagi et al., 2010; Miyagi et al., 2007; Schumacher et al., 2008). Second, the seven human *A3* genes share high levels of nucleotide identity, which has hindered the development of gene-specific quantitative real-time (Q)-PCR assays and knockdown

reagents. This problem is in the midst of being overcome with the development of robust Q-PCR assays (Koning et al., 2009; Refsland et al., 2010) and the creation of gene-specific knockdown constructs (this study and (Berger et al., 2011; Kamata et al., 2009; Wissing et al., 2011)). Finally, the field has yet to benefit from a robust genetic system, because HIV does not replicate in mouse models and the vast majority of human somatic cell lines are polyploid and/or difficult to engineer.

In this study, we use gene targeting and knockdown experiments to systematically interrogate the impact of the endogenous *A3* repertoire on Vif-deficient HIV replication in the near-diploid T cell line CEM2n. Null clones demonstrated that A3G is solely responsible for HIV 5'GG-to-AG hypermutations. *A3F*-null clones inflicted near-normal hypermutation patterns in Vif-deficient HIV, demonstrating that this enzyme alone is not responsible for 5'GA-to-AA hypermutations. Systematic depletion of other expressed *A3* mRNAs in the *A3F*-null background revealed an unanticipated, major role for A3D in generating 5'GA-to-AA hypermutations and restricting Vif-deficient HIV replication. We conclude that endogenous A3D and A3F combine to generate hallmark 5'GA-to-AA hypermutations and A3G generates the 5'GG-to-AG hypermutations. All three enzymes work together to suppress Vif-deficient HIV replication and generate the classical non-permissive phenotype.

RESULTS

A New Model System for T Cell Genetics

The human T cell line CEM was originally isolated from a 3 year-old female with acute leukemia (Foley et al., 1965). It is a model T cell system for HIV research that, along with its derivative CEM-SS, enabled the identification of A3G as a dominant HIV restriction factor (Sheehy et al., 2002). However, our CEM laboratory stock proved sub-ideal for gene targeting because it tested near tetraploid (**Figure 3-S1**). We noticed however that the original cytogenetic analyses of CEM had 47 chromosomes and a range of 41-95 (Foley et al., 1965). We therefore obtained an early stock of CEM (CCRF-CEM), generated subclones by limiting dilution, and measured DNA content by flow cytometry. In comparison to our original CEM line, several of these subclones had half the DNA content and were most likely diploid (**Figure 3-1A**).

One representative subclone, hereafter called CEM2n, was selected for further characterization. Flow cytometry showed that CEM2n expresses high levels of CD4 and the HIV co-receptor CXCR4 (**Figure 3-1B**). CEM2n still manifests the classic non-permissive phenotype by supporting wildtype HIV replication while restricting Vif-deficient HIV, similar to our tetraploid non-permissive line CEM (**Figure 3-1C**), as described originally (Gabuzda et al., 1992; Sheehy et al., 2002). As anticipated by this phenotype, Q-PCR demonstrated expression of multiple *A3* mRNAs (**Figure 3-S2** & below). Finally, karyotype analysis showed that CEM2n is near-diploid, with a total of 47 chromosomes, including three copies of chromosome 20 and a common T cell leukemia reciprocal translocation (**Figure 3-1D**). These characteristics indicated that CEM2n

would be an appropriate model system to delineate the endogenous A3s involved in HIV restriction.

Targeted Deletion of *APOBEC3G* in CEM2n

Over 100 reports support a role for A3G in Vif-deficient HIV restriction (reviewed by (Albin and Harris, 2010)). A3G shows a strong bias for 5'GG-to-AG hypermutation, but it also has a secondary preference for 5'GA-to-AA invoking the formal possibility that it alone could be responsible for both dinucleotide signatures (*e.g.*, (Harris et al., 2003; Yu et al., 2004b)). To address this possibility, we used two rounds of rAAV-mediated gene targeting to generate *A3G*-null derivatives. The *A3G* targeting construct replaces exon 3, which encodes the N-terminal zinc-coordinating deaminase domain, with a promoterless drug resistance cassette (**Figure 3-2A**). A correctly targeted *A3G* gene is expected to be null because transcripts originating at the *A3G* promoter will splice to an acceptor sequence within the 5' end of the cassette and then terminate with a polyA sequence at the 3' end of the cassette (*i.e.*, the C-terminal two-thirds of the mRNA and protein should never be expressed). CEM2n was transduced with rAAV-A3G::Neo and drug resistant clones were selected with G418. PCR showed that 6/103 (5.8%) clones were targeted (**Figure 3-S3** & **Table 3-1**).

To delete the remaining *A3G* allele, the drug resistance cassette was removed by transducing a representative clone with a Cre expressing adenovirus, and then subclones with a *loxP*-to-*loxP* recombination event were identified by PCR screening (**Figure 3-S3**). Next, the original rAAV-A3G::Neo construct was used for a second round of gene targeting. 2/86 drug resistant clones were null and 4/86 were retargeted, yielding a second

round targeting frequency of 7.0% (**Table 3-1**). The *A3G*-null clones had undetectable levels of *A3G* mRNA and protein and, importantly, the mRNA levels of all of the flanking *A3* genes and the A3F protein levels were largely unperturbed (**Figure 3-2B** & **C**). The parental CEM2n line and its *A3G*-null derivatives had similar morphologies and growth rates.

To explore the functional consequences of deleting the endogenous *A3G* gene, we performed single-cycle infectivity assays with VSV-G pseudotyped Vif-deficient HIV_{IIIB}. After one full round of replication, new viruses produced from *A3G*-null cells were used to infect CEM-GFP reporter cells and GFP fluorescence was measured 2 days later by flow cytometry to quantify infectivity. In comparison to the fully non-permissive parental line CEM2n, the *A3G*-null derivative lines produced viruses with approximately 10-fold improved infectivity (**Figure 3-2D**). However, these viruses were still 2-fold less infectious than Vif-deficient HIV produced in parallel using the related permissive T cell line, CEM-SS. We note that although CEM-SS is commonly accepted as permissive for Vif-deficient HIV replication, our recent work revealed that it expresses multiple A3s including low levels of A3G (Refsland et al., 2010). Thus, it is quite possible that a fully null derivative of CEM2n could produce even higher levels of infectious Vif-deficient HIV (further supported by experiments described below). Regardless, these data demonstrate that endogenous A3G is not the only factor contributing to Vif-deficient HIV restriction in CEM2n.

To gauge the gross level of G-to-A hypermutation in proviral DNA embedded in the genomes of the CEM-GFP reporter cells used above, we performed a series of differential DNA denaturation (3D-PCR) experiments (Hultquist et al., 2011; Suspène et

al., 2005b). A 511 bp region of the HIV *gag-pol* gene was amplified over a range of PCR denaturation temperatures from 77.2 to 85.5 ^{o}C and subjected to gel electrophoresis. As anticipated, Vif-deficient HIV proviruses derived from non-permissive T cell lines H9 and CEM2n yielded PCR products at low denaturation temperatures, down to 78.4 and 79.4^{o}C, respectively, indicative of high levels of G-to-A hypermutation (**Figure 3-2E**). Vif-deficient HIV proviruses derived from CEM-SS only amplified at high denaturation temperatures, also as expected. In contrast, Vif-deficient proviruses derived from *A3G*-null cells yielded PCR amplicons at temperatures as low as 80.4^{o}C, suggesting major levels of residual hypermutation but lower than those inflicted by the full A3 repertoire in CEM2n.

To examine mutational spectra, high temperature (unbiased) PCR amplicons were cloned and sequenced. As expected, Vif-deficient proviruses derived from the parental cell line CEM2n harbored extensive G-to-A hypermutations in two distinct dinucleotide contexts: 70% 5'GG and 30% 5'GA. This contrasted starkly to proviral sequences derived from *A3G*-null cells, in which the 5'GG-to-AG hypermutations nearly disappear (6%) (**Figure 3-2F**, **Table 3-2**, and **Figure 3-S4**). These results establish A3G as the major source of 5'GG-to-AG hypermutations, consistent with prior over-expression studies indicating that A3G is the only DNA deaminase that prefers minus strand 5'CC target sites (*e.g.*, (Harris et al., 2003; Mangeat et al., 2003; Yu et al., 2004b; Zhang et al., 2003). These mutation spectra also implicate at least one other DNA cytosine deaminase in restricting HIV and generating 5'GA-to-AA hypermutations.

Targeted Deletion of *APOBEC3F* in CEM2n

Q-PCR revealed that CEM2n cells express six of seven *A3* genes, *A3B*, *A3C*, *A3D*, *A3F*, *A3G*, and *A3H*, all of which have been implicated in catalyzing the 5'GA-to-AA hypermutation patterns (**Figure 3-S2**, **Discussion**, and examples (Bishop et al., 2004; Liddament et al., 2004; Wiegand et al., 2004; Zheng et al., 2004)). A3F was our top candidate because (i) *A3F* expression tracks with *A3G* in non-permissive T cell lines, primary lymphocytes, and secondary immune tissues (Hultquist et al., 2011; Liddament et al., 2004; Refsland et al., 2010), (ii) A3F is encapsidated into budding viruses and restricts Vif-deficient HIV when over-expressed in permissive T cell lines (Albin et al., 2010a), (iii) Vif targets A3F for degradation (Albin et al., 2010a; Wiegand et al., 2004; Zheng et al., 2004), (iv) A3F restriction capability and Vif counteraction activity is conserved with rhesus macaque A3F and SIV Vif (Hultquist et al., 2011; Virgen and Hatziioannou, 2007), and (v) Vif-deficient HIV isolates that regain the capacity to replicate on A3F expressing cells invariably restore Vif function (Albin et al., 2010a). However, despite this strong evidence favoring a role for A3F in HIV restriction and hypermutation, recent studies have questioned its importance (Miyagi et al., 2010; Mulder et al., 2010).

To determine the involvement of endogenous A3F in Vif-deficient HIV restriction and hypermutation, we generated *A3F*-null cell lines using two rounds of rAAV-mediated gene targeting as described above (**Figure 3-3A** and **Figure 3-S5**). *A3F*-null clones showed no detectable mRNA or protein, and the mRNA levels of all of the flanking *A3* genes and the A3G protein levels were largely unaffected (**Figure 3-3B** & **C**). Similar to the loss of *A3G*, the deletion of *A3F* resulted in cells semi-permissive for Vif-deficient

HIV (**Figure 3-3D**) as well as in a modest decrease in the overall level of mutation as gauged by 3D-PCR (**Figure 3-3E**). However, we were surprised to find that the hypermutation spectrum of Vif-deficient HIV produced in *A3F*-null cells was indistinguishable from that of the same virus produced in the CEM2n parent, retaining a large percentage of 5'GA-to-AA mutations (**Figure 3-3F**, **Table 3-2**, and **Figure 3-S4**). Thus, the infectivity data showed that endogenous A3F does contribute to Vif-deficient HIV restriction, but the hypermutation data clearly implicate at least one other endogenous 5'TC deaminating A3.

APOBEC3D and APOBEC3F Combine to Cause 5'GA-to-AA Hypermutations

To identify the remaining source of the 5'GA-to-AA hypermutations, we developed a panel of short hairpin (sh)RNA reagents to systematically knockdown the expression of *A3B*, *A3C*, *A3D*, *A3G*, and *A3H* in the *A3F*-null background (**Figure 3-4A**). Knockdown efficiencies ranged from 50-80% and were specific to each intended *A3* mRNA target. These efficiencies may be even higher at the protein level due to the known effects of shRNA in triggering mRNA cleavage/degradation and in suppressing translation (Fabian et al., 2010), but this could only be confirmed for A3G due a lack of specific antibodies for the other A3s (**Figure 3-4B**). The parental CEM2n and *A3F*-null lines were transduced with non-silencing (shNS) and each of the aforementioned shA3 constructs. Pools of shRNA expressing cells were selected with puromycin (co-expressed from the same transducing virus), subjected to knockdown verification by Q-PCR, and infected as above with VSV-G pseudotyped Vif-deficient HIV to determine infectivity levels and hypermutation patterns.

Each individual *A3* knockdown had little impact on the infectivity or the 3D-PCR profile of Vif-deficient HIV produced in the CEM2n parental line (**Figure 3-4C**, **3-4D**). *A3F*-null cells expressing the non-silencing shRNA yielded a significant 7-fold increase in the infectivity of Vif-deficient HIV over parental CEM2n as above (**Figure 3-4C**). Similarly, *A3F*-null cells transduced with shA3B or shA3C knockdown constructs produced Vif-deficient viruses with 8- and 9-fold infectivity increases, indicating no further contribution to restriction from these A3 proteins (**Figure 3-4C**). However, in contrast, *A3F*-null cells transduced with shA3D, shA3G, and shA3H constructs yielded Vif-deficient viruses that were 16-, 36-, and 14-fold more infectious, respectively (**Figure 3-4C**). Remarkably, the *A3F*-null/*A3G*-knockdown cells produced higher levels of infectious Vif-deficient HIV than CEM-SS cells (**Figure 3-4C**).

Proviral DNA sequencing of Vif-deficient viruses produced in *A3F*-null cells in combination with non-specific shRNA (shNS), shA3B, shA3C, or shA3H showed no significant alteration in the fraction of hypermutations that occurred within 5'GA or 5'GG dinucleotides (**Figure 3-4E**, and **Table 3-2**). Over 15 G-to-A mutations were found for all experimental conditions, with one exception being viral DNA originally produced in *A3F*-null/*A3G* knockdown cells, which yielded only 8 mutations (4 GG-to-AG and 4 GA-to-AA). In contrast, *A3F*-null/*A3D* knockdown cells produced almost no hypermutations in the 5'GA-to-AA dinucleotide context: 3% 5'GA, 93% 5'GG, and 4% other (**Figure 3-4E**, **Table 3-2**, and **Figure 3-S4**). This major contribution from A3D was only observed in the absence of A3F and it was rather surprising because most other reports have ascribed modest or no antiretroviral activity to this enzyme ((Bishop et al., 2004; Dang et al., 2011; Dang et al., 2006; Haché et al., 2008); reviewed recently (Albin

and Harris, 2010)). We conclude that, in CEM2n cells, endogenous A3D and A3F combine to restrict Vif-deficient HIV (with A3G) and, importantly, work together (mostly without A3G) to inflict 5'GA-to-AA hypermutations.

Knockout and Knockdown Subclone Analyses

An unavoidable consequence of gene targeting is the clonal nature of the procedure required to generate biallelic knockout derivatives of a parental cell line. Additionally, the previously described single-cycle assays on *A3G*-null and *A3F*-null cell lines were done several months apart with virus stocks produced at different times. To minimize these potential effects, two independently derived subclones of each of the knockout lines, *A3G*-null and *A3F*-null, as well as two independent subclones from the *A3D* and *A3G* shRNA knockdown pools were generated and assayed in parallel. Q-PCR and immunoblotting was used to confirm knockdown efficiencies, 60-85% for *A3D* and 60-70% for *A3G* (**Figure 3-5A** & **B**). These cell lines were infected with VSV-G pseudotyped Vif-deficient HIV to determine infectivity levels after approximately one round of replication (**Figure 3-5C**). As above, deletion of either *A3G* or *A3F* increased viral infectivity an average of 38- or 20-fold, respectively. Knockdown of either *A3D* or *A3G* in the CEM2n background had little effect. In contrast, knockdown of *A3D* or *A3G* in the *A3F*-null background resulted in additional increases in Vif-deficient virus infectivity, averaging 27- and 49-fold higher than the baseline level in fully restrictive CEM2n cells expressing non-specific shRNA.

To assess the impact of *A3* deletion and knockdown on HIV replication over time, a series of parallel spreading infections were initiated on the same panel of cell lines

(**Figure 3-5A** & **B**). Cells were infected at a multiplicity of infection of 1% with either Vif-proficient or Vif-deficient HIV_{IIIB}. Every 2-3 days supernatants were removed to infect the reporter line CEM-GFP to assay live virus and, in parallel, to measure p24 levels (**Figure 3-5D** and **Figure 3-S6**). Vif-proficient virus replicated on every cell line tested with peaks of infection occurring on days 7 to 13. The precise reason(s) for this kinetic variation is unclear but it does not seem to correlate with *A3* genotype (compare with **Figure 3-S6**). As expected, Vif-deficient HIV replication was fully restricted in CEM2n cells but peaked readily in CEM-SS. *A3G*-null and *A3F*-null cells failed to produce detectable levels of infectious Vif-deficient virus, as monitored by the CEM-GFP (live virus) reporter system (although modest increases in p24 levels were detected in cell-free supernatants). In addition, knockdown of *A3D* or *A3G* in either the CEM2n background, or somewhat surprisingly, in the *A3F*-null background was unable to render cells fully permissive for Vif-deficient virus replication. Taken together, even though strong infectivity recoveries were evident in single round infections of cells completely lacking *A3G* or *A3F* or cells lacking *A3F* plus knocked-down *A3D* or *A3G*, the remaining endogenous A3s were still sufficient to restrict the spread of Vif-deficient virus (*i.e.*, those A3s that were not manipulated by knockout or knockdown). These data combined to suggest that the simultaneous elimination of A3D, A3F, and A3G (and possibly also A3H) would ultimately be necessary to render CEM2n fully permissive for Vif-deficient HIV replication.

DISCUSSION

Null mutations are a gold standard of genetics as they enable a definitive assessment of a given gene's function by comparing the phenotype of the wildtype parental state with that of an isogenic null derivative. Here, we report the identification of a new T cell line, CEM2n, derived from the common parental line CCRF-CEM. CEM2n is near-diploid, expresses CD4 and CXCR4, supports Vif-proficient but not Vif-deficient HIV replication, expresses a complex *A3* gene repertoire (similar to primary CD4+ T lymphocytes; **Figure 3-S2** & (Refsland et al., 2010)), and is amendable to genetic manipulation by multiple methods including RNAi and gene targeting. We defined the HIV-restrictive A3 repertoire in this cell line by constructing *A3G*- and *A3F*-null derivatives, systematically depleting each expressed *A3* with specific shRNA constructs, and performing a series of Vif-deficient HIV infectivity and proviral DNA hypermutation experiments. These data enable the conclusions that endogenous levels of A3D, A3F, and A3G combine to limit the infectivity of Vif-deficient HIV, that A3G acts alone to inflict 5'GG-to-AG hypermutations, and that A3D and A3F work together to elicit 5'GA-to-AA hypermutations. Since both 5'GG and 5'GA mutational patterns are common in patient-derived HIV proviral DNA sequences and a CEM2n-like *A3* repertoire is expressed in primary CD4+ T lymphocytes, we hypothesize that these three A3s will also be responsible for HIV restriction and hypermutation *in vivo* (*e.g.*, (Janini et al., 2001; Kieffer et al., 2005; Kijak et al., 2008; Land et al., 2008; Piantadosi et al., 2009)).

However, despite considerable efforts from our group and others, A3H remains a debatable factor in HIV restriction. This is due partly to the fact that some A3H haplotypes are more stable at the protein level (haplotypes II, V, VII >> I, III, IV, VI

(OhAinle et al., 2008; Ooms et al.; Wang et al.)). Over-expression studies are mostly consistent showing that A3H-hapII can restrict Vif-deficient HIV, that Vif can counteract this activity, and that these interactions are conserved between the A3H orthologs of other species and their cognate lentiviral Vif proteins (Hultquist et al., 2011; LaRue et al., 2010; Li et al., 2010; OhAinle et al., 2008; Ooms et al., 2010; Zhen et al., 2010). A3H is also expressed in primary CD4+ lymphocytes and induced upon HIV infection (Hultquist et al., 2011; LaRue et al., 2010). However, our present studies were not able to unambiguously address whether this enzyme contributes to HIV restriction because endogenous *A3H* mRNA levels in CEM2n are ~20-fold lower than those in primary CD4+ lymphocytes (Refsland et al., 2010) (**Figure 3-S2**). Even so, we observed a significant increase in Vif-deficient HIV infectivity in *A3F*-null/*A3H*-knockdown cells in comparison to the *A3F*-null line suggesting that even sub-physiological levels of endogenous A3H may be restrictive. This observation is additionally interesting in light of sequencing data showing that CEM2n is homozygous for the haplotype II genotype.

Our present studies will also help resolve literature conflicts on the topic of Vif-deficient HIV restriction by A3A, A3B, and A3C (*e.g.*, A3A: (Aguiar et al., 2008; Berger et al., 2011; Bishop et al., 2004; Bogerd et al., 2006a; Goila-Gaur et al., 2007; Marin et al., 2008; Wiegand et al., 2004); A3B: (Bishop et al., 2004; Doehle et al., 2005a; Haché et al., 2008; Hultquist et al., 2011; Yu et al., 2004a); A3C: (Bishop et al., 2004; Bogerd et al., 2006b; Langlois et al., 2005; Wiegand et al., 2004; Yu et al., 2004a; Zheng et al., 2004)). A3A is not expressed in CD4+ T lymphocytes or cell lines and is therefore unlikely to have a role (Aguiar et al., 2008; Koning et al., 2009; Refsland et al., 2010; Stenglein et al., 2010). It is also Vif-resistant, suggesting that it is of little threat to the

virus, and incapable of restricting Vif-deficient HIV even upon stable over-expression in permissive T cell lines (*e.g.*, (Hultquist et al., 2011)). A3B is expressed at low levels in CD4+ T lymphocytes and some T cell lines such as CEM, and its over-expression potently restricts HIV in the HEK293T model system (Bishop et al., 2004; Doehle et al., 2005a; Hultquist et al., 2011; Koning et al., 2009; Refsland et al., 2010; Yu et al., 2004a). However, A3B is also insensitive to HIV Vif and does not restrict Vif-deficient HIV or inflict hypermutations when stably over-expressed in permissive T cell lines (Bishop et al., 2004; Doehle et al., 2005a; Haché et al., 2008; Hultquist et al., 2011; OhAinle et al., 2008; Rose et al., 2005; Yu et al., 2004a). Taken together with data shown here that *A3B* knockdown in parental CEM2n or the *A3F*-null background has no impact on virus infectivity or hypermutation profiles, we conclude that endogenous A3B also does not contribute to HIV restriction. A3C is highly expressed in primary CD4+ T cells and T cell lines, exhibits a 5'GA-to-AA hypermutation preference (especially with SIV-based viral substrates), and is susceptible to HIV Vif (Langlois et al., 2005; Refsland et al., 2010; Yu et al., 2004a). However, when over-expressed in permissive T cell lines SupT11 or CEM-SS, A3C does not encapsidate or restrict Vif-deficient HIV replication (Hultquist et al., 2011). Moreover, since *A3C* knockdown in the parental CEM2n or the *A3F*-null background had no impact on virus infectivity in our present studies, we conclude that endogenous A3C is also unlikely to contribute to Vif-deficient HIV restriction.

As elaborated in the Introduction, many prior published reports have been susceptible to a major weakness by being dependent upon enforced cDNA over-expression. Here, we overcome this drawback by identifying a new genetic system for

host factor studies and systematically deleting and/or depleting endogenous *A3* genes. Our studies demonstrate that A3G is the sole source of 5'GG-to-AG hypermutations and, formally, that it is a minor (at best) contributor to 5'GA-to-AA hypermutations. Our studies are also consistent with the phenotypes of the molecular clones engineered to be preferentially susceptible to A3F or A3G (Russell and Pathak, 2007; Simon et al., 2005). However, our studies are additionally unique by providing an unanticipated demonstration of a major role for endogenous A3D in inflicting 5'GA-to-AA hypermutations. Thus, virtually all 5'GA-to-AA hypermutations can now be explained by overlapping A3D and A3F activities. Such an unambiguous assignment of function would not have been possible by simple over-expression studies because A3D and A3F have a similar 5'TC deamination preferences. Altogether, we conclude that endogenous levels of three enzymes – A3D, A3F, and A3G – combine to inflict signature G-to-A hypermutations and mediate HIV restriction.

MATERIALS AND METHODS

Cell Lines

All T cell lines were maintained in RPMI supplemented with 10% fetal bovine serum (FBS) and 0.5% penicillin/streptomycin (P/S). Our original CEM (4n) line was obtained from Michael Malim (Sheehy et al., 2002). The initial CCRF-CEM line used here was obtained from the ATCC (cat #CCL119), and it was subcloned by limiting dilution to obtain multiple daughter clones including CEM2n. The HIV infectivity indicator cell line

CEM-GFP was obtained from the AIDS Research and Reference Reagent Program. HEK293T cells were cultured in DMEM supplemented with 10% FBS and 0.5% P/S.

Karyotype Analyses

Giemsa banding and karyotype determination were performed at the University of Minnesota's Cytogenetics Laboratory.

rAAV Targeting Constructs

Generation of rAAV targeting vectors was performed as described (Rago et al., 2007). Homology arms were selected to avoid identity with other *A3* genes and repetitive sequences. CEM2n genomic DNA was prepared using Gentra Puregene Cell Kit (Qiagen) and 1 kb homology arms were amplified using high fidelity PCR (Platinum Taq HiFi, Invitrogen). Primer sequences and genomic location of arms are listed in **Table 3-S1**. Arms were cloned into pJet1.2 using CloneJet PCR cloning kit (Fermentas) and sequence verified. Left arms were digested with *Spe*I and *Not*I and right arms were digested with *Sal*I and *Not*I. pSEPT (Rago et al., 2007) plasmid was digested with *Spe*I and *Sal*I and pAAV-MCS containing the viral ITR was digested using *Not*I. rAAV vectors were then constructed via four-way ligation and purified using standard phenol:chloroform extraction followed by ethanol precipitation. 20 µl of each ligation mixture was used to transform 10 µl electro-competent DH10B cells. Plasmid DNA was harvested from ampicillin resistant colonies and verified by restriction digest and DNA sequencing.

AAV Virus Production, Infection of Target Cells

AAV-2 viral stocks were prepared by co-transfecting HEK293T cells at 50% confluency with sequence verified rAAV targeting vector, pHelper and pAAV-RC (Stratagene) (Trans-IT, Mirus). Three days after transfection, media and cells were collected and subjected to 3 cycles of freeze-thaw-vortex (30 min at -80^{o}C, 10 min thaw at 37^{o}C, 30 s vortex). Cellular debris was removed by centrifuging for 30 min at 12,000 rpm in a table-top centrifuge. rAAV was further purified and concentrated using an AAV Purification ViraKit (Virapur) per manufacturer's instructions. One million CEM2n recipient cells were infected with a range of volumes (2-100 µl) of virus. Three days after infection, cells were seeded into 96-well plates at a density of 1000 cells/well in G418 containing media (1 mg/ml). G418 resistant clones were expanded and harvested for genomic DNA. Clones were screened for homologous recombination events at the desired locus with PCR primers specific to upstream (5'), downstream (3') regions of the targeted allele, and the targeting vector. To recycle the targeting construct for the second allele, clones were infected with a Cre-expressing adenovirus (Ad-Cre-GFP, Vector Biolabs). Cells were cloned by limiting dilution and tested for drug sensitivity.

RNA Isolation, cDNA Synthesis, and Q-PCR

mRNA isolation, reverse transcription and Q-PCR were performed as described (Refsland et al., 2010). RNA from 5 x 10^6 cells was isolated using the RNeasy kit (Qiagen). 1 µg total RNA was used to synthesize cDNA with random hexameric primers (AMV RT; Roche). cDNA levels were quantified using established procedures, primers, and probes (Refsland et al., 2010) using a Roche LightCycler 480. All reactions were

done in triplicate and *A3* levels were normalized to the housekeeping gene *TBP* and presented relative to values for the parental CEM2n line.

Viral Constructs

Vif-proficient (Genbank EU541617) and Vif-deficient (X26, X27) HIV_{IIIB} A200C proviral constructs have been described (Albin et al., 2010a; Haché et al., 2008).

Spreading Infections

Vif-proficient and Vif-deficient HIV spreading infections were performed as previously described (Albin et al., 2010a). For p24 ELISA, anti-p24 mAb (183-H12-5C, NIH ARRRP) coated 96-well plates (maxisorp, Nunc) were incubated with supernatants from infected cultures for 1 hr. Following 3 washes with PBS 0.1% Tween 20 (PBS-T), a second 1 hr incubation with a different anti-p24 mAb (9725, N. Somia) was used to 'sandwich' the p24 antigen. After an additional 3 PBS-T wash steps, p24 was quantified by 0.5 hr incubation with an enzyme-linked secondary goat anti-mouse IgG-2A/HRP followed by 3 PBS-T wash steps and incubation with 3,3',5,5' tetramethybenzidine (TMB) for 6 min. The reaction was stopped upon addition of 1 M H_2SO_4 and absorbance at 450 nm was quantified on a microplate reader (Synergy MX, Biotek).

Single Cycle Infections

HEK293T cells were co-transfected at 50% confluence with a Vif-deficient HIV proviral expression construct and a VSV-G expression construct. Viral containing supernatants were harvested, titered on CEM-GFP, and used to infect CEM2n and A3-null derivatives

at a 25% initial infection. 12 h post-infection cells were washed to remove remaining VSV-G virus and resuspended in RPMI growth media. 36 h later, viral containing supernatants were used to infect CEM-GFP reporter cells and cell and viral particle lysates were prepared.

Immunoblotting

Cell pellets were washed and directly lysed in 2.5x Laemmli sample buffer. Viral particles were isolated from filtered supernatants by centrifugation and then resuspended in 2.5x Laemmli sample buffer. SDS-PAGE and immobilization of protein to PVDF was carried out using the Criterion system (Bio-Rad).

Flow cytometry

Immunostaining with CD4-PC7 and CXCR4-PE (Beckman Coulter) was carried out per the manufacturer's instructions. CEM and CEM2n stained negative for CCR5 (data not shown). Infected CEM-GFP cells were fixed with 4% paraformaldehyde in 1X PBS. Fluorescence was measured on a FACS Canto II instrument (BD), and data were analyzed with FlowJo Flow Cytometry Analysis Software (Version 8.8.6).

shRNA Knockdown Vectors

pLKO.1 lentiviral vectors expressing short-hairpin RNA (shRNA) to A3B (TRCN00001420546), A3C (TRCN0000052102), A3D (TRCN0000154811), A3G (TRCN0000052191), and A3H (TRCN0000051799) were obtained from Open Biosystems. Lentivirus was produced by co-transfecting 50% confluent 293T cells with

an shRNA expression construct, an HIV *gag-pol* helper plasmid, a VSV-G expression construct and Trans-IT (Mirus). Two days after transfection, supernatant was collected and replaced with fresh media. The following day, media was pooled and virus was clarified by passing supernatants through a 0.45 μM PVDF filter. To concentrate the virus, clarified supernatants were centrifuged at 22,000 x g for 2 h. Virus containing pellets were resuspended in 1X PBS. One million CEM2n and *A3F*-null cells were infected with 50 μl of virus and two days post-infection, media was replaced with puromycin containing media (1μg/ml). Subclones were generated by limiting dilution in 96-well plates. Knockdown efficiency and specificity on transduced pools and subclones was assessed with Q-PCR (Refsland et al., 2010).

3D-PCR

3D-PCR was carried out as described (Hultquist et al., 2011). A 511 bp amplicon from the *gag-pol* genes of integrated proviruses was amplified with degenerate primers and quantified (Roche, LightCycler 480). Normalized amounts of integrated provirus were used for a second round of PCR over a range of denaturation temperatures using a gradient thermocycler (Eppendorf).

Mutational Spectra Analysis

48 h after infection of CEM-GFP cells with Vif-deficient HIV, genomic DNA was prepared and PCR was used to generate a 511 bp amplicon over the *gag-pol* region. Primer sequences are listed in **Table 3-S1**. This amplicon was cloned into pJet1.2 using

CloneJet PCR cloning kit (Fermentas) and sequenced. Duplicate sequences were discarded. Sequences were analyzed using Sequencher 4.6 (Gene Codes Corp).

ADDITIONAL CONTRIBUTIONS

We thank F. Bunz for the pSEPT plasmid, E.A. Hendrickson for the pAAV-MCS plasmid, L. Oseth and the UMN Cytogenetics Core for karyotype analysis, N. Somia for p24 mAb, and J. Albin and E.A. Hendrickson for comments on the manuscript. This work was supported by NIH R01 AI064046. EWR was supported in part by an Institute for Molecular Virology Training Program NIH T32 AI083196. JFH was supported in part by University of Minnesota Graduate Entrance and NSF Predoctoral Fellowships.

FIGURES

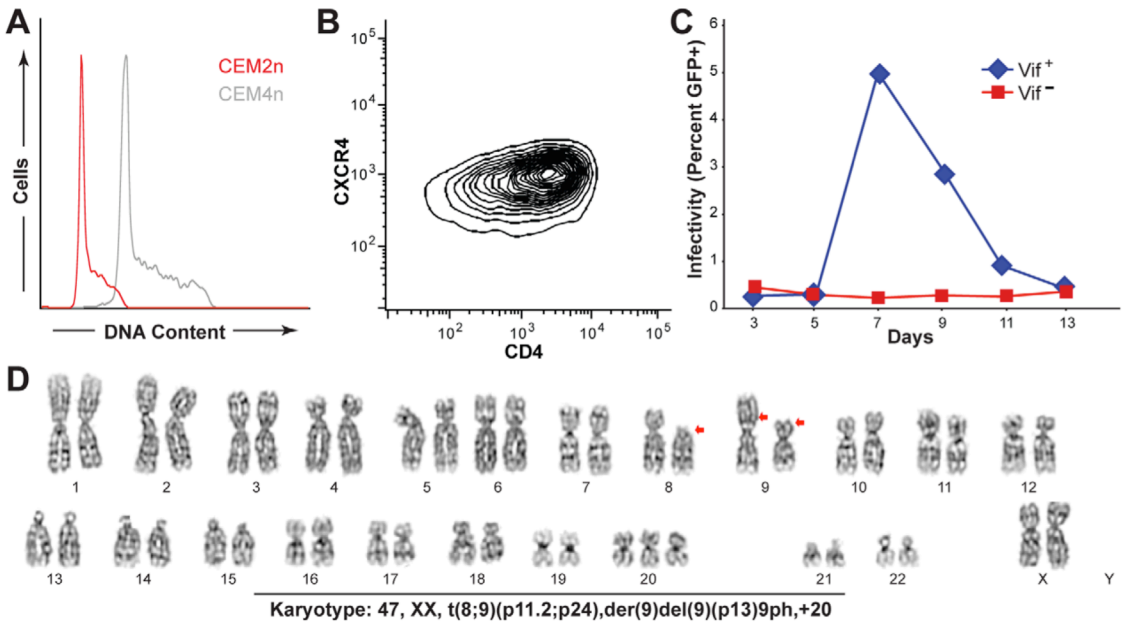


Figure 3-1. CEM2n is a Near-Diploid, Non-Permissive T Cell Line.

(A) Flow cytometric analysis of fixed CEM2n (red) and CEM4n (gray) cells stained with propidium iodide.

(B) Contour plot of CD4 and CXCR4 levels in CEM2n.

(C) HIV spreading infection profiles in CEM2n as monitored by periodic infection of an LTR-GFP reporter cell line, CEM-GFP.

(D) Giemsa-banding karyotype of a representative CEM2n metaphase spread. Red arrows indicate typical lesions in lymphoblastic leukemia.

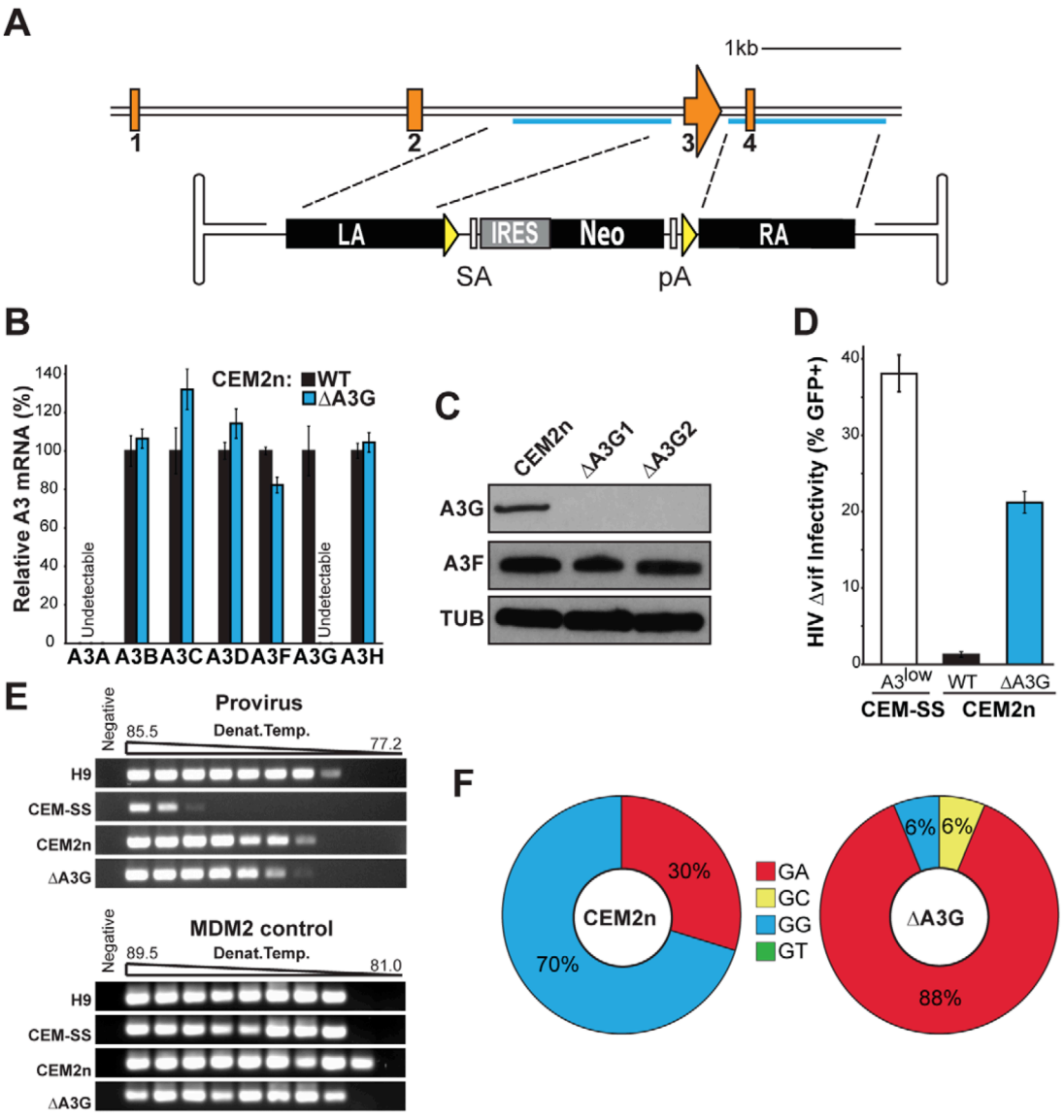

A
1kb
1
2
3
4
LA
IRES
Neo
RA
SA
pA
B
CEM2n: WT ΔA3G
Relative A3 mRNA (%)
Undetectable
Undetectable
A3A A3B A3C A3D A3F A3G A3H
C
CEM2n
ΔA3G1
ΔA3G2
A3G
A3F
TUB
D
HIV Δvif Infectivity (% GFP+)
A3low
WT
ΔA3G
CEM-SS
CEM2n
E
Provirus
Negative
Denat.Temp.
85.5
77.2
H9
CEM-SS
CEM2n
ΔA3G
MDM2 control
Negative
Denat.Temp.
89.5
81.0
H9
CEM-SS
CEM2n
ΔA3G
F
CEM2n
30%
70%
GA
GC
GG
GT
ΔA3G
6%
6%
88%

Figure 3-2. Construction and Characterization of *A3G*-Null CEM2n Cells.

(A) *A3G* exon 3 targeting strategy. LA, left homology arm; SA, splice acceptor; IRES, internal ribosomal entry site; Neo, G418 resistance gene; pA, poly adenylation signal; RA, right homology arm; yellow triangles, *loxP* sites.

(B) *A3* mRNA expression profiles of the indicated cells relative to parental CEM2n (mean and s.d. shown for triplicate experiments).

(C) Immunoblots of A3G, A3F, and tubulin (TUB) in the indicated cells.

(D) Infectivity of Vif-deficient HIV produced using the indicated cell lines following a single replicative cycle (mean and s.d. shown for p24-normalized triplicate experiments).

(E) 3D-PCR profiles of HIV *gag-pol* and cellular *MDM2* targets within genomic DNA of infected CEM-GFP reporter cells.

(F) HIV G-to-A mutation profiles of proviruses originating in the indicated cell types. The mutation frequency at each dinucleotide is illustrated as a pie chart wedge ($n \geq 15$ kb per condition).

Gene	Allele 1	Allele 2	Allele 1 retargeted	Total clones analyzed	Overall targeting frequency (%)
APOBEC3G	6	2	4	189	6.3
APOBEC3F	7	3	2	170	7.1
Totals:		24		359	6.7

Table 3-1. Gene targeting statistics in CEM2n.

Cell line	Expt. no.	Number of 511bp *gag-pol* amplicons	Total kb sequenced	Total GG-to-AG mutations	Total GA-to-AA mutations	Total GY-to-AY mutations	Total other mutations
CEM2n	1	30	15.3	40	17	0	8
CEM2n	2	30	15.3	21	30	0	3
ΔG	1	30	15.3	3	42	3	2
ΔF	2	30	15.3	18	21	0	6
ΔF shNS	3	30	15.3	31	30	1	5
ΔF shA3B	3	30	15.3	24	39	2	10
ΔF shA3C	3	30	15.3	10	7	1	9
ΔF shA3D	3	30	15.3	81	3	3	9
ΔF shA3G	3	30	15.3	4	4	0	19
ΔF shA3H	3	30	15.3	17	18	1	6

Table 3-2. Mutation Summary.

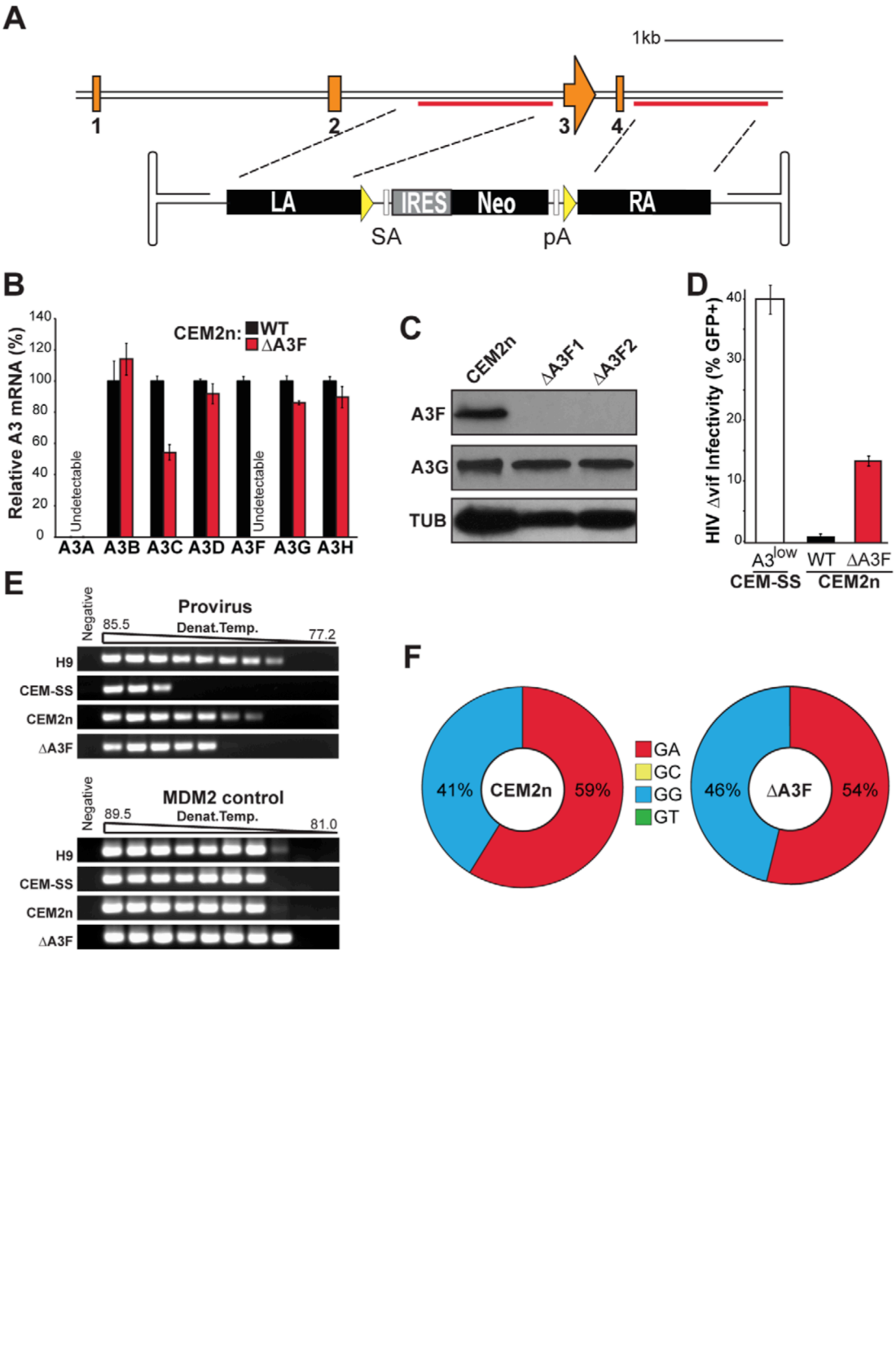

A
1kb
1
2
3
4
LA
IRES
Neo
RA
SA
pA
B
CEM2n: WT ΔA3F
Relative A3 mRNA (%)
Undetectable
Undetectable
A3A A3B A3C A3D A3F A3G A3H
C
CEM2n
ΔA3F1
ΔA3F2
A3F
A3G
TUB
D
HIV Δvif Infectivity (% GFP+)
A3low
WT
ΔA3F
CEM-SS
CEM2n
E
Provirus
Negative
85.5
Denat.Temp.
77.2
H9
CEM-SS
CEM2n
ΔA3F
MDM2 control
Negative
89.5
Denat.Temp.
81.0
H9
CEM-SS
CEM2n
ΔA3F
F
41%
CEM2n
59%
GA
GC
GG
GT
46%
ΔA3F
54%

Figure 3-3. Construction and Characterization of *A3F*-Null CEM2n Cells.

(A) *A3F* exon 3-4 targeting strategy. LA, left homology arm; SA, splice acceptor; IRES, internal ribosomal entry site; Neo, G418 resistance gene; pA, poly adenylation signal; RA, right homology arm; yellow triangles, *loxP* sites.

(B) *A3* mRNA expression profiles of the indicated cells relative to parental CEM2n (mean and s.d. shown for triplicate experiments).

(C) Immunoblots of A3F, A3G, and tubulin (TUB) in the indicated cells.

(D) Infectivity of Vif-deficient HIV produced using the indicated cell lines following a single replicative cycle (mean and s.d. shown for p24-normalized triplicate experiments).

(E) 3D-PCR profiles of HIV *gag-pol* and cellular *MDM2* targets within genomic DNA of infected CEM-GFP reporter cells.

(F) HIV G-to-A mutation profiles of proviruses originating in the indicated cell types. The mutation frequency at each dinucleotide is illustrated as a pie chart wedge ($n \geq 15$ kb per condition).

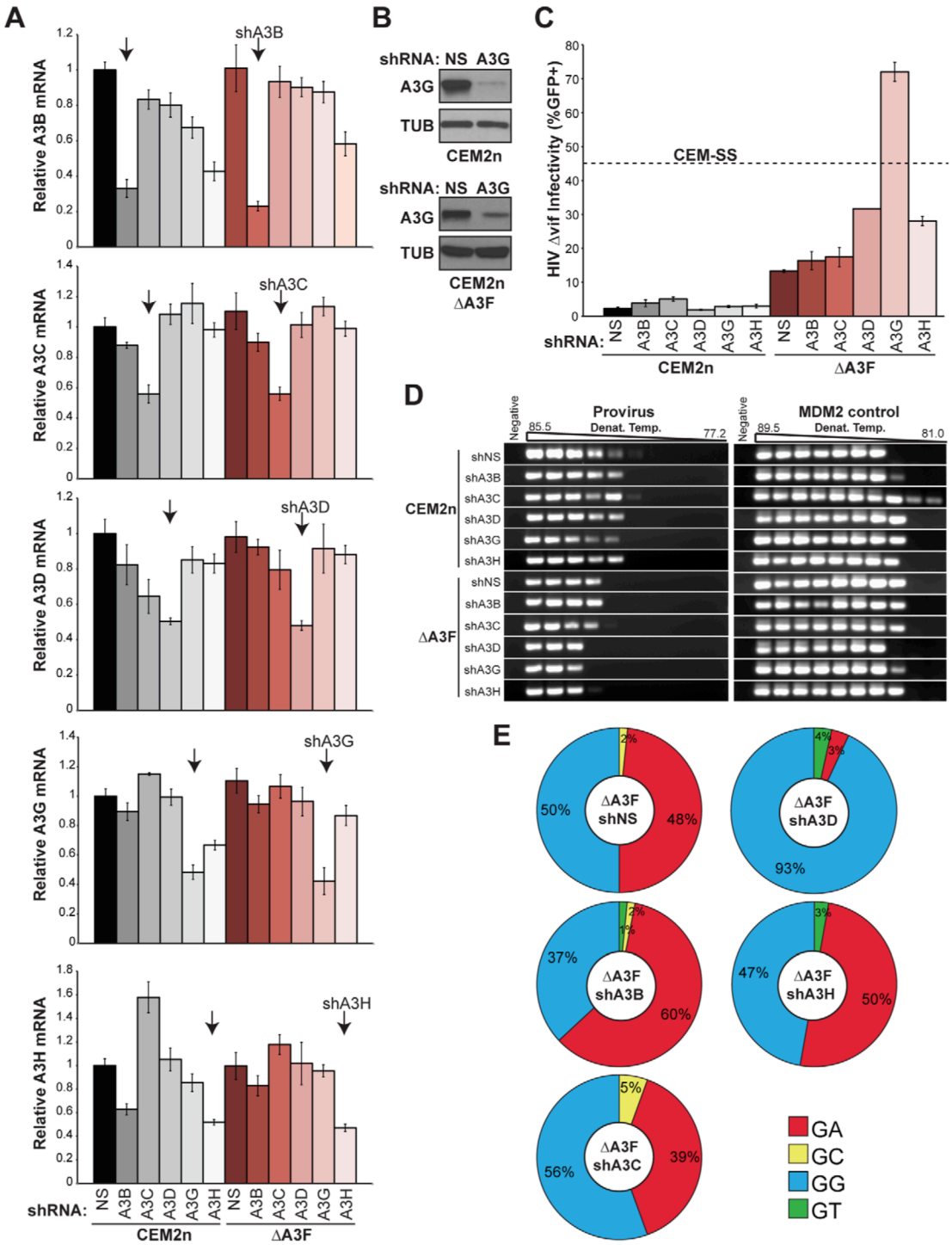

A
shA3B
Relative A3B mRNA
shA3C
Relative A3C mRNA
shA3D
Relative A3D mRNA
shA3G
Relative A3G mRNA
shA3H
Relative A3H mRNA
shRNA: NS A3B A3C A3D A3G A3H NS A3B A3C A3D A3G A3H
CEM2n
ΔA3F
B
shRNA: NS A3G
A3G
TUB
CEM2n
shRNA: NS A3G
A3G
TUB
CEM2n ΔA3F
C
HIV Δvif Infectivity (%GFP+)
CEM-SS
shRNA: NS A3B A3C A3D A3G A3H NS A3B A3C A3D A3G A3H
CEM2n
ΔA3F
D
Provirus
Denat. Temp.
85.5
77.2
Negative
MDM2 control
89.5
81.0
CEM2n
ΔA3F
shNS
shA3B
shA3C
shA3D
shA3G
shA3H
E
ΔA3F shNS
50%
48%
2%
ΔA3F shA3D
93%
4%
3%
ΔA3F shA3B
37%
60%
2%
1%
ΔA3F shA3H
47%
50%
3%
ΔA3F shA3C
56%
39%
5%
GA
GC
GG
GT

Figure 3-4. Construction and Characterization of *A3F*-Null/*A3*-Knockdown CEM2n Cells.

(A) Levels of each indicated *A3* mRNA in CEM2n or *A3F*-null cells transduced with shNS, shA3B, shA3C, shA3D, shA3G, or shA3H constructs (mean and s.d. shown for triplicate experiments).

(B) Immunoblots of A3G and tubulin (TUB) in CEM2n or *A3F*-null cells stably transduced with the indicated shRNA-expressing lentivirus.

(C) Infectivity of Vif-deficient HIV produced using the indicated transduced cell pool and reported using the CEM-GFP system (mean and s.d. shown for p24-normalized triplicate experiments; in some instances, the error is nearly indistinguishable from the histogram bar outline).

(D) 3D-PCR profiles of HIV *gag-pol* and cellular *MDM2* targets within genomic DNA of infected CEM-GFP reporter cells.

(E) HIV G-to-A mutation profiles of proviruses originating in the indicated cell types. The mutation frequency at each dinucleotide is illustrated as a pie chart wedge ($n \geq 15$ kb per condition). Pie charts were generated for those conditions with ≥ 1 mutation per kb analyzed. Mutation numbers for all conditions can be found in Table 2 and Table S1.

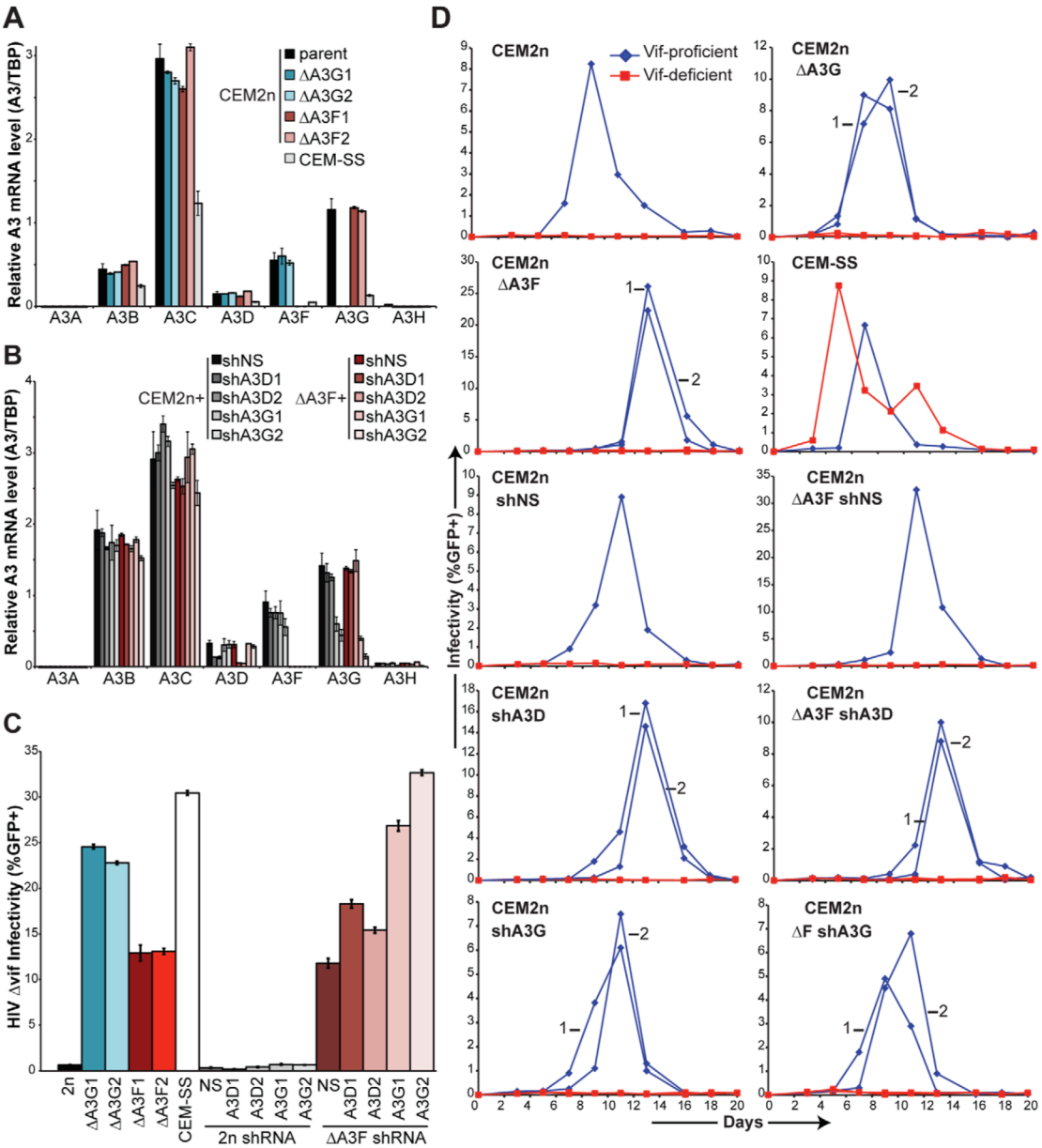

A
Relative A3 mRNA level (A3/TBP)
CEM2n
parent
ΔA3G1
ΔA3G2
ΔA3F1
ΔA3F2
CEM-SS
A3A
A3B
A3C
A3D
A3F
A3G
A3H
B
Relative A3 mRNA level (A3/TBP)
CEM2n+
ΔA3F+
shNS
shA3D1
shA3D2
shA3G1
shA3G2
C
HIV Δvif Infectivity (%GFP+)
2n
ΔA3G1
ΔA3G2
ΔA3F1
ΔA3F2
CEM-SS
NS
A3D1
A3D2
A3G1
A3G2
2n shRNA
ΔA3F shRNA
D
Infectivity (%GFP+)
Vif-proficient
Vif-deficient
CEM2n
CEM2n ΔA3G
CEM2n ΔA3F
CEM-SS
CEM2n shNS
CEM2n ΔA3F shNS
CEM2n shA3D
CEM2n ΔA3F shA3D
CEM2n shA3G
CEM2n ΔF shA3G
Days

Figure 3-5. Characterization of Independent Knockout and Knockdown Clones.

(A) Levels of each indicated *A3* mRNA in CEM2n, *A3G*- and *A3F*-null derivatives, and CEM-SS (relative to *TBP*; mean and s.d. shown for triplicate experiments).

(B) Levels of each indicated *A3* mRNA in CEM2n or *A3F*-null cells transduced with shNS, shA3D, or shA3G constructs (relative to *TBP*; mean and s.d. shown for triplicate experiments).

(C) Single-cycle infectivity of Vif-deficient HIV produced in parallel in the indicated cell lines (mean and s.d. shown for p24-normalized triplicate experiments).

(D) The kinetics of Vif-proficient (blue diamonds) and Vif-deficient (red squares) HIV spreading infection in the indicated cell lines. Numbers distinguish independent clones.

Gene Symbol	mRNA accession	Genomic location[a]	5' Primer	Seq (5'-3')[c]	3' Primer	Seq (5'-3')[c]
APOBEC3G	NM_021822					
Left arm		18865680-18866712	RSH4793	ATACATACGCGGCCGCACTTTGCAGGCAGGAGCTAA	RSH4794	NNNACTAGTACATGGCCAAGGGACTACTG
Right arm		18867801-18868826	RSH4178	NNNGAATTCGGTGAGAAGTGGGAGGTTCA	RSH4179	ATACATACGCGGCCGCGGACATTTGGAGGCAGAGAG
Primer set D[b]		18865306-18869303	RSH5670	GACGGAATTTCGCTCTTGTC	RSH5671	CCTTCACTGTTTCTCCCCAA
APOBEC3F	NM_145298					
Left arm		18830288-18831271	RSH4296	ATACATACGCGGCCGCCTTTGAAATGAGCTGTGTAGGC	RSH4297	NNNACTAGTGCACCTTTGTGGGTGCTACT
Right arm		18832500-18833555	RSH4676	NNNGTCGACCGACCTGAGTCCCTGAGAAG	RSH4677	ATACATACGCGGCCGCGTGCTCAGTTCCCGTCTCTC
Primer set D[b]		18829573-18833949	RSH5668	TTTCTCGTCGGAATACCGTC	RSH5669	CCATCTGAGTGCAGGTCTCA
Control & Diagnostic	na					
Primer set C[b]		na	RSH3561	CCAGCTGTTGGGGTGAGTA	RSH3564	CGAAGTTATTCCCAACTCATCC
Primer set B[b]		na	RSH3563	GGGGGATCCAGACATGATAA		
Primer set A[b]		na			RSH3870	GCCCTCACATTGCCAAAAG

(a) Based on build 37.2 human chromosome 22 sequence.
(b) See **Fig. S3** & **S5**.
(c) Restriction sites required for rAAV construction are underlined. See reference (59).

Table 3-S1. Primer Sequences.

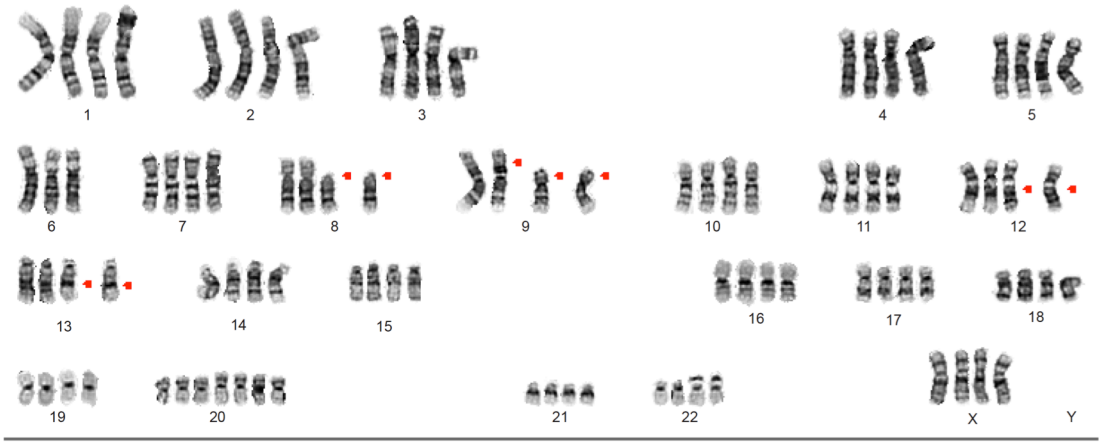


Figure 3-S1. CEM is a Near-Tetraploid T Cell Line.

Giemsa-banding karyotype of a representative CEM metaphase spread. Red arrows indicate typical lesions in lymphoblastic leukemia.

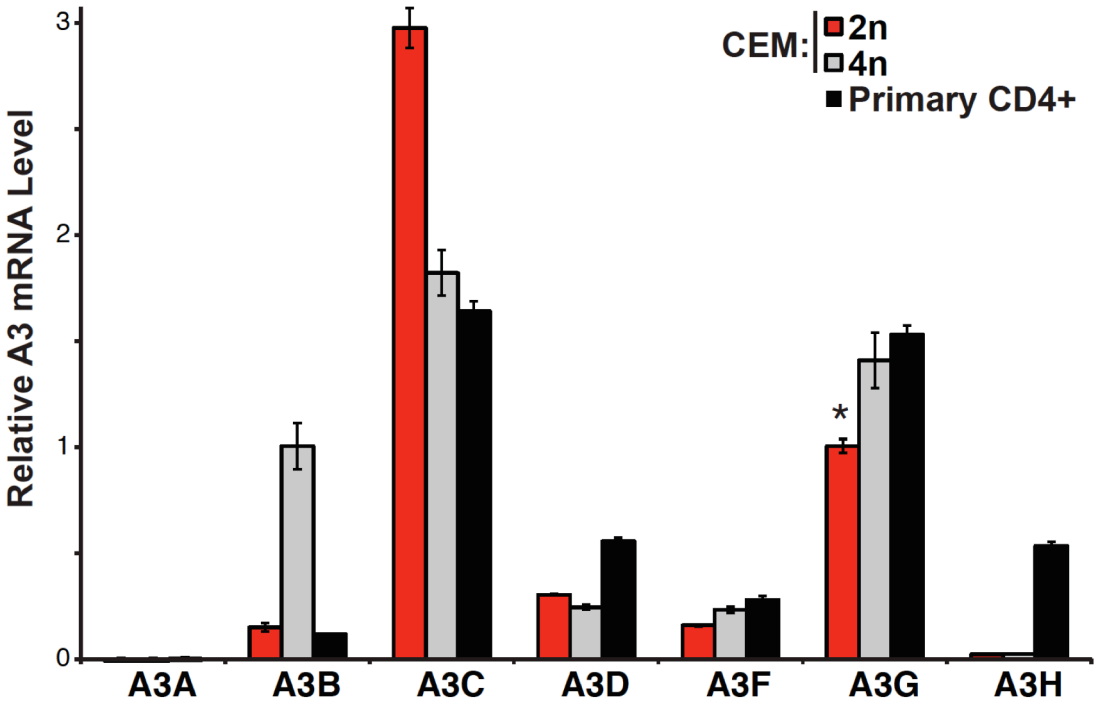


Figure 3-S2. A Comparison of *A3* mRNA Levels in CEM2n, CEM(4n), and Primary CD4+ Lymphocytes.

Total RNA was produced from the indicated sources, 1 µg was converted to cDNA, and one-twentieth of each cDNA was used to prime the indicated Q-PCR reactions. Levels of *A3G* in CEM2n are set to 1 to facilitate comparison. See methods for additional details.

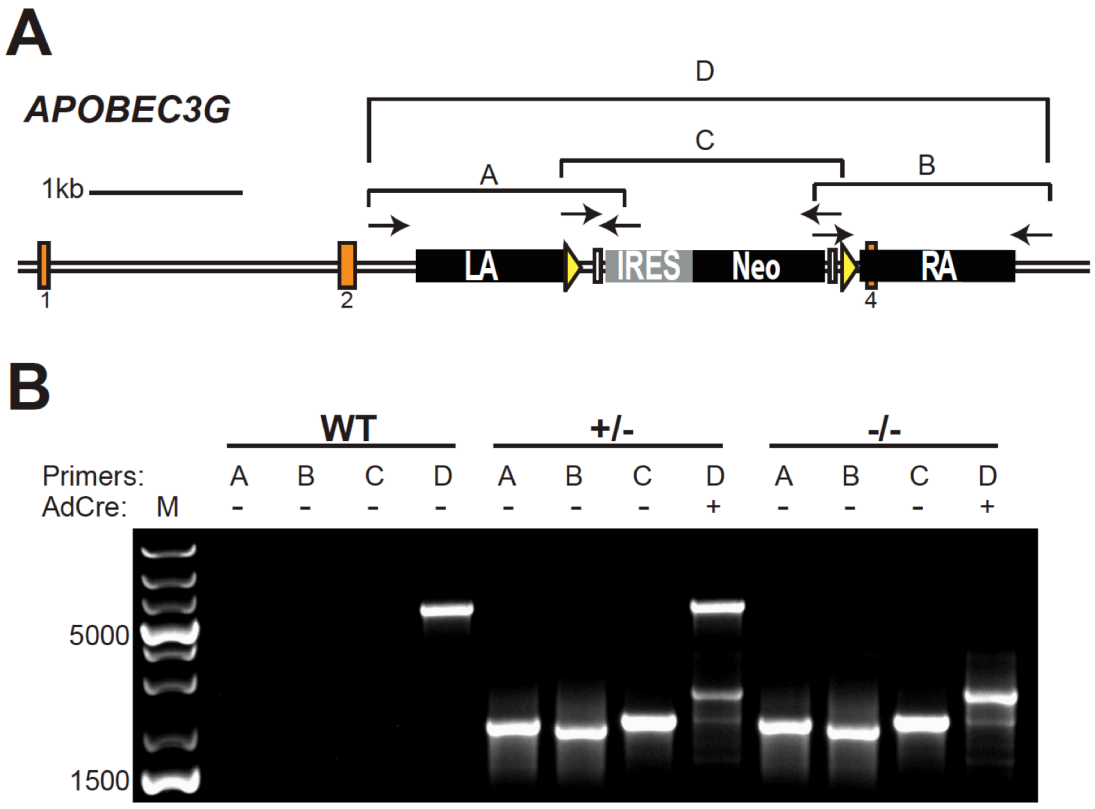


Figure 3-S3. PCR Reactions to Detect *A3G* Targeting Events.

(A) Schematic of a correctly targeted *A3G* locus. Diagnostic PCR reactions enable detection of the IRES-Neo cassette adjacent to the left targeting arm (PCR A = 1.8 kb) and adjacent to the right targeting arm (PCR B = 1.8 kb). An internal primer set enables detection of any drug resistant clone (PCR C = 1.9 kb). The flanking primers also enable detection of the *loxP*-to-*loxP* deletion product following Cre-mediated recombination (PCR D = 2.9 kb).

(B) Agarose gel image of PCR products produced from the genomic DNA of CEM2n (WT), an *A3G* heterozygote, and an *A3G*-null clone.

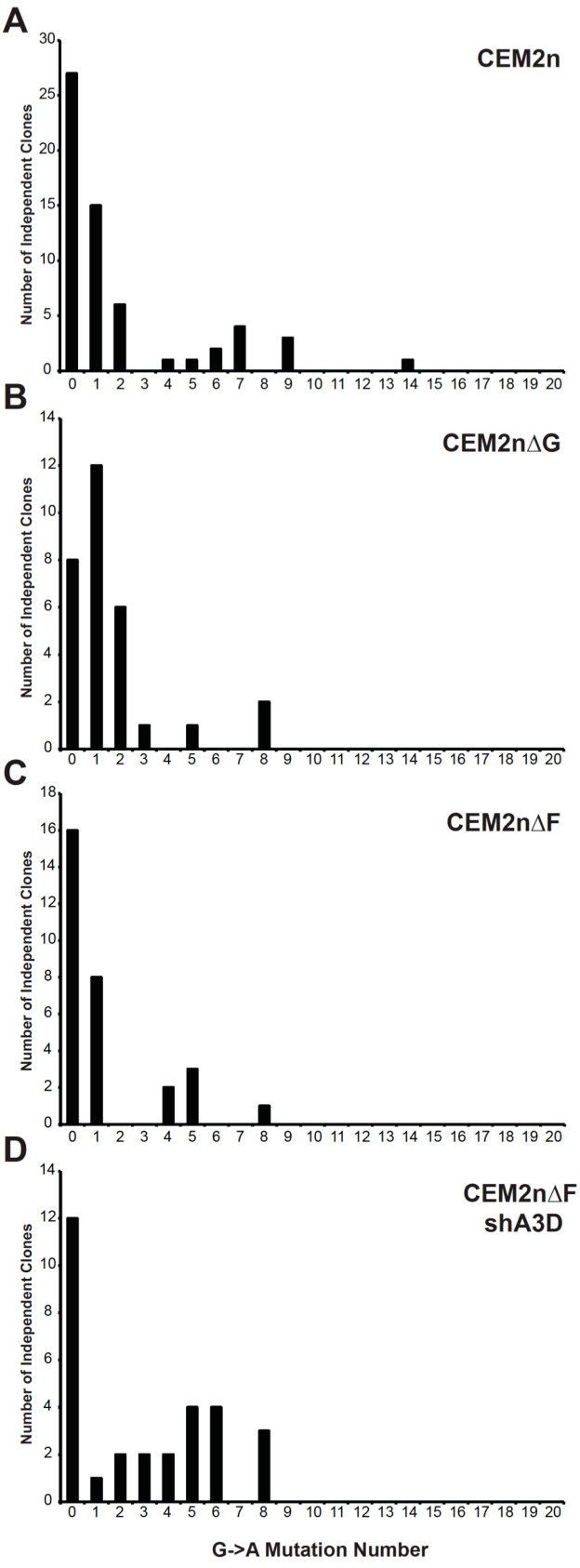

A
CEM2n
Number of Independent Clones
0 5 10 15 20 25 30
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
B
CEM2nΔG
Number of Independent Clones
0 2 4 6 8 10 12 14
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
C
CEM2nΔF
Number of Independent Clones
0 2 4 6 8 10 12 14 16 18
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
D
CEM2nΔF
shA3D
Number of Independent Clones
0 2 4 6 8 10 12 14
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
G->A Mutation Number

Figure 3-S4. G-to-A Mutation Loads for Vif-deficient HIV Produced in CEM2n and Key Derivatives.

The G-to-A mutation number per independent amplicon is shown in histogram format for Vif-deficient HIV recovered from single-round viral infections of the indicated cell lines.

(A) CEM2n experiments 1 and 2 are combined into a single histogram.

(B) *A3G*-null CEM2n.

(C) *A3F*-null CEM2n.

(D) *A3F*-null/*A3D*-knockdown CEM2n. The corresponding dinucleotide preferences are shown in pie format in **Figures 3-2F**, **3-3F**, and **3-4E**. The G-to-A mutations for all experimental conditions are summarized in **Table 3-2**. A trend toward higher G-to-A mutation loads and a bimodal mutation distribution was observed in viruses produced in parental CEM2n and in the *A3F*-null/*A3D*-knockdown derivative, suggesting that endogenous A3G may be more processive (not necessarily more anti-viral) than A3F or A3D.

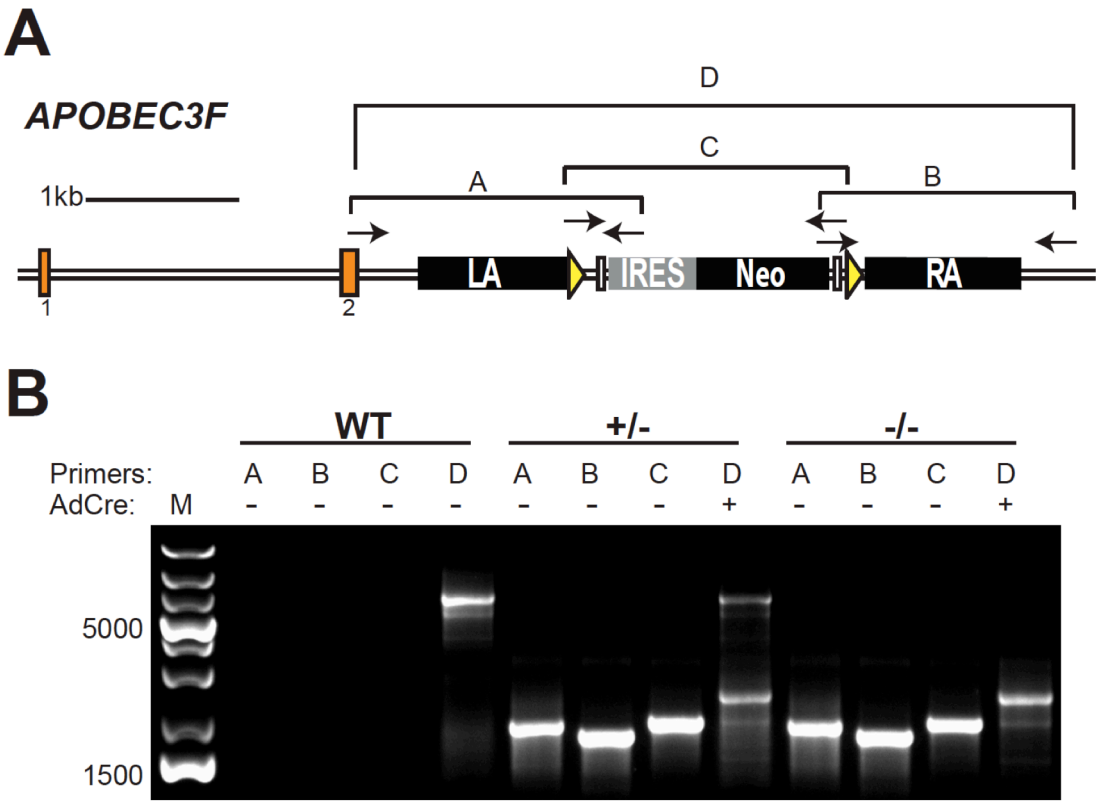


Figure 3-S5. PCR Reactions to Detect *A3F* Targeting Events.

(A) Schematic of a correctly targeted *A3F* locus. Diagnostic PCR reactions enable detection of the IRES-Neo cassette adjacent to the left targeting arm (PCR A = 2.1 kb) and adjacent to the right targeting arm (PCR B = 1.7 kb). An internal primer set enables detection of any drug resistant clone (PCR C = 1.9 kb). The flanking primers also enable detection of the *loxP*-to-*loxP* deletion product following Cre-mediated recombination (PCR D = 2.8 kb).

(B) Agarose gel image of PCR products produced from the genomic DNA of CEM2n (WT), an *A3F* heterozygote, and an *A3F*-null clone.

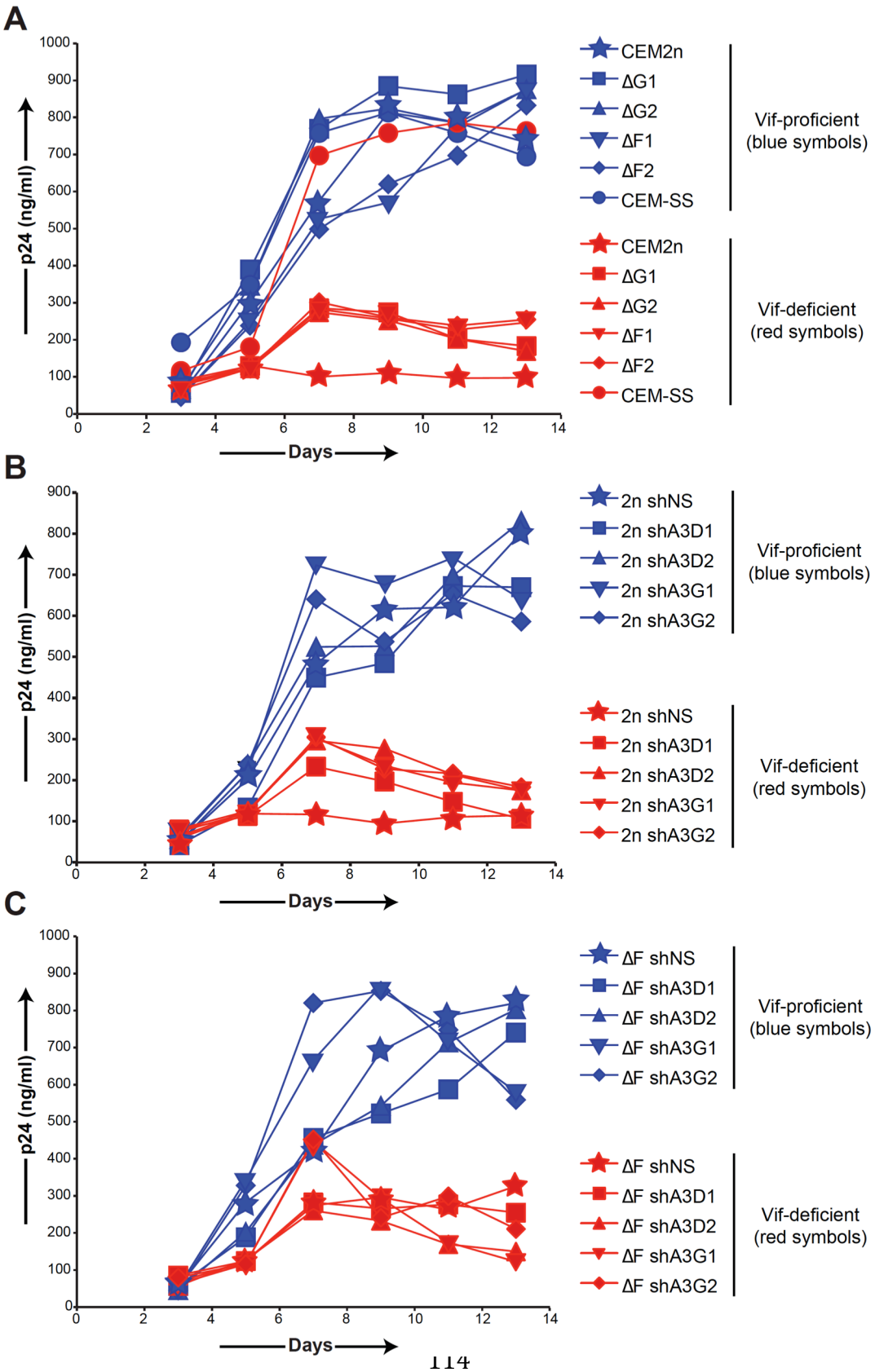

A
p24 (ng/ml)
Days
CEM2n
ΔG1
ΔG2
ΔF1
ΔF2
CEM-SS
Vif-proficient (blue symbols)
Vif-deficient (red symbols)
B
2n shNS
2n shA3D1
2n shA3D2
2n shA3G1
2n shA3G2
C
ΔF shNS
ΔF shA3D1
ΔF shA3D2
ΔF shA3G1
ΔF shA3G2

Figure 3-S6. Vif-Proficient and Vif-Deficient HIV Replication Kinetics.

Quantification of supernatant p24 levels at the indicated time points in (A) CEM2n and the indicated *A3F*- or *A3G*- null derivatives,

(B) CEM2n subclones expressing the indicated shRNA constructs, and

(C) CEM2n *A3F*-null subclones expressing the indicated shRNA constructs. The spreading infections in panels A, B, and C were done in parallel, and they are therefore directly comparable. Vif-proficient and Vif-deficient viral replication was done in parallel in CEM-SS (the only non-isogenic condition) and presented in panel A for comparison.

CHAPTER 4

Stable APOBEC3H Inhibits HIV-1 Replication in Primary T Lymphocytes

This chapter is adapted from a manuscript submitted for publication that is awaiting peer review: Eric W. Refsland, Judd F. Hultquist, Elizabeth M. Luengas, Terumasa Ikeda, Nadine M. Shaban, Emily K. Law, William L. Brown, Michael Emerman, and Reuben S. Harris. In review, PLOS Genetics.

Authors' Contributions

EWR performed all of the experiments, wrote and revised the manuscript, and designed all of the figures except as follows. JFH aided in the experimental design, primary cell isolations, HIV infection experiments and revised the manuscript. EML aided in the primary cell isolations, monoclonal APOBEC3H antibody characterization, and APOBEC3H genotyping. TI performed the APOBEC3H adaptation experiment. NMS aided in the HIV Vif surface modeling. EKL and WLB aided in the APOBEC3H monoclonal antibody characterization. ME provided the APOBEC3H monoclonal hybridomas and revised the manuscript. RSH aided in the design of the experiments and in manuscript preparation/revision.

FOREWORD

The APOBEC3 enzymes protect cells by inhibiting the spread of retroelements, including HIV-1, by blocking reverse transcription and mutating cytosines in single-stranded DNA replication intermediates. HIV-1 Vif counteracts restriction by marking APOBEC3 proteins for proteasomal degradation. APOBEC3H is the most diverse member of this protein family. Humans have seven distinct *APOBEC3H* haplotypes with three producing stable and four producing unstable proteins upon forced overexpression. Here, we examine the stability phenotype of endogenous APOBEC3H in donors with different haplotypes and address how these stability differences, as well as natural viral diversity, combine to determine HIV-1 infectivity. We found that endogenous *APOBEC3H* haplotypes yield stable or unstable proteins and that stable APOBEC3H is induced during viral infection and restricts the replication of isolates with naturally occurring hypo-functional but not hyper-functional Vif alleles. We also found that the global distribution of stable APOBEC3H alleles correlates with the prevalence of HIV-1 Vif alleles capable of mediating its degradation, strongly suggesting that the viral Vif protein is capable of adapting to the APOBEC3H restriction potential of an infected individual. Thus, the combination of human APOBEC3H haplotypes and virus Vif alleles may help account for some of the observed disparities in disease progression and virus transmission.

SUMMARY

The Vif protein of HIV-1 allows virus replication by degrading several members of the host-encoded APOBEC3 family of DNA cytosine deaminases. Polymorphisms in both host *APOBEC3* genes and the viral *vif* gene have the potential to impact the extent of virus replication among individuals. The most genetically diverse of the seven human *APOBEC3* genes is *APOBEC3H* with seven known haplotypes. Overexpression studies have shown that a subset of these variants express stable and active proteins whereas the others encode proteins with a short half-life and little, if any, antiviral activity. We demonstrate that these stable/unstable phenotypes are an intrinsic property of endogenous APOBEC3H proteins in primary CD4+ T lymphocytes and confer differential resistance to HIV-1 infection in a manner that depends on natural variation in the Vif protein of the infecting virus. HIV-1 with a Vif protein hypo-functional for APOBEC3H degradation, yet fully able to counteract APOBEC3D, APOBEC3F, and APOBEC3G, was susceptible to restriction and hypermutation in stable APOBEC3H expressing lymphocytes, but not in unstable APOBEC3H expressing lymphocytes. In contrast, HIV-1 with hyper-functional Vif counteracted stable APOBEC3H proteins as well as all other endogenous APOBEC3s and replicated to high levels. We also found that APOBEC3H protein levels are induced over 10-fold by infection. Finally, we found that the global distribution of stable APOBEC3H haplotypes correlates with the distribution a critical hyper-functional Vif amino acid residue. These data combine to strongly suggest that stable APOBEC3H haplotypes present as *in vivo* barriers to HIV-1 replication, that Vif is capable of adapting to these restrictive pressures, and that an evolutionary equilibrium has yet to be reached.

INTRODUCTION

The human APOBEC3 (A3) family of DNA cytosine deaminases is encoded by seven genes arranged in tandem on chromosome 22 (reviewed by (Conticello, 2008; Refsland and Harris, 2013)). These proteins inhibit the replication of a broad number of parasitic elements, including many retroviruses, some DNA viruses, and several endogenous retroelements and retrotransposons by both deaminase-dependent and -independent mechanisms (reviewed by (Desimmie et al., 2014; Malim and Bieniasz, 2012; Refsland and Harris, 2013)). Several lines of evidence indicate that A3D, A3F, A3G, and A3H contribute to HIV-1 restriction by packaging into assembling virus particles and, upon virus entry into new target cells, deaminating viral cDNA cytosines to uracils and impeding the progression of reverse transcription (reviewed by (Desimmie et al., 2014; Malim and Bieniasz, 2012; Refsland and Harris, 2013)). The cDNA uracil lesions template the insertion of genomic strand adenines during second strand reverse transcription and ultimately manifest as G-to-A mutations which destroy virus infectivity by the introduction of deleterious missense and nonsense mutations within viral open reading frames.

Mammalian A3 proteins pose such a threat to lentivirus replication that an ancestor of present day HIV-1 evolved an accessory protein called Vif (virion infectivity factor) to inhibit these restriction factors. HIV-1 Vif assembles an E3 ligase complex comprised of CBFβ, ELOB, ELOC, CUL5, and RBX2 to mediate the poly-ubiquitination and proteasomal degradation of restrictive A3s ((Jäger et al., 2012b; Zhang et al., 2012) and references therein). This process enables HIV-1 to replicate in its main target cell, CD4+ T lymphocytes, which express multiple A3s and would otherwise be non-

permissive for viral replication (Hultquist et al., 2011; Koning et al., 2009; Refsland et al., 2010). However, this process is less than 100% effective, as G-to-A mutations are commonly observed in viral sequences from clinical specimens and, depending on the patient, may be biased toward a GG-to-AG dinucleotide context indicative of A3G or a GA-to-AA characteristic of A3D, A3F, and/or A3H (Albin and Harris, 2010; Caride et al., 2002; Fitzgibbon et al., 1993; Gandhi et al., 2008; Hultquist et al., 2011; Janini et al., 2001; Kieffer et al., 2005; Kijak et al., 2008; Koning et al., 2009; Land et al., 2008; Pace et al., 2006; Piantadosi et al., 2009; Refsland et al., 2010; Suspène et al., 2006; Ulenga et al., 2008; Vartanian et al., 1994; Wood et al., 2009).

The present day human *A3* locus is a result of multiple gene duplication events during evolution from an ancestral mammalian locus (LaRue et al., 2008). Unlike the Z1 domain (*e.g. A3A*) or Z2 domain (*e.g. A3C*) deaminase genes, which show considerable copy number variation between mammalian phylogenetic tree branches, the Z3 domain gene (*e.g. A3H*) exists in only one copy in all mammalian genomes sequenced to date (LaRue et al., 2010). Overexpression studies have shown that A3H and the orthologous Z3 domain deaminases from several mammalian species have a conserved capacity to restrict retrovirus replication, and likewise are neutralized by the Vif proteins of each species' lentivirus (Bogerd et al., 2004; LaRue et al., 2010; Mangeat et al., 2004; Mariani et al., 2003a; Schröfelbauer et al., 2004; Virgen and Hatziioannou, 2007; Xu et al., 2004; Zielonka et al., 2009; Zielonka et al., 2010). However, a role for A3H *in vivo* has not yet been elucidated since endogenous expression of the protein has been difficult to detect (Li et al., 2010).

A3H is the most polymorphic of the human *A3* genes due to circulation of at least 7 distinct *A3H* haplotypes (OhAinle et al., 2008; Wang et al., 2011). These haplotypes are comprised of various combinations of 5 single nucleotide polymorphisms located in exons 2, 3 and 4 that range in allele frequencies globally from 6 to 87% (Duggal et al., 2013). Previous overexpression and pulse-chase experiments have shown that 3 *A3H* haplotypes yield proteins with relatively long half-lives (stable), 1 produces a protein with weak stability, and another 3 make completely unstable proteins (Dang et al., 2008; OhAinle et al., 2008; Wang et al., 2011). For example, transient transfection of cells with *A3H* haplotype II cDNA with SNPs at residues 15, 18, 105, 121, and 178 (NRRDD) yields a protein readily detectable by immunoblotting, whereas *A3H* haplotypes I (NRGKE) and III (ΔRRDD) produce weakly expressed or undetectable proteins, respectively (OhAinle et al., 2008; Wang et al., 2011). However, it is not yet known whether these dramatic haplotype-associated stability/instability phenotypes also manifest for endogenous A3H proteins expressed in primary immune cells. This distinction is important because only stably expressed haplotypes would be predicted to exert selective pressure on HIV-1 replication and potentially contribute to virus restriction and diversification *in vivo*. Indeed, a recent paper linked putatively stable A3H haplotype II to enhanced restriction, higher frequencies of G-to-A mutation, and more favorable clinical phenotypes (lower viral loads and higher patient CD4+ T cell numbers) (Ooms et al., 2013a).

Here, we test the hypothesis that HIV-1 infectivity will be influenced by both host *A3H* haplotype as well as viral *vif* genotype. We detect the expression and induction of endogenous A3H protein from stable *A3H* alleles following HIV-1 infection. By

constructing a set of molecular clones encoding Vif proteins of varying abilities to antagonize stable A3H, we showed that expression of one allele of stable *A3H* is sufficient to inhibit hypo-functional Vif virus replication and promote G-to-A mutagenesis in the expected GA-to-AA dinucleotide context. In contrast, a hyper-functional Vif variant effectively counteracted stable A3H activity. Virus adaptation to stable A3H may be occurring on a global scale because the geographic distribution of a key hyper-Vif amino acid correlates with the distribution of stable A3H haplotypes. Taken together, these data demonstrate that stable A3H haplotypes affect HIV-1 replication in the primary target cells of the virus, and further suggest that viral evolution requires the Vif protein to adapt to counteract the restrictive pressure imposed by these enzymes. Therefore, the combination of *A3H* haplotype and viral *vif* variation may be critical and linked factors in HIV-1 adaptation to human populations.

RESULTS

At least two A3H haplotypes encode stable proteins in primary T lymphocytes

To address whether the dramatic protein stability phenotypes observed previously in overexpression studies ((Dang et al., 2008; OhAinle et al., 2006; Wang et al., 2011); **Figure 1A**) extend to endogenous *A3H* haplotypes in primary T cells, we used our recently developed anti-A3H monoclonal antibodies (Li et al., 2010) as well as a new commercial anti-A3H polyclonal antibody to probe protein expression in CD4+ T lymphocytes from healthy donors with different haplotypes (**Materials and Methods**, **Figure 1B & Figure S1**). Primary CD4+ lymphocytes were isolated by negative selection, stimulated with IL2 and PHA for 72 hrs, and subjected to immunoblotting.

Three of 24 donors were heterozygous for a predicted stable A3H haplotype and none were homozygous (6% allele frequency; **Table S1**). Anti-A3H immunoblots demonstrated clear differences in steady-state protein levels, with robust detection of haplotypes II and V, weaker detection of haplotype I, and no detection of haplotypes III and IV (**Figure 1B**). In comparison, A3G and HSP90 protein levels were expressed similarly between donors.

To confirm the prediction from previous overexpression studies that the observed difference in A3H protein levels is really at the protein and not the RNA level, RT-qPCR (Refsland et al., 2010) was used to quantify *A3* mRNA levels in primary CD4+ T lymphocytes. Similar *A3H* mRNA levels were observed regardless of haplotype (**Figure 1C**). The mRNA levels of the other *A3* genes were also unrelated to *A3H* haplotype, with some minor variation observed between donors (typically less than 2-fold, as reported (Koning et al., 2009; Refsland et al., 2010)). Thus, consistent with prior overexpression and pulse-chase experiments to measure protein half-life (Dang et al., 2008; OhAinle et al., 2008; Wang et al., 2011), the large difference between the various *A3H* haplotypes in primary CD4+ cells occurs at the protein level.

Identification of Vif residues that interact with stable A3H and still antagonize the other A3 proteins expressed in primary T cells

Primary CD4+ T lymphocytes express 6 *A3* family members (Koning et al., 2009; Refsland et al., 2010). We asked whether or not the ability of Vif to antagonize A3H could be separated from its ability to antagonize the other A3 proteins expressed in primary CD4+ T cells. Previous work indicated that Vif proteins of different HIV-1

isolates show varying capacities to counteract stable A3H (NL4-3/IIIB and LAI (Binka et al., 2012; Hultquist et al., 2011; LaRue et al., 2010; Li et al., 2010; Ooms et al., 2013b)), but these studies did not address spreading infections in T cells where A3D, A3F, and especially A3G selective pressures might influence Vif function. To help inform this construction, we performed a series of HIV-1 IIIB N48H adaptation experiments in which this lab-derived strain was subjected by stepwise passage to increasing ratios and levels of stable A3H haplotype II expressed in SupT11 cells (**Figure 2A & B**). An N48H variant was used to initiate these experiments because prior work had already demonstrated that this natural variation improved the A3H counteraction capacity of the Vif protein of IIIB/NL4-3 (Ooms et al., 2013b). These experiments were initiated by infecting (MOI 0.05) uniform pools of stable A3H-low expressing cells (weakly selective) and, after an 8 day incubation, viral supernatants were transferred to the next highest selective condition, and this process was repeated until 100% stable A3H-high conditions were reached (**Materials and Methods**).

In comparison to the starting virus, HIV-1 IIIB N48H, which could only replicate in SupT11 control vector and A3H-low expressing cells, one of the adapted viruses showed rapid and nearly indistinguishable replication kinetics in SupT11 cells expressing low, intermediate, and high levels of stable A3H (**Figure 2C**). The majority (5/9) of Vif sequences from this adapted population had a K63E amino acid substitution. Furthermore, we noticed that the K63E substitution recovered in these adaptation experiments is part of a cluster of 4 amino acids that distinguishes the IIIB lab strain from an isolate recovered from a homozygous stable A3H haplotype II patient (Li et al., 2010). The Vif protein from this particular isolate is even better than LAI Vif at counteracting

stable A3H haplotype II (Li et al., 2010). We therefore predicted that the combination of N48H (to be more LAI Vif-like (Ooms et al., 2013b)) and GDAK60-63 to EKGE60-63 (to be more haplotype II patient Vif-like (Li et al., 2010)) would result in a hyper-functional Vif protein capable of fully antagonizing stable A3H haplotype II (hyper-Vif; **Figure 3A**). In addition, the separation-of-function approach required the generation of a Vif variant even more sensitive than lab-Vif to restriction by stable A3H haplotype II. This was done by introducing F39V into the HIV-1 IIIB lab strain, which is a substitution that confers sensitivity to stable A3H haplotype II (Binka et al., 2012) (**Figure 3A**).

The A3 neutralization activities of the hyper-, lab-, and hypo-Vif HIV-1 IIIB variants were assessed by comparing spreading infection kinetics in SupT11 cell lines stably expressing a control vector, A3D-HA, A3F-HA, A3G-HA, and low, intermediate, and high levels of untagged A3H haplotype II (**Figure 3B & C**). The A3 expression level in each SupT11 clone was sufficient to delay (A3D) or fully suppress (A3F, A3G, A3H haplotype II) replication of a Vif-null HIV-1 IIIB control virus, consistent with our original studies using these stable cell lines (Hultquist et al., 2011). Importantly, the hyper-, lab-, and hypo-Vif HIV-1 IIIB variants spread with nearly identical efficiencies in the SupT11 cell lines expressing the control vector, A3D-HA, A3F-HA, and A3G-HA, demonstrating fully intact capacities to neutralize these restriction factors. In contrast, each isolate exhibited differential replication kinetics on the SupT11 lines expressing low, intermediate, and high levels of stable A3H haplotype II. The hyper-Vif isolate replicated with strong kinetics under all conditions including the highest stable A3H haplotype II expression levels, the lab-Vif isolate was delayed under intermediate and high A3H expression conditions, and the hypo-Vif isolate showed restricted replication

phenotypes almost indistinguishable from the Vif-deficient control. These data demonstrate that naturally occurring amino acid residues that positively or negatively influence Vif-mediated neutralization of A3H can do so without impacting Vif activity against A3D, A3F, and A3G and therefore allow for a direct interrogation of endogenous A3H function in primary T lymphocytes.

HIV-1 restriction by stable A3H in primary T lymphocytes

In order to determine the relative importance of the polymorphisms of A3H and Vif on HIV-1 replication in primary cells, a series of spreading infection experiments was initiated in CD4+ cells from individuals who encode different haplotypes of A3H with viruses encoding the hyper-, lab-, and hypo-Vif alleles that differentially antagonize A3H. We found that all 3 viruses replicated with similar kinetics in primary CD4+ T cells from donors homozygous for unstable A3H haplotype I or heterozygous for unstable A3H haplotypes III and IV. Indicating that hyper-, lab-, and hypo-Vif viruses are all capable of neutralizing the restrictive levels of A3D, A3F, and A3G expression in primary cells (**Figure 4A & S2**). In contrast, only the hyper-Vif virus replicated with robust kinetics in CD4+ T cells from donors heterozygous for stable A3H haplotypes II or V (**Figure 4B & S2**). The lab-Vif and hypo-Vif viruses showed delayed and strongly inhibited replication kinetics, respectively, in CD4+ T cells from donors heterozygous for stable A3H haplotypes (**Figure 4B & S2**). All of the heterozygous stable A3H haplotype II donors also had a copy of unstable A3H haplotype I, so these large differences in replication kinetics are due to only a single stable A3H allele. Thus, these experiments

demonstrate that even a single allele of endogenous stable A3H constitutes a formidable barrier to replication of susceptible HIV-1 isolates.

The APOBEC3 proteins must be packaged into viral particles to restrict HIV-1 replication. To confirm that stable, endogenous A3H acts by this mechanism, aliquots of cells and virus-containing supernatants were taken on days 3, 5, 7 and 9 post-infection for each donor/virus condition and analyzed by immunoblotting (**Figure 4C**). First, we observed that A3H is induced at the protein level over the course of infection. This is most evident in cells infected with the hypo-Vif virus, most likely because this Vif variant fails to bind endogenous A3H and trigger its degradation. The level of A3H induction observed here at the protein level corresponds roughly to the level of induction observed previously at the mRNA level (>10-fold (Hultquist et al., 2011)). Second, there is a strong correlation between cellular Vif, cellular A3H, and packaged A3H levels. This is most evident upon direct comparison of the hyper-Vif and hypo-Vif infections. Vif disappears, cellular A3H is induced weakly, and very little viral A3H is observed in particles from hyper-Vif infection of stable haplotype II cells. In contrast, Vif decreases modestly, cellular A3H is induced strongly, and large amounts of A3H are observed in particles from the hypo-Vif infection of stable haplotype II cells. As expected from the SupT11 experiments described above, the lab-Vif spreading infection elicited intermediate phenotypes. Most importantly, levels of A3H in viral particles correlated with virus replication kinetics, with the hyper-Vif virus spreading robustly and the hypo-Vif virus being restricted almost completely. Thus, HIV-1 replication, as well as the induction and encapsidation of A3H, are influenced by both host *A3H* haplotype and the viral *vif* genotype.

G-to-A mutation spectra of Vif separation-of-function isolates subjected to different endogenous A3H haplotypes

The hallmark activity of HIV-1-restrictive A3 family members is viral cDNA deamination of C-to-U, with the uracil lesions ultimately immortalizing as genomic strand G-to-A mutations (reviewed by (Desimmie et al., 2014; Duggal and Emerman, 2012; Harris et al., 2012; Malim and Bieniasz, 2012)). To determine the correlation between mutagenesis and the relative restriction incurred by each of the separation-of-function Vif variants in combination with different A3H haplotypes, the viruses from each stable and unstable A3H experimental condition were subjected to analysis by differential DNA denaturation (3D)-PCR (Hultquist et al., 2011; Refsland et al., 2012; Suspène et al., 2005b). This technique depends on the fact that every PCR amplicon has a characteristic denaturation temperature that must be reached in order to yield visible PCR product, and G-to-A mutations within the amplicon will result in the appearance of product at significantly lower temperatures. After 15 days of spreading infection, viral supernatants were removed and used to infect CEM-GFP reporter cells. After allowing 2 days for infection and integration, total cellular DNA was prepared, quantified, and subjected to 3D-PCR analysis of a 564 bp HIV-1 *pol* region (**Figure 4D**). All viruses from unstable A3H haplotype III/IV expressing cells showed no evidence for G-to-A hypermutation, as the lowest denaturation temperature with visible PCR product was the same as that of the original molecular clone. The level of G-to-A mutation in hypo-Vif viruses from stable A3H haplotype II/I cells was clearly the highest, but levels for lab-Vif and hyper-Vif viruses were also significantly above the background level established by

the amplification cut-off for the original molecular clone. Interestingly, a significant level of G-to-A mutation was also evident in hypo-Vif and lab-Vif viruses (but not in hyper-Vif virus) from unstable haplotype I/I cells, indicating that A3H haplotype I protein is capable of lower levels of viral cDNA deamination and may contribute to HIV-1 mutagenesis, though not to obvious infectivity decreases.

To analyze the mutation spectrum and local dinucleotide mutation contexts, we used normal high-fidelity PCR (98°C denaturation temperature) and Sanger sequencing to analyze the *pol* region of proviral DNA from an independent experiment (*i.e.*, viruses from day 15 of the experiment shown in **Figure 4A-C**). Consistent with the 3D-PCR experiment discussed above, the highest level of G-to-A mutation occurred in hypo-Vif viruses from cells expressing stable A3H haplotype II, and the lowest level of G-to-A mutation occurred in hyper-Vif viruses from cells expressing unstable A3H haplotype I (5.0 G-to-A mutations per kb versus 0.2 G-to-A mutations per kb; **Figure 4E**). Most (76%) of the hypo-Vif viruses' G-to-A mutations occurred in a GA-to-AA context, consistent with the 5'TC deamination preference of A3H shown previously (Harari et al., 2009; Hultquist et al., 2011; Ooms et al., 2013a). In contrast, few G-to-A mutations were observed in a GG-to-AG context characteristic of A3G, further supporting the SupT11 experiments shown above that demonstrate that each of the Vif separation-of-function variants retains equivalent and strong A3G antagonism (**Figure 4E & S3**).

Global correlations between HIV-1 Vif genotypes and human A3H haplotypes

If A3H exerts selective pressure on HIV-1, then Vif variants most capable of neutralizing its activity should predominate in areas where stable A3H haplotypes also

predominate. A geographic breakdown of 9713 HIV-1 isolates represented in the Los Alamos database revealed that F39 (an A3H resistance residue) predominates in Africa, whereas V39 (an A3H susceptibility residue) predominates in Asia (**Figure 5A**). In addition, the N48 residue in Vif that confers sensitivity to high levels of stable A3H haplotype II (Binka et al., 2012; Hultquist et al., 2011; Ooms et al., 2013b) in HIV-1 IIIB and NL4-3, is present in only 9% of African isolates but in 30% of isolates from Asia. Interestingly, these Vif allelic distributions correspond to the worldwide estimates for A3H haplotypes from the 1000 genomes project previously reported (Genomes Project et al., 2012; OhAinle et al., 2008), with stable haplotypes predominating in Africa and unstable haplotypes in Asia (**Figure 5B**). Thus, HIV-1 Vif appears to have the capacity to adapt to the A3H haplotype of a population. A major prediction is therefore that stable A3H may be a protective factor in HIV-1 acquisition, especially in instances in which a stable A3H haplotype individual is exposed to a hypo-Vif inoculum from an unstable A3H haplotype patient (**Figure 5C**).

A3H-altering Vif residues define an interaction surface

All current evidence indicates that Vif heterodimerizes with CBFβ, directly binds restrictive A3s, and recruits an E3-ligase ubiquitin complex to target them for proteasomal degradation (Desimmie et al., 2014; Duggal and Emerman, 2012; Harris et al., 2012; Malim and Bieniasz, 2012). Recently, a crystal structure of the Vif-CBFβ-ELOB-ELOC-CUL5 complex was determined, providing long-awaited details of the molecular architecture of Vif (Guo et al., 2014). The separation-of-function Vif variants described here provide an opportunity to define the A3H interaction surface.

Interestingly, all of the amino acids that comprise these variants map to the same solvent-exposed surface of Vif (**Figure 6A**). The critical hypo- and hyper-Vif residues, F39 and GDAK60-63, are particularly close together. Moreover, the delayed and differential spreading infection phenotypes of HIV-1 isolates with each of these single amino acid substitutions in SupT11 T cells expressing high levels of stable A3H haplotype II indicates that each of these residues makes partial contribution to the overall hyper-Vif phenotype (**Figure 6B**).

Analogous but physically distinct separation-of-function Vif mutants have been described for A3F and A3G. For instance, the DRMR14-17 motif is required for A3F degradation, but does not affect A3G neutralization (Russell and Pathak, 2007; Schröfelbauer et al., 2006). Furthermore, another motif, YRHHY40-44, specifically compromises the A3G interaction (Russell and Pathak, 2007). The fact that the A3H interaction surface defined above and these A3F- and A3G-specific motifs map to distinct solvent exposed surfaces strongly indicates that Vif may use different binding modes in order to neutralize each of these restriction factors.

DISCUSSION

A3H is the most polymorphic A3 family member in the human population, and here we provide the first protein-level demonstration that the differential stability and inducibility of endogenous A3H haplotypes in primary T cells impacts HIV-1 replication in its normal target cells. We also report that natural variation in HIV-1 Vif results in the differentially neutralization of stable A3H haplotype II and influences virus replication

and hypermutation activities in primary CD4+ lymphocytes. The hyper-Vif virus showed strong replication kinetics and low levels of G-to-A mutation under all conditions tested, whereas the hypo-Vif virus was only restricted and hypermutated in primary CD4+ T lymphocytes expressing endogenous levels of stable A3H haplotype II. Because the separation-of-function variants were based on naturally occurring HIV-1 amino acid residues, our data combine to indicate that stable haplotypes of endogenous A3H may constitute a replication barrier to a significant subset of circulating HIV-1 strains. This conclusion is further supported by strong correlations between A3H haplotype stability and predicted Vif functionality in worldwide human and HIV-1 genotype information (**Figure 5A & B**).

Homology to other A3s with high-resolution structural information indicates that N15 is a highly conserved residue near the end of α-helix 1 and therefore most likely required for A3H structural integrity. NMR and mutagenesis studies with the catalytic domain of A3G have shown that this helix is essential for the stability and integrity of the entire deaminase domain, as it makes several essential contacts with internal residues (Chen et al., 2008; Harjes et al., 2009). For instance, an alanine substitution of the homologous A3G residue N208 renders the enzyme catalytically dead (Chen et al., 2008). In contrast, the glycine at position 105 is predicted to be in the middle of β-strand 4, located far from N15 and on the backside of the enzyme relative to the catalytic zinc-coordinating active site. This location and the observed intermediate stability of A3H haplotype I protein combine to suggest a different mechanism, possibly by altering an interaction with a cellular binding partner (Li et al., 2010). Consistent with these predictions, all evidence indicates that *A3H* haplotypes III and IV (ΔN15) encode

complete loss-of-function proteins, whereas haplotype I (G105) encodes a hypofunctional variant because higher-than-background levels of G-to-A mutations are observed in the hypo-Vif virus in haplotype I expressing T cells.

Our data agree with the main conclusion of a study published recently by the Simon group (Ooms et al., 2013a). They concluded that A3H haplotype II is a relevant HIV-1 restriction factor by showing more G-to-A mutations accumulating in viruses replicating in A3H haplotype II donor PBMC, in comparison to the same viruses replicating in haplotype I donor PMBC (Ooms et al., 2013a). Our studies help explain their results by showing that stable A3H haplotypes are indeed expressed at the protein level in primary HIV-1 target cells and that, depending upon the functionality of the infecting viruses Vif protein, the outcomes can range from aphenotypic (hyper-Vif) to strong infectivity restriction and hypermutation (hypo-Vif). Importantly, the wide variation in G-to-A hypermutation levels observed in these controlled experiments may help explain the similarly wide variation reported in patient derived HIV-1 sequences (Albin and Harris, 2010; Caride et al., 2002; Fitzgibbon et al., 1993; Gandhi et al., 2008; Hultquist et al., 2011; Janini et al., 2001; Kieffer et al., 2005; Kijak et al., 2008; Koning et al., 2009; Land et al., 2008; Pace et al., 2006; Piantadosi et al., 2009; Refsland et al., 2010; Suspène et al., 2006; Ulenga et al., 2008; Vartanian et al., 1994; Wood et al., 2009). Variation occurs for both overall G-to-A mutation frequency and for local dinucleotide preferences with, in many instances, GA-to-AA mutations predominating (*e.g.* (Janini et al., 2001; Land et al., 2008)). These *in vivo* biases could be due to the combination of stable A3H haplotypes and infection by hypo-functional Vif isolates.

The results presented here, together with recent data from the Simon group (Ooms et al., 2013a), combine to strongly suggest that stable A3H haplotypes may be functioning as contemporary HIV-1 restriction factors. Stable A3H haplotypes may contribute to disease progression by limiting HIV-1 replication within an infected individual, as indicated by statistically lower viral loads and higher CD4+ T cell counts in haplotype II patients in comparison to haplotype I patients (Ooms et al., 2013a). However, stable A3H haplotypes may also affect rates of transmission. In particular, our studies predict lower rates of virus acquisition when an infected patient has an unstable haplotype and their uninfected partner has a stable A3H haplotype (**Figure 5C**). Moreover, in instances where HIV-1 breaches this transmission barrier, our studies predict that the Vif protein will have to adapt in order to effectively counteract stable A3H. Thus, HIV-1 may need to adapt differently to the A3 repertoire in different humans, supported by correlations between worldwide Vif and A3H genotypes (**Figure 5A & B**). Overall, polymorphisms in human A3H and HIV-1 Vif appear to be combining to actively limit HIV-1 pathogenesis.

MATERIALS AND METHODS

Virus constructs

Vif-proficient (GenBank EU541617) and Vif-deficient (X26, X27) HIVIIIB A200C proviral constructs have been described (Albin et al., 2010a; Haché et al., 2008). Substitutions to construct the hypo-Vif and hyper-Vif variants were introduced with site-directed mutagenesis using primers RSH6951 5'TCA-AGG-AAA-GCT-AAG-GAC-

TGG-GTT-TAT-AGA-CAT-CAC-TAT-GAA-AG & RSH6952 5'CTT-TCA-TAG-TGA-TGT-CTA-TAA-ACC-CAG-TCC-TTA-GCT-TTC-CTT-GA for the F39V substitution, RSH6949 5'TTT-ATA-GAC-ATC-ACT-ATG-AAA-GTA-CTC-ATC-CAA-AAA-TAA-GTT-CAG-AAG-TAC-AC & RSH6950 5'GTG-TAC-TTC-TGA-ACT-TAT-TTT-TGG-ATG-AGT-ACT-TTC-ATA-GTG-ATG-TCT-ATA-AA for the N48H substitution, and RSH6965 5'TCC-AAA-AAT-AAG-TTC-AGA-AGT-ACA-CAT-CCC-ACT-AGA-GAA-GGG-CGA-GTT-AGT-AAT-AAC-AAC-ATA-TTG-GGG-TCT-GCA-TAC-AGG & RSH6966 5'CCT-GTA-TGC-AGA-CCC-CAA-TAT-GTT-GTT-ATT-ACT-AAC-TCG-CCC-TTC-TCT-AGT-GGG-ATG-TGT-ACT-TCT-GAA-CTT-ATT-TTT-GGA for the GDAK60-63EKGE substitutions.

Cell lines

SupT11 and CEM-GFP T cells were maintained in RPMI supplemented with 10% fetal bovine serum (FBS) and 0.5% penicillin/streptomycin (P/S). 293T cells were cultured in DMEM supplemented with 10% FBS and 0.5% P/S. The generation and characterization of the SupT11 panel stably expressing vector, A3D-HA, A3F-HA, A3G-HA, and untagged A3H haplotype II have been described (Hultquist et al., 2011).

APOBEC3H antibodies and immunoblotting

APOBEC3H monoclonal antibodies P1H6-1, P1D8-1, P3A1-1, and P3A3-A10 were generated previously (Li et al., 2010) and purified by ammonium sulfate precipitation. We presume that the increased sensitivity of these antibodies over the original description (Li et al., 2010) is due to purification and increased concentration relative to the

unpurified culture supernatants used previously. Aliquots of two of these monoclonal antibodies have been made available through the NIH AIDS Reagent Program as #12155 (P3A3-A10) and 12156 (P1H6-1). Specificity was demonstrated by immunoblotting lysates from 293T cells transiently transfected with HA-tagged A3A, A3B, A3C, A3D, A3F, A3G, and A3H haplotype II (**Figure S1A**). Epitopes were mapped by immunoblotting lysates from 293T cells transiently transfected with a panel of chimeric human/cow A3H/A3Z3 constructs (**Figure S1C**). A polyclonal antibody raised against APOBEC3H residues 45-183 was used according to the manufacturer's instructions (NBP1-91682, Novus Biologicals). Cells were pelleted, washed, and then directly lysed in 2.5X Laemmli sample buffer. Virus containing supernatants were filtered and virus-like particles isolated by centrifugation through a 20% sucrose cushion and resuspended in 2.5X Laemmli sample buffer. Lysates were subjected to SDS-PAGE and protein transfer to PVDF using the Criterion system (Bio-Rad).

RT-qPCR

Isolation of total RNA, reverse transcription, and qPCR were performed as described (Refsland et al., 2010). Briefly, total RNA was isolated from cells using the RNeasy kit (Qiagen). 1 μg of RNA was used to generate cDNA with Transcriptor reverse transcriptase (Roche) according to the manufacturer's instructions using random hexameric primers. Quantitative PCR was performed using a LightCycler480 instrument (Roche). All reactions were done in triplicate and *A3* levels were normalized to the housekeeping gene *Tata Binding Protein* (*TBP*).

Selection of APOBEC3H adapted viruses

The procedures to adapt viruses in culture to an APOBEC3 challenge are based on prior reports (Albin et al., 2010a; Haché et al., 2008). HIV-1 IIIB Vif N48H was generated as described above. The selection experiments were initiated at a MOI of 0.05 by infecting 150,000 cells, 100% of which were SupT11 expressing low levels of A3H haplotype II in a total volume of 1 ml in one well of a 24 well plate. Infections were monitored by removing supernatants and infecting the reporter cell line CEM-GFP every 3-4 days. The infected cells were split and the media was replenished at 4 days post-infection to prevent overgrowth of the culture. On day 8 of the infection, 250 µl of the virus containing supernatant from the infected wells was passaged to 150,000 SupT11 cells expressing higher levels of A3H haplotype II (50% low/50% intermediate). The virus supernatants were passaged in a stepwise manner 4 additional times into cultures with increasing levels of A3H haplotype II expression: 1:2:1 low:intermediate:high; 1:1 int:high; 1:3 int:high; 100% high. To purify emerging adapted viruses, a culture of 100% SupT11 expressing high levels of A3H haplotype II was infected at a MOI of 0.05 an additional 3 times. Finally, viruses were used to infect CEM-GFP and genomic DNA was isolated using the Puregene reagents (Qiagen). Proviral DNA encoding a 1287-bp fragment of *pol-vif-vpr* amplified with primers RSH1438 5'CCC-TAC-AAT-CCC-CAA-AGT-CA and RSH1454 5' CAA-ACT-TGG-CAA-TGA-AAG-CA. Amplicons were cloned into CloneJet (ThermoScientific) and Sanger sequenced using flanking plasmid-specific primers.

HIV-1 spreading infections

Vif-proficient and Vif-deficient HIV-1 spreading infections were performed as previously described (Albin et al., 2010a). Viruses were generated by transfecting 10 µg of proviral expression construct into 293T cells. Titers of the viruses were assessed using the CEM-GFP reporter cell line. SupT11 spreading infections were initiated at a 0.01 MOI and the infection was monitored every 2-3 days using the CEM-GFP reporter line (Gervaix et al., 1997). Primary CD4+ T cells were infected at a 0.02 MOI and culture supernatants were collected every 2-3 days and analyzed using an in-house p24 ELISA. Supernatants from infected cultures were incubated with for 1 hr on anti-p24 mAb (183-H12-5C, NIH ARRRP) coated 96-well plates (Nunc). Following 4 washes with PBS 0.1% Tween 20 (PBS-T), a second 1 hr incubation anti-p24 mAb (9725) was used to 'sandwich' the p24 antigen. Wells were again washed 4 times, p24 was quantified by 0.5 hr incubation with an enzyme-linked secondary goat anti-mouse IgG-2A/HRP followed by 4 PBS-T wash steps and incubation with 3,3',5,5' tetramethybenzidine (TMB) for 6 min. The reaction was stopped upon addition of 1 M H_2SO_4 and absorbance at 450 nm was quantified on a microplate reader (Synergy MX, Biotek).

Primary cells

Peripheral blood mononuclear cells were isolated from whole blood (Memorial Blood Center, St. Paul, MN) by ficoll gradient centrifugation as previously described (Hultquist et al., 2011; Refsland et al., 2010). Naïve CD4+ T lymphocytes were purified by negative selection according to the manufacturer's instructions (Miltenyi Biotech). Cells were stimulated and maintained in RPMI with 20 U/mL human interleukin-2 (Miltenyi

Biotech) and 10 μg/ml phytohemagglutinin (Thermo Scientific). Cells were stimulated for 72 hours prior to infection or harvesting for total RNA and protein lysate. Purity (> 95%) and activation were confirmed by staining with an anti-CD4 antibody or anti-CD25 antibody respectively (Miltenyi Biotech).

Ethics Statement

This study has been reviewed and exempted by the University of Minnesota Institutional Review Board (#0503E68687). All PBMCs were isolated from healthy and de-identified individuals.

***APOBEC3H* genotyping**

Primers and PCR conditions for polymorphisms (N15/ΔN15A3H, rs140936762; R18/L18, rs139293; G105/R105, rs139297; D121/K121, rs139299 & rs139298; E178/D178, rs139302) genotyping have been described (Wang et al., 2011). Amplicons were cloned into CloneJet (ThermoScientific) and Sanger sequenced. When possible, linkage of heterozygous polymorphic sites was established by Sanger sequencing of amplicons from cDNA following reverse transcription of total RNA. Healthy donors were assessed for A3H haplotype by preliminary screening of PBMCs by immunoblot for A3H protein expression and a diagnostic PCR for the N15/ΔN15 polymorphism. CD4+ T cells were isolated from the selected donors (**Table S1**) as described above.

3D-PCR

To assess global hypermutation semi-quantitatively, 3D-PCR was performed as described

(Hultquist et al., 2011). 15 days postinfection virus containing supernatants were used to infect 25,000 CEM-GFP cells. 48 hours later, genomic DNA was isolated (Qiagen) and an 876 bp region of *pol* was amplified from proviral DNA. The relative amount of this amplicon in each sample was assessed with qPCR (Roche). Normalized amounts of proviral DNA was then used to generate a smaller 564 bp inner amplicon over a range of denaturation temperatures (77.0-86.2° C). The resulting PCR products were fractionated using agarose gels and visualized following ethidium bromide staining.

Proviral sequencing

Hypermutation spectra analysis was performed as described (Hultquist et al., 2011; Refsland et al., 2012). The 876 bp outer amplicon generated from proviral *pol* DNA was used to seed a second, inner PCR reaction with a uniform denaturation temperature of 98°C. The resulting 564 bp amplicon was cloned into CloneJet (ThermoScientific). A total of 10 independent clones (>5 kb) were Sanger sequenced for each condition. Clones with identical mutations were eliminated.

ADDITIONAL CONTRIBUTIONS

N. Somia kindly provided the anti-p24 antibody. This work was supported by grants from the National Institutes of Health (R01 AI064046 and P01 GM091743 to RSH; F31 DA033186 to EWR; R01 AI030927 to ME).

FIGURES

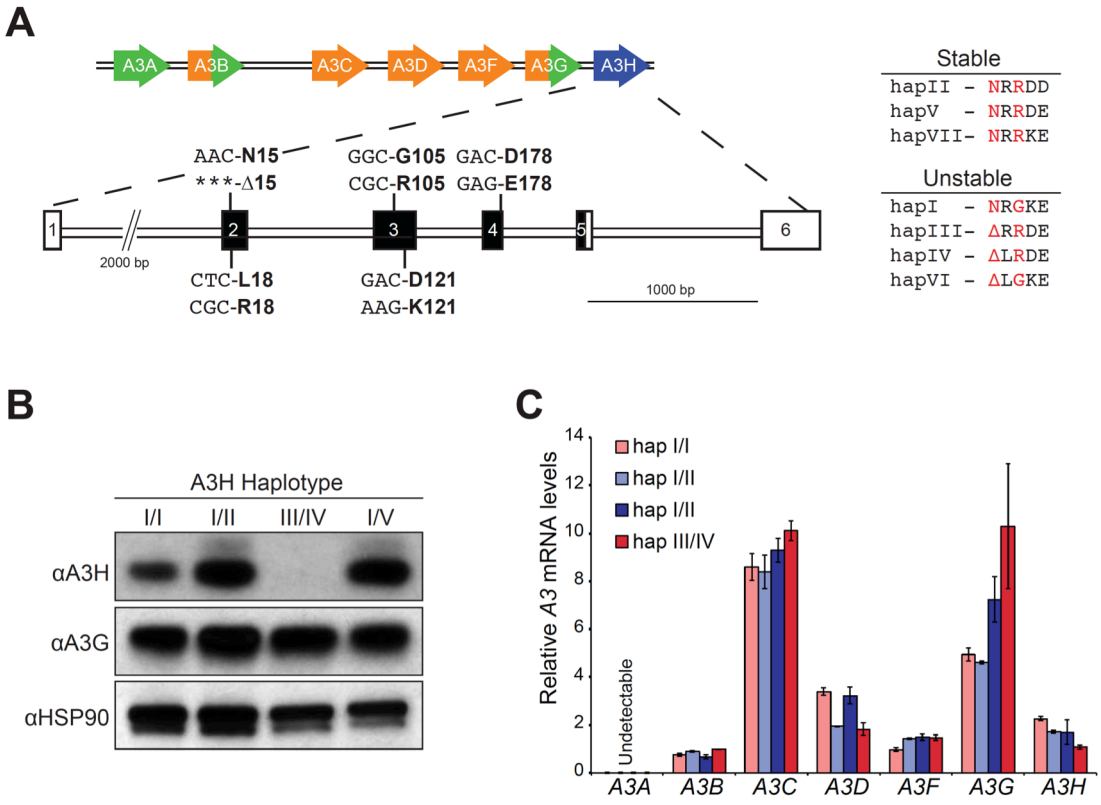

A
A3A
A3B
A3C
A3D
A3F
A3G
A3H
AAC-N15
***-Δ15
GGC-G105 GAC-D178
CGC-R105 GAG-E178
1
2
3
4
5
6
2000 bp
CTC-L18
CGC-R18
GAC-D121
AAG-K121
1000 bp
Stable
hapII - NRRDD
hapV - NRRDE
hapVII- NRRKE
Unstable
hapI - NRGKE
hapIII- ΔRRDE
hapIV - ΔLRDE
hapVI - ΔLGKE
B
A3H Haplotype
I/I
I/II
III/IV
I/V
αA3H
αA3G
αHSP90
C
Relative A3 mRNA levels
hap I/I
hap I/II
hap I/II
hap III/IV
Undetectable
A3A
A3B
A3C
A3D
A3F
A3G
A3H

Figure 4-1. Endogenous APOBEC3H stability/instability occurs at the protein level.

A) A schematic of the 7 gene *A3* locus and the 5 polymorphisms in *A3H* exons 2, 3, and 4 that combine to produce 7 different haplotypes. On the right, a summary of the 7 different A3H haplotypes based on observed protein stability or instability in overexpression studies (OhAinle et al., 2008; Wang et al., 2011) and as defined here, in primary CD4+ T lymphocytes. The two residues that underlie the stable/unstable phenotypes are highlighted in red (N15/Δ15 and R105/G105).

B) Immunoblots showing endogenous A3H, A3G, and HSP90 protein levels in stimulated primary T lymphocytes from 4 donors with the indicated *A3H* haplotypes (donors 10, 11, 12, and 18). In this experiment endogenous A3H is detected with a polyclonal rabbit antibody.

C) *A3* mRNA levels in primary T lymphocytes from 4 donors with the indicated *A3H* haplotypes (donors 4, 10, 12, and 18). Expression levels are shown relative to the housekeeping gene *TBP*.

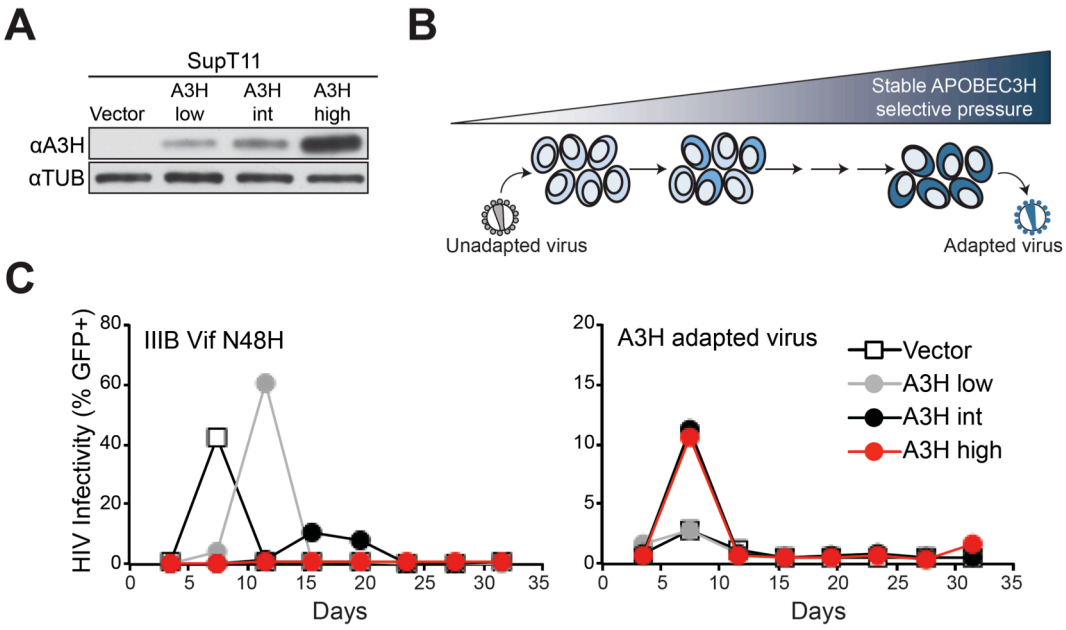


Figure 4-2. APOBEC3H adaptation studies.

A) Immunoblots of A3H and tubulin in SupT11 cell lines stably expressing low, intermediate (int), and high levels of stable haplotype II protein. In this experiment A3H is detected with the mouse monoclonal antibody P3A3-A10.

B) Schematic of the stepwise A3H adaptation procedure (see text and **Materials and Methods** for details).

C) Spreading infection kinetics of the parental and adapted viruses on the A3H-expressing SupT11 lines shown in A. The parental virus is completely inhibited in cells expressing high A3H levels, whereas adapted virus replication is robust.

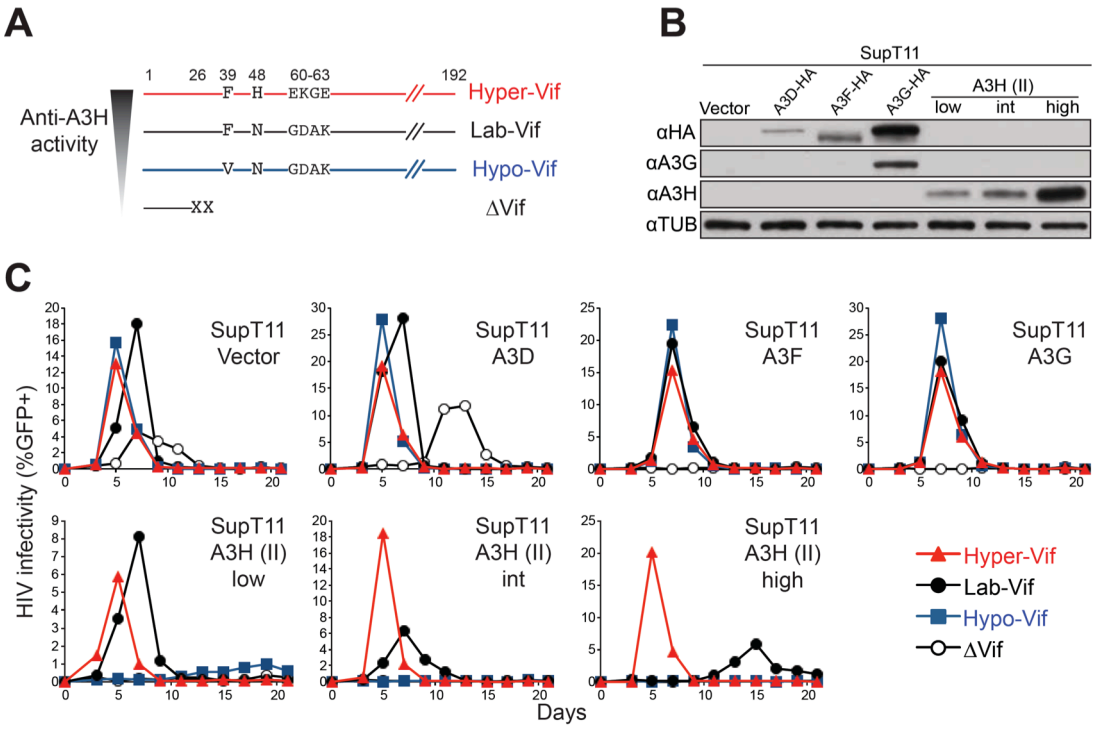

A
Anti-A3H activity
1
26
39
48
60-63
192
F—H—EKGE
Hyper-Vif
F—N—GDAK
Lab-Vif
V—N—GDAK
Hypo-Vif
XX
ΔVif
B
SupT11
Vector
A3D-HA
A3F-HA
A3G-HA
A3H (II)
low
int
high
αHA
αA3G
αA3H
αTUB
C
HIV infectivity (%GFP+)
SupT11 Vector
SupT11 A3D
SupT11 A3F
SupT11 A3G
SupT11 A3H (II) low
SupT11 A3H (II) int
SupT11 A3H (II) high
Days
Hyper-Vif
Lab-Vif
Hypo-Vif
ΔVif

Figure 4-3. Generation and validation of HIV-1 Vif separation-of-function molecular/viral probes.

A) A schematic of the Vif protein encoded by each HIV-1 molecular clone showing amino acid differences responsible for the hyper- and hypo-Vif functionality relative to lab-Vif (HIV-1 IIIB/NL4-3) against stable A3H haplotype II.

B) Immunoblots showing the expression levels of the indicated A3 proteins stably expressed in SupT11 cells. In this experiment untagged A3H is detected with the mouse monoclonal antibody P3A3-A10.

C) HIV-1 spreading infection kinetics for the indicated viruses on A3-expressing SupT11 cells lines described in panel B. The hyper-, lab-, and hypo-Vif isolates spread with similar kinetics on cells expressing a control vector, A3D, A3F, or A3G, but showed clear phenotypic differences on cells expressing low, intermediate (int), and high levels of stable A3H haplotype II.

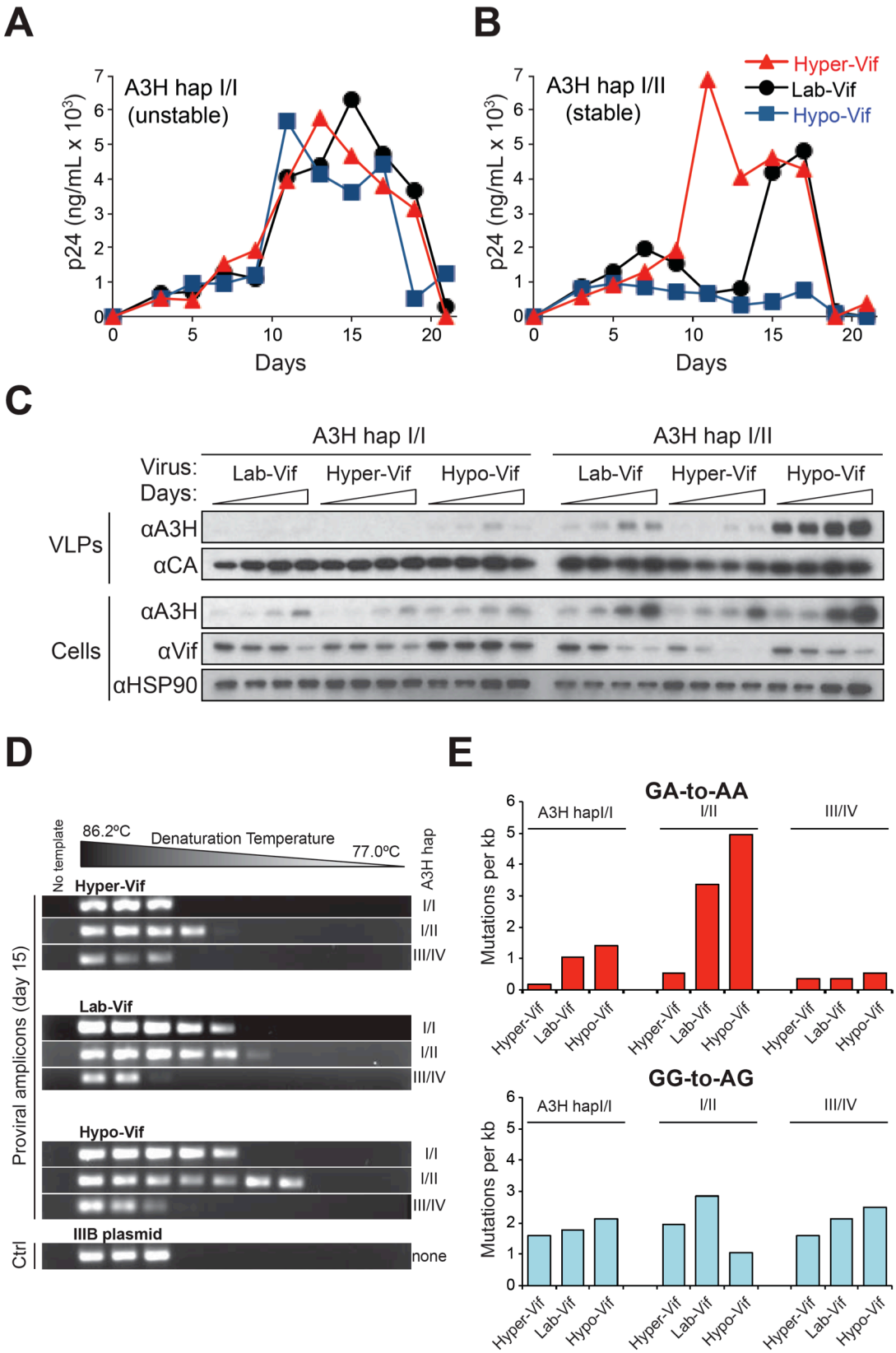

A
A3H hap I/I
(unstable)
p24 (ng/mL x 10^3)
Days
B
A3H hap I/II
(stable)
Hyper-Vif
Lab-Vif
Hypo-Vif
C
A3H hap I/I
A3H hap I/II
Virus:
Days:
Lab-Vif
Hyper-Vif
Hypo-Vif
VLPs
αA3H
αCA
Cells
αA3H
αVif
αHSP90
D
No template
86.2ºC
Denaturation Temperature
77.0ºC
A3H hap
Hyper-Vif
Lab-Vif
Hypo-Vif
I/I
I/II
III/IV
Proviral amplicons (day 15)
Ctrl
IIIB plasmid
none
E
GA-to-AA
GG-to-AG
Mutations per kb
A3H hapI/I
I/II
III/IV
Hyper-Vif
Lab-Vif
Hypo-Vif

Figure 4-4. Stable APOBEC3H inhibits HIV-1 replication in primary T lymphocytes and inflicts GA-to-AA hypermutations.

A) HIV-1 replication kinetics of the hyper-, lab-, and hypo-Vif variants in CD4+ T lymphocytes from a representative healthy donor encoding unstable A3H haplotype I/I (donor 2).

B) HIV-1 replication kinetics of the hyper-, lab-, and hypo-Vif variants in CD4+ T lymphocytes from a representative healthy donor encoding one allele of stable A3H haplotype II and one allele of unstable A3H haplotype I (donor 4).

C) Immunoblots of the indicated proteins in virus-like particles (VLPs) and in cells from days 3, 5, 7, and 9 of the spreading experiments shown in panels A & B.

D) 3D-PCR amplicons generated from proviral DNA of the indicated viruses isolated on day 15 of a spreading infection of unstable A3H (haplotypes I/I and III/IV) or stable A3H (haplotype I/II) donor cells (donors 2, 4, 12). This experiment is representative and performed independently of those shown in panels A-C.

E) Histograms depicting the frequencies of GA-to-AA and GG-to-AG mutations under the indicated spreading infection conditions (complementary to the experiment shown in panel D with all sequences derived from independent 98°C high-fidelity PCR amplifications). A minimum of 10 clones were sequenced for each condition (≥5 kb).

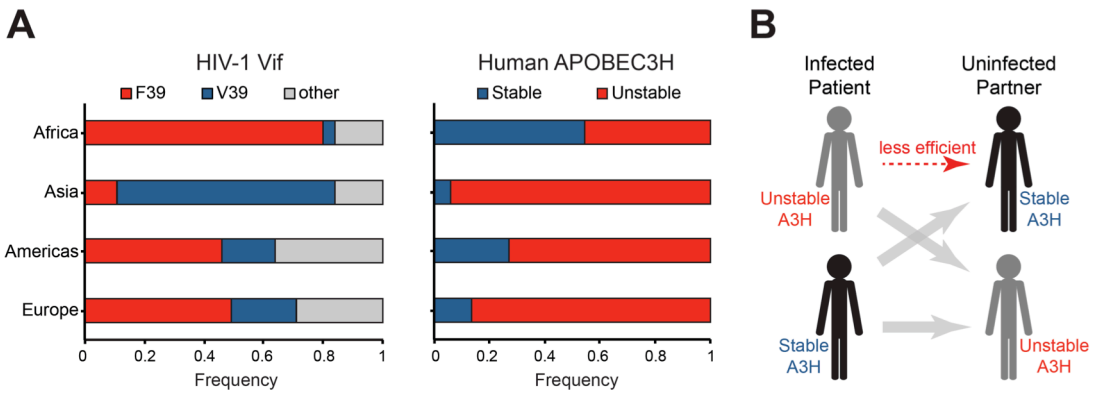


Figure 4-5. Correlations between the global distributions of HIV-1 hyper-Vif alleles and human A3H haplotypes.

A) The left histogram depicts the frequency of HIV-1 isolates encoding a phenylalanine or valine at Vif residue 39 from the indicated geographic regions (n=9713; www.hiv.lanl.gov). The right histogram shows the frequency of stable versus unstable *A3H* alleles from the same geographic regions (n=1092; www.1000Genomes.org).

B) A model depicting the anticipated relative transmission efficiencies between infected patients and uninfected individuals with equivalent or different *A3H* haplotypes.

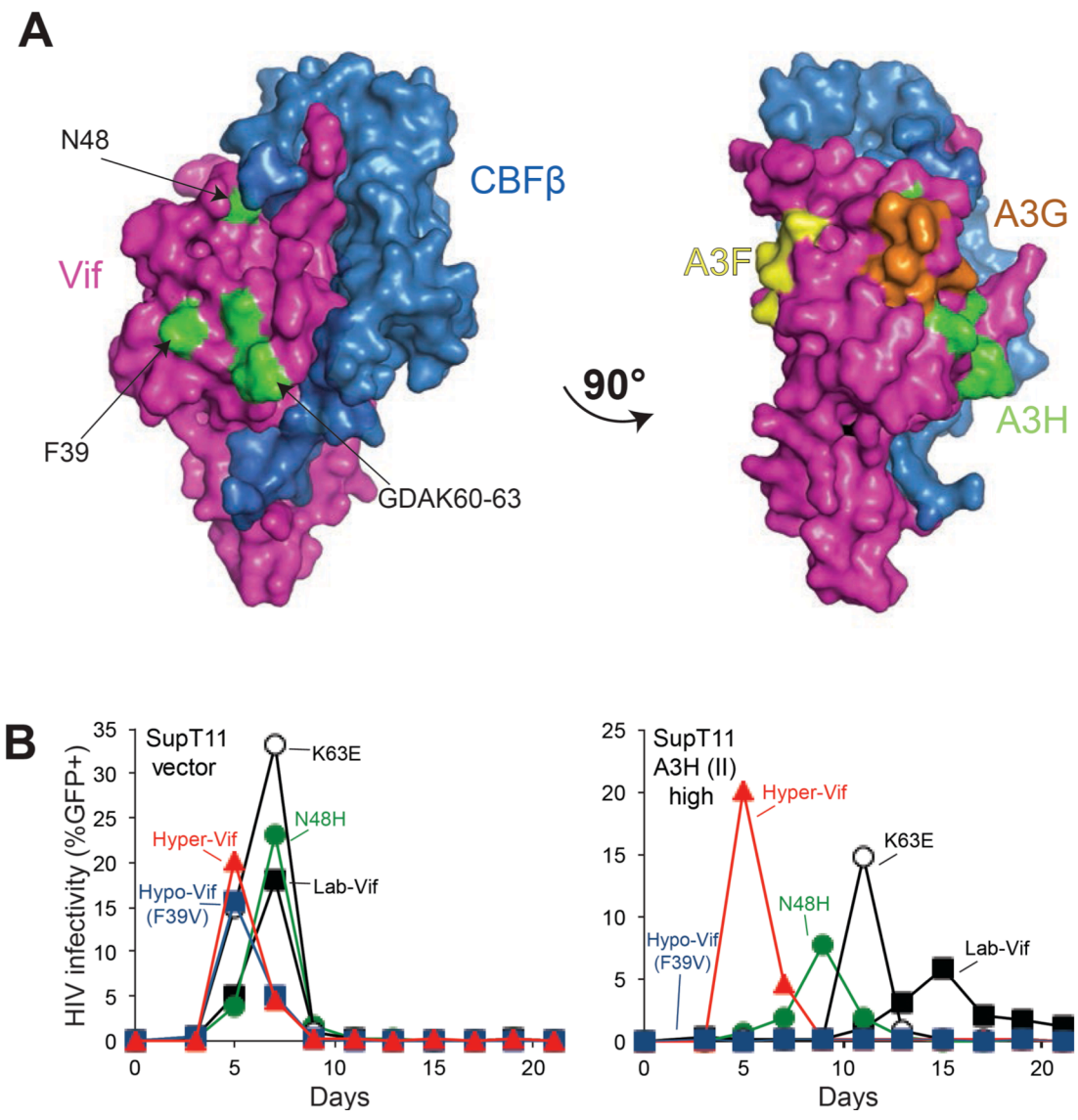

A
N48
CBFβ
Vif
F39
GDAK60-63
90°
A3G
A3F
A3H
B
HIV infectivity (%GFP+)
SupT11
vector
K63E
N48H
Hyper-Vif
Lab-Vif
Hypo-Vif
(F39V)
Days
SupT11
A3H (II)
high
Hyper-Vif
K63E
N48H
Hypo-Vif
(F39V)
Lab-Vif
Days

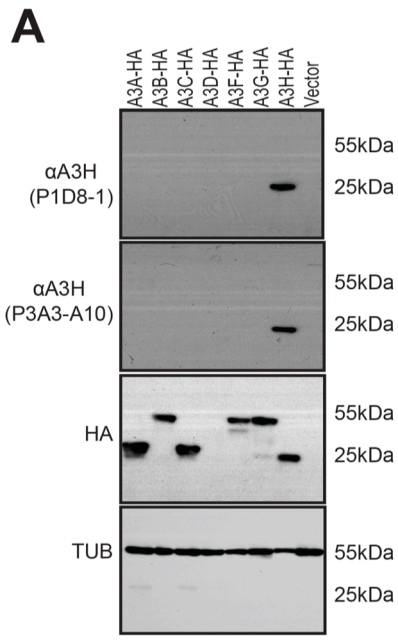
Figure 4-6. Vif separation-of-function substitutions define the likely APOBEC3H interaction surface.

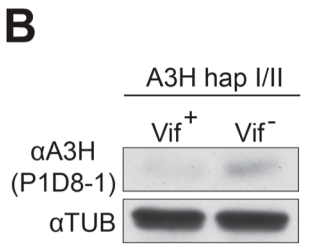
A) A surface representation of Vif and CBFβ (pdb 4N9F). The side-chains of the amino acid residues conferring hypo-Vif (residue 39) and hyper-Vif phenotypes (residues 48 and 60-63) are shaded green and located on a common solvent-exposed surface. A 90° rotation reveals distinct Vif separation-of-function residues implicated in the interactions with A3F (yellow) and A3G (orange). See main text for details.

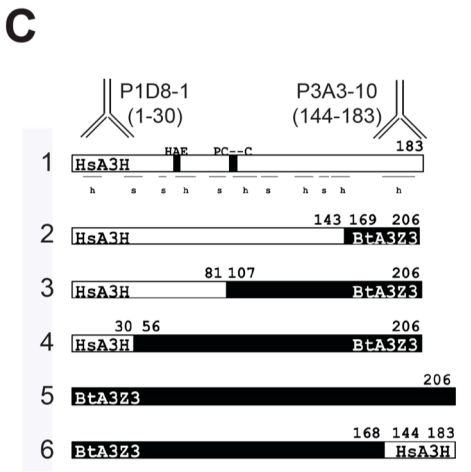
B) Spreading infection kinetics of the indicated HIV-1 Vif variants on SupT11 cells stably expressing a vector control (left) or high levels of A3H haplotype II (right). Figure 4C data for hyper-, lab-, and hypo-Vif are shown again here to facilitate comparisons.

	A3H Polymorphisms						
Donor	**15**	**18**	**105**	**121**	**178**	**A3H Haplotype**	**Expression**
1	N/Δ	R/L	G/R	K/D	E/E	I/IV	Unstable
2	N/N	R/R	G/G	K/K	E/E	I/I	Unstable
4	N/N	R/R	G/R	K/D	E/D	I/II	Stable
5	N/N	R/R	G/G	K/K	E/E	I/I	Unstable
8	N/N	R/R	G/G	K/K	E/E	I/I	Unstable
10	N/N	R/R	G/R	K/D	E/D	I/II	Stable
11	N/N	R/R	G/R	K/D	E/E	I/V	Stable
12	Δ/Δ	R/L	R/R	D/D	E/E	III/IV	Unstable
18	N/N	R/R	G/G	K/K	E/E	I/I	Unstable

Table 4-S1. Selected donor genotyping results.


A
A3A-HA
A3B-HA
A3C-HA
A3D-HA
A3F-HA
A3G-HA
A3H-HA
Vector
αA3H
(P1D8-1)
αA3H
(P3A3-A10)
HA
TUB
55kDa
25kDa


B
A3H hap I/II
Vif+
Vif-
αA3H
(P1D8-1)
αTUB


C
P1D8-1
(1-30)
P3A3-10
(144-183)
HAE
PC--C
HsA3H
BtA3Z3
183
143 169 206
81 107 206
30 56 206
206
168 144 183

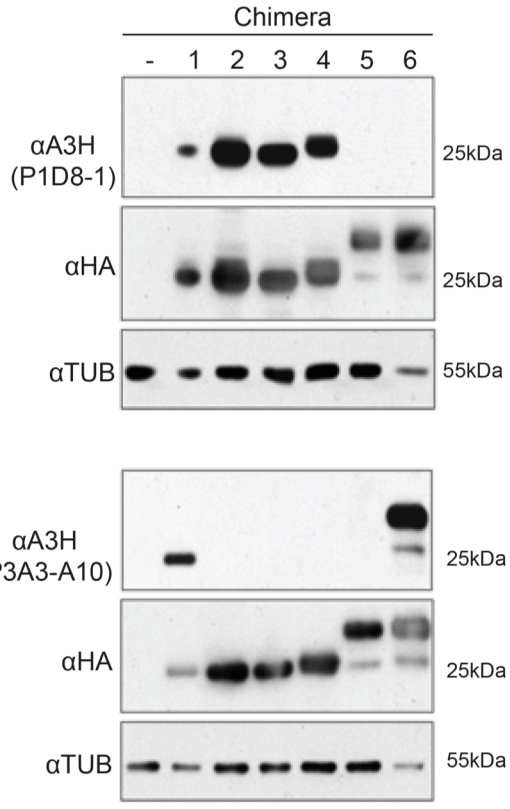

Chimera
- 1 2 3 4 5 6
αA3H
(P1D8-1)
αHA
αTUB
αA3H
(P3A3-A10)
αHA
αTUB
25kDa
55kDa

Figure 4-S1. Characterization of mouse monoclonal antibodies specific to human APOBEC3H.

A) Immunoblots demonstrating the specificity of the A3H monoclonal antibodies P1D8-1 and P3A3-A10 in 293T cells transiently expressing the indicated A3-HA protein.

B) Immunoblots of A3H and tubulin expression levels in primary T lymphocytes 9 days after infection with Vif proficient or deficient viruses. A3D-HA expressed poorly in this experiment (donor 25).

C) Schematics of human A3H hap II (open box), cow A3Z3 (black box), and A3H/A3Z3 chimeric derivatives. The epitopes for mouse monoclonal antibodies P1D8-1 and P3A3-A10 are shown. Immunoblots of 293T cells transiently transfected with the indicated human/cow chimeric A3H/A3Z3 constructs. Monoclonal antibodies P1D8-1 and P3A3-A10 recognize distinct N- and C-terminal epitopes, respectively.

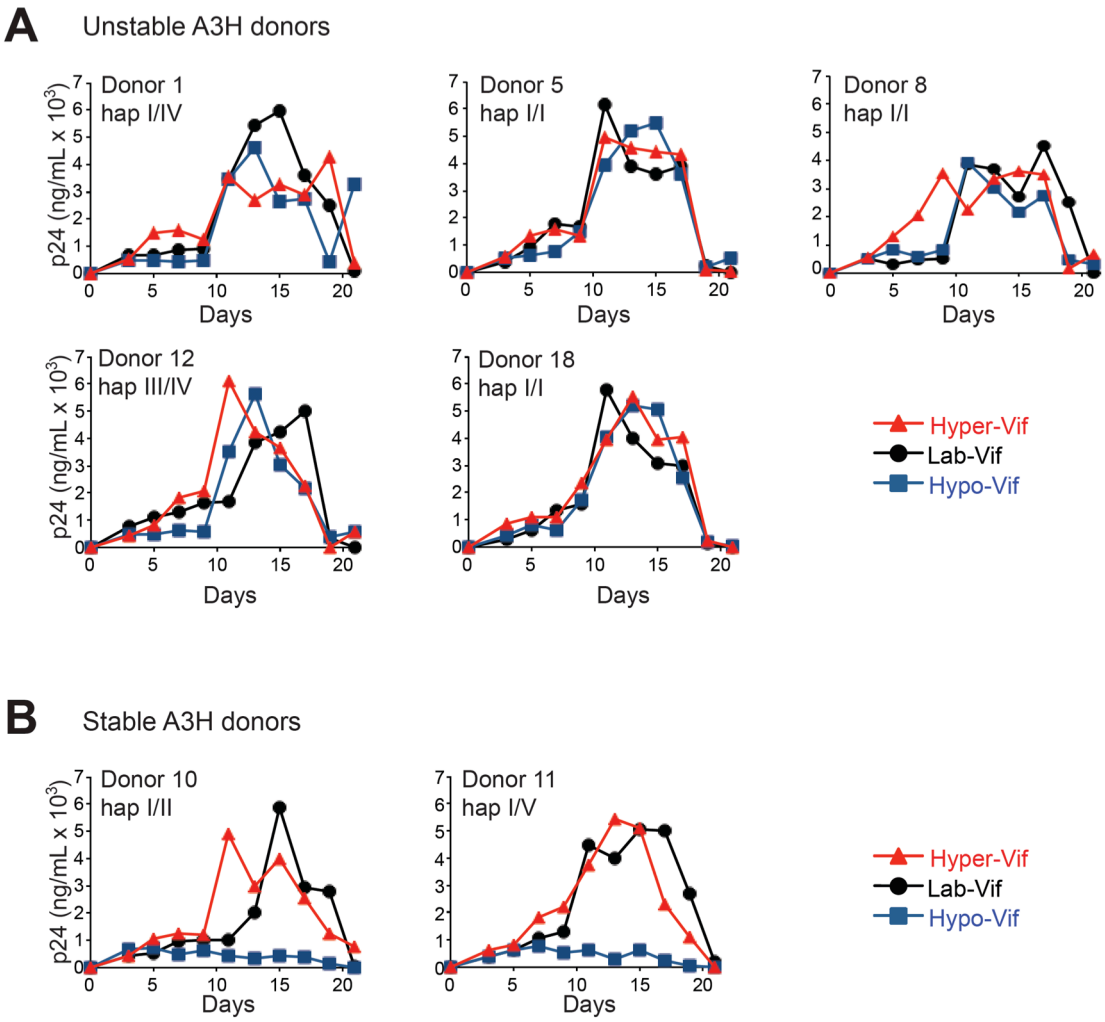

A
Unstable A3H donors
Donor 1
hap I/IV
Donor 5
hap I/I
Donor 8
hap I/I
Donor 12
hap III/IV
Donor 18
hap I/I
p24 (ng/mL x 10^3)
Days
Hyper-Vif
Lab-Vif
Hypo-Vif
B
Stable A3H donors
Donor 10
hap I/II
Donor 11
hap I/V
p24 (ng/mL x 10^3)
Days
Hyper-Vif
Lab-Vif
Hypo-Vif

Figure 4-S2. Stable APOBEC3H inhibits HIV-1 replication in primary T lymphocytes.

A) HIV-1 replication kinetics of the hyper-, lab-, and hypo-Vif variants in CD4+ T lymphocytes from 5 healthy donors encoding the unstable A3H haplotype indicated (donors 1, 5, 8, 12, and 18).

B) HIV-1 replication kinetics of the hyper-, lab-, and hypo-Vif variants in CD4+ T lymphocytes from 2 healthy donors heterozygous for the indicated allele of stable A3H (donors 10 and 11).

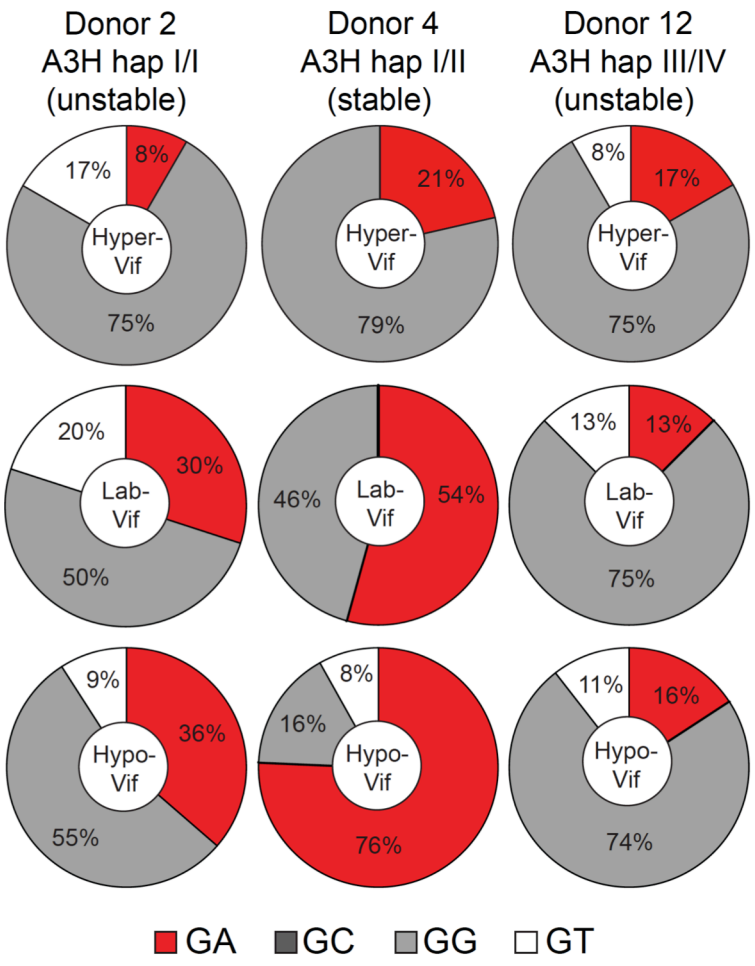


Figure 4-S3. Stable APOBEC3H alleles inflict GA-to-AA hypermutations in viruses encoding hypo-Vif variants.

HIV-1 G-to-A mutation profiles of the hyper-, lab-, and hypo-Vif proviruses originating from primary T lymphocytes with the indicated A3H haplotype (donors 2, 4, and 12). GA-to-AA mutations characteristic of A3H activity are shown in red.

CHAPTER 5

CONCLUSIONS AND DISCUSSION – APOBEC3 Expression and the HIV-Restrictive APOBEC3 Repertoire

FOREWORD

HIV pathogenesis requires the hijacking or neutralization of a significant portion of human proteins. These proteins are collectively referred to as either dependency or restriction factors. While the current treatments for HIV infection include drugs that target the enzymes encoded by the HIV genome, an attractive alternative target is the interaction between other viral proteins and these dependency or restriction factors. Especially appealing are the HIV restriction factors of the APOBEC3 family of DNA cytosine deaminases. These enzymes directly inhibit viral replication by inflicting lethal levels of mutagenesis following the first round of reverse transcription. These potent innate immune factors are neutralized by HIV Vif, which nucleates the formation of an ubiquitin ligase complex and binds the APOBEC3 proteins to initiate their degradation. Successfully disrupting this interaction, either the APOBEC3/Vif interaction, or Vif with one of its essential cellular co-factors, necessitates the identification of all the relevant participants.

To determine which *APOBEC3* genes are expressed in CD4+ T cells and positioned to restrict HIV replication, my colleagues and I developed first-in-their class qPCR assays specific for each of the 7 *APOBEC3* genes. These assays where used to compare *APOBEC3* expression in human T cell lines, primary T lymphocytes, and 22 normal human tissues. These data demonstrated that multiple *APOBEC3* genes are expressed in most types of cells, positioned together to impede attacks from retroelements, and that this expression can be induced by transcriptional activation.

To begin to address which individual APOBEC3 proteins restrict HIV replication *in vivo*, we generated isogenic human cell lines lacking only APOBEC3F or APOBEC3G

expression. We concluded from Vif deficient HIV single-cycle and spreading infection experiments that APOBEC3D, APOBEC3F, and APOBEC3G combine to restrict HIV and are responsible for nearly all of the G-to-A hypermutation observed in the human cell line CEM2n.

A fourth APOBEC3, APOBEC3H has been implicated in HIV restriction but its contribution remains somewhat ambiguous partly due to a high degree of genetic diversity. To gain an appreciation for how this genetic variation impacts the expression and HIV restriction activity of APOBEC3H, we used a series of primary cell experiments and HIV spreading infections to probe the 7 distinct haplotypes encoded by the human genome. Three of these haplotypes encode a protein that expresses well and is able to restrict circulating HIV variants that encode hypofunctional Vif proteins unable to mount a counter-defense against this subset of *APOBEC3H* alleles. We conclude from these data that APOBEC3H constitutes a natural barrier to HIV replication of isolates that may have lost or never developed the ability to antagonize this APOBEC3.

Taken together, my research demonstrates that multiple APOBEC3 are expressed constitutively in HIV target cells. This expression increases upon cellular activation and further increases upon HIV infection. By inflicting catastrophic levels of mutation in the genetic information of the virus, 4 of these proteins, APOBEC3D, APOBEC3F, APOBEC3G and stable APOBEC3H haplotypes restrict virus replication. These insights will forward the pursuit of targeted innate immune therapies that can free the swarm of HIV-restrictive APOBEC3s and extinguish viral replication.

APOBEC3 Expression

The APOBEC3 (A3; A3A-H) proteins inhibit a variety of retroelements including endogenous retroviruses, retrotransposons, DNA viruses, and exogenous retroviruses including HIV [reviewed in (Chiu and Greene, 2008; Goila-Gaur and Strebel, 2008; Hultquist and Harris, 2009)]. However, the individual A3 proteins that function *in vivo* against each of these substrates is poorly defined. For example, both A3F and A3G strongly restrict HIV *in vitro* and likely contribute to HIV restriction *in vivo*, but all other human A3 proteins have also been implicated [reviewed by (Albin and Harris, 2010)].

The APOBEC3 locus in primates is the result of multiple gene duplication events during evolution from an ancestral mammalian locus (LaRue et al., 2008). As a result, all of the *APOBEC3* genes and gene products share extensive homology including long stretches of hundreds of nucleotides with >96% sequence identity (**Figure 2-1A**). To no surprise then, the development of specific reagents to detect and quantify expression levels has been hindered. To overcome this, we have generated and characterized a full panel of qPCR primers and probes specific to each of the 7 A3s. We used these assays to interrogate *A3* mRNA expression in multiple T cell lines, primary T lymphocytes, and a panel of 20 normal human tissues. Human T cell lines express multiple *A3s* and primary T lymphocytes express 6 of the 7. This is consistent with our hypothesis that multiple A3s target HIV. *A3A* was not expressed in any T cell we examined and we concluded that it does not contribute to restriction in this cell type. *A3* expression was widespread and not confined to the immune compartments analyzed or tissues with high numbers of infiltrating immune cells (*e.g.* the spleen, thymus, adipose, and lung) suggesting a

possible functional role for these enzymes in tissues including colon, cervix, bladder, and heart, a matter deserving of further investigation.

In addition to constitutive expression, we analyzed the A3 expression following transcriptional activation. Multiple prior studies have examined the changes in A3 (most often A3G) expression following type I interferon (IFN) treatment with little consensus emerging (Argyris et al., 2007; Bonvin et al., 2006; Chen et al., 2006; Koning et al., 2009; Peng et al., 2006; Stenglein et al., 2010; Tanaka et al., 2006; Wang et al., 2008). Using our qPCR assays we did not detect any transcriptional increase of any *A3* in primary CD4+ lymphocytes upon IFN treatment. This may indicate that IFN stimulation of *A3* expression is confined to other blood cell lineages (*e.g.* monocytes and macrophages (Chen et al., 2006; Koning et al., 2009; Stenglein et al., 2010). In contrast, 6 of 7 *A3s* were induced upon addition of T cell activators IL-2 and PHA demonstrating the induciblity of *A3* expression in T cells and the likelihood of multiple mechanisms of regulation. Consistent with this idea is an experiment I performed with another graduate student measuring A3 mRNA expression in primary CD4+ T lymphocytes during the course of an HIV spreading infection (Hultquist et al., 2011). The expression of three A3s – A3C, A3G, and A3H dramatically increased over 9 days post-infection indicating an as yet unknown mechanism of transcriptional control imparted by an HIV infection.

It is also important to note that these qPCR assays have added benefits beyond defining the HIV-restrictive repertoire in T cells. Recently they have been applied to cancer cell lines and primary tumor samples (Burns et al., 2013; Leonard et al., 2013). These papers detail the overexpression of A3B in both breast and ovarian cancers and its potential role in mutagenizing the genome and initiating tumorigenesis or in the

development of drug resistance. These studies highlight the importance of our qPCR assays in attributing expression of a specific A3 in any cell type or tissue.

The HIV-Restrictive APOBEC3 Repertoire

In addition to multiple A3s being expressed in the relevant cell types and positioned to target HIV, another clue supporting the role of multiple A3s in HIV restriction is the mutation pattern observed in proviral DNA from infected patients. Two distinct dinucleotide patterns of G-to-A hypermutation emerge, 5'GG-to-AG and 5'GA-to-AA (Janini et al., 2001; Kieffer et al., 2005; Land et al., 2008; Piantadosi et al., 2009). These mutations in the viral coding strand are complementary to the two intrinsic dinucleotide preferences (5'CC and 5'TC) of the A3s on the non-coding strand [reviewed in (Albin and Harris, 2010)]. While our qPCR assays were informative for the reasons I just highlighted above, they were unable to sufficiently narrow the list of A3s potentially relevant to HIV restriction to less than 6 of 7. Of these, A3G has a unique preference for the 5'CC dinucleotide and all of the remaining 5 have a reported preference for 5'TC (Albin and Harris, 2010).

To examine the individual contribution of an A3 on HIV restriction, prior studies have relied on forced transient overexpression in heterologous cell types to define whether or not an A3 *can* restrict HIV replication. Taking an alternative genetic approach, we systematically deleted all copies of a target *A3* gene in a near-diploid human CD4+ T cell line. This generation of an *A3*-null cell lines had four distinct advantages. One, the cell type we manipulated was a relevant target of HIV. Two, the transgene expression was driven by a different promoter, often viral, outside its normal

genomic context. Three, the expression levels were difficult to correlate with endogenous levels. Finally, genetic ablation of an *A3* and creation of an isogenic set of cell lines allowed for a definitive determination of whether or not the A3 *does* restrict HIV replication in that cell type.

A3G is the most studied and best-described A3 family member with potent activity against Vif-deficient HIV (Albin and Harris, 2010; Chiu and Greene, 2008; Harris et al., 2012; Malim, 2009). Using an rAAV-mediated gene targeting approach, we deleted the two copies of *A3G* from the genome of the human T cell line CEM2n. Consistent with prior reports, single-cycle infectivity experiments and proviral DNA sequencing analysis demonstrated that while A3G is the major source for the 5'GG-to-AG hypermutations, it is not the only factor contributing to Vif-deficient HIV restriction. A3F is perhaps the next most potent on the list of A3 restriction candidates (Albin and Harris, 2010). Using a similar approach we deleted both copies of *A3F* from CEM2n. Like *A3G*, deletion of *A3F* did not result in a complete restoration of Vif-deficient HIV replication, presumably due to the presence of A3G in the cells. We expected to see a similar reduction in the levels of 5'GA-to-AA mutations (the dinucleotide preference of A3F) in proviral DNA from viruses spread on *A3F*-null cells. However, the levels of 5'GA-to-AA mutation were virtually unchanged, strongly suggesting another A3 in the generation of this mutation type. Only upon addition of an shRNA directed against A3D did the 5'GA-to-AA mutations vanish. We concluded from these experiments that in CEM2n T cells, 3 A3s – A3D, A3F, and A3G combine to restrict Vif-deficient HIV. In addition, these 3 A3s likely contribute to the 2 major G-to-A mutation patterns observed *in vivo*.

In a set of complimentary experiments performed in parallel, another graduate student and I performed a series of single-cycle and spreading infections on the T cell line SupT11 stably expressing each of the 7 A3s from humans and rhesus macaques (Hultquist et al., 2011). The results of these experiments were consistent with our deletion studies, demonstrating a conserved ability of A3D, A3F, A3G, and A3H to encapsidate into viral particles, to restrict the replication of Vif-deficient HIV and mutagenize proviral DNA sequences. Taken together, these two studies utilizing two different but complimentary approaches arrived at the same conclusion, that T cells are armed with multiple A3s that form an imposing barrier to HIV lacking Vif by inflicting catastrophic levels of mutation of the virus' genome.

An unfortunate difference between the deletion and stable expression experiments discussed above was the former's inability to adequately interrogate A3H. CEM2n cells do not express levels of *A3H* mRNA above background making a definitive assessment of its role in HIV restriction in this system challenging (Refsland et al., 2012). *A3H* is the most genetically diverse of all of the *A3s*. In humans, at least 7 haplotypes are known and the differences between haplotypes manifest in varying degrees of protein stability. Two phenotypic groupings have been assigned to the haplotypes based on prior overexpression analyses, with one group (haplotypes II, V, VII) capable of robust stable expression and the other (haplotypes I, II, IV, VI) either expressed weakly or not at all (OhAinle et al., 2008; Wang et al., 2011). How these haplotype differences affect endogenous A3H protein expression and what impact this genetic diversity has on its HIV restriction activity remain under studied.

To better understand how endogenous A3H haplotypes dictate protein stability and anti-HIV activity, we performed a series of primary cell experiments and *ex vivo* spreading infections using cells isolated from healthy donors with different A3H haplotypes. Using novel antibodies to detect endogenous A3H for the first time, alleles of haplotypes II and V were readily detected by immunoblot. Haplotypes I, III, and IV however, were either weakly detected or undetectable. These results were consistent with prior overexpression studies and demonstrated that humans differentially express A3H. To avoid the confounding and possibly dominant effects of the expression of A3D, A3F, and A3G in primary CD4+ lymphocytes, naturally occurring A3H-specific Vif separation-of-function variants were generated that had either increased capacity (hyper-Vif) or decreased capacity (hypo-Vif) to neutralize A3H. Importantly, these molecular probes retained their ability to counteract A3D, A3F, and A3G. Using these viruses, a single stable allele of A3H was enough to encapsidate, restrict, and hypermutate a hypo-Vif virus, demonstrating that stable A3H alleles are an endogenous barrier to infection by some HIV variants. These data also predict that an individual's A3H haplotype may be used to predict the transmission efficiency of some circulating HIV variants. Large clinical datasets will be required to test this prediction.

But why would humans express an unstable A3H? Three possibilities relate to its function as a restriction factor. First, it's tempting to hypothesize that populations encoding heterogeneity in a potent host restriction factor could limit the possibility of devastating viral sweeps. In other words, decreasing the selective pressure in certain subsets of the population only to reintroduce it upon infection to an individual or subset

of the population that encodes the active form of the restriction factor could be one strategy the slower evolving host immune system adopts to limit pathogen spread.

The second explanation is related to the role the A3 family plays in limiting endogenous retrovirus replication and retrotransposition (Esnault et al., 2005; MacDuff et al., 2009; Schumacher et al., 2005). These retroelements have been implicated in generating diversity, and adaptation within a species (Carmi et al., 2011; Kazazian, 2004). By inactivating a large number of alleles in a population of a retroelement restriction factor, it may permit the slow accumulation of genetic variation driving evolution of a host's genome.

The third explanation is that the ancient pathogenic agent that necessitated the evolution of an A3H deaminase has since been eradicated and the benefit of A3H expression is outweighed by the costs associated with its now unnecessary activity. If this were the case, the ancient pathogen would be very ancient indeed as an A3H-like (or A3Z3-like) single domain deaminase is encoded in every mammal sequenced to date (LaRue et al., 2009; LaRue et al., 2008; LaRue et al., 2010). This implies the original A3H target pathogen would likely be more than 100 million years old, the date around which mammals started to diverge into different clades. But while there is one A3Z3-like deaminase encoded in every mammalian genome, there is only one. Unlike the often duplicated A3Z1 and A3Z2 demainase domains that constitute human A3A-G, A3Z3 (A3H) is only present in a single copy, suggesting that the dosage of this A3 is tightly regulated. Consistent with this idea, another A3 family member A3B, recently implicated in cancer mutagenesis by our laboratory and others, has a similar inactivating genotype in a subset of the human population (Burns et al., 2013; Kidd et al., 2007; Leonard et al.,

2013; Nik-Zainal et al., 2012; Roberts et al., 2013). This raises the possibility that another A3 deaminase, A3H, may have pathological consequences when misregulated and is being selected against in spite of its innate immune functions.

Finally, consideration must be given to the question why primates including human encode 7 A3 proteins when many have similar functions. The most straightforward and likely reason is that immune systems favor redundancy. Multiple A3s (as with other arms of the innate and adaptive immune systems) with overlapping functions constitute a much more formidable barrier to the zoonotic transmissions that underlie the emergence of many of the most pathogenic agents (akin to an invading army encountering a castle it would like to overtake). The castle is far more likely to fend off the invasion if, in addition to its high walls, there is a moat and draw bridge, archers on the towers, and a catapult defending it. In addition, this redundancy may provide the A3 defenses the flexibility required for an increased rate of evolutionary change, an important feature for any immunity factor challenged with keeping up with the extraordinary adaptation rates of viruses. Consistent with this idea is that many of the A3s are under positive selection either from individual challenges or shared pressures to maintain their overlapping functions (Duggal and Emerman, 2012). Thus, while the experiments we have performed here indicate a single A3 is sufficient to restrict Vif-deficient HIV, this may not be the case for other pathogens, contemporary or ancient.

It is easy to forget that HIV-1, while a formidable pathogenic agent, is a very recent addition to the long list of pathogens humans have encountered. It is very likely the root causes of the primate expansion of the *A3* locus some 95 million years ago, was initiated by threats long since extinct. The challenges facing humans today are very

different, reflecting the dispersal of our species and the mobility of modern societies. Our immune defenses are up to this task. Taken together, my research highlights the expression and HIV-restrictive capacity of 4 members- A3D, A3F, A3G, and A3H- of a unique family of DNA cytosine deaminases. I present evidence that one of member, A3H, may in fact constitute a contemporary HIV-1 restriction factor, and its differential protein stability phenotypes may represent the latest move in an ongoing evolutionary conflict. These enzymes, capable of directly inhibiting HIV replication, represent one of the most promising avenues for the development of targeted innate immune therapies.

BIBLIOGRAPHY

Abudu, A., Takaori-Kondo, A., Izumi, T., Shirakawa, K., Kobayashi, M., Sasada, A., Fukunaga, K., and Uchiyama, T. (2006). Murine retrovirus escapes from murine APOBEC3 via two distinct novel mechanisms. Curr Biol *16*: 1565-1570.

Agranat, L., Raitskin, O., Sperling, J., and Sperling, R. (2008). The editing enzyme ADAR1 and the mRNA surveillance protein hUpf1 interact in the cell nucleus. Proc Natl Acad Sci U S A *105*, 5028-5033.

Aguiar, R. S., Lovsin, N., Tanuri, A., and Peterlin, B. M. (2008). Vpr.A3A chimera inhibits HIV replication. J Biol Chem *283*, 2518-2525.

Albin, J. S., Haché, G., Hultquist, J. F., Brown, W. L., and Harris, R. S. (2010a). Long-term restriction by APOBEC3F selects human immunodeficiency virus type 1 variants with restored Vif function. J Virol *84*, 10209-10219.

Albin, J. S., and Harris, R. S. (2010). Interactions of host APOBEC3 restriction factors with HIV-1 in vivo: implications for therapeutics. Expert Rev Mol Med *12*, e4.

Albin, J. S., LaRue, R. S., Weaver, J. A., Brown, W. L., Shindo, K., Harjes, E., Matsuo, H., and Harris, R. S. (2010b). A single amino acid in human APOBEC3F alters susceptibility to HIV-1 Vif. J Biol Chem *285*, 40785-40792.

Alce, T. M., and Popik, W. (2004). APOBEC3G is incorporated into virus-like particles by a direct interaction with HIV-1 Gag nucleocapsid protein. J Biol Chem *279*, 34083-34086.

Ali, A., Wang, J., Nathans, R. S., Cao, H., Sharova, N., Stevenson, M., and Rana, T. M. (2012). Synthesis and structure-activity relationship studies of HIV-1 virion infectivity factor (Vif) inhibitors that block viral replication. Chem Med Chem *7*, 1217-1229.

Anwar, F., Davenport, M. P., and Ebrahimi, D. (2013). Footprint of APOBEC3 on the genome of human retroelements. J Virol *87*, 8195-8204.

Argyris, E. G., Acheampong, E., Wang, F., Huang, J., Chen, K., Mukhtar, M., and Zhang, H. (2007). The interferon-induced expression of APOBEC3G in human blood-brain barrier exerts a potent intrinsic immunity to block HIV-1 entry to central nervous system. Virology *367*, 440-451.

Berger, G., Durand, S., Fargier, G., Nguyen, X. N., Cordeil, S., Bouaziz, S., Muriaux, D., Darlix, J. L., and Cimarelli, A. (2011). APOBEC3A is a specific inhibitor of the early phases of HIV-1 infection in myeloid cells. PLOS Pathog *7*, e1002221.

Betts, S. D., Hachigian, T. M., Pichersky, E., and Yocum, C. F. (1994). Reconstitution of the spinach oxygen-evolving complex with recombinant Arabidopsis manganese-stabilizing protein. Plant molecular biology *26*, 117-130.

Binka, M., Ooms, M., Steward, M., and Simon, V. (2012). The activity spectrum of Vif from multiple HIV-1 subtypes against APOBEC3G, APOBEC3F, and APOBEC3H. J Virol *86*, 49-59.

Bishop, K. N., Holmes, R. K., and Malim, M. H. (2006). Antiviral potency of APOBEC proteins does not correlate with cytidine deamination. J Virol *80*, 8450-8458.

Bishop, K. N., Holmes, R. K., Sheehy, A. M., Davidson, N. O., Cho, S. J., and Malim, M. H. (2004). Cytidine deamination of retroviral DNA by diverse APOBEC proteins. Curr Biol *14*, 1392-1396.

Bishop, K. N., Verma, M., Kim, E. Y., Wolinsky, S. M., and Malim, M. H. (2008). APOBEC3G inhibits elongation of HIV-1 reverse transcripts. PLOS pathogens *4*, e1000231.

Bogerd, H. P., and Cullen, B. R. (2008). Single-stranded RNA facilitates nucleocapsid: APOBEC3G complex formation. RNA *14*, 1228-1236.

Bogerd, H. P., Doehle, B. P., Wiegand, H. L., and Cullen, B. R. (2004). A single amino acid difference in the host APOBEC3G protein controls the primate species specificity of HIV type 1 virion infectivity factor. Proc Natl Acad Sci U S A *101*, 3770-3774.

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

Bogerd, H. P., Wiegand, H. L., Hulme, A. E., Garcia-Perez, J. L., O'Shea, K. S., Moran, J. V., and Cullen, B. R. (2006b). Cellular inhibitors of long interspersed element 1 and Alu retrotransposition. Proc Natl Acad Sci U S A *103*, 8780-8785.

Bohn, M.-F., S.M.D. Shandilya, J.S. Albin, T. Kouno, B.D. Anderson, R.M. McDougle, M.A. Carpenter, A. Rathore, A.N. Davis, J. Zhang, Y. Lu, M. Somasundaran, H. Matsuo, R.S. Harris & C.A. Schiffer (2013). Crystal structure of the HIV-1 Vif binding and catalytically active domain of APOBEC3F. Structure *21*:1042-50.

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

Bransteitter, R., Pham, P., Scharff, M. D., and Goodman, M. F. (2003). Activation-induced cytidine deaminase deaminates deoxycytidine on single-stranded DNA but requires the action of RNase. Proc Natl Acad Sci U S A *100*, 4102-4107.

Brass, A. L., Dykxhoorn, D. M., Benita, Y., Yan, N., Engelman, A., Xavier, R. J., Lieberman, J., and Elledge, S. J. (2008). Identification of host proteins required for HIV infection through a functional genomic screen. Science *319*, 921-926.

Britan-Rosich, E., Nowarski, R., and Kotler, M. (2011). Multifaceted counter-APOBEC3G mechanisms employed by HIV-1 Vif. J Mol Biol *410*, 1065-1076.

Browne, E. P., Allers, C., and Landau, N. R. (2009). Restriction of HIV-1 by APOBEC3G is cytidine deaminase-dependent. Virology *387*, 313-321.

Burnett, A., and Spearman, P. (2007). APOBEC3G multimers are recruited to the plasma membrane for packaging into human immunodeficiency virus type 1 virus-like particles in an RNA-dependent process requiring the NC basic linker. J Virol *81*, 5000-5013.

Burns, M. B., Lackey, L., Carpenter, M. A., Rathore, A., Land, A. M., Leonard, B., Refsland, E. W., Kotandeniya, D., Tretyakova, N., Nikas, J. B.*, et al.* (2013). APOBEC3B is an enzymatic source of mutation in breast cancer. Nature *494*, 366-370.

Caride, E., Brindeiro, R. M., Kallas, E. G., de Sa, C. A., Eyer-Silva, W. A., Machado, E., and Tanuri, A. (2002). Sexual transmission of HIV-1 isolate showing G-->A hypermutation. J Clin Virol *23*, 179-189.

Carmi, S., Church, G. M., and Levanon, E. Y. (2011). Large-scale DNA editing of retrotransposons accelerates mammalian genome evolution. Nat Comm *2*, 519.

Carpenter, M. A., Rajagurubandara, E., Wijesinghe, P., and Bhagwat, A. S. (2010). Determinants of sequence-specificity within human AID and APOBEC3G. DNA Repair *9*, 579-587.

Cen, S., Guo, F., Niu, M., Saadatmand, J., Deflassieux, J., and Kleiman, L. (2004). The interaction between HIV-1 Gag and APOBEC3G. J Biol Chem *279*, 33177-33184.

Chan, L. (1992). Apolipoprotein B, the major protein component of triglyceride-rich and low density lipoproteins. J Biol Chem *267*, 25621-25624.

Chen, H., Lilley, C. E., Yu, Q., Lee, D. V., Chou, J., Narvaiza, I., Landau, N. R., and Weitzman, M. D. (2006). APOBEC3A is a potent inhibitor of adeno-associated virus and retrotransposons. Curr Biol *16*, 480-485.

Chen, K. M., Harjes, E., Gross, P. J., Fahmy, A., Lu, Y., Shindo, K., Harris, R. S., and Matsuo, H. (2008). Structure of the DNA deaminase domain of the HIV-1 restriction factor APOBEC3G. Nature *452*, 116-119.

Chiu, Y. L., and Greene, W. C. (2008). The APOBEC3 cytidine deaminases: an innate defensive network opposing exogenous retroviruses and endogenous retroelements. Ann Rev Immunol *26*, 317-353.

Chiu, Y. L., Witkowska, H. E., Hall, S. C., Santiago, M., Soros, V. B., Esnault, C., Heidmann, T., and Greene, W. C. (2006). High-molecular-mass APOBEC3G complexes restrict Alu retrotransposition. Proc Natl Acad Sci U S A *103*, 15588-15593.

Conticello, S. G. (2008). The AID/APOBEC family of nucleic acid mutators. Genome Biol *9*, 229.

Conticello, S. G., Harris, R. S., and Neuberger, M. S. (2003). The Vif protein of HIV triggers degradation of the human antiretroviral DNA deaminase APOBEC3G. Curr Biol *13*, 2009-2013.

Conticello, S. G., Langlois, M. A., and Neuberger, M. S. (2007a). Insights into DNA deaminases. Nat Struct Mol Biol *14*, 7-9.

Conticello, S. G., Langlois, M. A., Yang, Z., and Neuberger, M. S. (2007b). DNA deamination in immunity: AID in the context of its APOBEC relatives. Adv Immunol *94*, 37-73.

Conticello, S. G., Thomas, C. J., Petersen-Mahrt, S. K., and Neuberger, M. S. (2005). Evolution of the AID/APOBEC family of polynucleotide (deoxy)cytidine deaminases. Mol Biol Evol *22*, 367-377.

Dang, Y., Abudu, A., Son, S., Harjes, E., Spearman, P., Matsuo, H., and Zheng, Y. H. (2011). Identification of a single amino acid required for APOBEC3 antiretroviral cytidine deaminase activity. J Virol *85*, 5691-5695.

Dang, Y., Siew, L. M., Wang, X., Han, Y., Lampen, R., and Zheng, Y. H. (2008). Human cytidine deaminase APOBEC3H restricts HIV-1 replication. J Biol Chem *283*, 11606-11614.

Dang, Y., Wang, X., Esselman, W. J., and Zheng, Y. H. (2006). Identification of APOBEC3DE as another antiretroviral factor from the human APOBEC family. J Virol *80*, 10522-10533.

Daniele, R. P., Dauber, J. H., Altose, M. D., Rowlands, D. T., Jr., and Gorenberg, D. J. (1977). Lymphocyte studies in asymptomatic cigarette smokers. A comparison between lung and peripheral blood. Amer Rev Respir Dis *116*, 997-1005.

Delebecque, F., Suspene, R., Calattini, S., Casartelli, N., Saib, A., Froment, A., Wain-Hobson, S., Gessain, A., Vartanian, J. P., and Schwartz, O. (2006). Restriction of foamy viruses by APOBEC cytidine deaminases. J Virol *80*, 605-614.

Derse, D., Hill, S. A., Princler, G., Lloyd, P., and Heidecker, G. (2007). Resistance of human T cell leukemia virus type 1 to APOBEC3G restriction is mediated by elements in nucleocapsid. Proc Natl Acad Sci U S A *104*, 2915-2920.

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

Di Noia, J., and Neuberger, M. S. (2002). Altering the pathway of immunoglobulin hypermutation by inhibiting uracil-DNA glycosylase. Nature *419*, 43-48.

Di Noia, J. M., and Neuberger, M. S. (2007). Molecular mechanisms of antibody somatic hypermutation. Annu Rev Biochem *76*, 1-22.

Doehle, B. P., Schäfer, A., and Cullen, B. R. (2005a). Human APOBEC3B is a potent inhibitor of HIV-1 infectivity and is resistant to HIV-1 Vif. Virology *339*, 281-288.

Doehle, B. P., Schafer, A., Wiegand, H. L., Bogerd, H. P., and Cullen, B. R. (2005b). Differential sensitivity of murine leukemia virus to APOBEC3-mediated inhibition is governed by virion exclusion. J Virol *79*, 8201-8207.

Duggal, N. K., and Emerman, M. (2012). Evolutionary conflicts between viruses and restriction factors shape immunity. Nat Rev Immunol *12*, 687-695.

Duggal, N. K., Fu, W., Akey, J. M., and Emerman, M. (2013). Identification and antiviral activity of common polymorphisms in the APOBEC3 locus in human populations. Virology *443*, 329-337.

Dutko, J. A., Schafer, A., Kenny, A. E., Cullen, B. R., and Curcio, M. J. (2005). Inhibition of a yeast LTR retrotransposon by human APOBEC3 cytidine deaminases. Curr Biol *15*, 661-666.

Esnault, C., Heidmann, O., Delebecque, F., Dewannieux, M., Ribet, D., Hance, A. J., Heidmann, T., and Schwartz, O. (2005). APOBEC3G cytidine deaminase inhibits retrotransposition of endogenous retroviruses. Nature *433*, 430-433.

Esnault, C., Millet, J., Schwartz, O., and Heidmann, T. (2006). Dual inhibitory effects of APOBEC family proteins on retrotransposition of mammalian endogenous retroviruses. Nucleic Acids Res *34*, 1522-1531.

Fabian, M. R., Sonenberg, N., and Filipowicz, W. (2010). Regulation of mRNA translation and stability by microRNAs. Annu Rev Biochem *79*, 351-379.

Fisher, A. G., Ensoli, B., Ivanoff, L., Chamberlain, M., Petteway, S., Ratner, L., Gallo, R. C., and Wong-Staal, F. (1987). The sor gene of HIV-1 is required for efficient virus transmission in vitro. Science *237*, 888-893.

Fitzgibbon, J. E., Mazar, S., and Dubin, D. T. (1993). A new type of G-->A hypermutation affecting human immunodeficiency virus. AIDS Res Hum Retroviruses *9*, 833-838.

Foley, G. E., Lazarus, H., Farber, S., Uzman, B. G., Boone, B. A., and McCarthy, R. E. (1965). Continuous Culture of Human Lymphoblasts from Peripheral Blood of a Child with Acute Leukemia. Cancer *18*, 522-529.

Furukawa, A., Nagata, T., Matsugami, A., Habu, Y., Sugiyama, R., Hayashi, F., Kobayashi, N., Yokoyama, S., Takaku, H., and Katahira, M. (2009). Structure, interaction and real-time monitoring of the enzymatic reaction of wild-type APOBEC3G. EMBO J *28*, 440-451.

Gabuzda, D. H., Lawrence, K., Langhoff, E., Terwilliger, E., Dorfman, T., Haseltine, W. A., and Sodroski, J. (1992). Role of vif in replication of human immunodeficiency virus type 1 in CD4+ T lymphocytes. J Virol *66*, 6489-6495.

Gandhi, S. K., Siliciano, J. D., Bailey, J. R., Siliciano, R. F., and Blankson, J. N. (2008). Role of APOBEC3G/F-mediated hypermutation in the control of human immunodeficiency virus type 1 in elite suppressors. J Virol *82*, 3125-3130.

Genomes Project, C., Abecasis, G. R., Auton, A., Brooks, L. D., DePristo, M. A., Durbin, R. M., Handsaker, R. E., Kang, H. M., Marth, G. T., and McVean, G. A. (2012). An integrated map of genetic variation from 1,092 human genomes. Nature *491*, 56-65.

Gerber, A. P., and Keller, W. (1999). An adenosine deaminase that generates inosine at the wobble position of tRNAs. Science *286*, 1146-1149.

Gervaix, A., West, D., Leoni, L. M., Richman, D. D., Wong-Staal, F., and Corbeil, J. (1997). A new reporter cell line to monitor HIV infection and drug susceptibility in vitro. Proc Natl Acad Sci U S A *94*, 4653-4658.

Goila-Gaur, R., Khan, M. A., Miyagi, E., Kao, S., Opi, S., Takeuchi, H., and Strebel, K. (2008). HIV-1 Vif promotes the formation of high molecular mass APOBEC3G complexes. Virology *372*, 136-146.

Goila-Gaur, R., Khan, M. A., Miyagi, E., Kao, S., and Strebel, K. (2007). Targeting APOBEC3A to the viral nucleoprotein complex confers antiviral activity. Retrovirology *4*, 61.

Goila-Gaur, R., and Strebel, K. (2008). HIV-1 Vif, APOBEC, and intrinsic immunity. Retrovirology *5*, 51.

Grosjean, H. (2009). DNA and RNA modification enzymes : structure, mechanism, function and evolution, (Austin, Tex.: Landes Bioscience).

Guo, Y., Dong, L., Qiu, X., Wang, Y., Zhang, B., Liu, H., Yu, Y., Zang, Y., Yang, M., and Huang, Z. (2014). Structural basis for hijacking CBF-beta and CUL5 E3 ligase complex by HIV-1 Vif. Nature *505*, 229-233.

Hache, G., and Harris, R. S. (2009). CEM-T4 cells do not lack an APOBEC3G cofactor. PLOS Pathog *5*, e1000528.

Haché, G., Liddament, M. T., and Harris, R. S. (2005). The retroviral hypermutation specificity of APOBEC3F and APOBEC3G is governed by the C-terminal DNA cytosine deaminase domain. J Biol Chem *280*, 10920-10924.

Haché, G., Mansky, L. M., and Harris, R. S. (2006). Human APOBEC3 proteins, retrovirus restriction, and HIV drug resistance. AIDS Rev *8*, 148-157.

Haché, G., Shindo, K., Albin, J. S., and Harris, R. S. (2008). Evolution of HIV-1 isolates that use a novel Vif-independent mechanism to resist restriction by human APOBEC3G. Curr Biol *18*, 819-824.

Harari, A., Ooms, M., Mulder, L. C., and Simon, V. (2009). Polymorphisms and splice variants influence the antiretroviral activity of human APOBEC3H. J Virol *83*, 295-303.

Harjes, E., Gross, P. J., Chen, K. M., Lu, Y., Shindo, K., Nowarski, R., Gross, J. D., Kotler, M., Harris, R. S., and Matsuo, H. (2009). An extended structure of the APOBEC3G catalytic domain suggests a unique holoenzyme model. J Mol Biol *389*, 819-832.

Harris, R. S. (2008). Enhancing immunity to HIV through APOBEC. Nat Biotech *26*, 1089-1090.

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

Harris, R. S., Hultquist, J. F., and Evans, D. T. (2012). The restriction factors of human immunodeficiency virus. J Biol Chem *287*, 40875-40883.

Harris, R. S., and Liddament, M. T. (2004). Retroviral restriction by APOBEC proteins. Nat Rev Immunol *4*, 868-877.

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

Henry, M., Guetard, D., Suspene, R., Rusniok, C., Wain-Hobson, S., and Vartanian, J. P. (2009). Genetic editing of HBV DNA by monodomain human APOBEC3 cytidine deaminases and the recombinant nature of APOBEC3G. PLOS One *4*, e4277.

Holden, L. G., Prochnow, C., Chang, Y. P., Bransteitter, R., Chelico, L., Sen, U., Stevens, R. C., Goodman, M. F., and Chen, X. S. (2008). Crystal structure of the anti-viral APOBEC3G catalytic domain and functional implications. Nature *456*, 121-124.

Holmes, R. K., Koning, F. A., Bishop, K. N., and Malim, M. H. (2007). APOBEC3F can inhibit the accumulation of HIV-1 reverse transcription products in the absence of hypermutation. Comparisons with APOBEC3G. J Biol Chem *282*, 2587-2595.

Huang, W., Zuo, T., Jin, H., Liu, Z., Yang, Z., Yu, X., and Zhang, L. (2013a). Design, synthesis and biological evaluation of indolizine derivatives as HIV-1 VIF-ElonginC interaction inhibitors. Molecular Diversity.

Huang, W., Zuo, T., Luo, X., Jin, H., Liu, Z., Yang, Z., Yu, X., and Zhang, L. (2013b). Indolizine derivatives as HIV-1 VIF-ElonginC interaction inhibitors. Chem Biol & Drug Des.

Hultquist, J. F., and Harris, R. S. (2009). Leveraging APOBEC3 proteins to alter the HIV mutation rate and combat AIDS. Future Virol *4*, 605.

Hultquist, J. F., Lengyel, J. A., Refsland, E. W., LaRue, R. S., Lackey, L., Brown, W. L., and Harris, R. S. (2011). Human and rhesus APOBEC3D, APOBEC3F, APOBEC3G, and APOBEC3H demonstrate a conserved capacity to restrict Vif-deficient HIV-1. J Virol *85*, 11220-11234.

Ikeda, T., Abd El Galil, K. H., Tokunaga, K., Maeda, K., Sata, T., Sakaguchi, N., Heidmann, T., and Koito, A. (2011). Intrinsic restriction activity by apolipoprotein B mRNA editing enzyme APOBEC1 against the mobility of autonomous retrotransposons. Nucleic Acids Res *39*, 5538-5554.

Ikeda, T., Ohsugi, T., Kimura, T., Matsushita, S., Maeda, Y., Harada, S., and Koito, A. (2008). The antiretroviral potency of APOBEC1 deaminase from small animal species. Nucleic Acids Res *36*, 6859-6871.

Ireton, G. C., Black, M. E., and Stoddard, B. L. (2003). The 1.14 A crystal structure of yeast cytosine deaminase: evolution of nucleotide salvage enzymes and implications for genetic chemotherapy. Structure *11*, 961-972.

Iwatani, Y., Chan, D. S., Wang, F., Maynard, K. S., Sugiura, W., Gronenborn, A. M., Rouzina, I., Williams, M. C., Musier-Forsyth, K., and Levin, J. G. (2007). Deaminase-independent inhibition of HIV-1 reverse transcription by APOBEC3G. Nucleic Acids Res *35*, 7096-7108.

Iwatani, Y., Takeuchi, H., Strebel, K., and Levin, J. G. (2006). Biochemical activities of highly purified, catalytically active human APOBEC3G: correlation with antiviral effect. J Virol *80*, 5992-6002.

Jäger, S., Cimermancic, P., Gulbahce, N., Johnson, J. R., McGovern, K. E., Clarke, S. C., Shales, M., Mercenne, G., Pache, L., Li, K.*, et al.* (2012a). Global landscape of HIV-human protein complexes. Nature *481*, 365-370.

Jäger, S., Kim, D. Y., Hultquist, J. F., Shindo, K., LaRue, R. S., Kwon, E., Li, M., Anderson, B. D., Yen, L., Stanley, D.*, et al.* (2012b). Vif hijacks CBF-beta to degrade APOBEC3G and promote HIV-1 infection. Nature *481*, 371-375.

Janini, M., Rogers, M., Birx, D. R., and McCutchan, F. E. (2001). 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 *75*, 7973-7986.

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

Jern, P., and Coffin, J. M. (2008). Effects of retroviruses on host genome function. Annu Rev Genet *42*, 709-732.

Johansson, E., Mejlhede, N., Neuhard, J., and Larsen, S. (2002). Crystal structure of the tetrameric cytidine deaminase from Bacillus subtilis at 2.0 A resolution. Biochemistry *41*, 2563-2570.

Johansson, E., Neuhard, J., Willemoes, M., and Larsen, S. (2004). Structural, kinetic, and mutational studies of the zinc ion environment in tetrameric cytidine deaminase. Biochemistry *43*, 6020-6029.

Jonsson, S. R., LaRue, R. S., Stenglein, M. D., Fahrenkrug, S. C., Andresdottir, V., and Harris, R. S. (2007). The restriction of zoonotic PERV transmission by human APOBEC3G. PLOS One *2*, e893.

Jost, S., Turelli, P., Mangeat, B., Protzer, U., and Trono, D. (2007). Induction of antiviral cytidine deaminases does not explain the inhibition of hepatitis B virus replication by interferons. J Virol *81*, 10588-10596.

Kaiser, S. M., and Emerman, M. (2006). Uracil DNA glycosylase is dispensable for human immunodeficiency virus type 1 replication and does not contribute to the antiviral effects of the cytidine deaminase Apobec3G. J Virol *80*, 875-882.

Kamata, M., Nagaoka, Y., and Chen, I. S. (2009). Reassessing the role of APOBEC3G in human immunodeficiency virus type 1 infection of quiescent CD4+ T-cells. PLOS Pathog *5*, e1000342.

Kan, N. C., Franchini, G., Wong-Staal, F., DuBois, G. C., Robey, W. G., Lautenberger, J. A., and Papas, T. S. (1986). Identification of HTLV-III/LAV sor gene product and detection of antibodies in human sera. Science *231*, 1553-1555.

Kao, S., Khan, M. A., Miyagi, E., Plishka, R., Buckler-White, A., and Strebel, K. (2003). The human immunodeficiency virus type 1 Vif protein reduces intracellular expression and inhibits packaging of APOBEC3G (CEM15), a cellular inhibitor of virus infectivity. J Virol *77*, 11398-11407.

Kao, S., Miyagi, E., Khan, M. A., Takeuchi, H., Opi, S., Goila-Gaur, R., and Strebel, K. (2004). Production of infectious human immunodeficiency virus type 1 does not require depletion of APOBEC3G from virus-producing cells. Retrovirology *1*, 27.

Kazazian, H. H., Jr. (2004). Mobile elements: drivers of genome evolution. Science *303*, 1626-1632.

Khan, M. A., Goila-Gaur, R., Opi, S., Miyagi, E., Takeuchi, H., Kao, S., and Strebel, K. (2007). Analysis of the contribution of cellular and viral RNA to the packaging of APOBEC3G into HIV-1 virions. Retrovirology *4*, 48.

Khan, M. A., Kao, S., Miyagi, E., Takeuchi, H., Goila-Gaur, R., Opi, S., Gipson, C. L., Parslow, T. G., Ly, H., and Strebel, K. (2005). Viral RNA is required for the association of APOBEC3G with human immunodeficiency virus type 1 nucleoprotein complexes. J Virol *79*, 5870-5874.

Kidd, J. M., Newman, T. L., Tuzun, E., Kaul, R., and Eichler, E. E. (2007). Population stratification of a common APOBEC gene deletion polymorphism. PLOS Genet *3*, e63.

Kieffer, T. L., Kwon, P., Nettles, R. E., Han, Y., Ray, S. C., and Siliciano, R. F. (2005). 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 *79*, 1975-1980.

Kijak, G. H., Janini, L. M., Tovanabutra, S., Sanders-Buell, E., Arroyo, M. A., Robb, M. L., Michael, N. L., Birx, D. L., and McCutchan, F. E. (2008). Variable contexts and levels of hypermutation in HIV-1 proviral genomes recovered from primary peripheral blood mononuclear cells. Virology *376*, 101-111.

Kim, D. Y., Kwon, E., Hartley, P. D., Crosby, D. C., Mann, S., Krogan, N. J., and Gross, J. D. (2013). CBFbeta Stabilizes HIV Vif to Counteract APOBEC3 at the Expense of RUNX1 Target Gene Expression. Molecular cell *49,* 632-44.

Kim, U., Wang, Y., Sanford, T., Zeng, Y., and Nishikura, K. (1994). Molecular cloning of cDNA for double-stranded RNA adenosine deaminase, a candidate enzyme for nuclear RNA editing. Proc Natl Acad Sci U S A *91*, 11457-11461.

Kinomoto, M., Kanno, T., Shimura, M., Ishizaka, Y., Kojima, A., Kurata, T., Sata, T., and Tokunaga, K. (2007). All APOBEC3 family proteins differentially inhibit LINE-1 retrotransposition. Nucleic Acids Res *35*, 2955-2964.

Kitamura, S., Ode, H., Nakashima, M., Imahashi, M., Naganawa, Y., Kurosawa, T., Yokomaku, Y., Yamane, T., Watanabe, N., Suzuki, A*., et al.* (2012a). The APOBEC3C

crystal structure and the interface for HIV-1 Vif binding. Nat Struct Mol Biol *19*, 1005-1010.

Kitamura, S., Ode, H., Nakashima, M., Imahashi, M., Naganawa, Y., Kurosawa, T., Yokomaku, Y., Yamane, T., Watanabe, N., Suzuki, A.*, et al.* (2012b). The APOBEC3C crystal structure and the interface for HIV-1 Vif binding. Nat Struct Mol Biol *19*, 1005-1010.

Klemm, L., Duy, C., Iacobucci, I., Kuchen, S., von Levetzow, G., Feldhahn, N., Henke, N., Li, Z., Hoffmann, T. K., Kim, Y. M.*, et al.* (2009). The B cell mutator AID promotes B lymphoid blast crisis and drug resistance in chronic myeloid leukemia. Cancer Cell *16*, 232-245.

Ko, T. P., Lin, J. J., Hu, C. Y., Hsu, Y. H., Wang, A. H., and Liaw, S. H. (2003). Crystal structure of yeast cytosine deaminase. Insights into enzyme mechanism and evolution. J Biol Chem *278*, 19111-19117.

Kobayashi, M., Takaori-Kondo, A., Shindo, K., Abudu, A., Fukunaga, K., and Uchiyama, T. (2004). APOBEC3G targets specific virus species. J Virol *78*, 8238-8244.

Kock, J., and Blum, H. E. (2008). Hypermutation of hepatitis B virus genomes by APOBEC3G, APOBEC3C and APOBEC3H. J Gen Virol *89*, 1184-1191.

Kohli, R. M., Abrams, S. R., Gajula, K. S., Maul, R. W., Gearhart, P. J., and Stivers, J. T. (2009). A portable hot spot recognition loop transfers sequence preferences from APOBEC family members to activation-induced cytidine deaminase. J Biol Chem *284*, 22898-22904.

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

Konig, R., Zhou, Y., Elleder, D., Diamond, T. L., Bonamy, G. M., Irelan, J. T., Chiang, C. Y., Tu, B. P., De Jesus, P. D., Lilley, C. E.*, et al.* (2008). Global analysis of host-pathogen interactions that regulate early-stage HIV-1 replication. Cell *135*, 49-60.

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

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

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

Land, A. M., Ball, T. B., Luo, M., Pilon, R., Sandstrom, P., Embree, J. E., Wachihi, C., Kimani, J., and Plummer, F. A. (2008). Human immunodeficiency virus (HIV) type 1 proviral hypermutation correlates with CD4 count in HIV-infected women from Kenya. J Virol *82*, 8172-8182.

Langlois, M. A., Beale, R. C., Conticello, S. G., and Neuberger, M. S. (2005). 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 Res *33*, 1913-1923.

LaRue, R. S., Andrésdóttir, V., Blanchard, Y., Conticello, S. G., Derse, D., Emerman, M., Greene, W. C., Jónsson, S. R., Landau, N. R., Löchelt, M*., et al.* (2009). Guidelines for naming nonprimate APOBEC3 genes and proteins. J Virol *83*, 494-497.

LaRue, R. S., Jonsson, S. R., Silverstein, K. A., Lajoie, M., Bertrand, D., El-Mabrouk, N., Hotzel, I., Andresdottir, V., Smith, T. P., and Harris, R. S. (2008). The artiodactyl APOBEC3 innate immune repertoire shows evidence for a multi-functional domain organization that existed in the ancestor of placental mammals. BMC Mol Biol *9*, 104.

LaRue, R. S., Lengyel, J., Jónsson, S. R., Andrésdóttir, V., and Harris, R. S. (2010). Lentiviral Vif degrades the APOBEC3Z3/APOBEC3H protein of its mammalian host and is capable of cross-species activity. J Virol *84*, 8193-8201.

Lecossier, D., Bouchonnet, F., Clavel, F., and Hance, A. J. (2003). Hypermutation of HIV-1 DNA in the absence of the Vif protein. Science *300*, 1112.

Lee, T. H., Coligan, J. E., Allan, J. S., McLane, M. F., Groopman, J. E., and Essex, M. (1986). A new HTLV-III/LAV protein encoded by a gene found in cytopathic retroviruses. Science *231*, 1546-1549.

Lee, Y. N., and Bieniasz, P. D. (2007). Reconstitution of an infectious human endogenous retrovirus. PLOS Pathog *3*, e10.

Lee, Y. N., Malim, M. H., and Bieniasz, P. D. (2008). Hypermutation of an ancient human retrovirus by APOBEC3G. J Virol *82*, 8762-8770.

Leonard, B., Hart, S. N., Burns, M. B., Carpenter, M. A., Temiz, N. A., Rathore, A., Vogel, R. I., Nikas, J. B., Law, E. K., Brown, W. L.*, et al.* (2013). APOBEC3B upregulation and genomic mutation patterns in serous ovarian carcinoma. Cancer Res *73*, 7222-7231.

Levanon, E. Y., Eisenberg, E., Yelin, R., Nemzer, S., Hallegger, M., Shemesh, R., Fligelman, Z. Y., Shoshan, A., Pollock, S. R., Sztybel, D.*, et al.* (2004). Systematic identification of abundant A-to-I editing sites in the human transcriptome. Nat Biotech *22*, 1001-1005.

Lever, A. M. L., Jeang, K.-T., and Berkhout, B. (2010). Recent advances in human retroviruses : principles of replication and pathogenesis : advances in retroviral research, (Singapore ; Hackensack, NJ: World Scientific).

Li, J., Chen, Y., Li, M., Carpenter, M. A., McDougle, R. M., Luengas, E. M., Macdonald, P. J., Harris, R. S., and Mueller, J. D. (2014). APOBEC3 multimerization correlates with HIV-1 packaging and restriction activity in living cells. J Mol Biol *426*, 1296-1307.

Li, J. B., Levanon, E. Y., Yoon, J. K., Aach, J., Xie, B., Leproust, E., Zhang, K., Gao, Y., and Church, G. M. (2009). Genome-wide identification of human RNA editing sites by parallel DNA capturing and sequencing. Science *324*, 1210-1213.

Li, M., Shandilya, S. M., Carpenter, M. A., Rathore, A., Brown, W. L., Perkins, A. L., Harki, D. A., Solberg, J., Hook, D. J., Pandey, K. K.*, et al.* (2012). First-in-class small molecule inhibitors of the single-strand DNA cytosine deaminase APOBEC3G. ACS Chem Biol *7*, 506-517.

Li, M. M., Wu, L. I., and Emerman, M. (2010). The range of human APOBEC3H sensitivity to lentiviral Vif proteins. J Virol *84*, 88-95.

Liddament, M. T., Brown, W. L., Schumacher, A. J., and Harris, R. S. (2004). APOBEC3F properties and hypermutation preferences indicate activity against HIV-1 in vivo. Curr Biol *14*, 1385-1391.

Lindahl, T. (2000). Suppression of spontaneous mutagenesis in human cells by DNA base excision-repair. Mutation Res *462*, 129-135.

Lochelt, M., Romen, F., Bastone, P., Muckenfuss, H., Kirchner, N., Kim, Y. B., Truyen, U., Rosler, U., Battenberg, M., Saib, A.*, et al.* (2005). The antiretroviral activity of APOBEC3 is inhibited by the foamy virus accessory Bet protein. Proc Natl Acad Sci U S A *102*, 7982-7987.

Loeb, L. A., Essigmann, J. M., Kazazi, F., Zhang, J., Rose, K. D., and Mullins, J. I. (1999). Lethal mutagenesis of HIV with mutagenic nucleoside analogs. Proc Natl Acad Sci U S A *96*, 1492-1497.

Longerich, S., Basu, U., Alt, F., and Storb, U. (2006). AID in somatic hypermutation and class switch recombination. Curr Opin Immunol *18*, 164-174.

Losey, H. C., Ruthenburg, A. J., and Verdine, G. L. (2006). Crystal structure of Staphylococcus aureus tRNA adenosine deaminase TadA in complex with RNA. Nat Struct Mol Biol *13*, 153-159.

Luo, K., Liu, B., Xiao, Z., Yu, Y., Yu, X., Gorelick, R., and Yu, X. F. (2004). Amino-terminal region of the human immunodeficiency virus type 1 nucleocapsid is required for human APOBEC3G packaging. J Virol *78*, 11841-11852.

MacDuff, D. A., Demorest, Z. L., and Harris, R. S. (2009). AID can restrict L1 retrotransposition suggesting a dual role in innate and adaptive immunity. Nucleic Acids Res *37*, 1854-1867.

Maksakova, I. A., Romanish, M. T., Gagnier, L., Dunn, C. A., van de Lagemaat, L. N., and Mager, D. L. (2006). Retroviral elements and their hosts: insertional mutagenesis in the mouse germ line. PLOS Genet *2*, e2.

Malim, M. H. (2009). APOBEC proteins and intrinsic resistance to HIV-1 infection. Philos Trans R Soc Lond B Biol Sci *364*, 675-687.

Malim, M. H., and Bieniasz, P. D. (2012). HIV Restriction Factors and Mechanisms of Evasion. Cold Spring Harb Perspect Med *2*, a006940.

Malim, M. H., and Emerman, M. (2008a). HIV-1 accessory proteins--ensuring viral survival in a hostile environment. Cell Host Microbe *3*, 388-398.

Malim, M. H., and Emerman, M. (2008b). HIV-1 accessory proteins--ensuring viral survival in a hostile environment. Cell Host Microbe *3*, 388-398.

Mangeat, B., Turelli, P., Caron, G., Friedli, M., Perrin, L., and Trono, D. (2003). Broad antiretroviral defence by human APOBEC3G through lethal editing of nascent reverse transcripts. Nature *424*, 99-103.

Mangeat, B., Turelli, P., Liao, S., and Trono, D. (2004). A single amino acid determinant governs the species-specific sensitivity of APOBEC3G to Vif action. J Biol Chem *279*, 14481-14483.

Mariani, R., Chen, D., Schrofelbauer, B., Navarro, F., Konig, R., Bollman, B., Munk, C., Nymark-McMahon, H., and Landau, N. R. (2003a). Species-specific exclusion of APOBEC3G from HIV-1 virions by Vif. Cell *114*, 21-31.

Mariani, R., Chen, D., Schröfelbauer, B., Navarro, F., König, R., Bollman, B., Münk, C., Nymark-McMahon, H., and Landau, N. R. (2003b). Species-specific exclusion of APOBEC3G from HIV-1 virions by Vif. Cell *114*, 21-31.

Marin, M., Golem, S., Rose, K. M., Kozak, S. L., and Kabat, D. (2008). Human immunodeficiency virus type 1 Vif functionally interacts with diverse APOBEC3 cytidine deaminases and moves with them between cytoplasmic sites of mRNA metabolism. J Virol *82*, 987-998.

Marin, M., Rose, K. M., Kozak, S. L., and Kabat, D. (2003). HIV-1 Vif protein binds the editing enzyme APOBEC3G and induces its degradation. Nat Med *9*, 1398-1403.

Martin, K. L., Johnson, M., and D'Aquila, R. T. (2011). APOBEC3G complexes decrease human immunodeficiency virus type 1 production. J Virol *85*, 9314-9326.

Mbisa, J. L., Barr, R., Thomas, J. A., Vandegraaff, N., Dorweiler, I. J., Svarovskaia, E. S., Brown, W. L., Mansky, L. M., Gorelick, R. J., Harris, R. S.*, et al.* (2007). Human

immunodeficiency virus type 1 cDNAs produced in the presence of APOBEC3G exhibit defects in plus-strand DNA transfer and integration. J Virol *81*, 7099-7110.

Mehle, A., Strack, B., Ancuta, P., Zhang, C., McPike, M., and Gabuzda, D. (2004). Vif overcomes the innate antiviral activity of APOBEC3G by promoting its degradation in the ubiquitin-proteasome pathway. J Biol Chem *279*, 7792-7798.

Miyagi, E., Brown, C. R., Opi, S., Khan, M., Goila-Gaur, R., Kao, S., Walker, R. C., Jr., Hirsch, V., and Strebel, K. (2010). Stably expressed APOBEC3F has negligible antiviral activity. J Virol *84*, 11067-11075.

Miyagi, E., Opi, S., Takeuchi, H., Khan, M., Goila-Gaur, R., Kao, S., and Strebel, K. (2007). Enzymatically active APOBEC3G is required for efficient inhibition of human immunodeficiency virus type 1. J Virol *81*, 13346-13353.

Morse, D. P., Aruscavage, P. J., and Bass, B. L. (2002). RNA hairpins in noncoding regions of human brain and Caenorhabditis elegans mRNA are edited by adenosine deaminases that act on RNA. Proc Natl Acad Sci U S A *99*, 7906-7911.

Muckenfuss, H., Hamdorf, M., Held, U., Perkovic, M., Lower, J., Cichutek, K., Flory, E., Schumann, G. G., and Münk, C. (2006). APOBEC3 proteins inhibit human LINE-1 retrotransposition. J Biol Chem *281*, 22161-22172.

Mulder, L. C., Ooms, M., Majdak, S., Smedresman, J., Linscheid, C., Harari, A., Kunz, A., and Simon, V. (2010). Moderate influence of human APOBEC3F on HIV-1 replication in primary lymphocytes. J Virol *84*, 9613-9617.

Münk, C., Willemsen, A., and Bravo, I. G. (2012). An ancient history of gene duplications, fusions and losses in the evolution of APOBEC3 mutators in mammals. BMC Evol Biol *12*, 71.

Muramatsu, M., Kinoshita, K., Fagarasan, S., Yamada, S., Shinkai, Y., and Honjo, T. (2000). Class switch recombination and hypermutation require activation-induced cytidine deaminase (AID), a potential RNA editing enzyme. Cell *102*, 553-563.

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

Narvaiza, I., Linfesty, D. C., Greener, B. N., Hakata, Y., Pintel, D. J., Logue, E., Landau, N. R., and Weitzman, M. D. (2009). Deaminase-independent inhibition of parvoviruses by the APOBEC3A cytidine deaminase. PLOS Pathog *5*, e1000439.

Nathans, R., Cao, H., Sharova, N., Ali, A., Sharkey, M., Stranska, R., Stevenson, M., and Rana, T. M. (2008). Small-molecule inhibition of HIV-1 Vif. Nat Biotech *26*, 1187-1192.

Newman, E. N., Holmes, R. K., Craig, H. M., Klein, K. C., Lingappa, J. R., Malim, M. H., and Sheehy, A. M. (2005). Antiviral function of APOBEC3G can be dissociated from cytidine deaminase activity. Curr Biol *15*, 166-170.

Nik-Zainal, S., Alexandrov, L. B., Wedge, D. C., Van Loo, P., Greenman, C. D., Raine, K., Jones, D., Hinton, J., Marshall, J., Stebbings, L. A.*, et al.* (2012). Mutational processes molding the genomes of 21 breast cancers. Cell *149*, 979-993.

Nishikura, K. (2010). Functions and regulation of RNA editing by ADAR deaminases. Annu Rev Biochem *79*, 321-349.

Nowarski, R., Britan-Rosich, E., Shiloach, T., and Kotler, M. (2008). Hypermutation by intersegmental transfer of APOBEC3G cytidine deaminase. Nat Struct Mol Biol *15*, 1059-1066.

OhAinle, M., Kerns, J. A., Li, M. M., Malik, H. S., and Emerman, M. (2008). Antiretroelement activity of APOBEC3H was lost twice in recent human evolution. Cell Host Microbe *4*, 249-259.

OhAinle, M., Kerns, J. A., Malik, H. S., and Emerman, M. (2006). Adaptive evolution and antiviral activity of the conserved mammalian cytidine deaminase APOBEC3H. J Virol *80*, 3853-3862.

Okazaki, I. M., Hiai, H., Kakazu, N., Yamada, S., Muramatsu, M., Kinoshita, K., and Honjo, T. (2003). Constitutive expression of AID leads to tumorigenesis. J Exp Med *197*, 1173-1181.

Olson, M. E., Li, M., Harris, R. S., and Harki, D. A. (2013). Small-molecule APOBEC3G DNA cytosine deaminase inhibitors based on a 4-amino-1,2,4-triazole-3-thiol scaffold. Chem Med Chem *8*, 112-117.

Ooms, M., Brayton, B., Letko, M., Maio, S. M., Pilcher, C. D., Hecht, F. M., Barbour, J. D., and Simon, V. (2013a). HIV-1 Vif adaptation to human APOBEC3H haplotypes. Cell Host Microbe *14*, 411-421.

Ooms, M., Letko, M., Binka, M., and Simon, V. (2013b). The resistance of human APOBEC3H to HIV-1 NL4-3 molecular clone is determined by a single amino acid in Vif. PLOS One *8*, e57744.

Ooms, M., Majdak, S., Seibert, C. W., Harari, A., and Simon, V. (2010). The localization of APOBEC3H variants in HIV-1 virions determines their antiviral activity. J Virol *84*, 7961-7969.

Opi, S., Kao, S., Goila-Gaur, R., Khan, M. A., Miyagi, E., Takeuchi, H., and Strebel, K. (2007). Human immunodeficiency virus type 1 Vif inhibits packaging and antiviral activity of a degradation-resistant APOBEC3G variant. J Virol *81*, 8236-8246.

Opi, S., Takeuchi, H., Kao, S., Khan, M. A., Miyagi, E., Goila-Gaur, R., Iwatani, Y., Levin, J. G., and Strebel, K. (2006). Monomeric APOBEC3G is catalytically active and has antiviral activity. J Virol *80*, 4673-4682.

Pace, C., Keller, J., Nolan, D., James, I., Gaudieri, S., Moore, C., and Mallal, S. (2006). Population level analysis of human immunodeficiency virus type 1 hypermutation and its relationship with APOBEC3G and vif genetic variation. J Virol *80*, 9259-9269.

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

Peirson, S. N., Butler, J. N., and Foster, R. G. (2003). Experimental validation of novel and conventional approaches to quantitative real-time PCR data analysis. Nucleic Acids Res *31*, e73.

Peng, G., Greenwell-Wild, T., Nares, S., Jin, W., Lei, K. J., Rangel, Z. G., Munson, P. J., and Wahl, S. M. (2007). Myeloid differentiation and susceptibility to HIV-1 are linked to APOBEC3 expression. Blood *110*, 393-400.

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

Perez-Duran, P., de Yebenes, V. G., and Ramiro, A. R. (2007). Oncogenic events triggered by AID, the adverse effect of antibody diversification. Carcinogenesis *28*, 2427-2433.

Petersen-Mahrt, S. K., Harris, R. S., and Neuberger, M. S. (2002). AID mutates E. coli suggesting a DNA deamination mechanism for antibody diversification. Nature *418*, 99-103.

Petit, V., Guetard, D., Renard, M., Keriel, A., Sitbon, M., Wain-Hobson, S., and Vartanian, J. P. (2009). Murine APOBEC1 is a powerful mutator of retroviral and cellular RNA in vitro and in vivo. J Mol Biol *385*, 65-78.

Pfaffl, M. W. (2001). A new mathematical model for relative quantification in real-time RT-PCR. Nucleic Acids Res *29*, e45.

Piantadosi, A., Humes, D., Chohan, B., McClelland, R. S., and Overbaugh, J. (2009). Analysis of the percentage of human immunodeficiency virus type 1 sequences that are hypermutated and markers of disease progression in a longitudinal cohort, including one individual with a partially defective Vif. J Virol *83*, 7805-7814.

Powell, L. M., Wallis, S. C., Pease, R. J., Edwards, Y. H., Knott, T. J., and Scott, J. (1987). A novel form of tissue-specific RNA processing produces apolipoprotein-B48 in intestine. Cell *50*, 831-840.

Rago, C., Vogelstein, B., and Bunz, F. (2007). Genetic knockouts and knockins in human somatic cells. Nat Protoc *2*, 2734-2746.

Ramiro, A. R., Jankovic, M., Eisenreich, T., Difilippantonio, S., Chen-Kiang, S., Muramatsu, M., Honjo, T., Nussenzweig, A., and Nussenzweig, M. C. (2004). AID is required for c-myc/IgH chromosome translocations in vivo. Cell *118*, 431-438.

Rathore, A., Carpenter, M. A., Demir, O., Ikeda, T., Li, M., Shaban, N. M., Law, E. K., Anokhin, D., Brown, W. L., Amaro, R. E., and Harris, R. S. (2013). The local dinucleotide preference of APOBEC3G can be altered from 5'-CC to 5'-TC by a single amino acid substitution. J Mol Biol *425*, 4442-4454.

Refsland, E. W., and Harris, R. S. (2013). The APOBEC3 family of retroelement restriction factors. Curr Top Microbiol Immunol *371*, 1-27.

Refsland, E. W., Hultquist, J. F., and Harris, R. S. (2012). Endogenous Origins of HIV-1 G-to-A Hypermutation and Restriction in the Nonpermissive T Cell Line CEM2n. PLOS Pathog *8*, e1002800.

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

Roberts, S. A., Lawrence, M. S., Klimczak, L. J., Grimm, S. A., Fargo, D., Stojanov, P., Kiezun, A., Kryukov, G. V., Carter, S. L., Saksena, G.*, et al.* (2013). An APOBEC cytidine deaminase mutagenesis pattern is widespread in human cancers. Nat Genet *45*, 970-976.

Roberts, S. A., Sterling, J., Thompson, C., Harris, S., Mav, D., Shah, R., Klimczak, L. J., Kryukov, G. V., Malc, E., Mieczkowski, P. A.*, et al.* (2012). Clustered mutations in yeast and in human cancers can arise from damaged long single-strand DNA regions. Mol Cell *46*, 424-435.

Rogozin, I. B., Iyer, L. M., Liang, L., Glazko, G. V., Liston, V. G., Pavlov, Y. I., Aravind, L., and Pancer, Z. (2007). Evolution and diversification of lamprey antigen receptors: evidence for involvement of an AID-APOBEC family cytosine deaminase. Nat Immunol *8*, 647-656.

Rose, K. M., Marin, M., Kozak, S. L., and Kabat, D. (2005). Regulated production and anti-HIV type 1 activities of cytidine deaminases APOBEC3B, 3F, and 3G. AIDS Res Hum Retroviruses *21*, 611-619.

Russell, R. A., and Pathak, V. K. (2007). Identification of two distinct human immunodeficiency virus type 1 Vif determinants critical for interactions with human APOBEC3G and APOBEC3F. J Virol *81*, 8201-8210.

Russell, R. A., Wiegand, H. L., Moore, M. D., Schafer, A., McClure, M. O., and Cullen, B. R. (2005). Foamy virus Bet proteins function as novel inhibitors of the APOBEC3 family of innate antiretroviral defense factors. J Virol *79*, 8724-8731.

Santa-Marta, M., da Silva, F. A., Fonseca, A. M., and Goncalves, J. (2005). HIV-1 Vif can directly inhibit apolipoprotein B mRNA-editing enzyme catalytic polypeptide-like 3G-mediated cytidine deamination by using a single amino acid interaction and without protein degradation. J Biol Chem *280*, 8765-8775.

Sanville, B., Dolan, M. A., Wollenberg, K., Yan, Y., Martin, C., Yeung, M. L., Strebel, K., Buckler-White, A., and Kozak, C. A. (2010). Adaptive evolution of Mus Apobec3 includes retroviral insertion and positive selection at two clusters of residues flanking the substrate groove. PLOS Pathog *6*, e1000974.

Sawyer, S. L., Emerman, M., and Malik, H. S. (2004). Ancient adaptive evolution of the primate antiviral DNA-editing enzyme APOBEC3G. PLOS Biol *2*, E275.

Sawyer, S. L., Emerman, M., and Malik, H. S. (2007). Discordant evolution of the adjacent antiretroviral genes TRIM22 and TRIM5 in mammals. PLOS Pathog *3*, e197.

Schafer, A., Bogerd, H. P., and Cullen, B. R. (2004). Specific packaging of APOBEC3G into HIV-1 virions is mediated by the nucleocapsid domain of the gag polyprotein precursor. Virology *328*, 163-168.

Schröfelbauer, B., Chen, D., and Landau, N. R. (2004). A single amino acid of APOBEC3G controls its species-specific interaction with virion infectivity factor (Vif). Proc Natl Acad Sci U S A *101*, 3927-3932.

Schröfelbauer, B., Senger, T., Manning, G., and Landau, N. R. (2006). Mutational alteration of human immunodeficiency virus type 1 Vif allows for functional interaction with nonhuman primate APOBEC3G. J Virol *80*, 5984-5991.

Schumacher, A. J., Haché, G., MacDuff, D. A., Brown, W. L., and Harris, R. S. (2008). The DNA deaminase activity of human APOBEC3G is required for Ty1, MusD, and human immunodeficiency virus type 1 restriction. J Virol *82*, 2652-2660.

Schumacher, A. J., Nissley, D. V., and Harris, R. S. (2005). APOBEC3G hypermutates genomic DNA and inhibits Ty1 retrotransposition in yeast. Proc Natl Acad Sci U S A *102*, 9854-9859.

Severi, F., Chicca, A., and Conticello, S. G. (2011). Analysis of reptilian APOBEC1 suggests that RNA editing may not be its ancestral function. Mol Biol Evol *28*, 1125-1129.

Shandilya, S. M., Nalam, M. N., Nalivaika, E. A., Gross, P. J., Valesano, J. C., Shindo, K., Li, M., Munson, M., Royer, W. E., Harjes, E.*, et al.* (2010). Crystal structure of the APOBEC3G catalytic domain reveals potential oligomerization interfaces. Structure *18*, 28-38.

Sheehy, A. M., Gaddis, N. C., Choi, J. D., and Malim, M. H. (2002). Isolation of a human gene that inhibits HIV-1 infection and is suppressed by the viral Vif protein. Nature *418*, 646-650.

Sheehy, A. M., Gaddis, N. C., and Malim, M. H. (2003). The antiretroviral enzyme APOBEC3G is degraded by the proteasome in response to HIV-1 Vif. Nat Med *9*, 1404-1407.

Shindo, K., Li, M., Gross, P. J., Brown, W. L., Harjes, E., Lu, Y., Matsuo, H., and Harris, R. S. (2012). A Comparison of Two Single-Stranded DNA Binding Models by Mutational Analysis of APOBEC3G. Biology *1*, 260-276.

Shlyakhtenko, L. S., Lushnikov, A. Y., Li, M., Lackey, L., Harris, R. S., and Lyubchenko, Y. L. (2011). Atomic force microscopy studies provide direct evidence for dimerization of the HIV restriction factor APOBEC3G. J Biol Chem *286*, 3387-3395.

Shlyakhtenko, L. S., Lushnikov, A. Y., Miyagi, A., Li, M., Harris, R. S., and Lyubchenko, Y. L. (2012). Nanoscale structure and dynamics of ABOBEC3G complexes with single-stranded DNA. Biochemistry *51*, 6432-6440.

Simon, V., Zennou, V., Murray, D., Huang, Y., Ho, D. D., and Bieniasz, P. D. (2005). Natural variation in Vif: differential impact on APOBEC3G/3F and a potential role in HIV-1 diversification. PLOS Pathog *1*, e6.

Sodroski, J., Goh, W. C., Rosen, C., Tartar, A., Portetelle, D., Burny, A., and Haseltine, W. (1986). Replicative and cytopathic potential of HTLV-III/LAV with sor gene deletions. Science *231*, 1549-1553.

Song, C., Sutton, L., Johnson, M. E., D'Aquila, R. T., and Donahue, J. P. (2012). Signals in APOBEC3F N-terminal and C-terminal deaminase domains each contribute to encapsidation in HIV-1 virions and are both required for HIV-1 restriction. J Biol Chem *287*, 16965-16974.

Soros, V. B., Yonemoto, W., and Greene, W. C. (2007). Newly synthesized APOBEC3G is incorporated into HIV virions, inhibited by HIV RNA, and subsequently activated by RNase H. PLOS Pathog *3*, e15.

Stenglein, M. D., Burns, M. B., Li, M., Lengyel, J., and Harris, R. S. (2010). APOBEC3 proteins mediate the clearance of foreign DNA from human cells. Nat Struct Mol Biol *17*, 222-229.

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

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

Stopak, K., de Noronha, C., Yonemoto, W., and Greene, W. C. (2003). HIV-1 Vif blocks the antiviral activity of APOBEC3G by impairing both its translation and intracellular stability. Mol Cell *12*, 591-601.

Stopak, K. S., Chiu, Y. L., Kropp, J., Grant, R. M., and Greene, W. C. (2007). Distinct patterns of cytokine regulation of APOBEC3G expression and activity in primary lymphocytes, macrophages, and dendritic cells. J Biol Chem *282*, 3539-3546.

Strebel, K., Daugherty, D., Clouse, K., Cohen, D., Folks, T., and Martin, M. A. (1987). The HIV 'A' (sor) gene product is essential for virus infectivity. Nature *328*, 728-730.

Suspène, R., Guetard, D., Henry, M., Sommer, P., Wain-Hobson, S., and Vartanian, J. P. (2005a). Extensive editing of both hepatitis B virus DNA strands by APOBEC3 cytidine deaminases in vitro and in vivo. Proc Natl Acad Sci U S A *102*, 8321-8326.

Suspène, R., Henry, M., Guillot, S., Wain-Hobson, S., and Vartanian, J. P. (2005b). Recovery of APOBEC3-edited human immunodeficiency virus G->A hypermutants by differential DNA denaturation PCR. J Gen Virol *86*, 125-129.

Suspène, R., Rusniok, C., Vartanian, J. P., and Wain-Hobson, S. (2006). Twin gradients in APOBEC3 edited HIV-1 DNA reflect the dynamics of lentiviral replication. Nucleic Acids Res *34*, 4677-4684.

Svarovskaia, E. S., Xu, H., Mbisa, J. L., Barr, R., Gorelick, R. J., Ono, A., Freed, E. O., Hu, W. S., and Pathak, V. K. (2004). Human apolipoprotein B mRNA-editing enzyme-catalytic polypeptide-like 3G (APOBEC3G) is incorporated into HIV-1 virions through interactions with viral and nonviral RNAs. J Biol Chem *279*, 35822-35828.

Swirski, F. K., Nahrendorf, M., Etzrodt, M., Wildgruber, M., Cortez-Retamozo, V., Panizzi, P., Figueiredo, J. L., Kohler, R. H., Chudnovskiy, A., Waterman, P.*, et al.* (2009). Identification of splenic reservoir monocytes and their deployment to inflammatory sites. Science *325*, 612-616.

Tan, L., Sarkis, P. T., Wang, T., Tian, C., and Yu, X. F. (2009). Sole copy of Z2-type human cytidine deaminase APOBEC3H has inhibitory activity against retrotransposons and HIV-1. FASEB J *23*, 279-287.

Tanaka, Y., Marusawa, H., Seno, H., Matsumoto, Y., Ueda, Y., Kodama, Y., Endo, Y., Yamauchi, J., Matsumoto, T., Takaori-Kondo, A.*, et al.* (2006). Anti-viral protein APOBEC3G is induced by interferon-alpha stimulation in human hepatocytes. Biochem Biophys Res Commun *341*, 314-319.

Tareen, S. U., Sawyer, S. L., Malik, H. S., and Emerman, M. (2009). An expanded clade of rodent Trim5 genes. Virology *385*, 473-483.

Taylor, M. W., Grosse, W. M., Schaley, J. E., Sanda, C., Wu, X., Chien, S. C., Smith, F., Wu, T. G., Stephens, M., Ferris, M. W.*, et al.* (2004). Global effect of PEG-IFN-alpha and ribavirin on gene expression in PBMC in vitro. J Interferon Cytokine Res *24*, 107-118.

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

Tian, C., Wang, T., Zhang, W., and Yu, X. F. (2007). Virion packaging determinants and reverse transcription of SRP RNA in HIV-1 particles. Nucleic Acids Res *35*, 7288-7302.

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

Ulenga, N. K., Sarr, A. D., Hamel, D., Sankale, J. L., Mboup, S., and Kanki, P. J. (2008). The level of APOBEC3G (hA3G)-related G-to-A mutations does not correlate with viral load in HIV type 1-infected individuals. AIDS Res Hum Retroviruses *24*, 1285-1290.

Unniraman, S., Zhou, S., and Schatz, D. G. (2004). Identification of an AID-independent pathway for chromosomal translocations between the Igh switch region and Myc. Nat Immunol *5*, 1117-1123.

Vartanian, J. P., Guetard, D., Henry, M., and Wain-Hobson, S. (2008). Evidence for editing of human papillomavirus DNA by APOBEC3 in benign and precancerous lesions. Science *320*, 230-233.

Vartanian, J. P., Meyerhans, A., Sala, M., and Wain-Hobson, S. (1994). G-->A hypermutation of the human immunodeficiency virus type 1 genome: evidence for dCTP pool imbalance during reverse transcription. Proc Natl Acad Sci U S A *91*, 3092-3096.

Vetter, M. L., Johnson, M. E., Antons, A. K., Unutmaz, D., and D'Aquila, R. T. (2009). Differences in APOBEC3G expression in CD4+ T helper lymphocyte subtypes modulate HIV-1 infectivity. PLOS Pathog *5*, e1000292.

Virgen, C. A., and Hatziioannou, T. (2007). Antiretroviral activity and Vif sensitivity of rhesus macaque APOBEC3 proteins. J Virol *81*, 13932-13937.

von Schwedler, U., Song, J., Aiken, C., and Trono, D. (1993). Vif is crucial for human immunodeficiency virus type 1 proviral DNA synthesis in infected cells. J Virol *67*, 4945-4955.

Wang, F. X., Huang, J., Zhang, H., Ma, X., and Zhang, H. (2008). APOBEC3G upregulation by alpha interferon restricts human immunodeficiency virus type 1 infection in human peripheral plasmacytoid dendritic cells. J Gen Virol *89*, 722-730.

Wang, M., Rada, C., and Neuberger, M. S. (2010). Altering the spectrum of immunoglobulin V gene somatic hypermutation by modifying the active site of AID. The J Exp Med *207*, 141-153.

Wang, T., Tian, C., Zhang, W., Luo, K., Sarkis, P. T., Yu, L., Liu, B., Yu, Y., and Yu, X. F. (2007). 7SL RNA mediates virion packaging of the antiviral cytidine deaminase APOBEC3G. J Virol *81*, 13112-13124.

Wang, X., Abudu, A., Son, S., Dang, Y., Venta, P. J., and Zheng, Y. H. (2011). Analysis of human APOBEC3H haplotypes and anti-human immunodeficiency virus type 1 activity. J Virol *85*, 3142-3152.

Wedekind, J. E., Dance, G. S., Sowden, M. P., and Smith, H. C. (2003). Messenger RNA editing in mammals: new members of the APOBEC family seeking roles in the family business. Trends Genet : TIG *19*, 207-216.

Weisberg, S. P., McCann, D., Desai, M., Rosenbaum, M., Leibel, R. L., and Ferrante, A. W., Jr. (2003). Obesity is associated with macrophage accumulation in adipose tissue. J Clin Invest *112*, 1796-1808.

Wiegand, H. L., Doehle, B. P., Bogerd, H. P., and Cullen, B. R. (2004). A second human antiretroviral factor, APOBEC3F, is suppressed by the HIV-1 and HIV-2 Vif proteins. EMBO J *23*, 2451-2458.

Wissing, S., Montano, M., Garcia-Perez, J. L., Moran, J. V., and Greene, W. C. (2011). Endogenous APOBEC3B restricts LINE-1 retrotransposition in transformed cells and human embryonic stem cells. J Biol Chem *286*, 36427-36437.

Wood, N., Bhattacharya, T., Keele, B. F., Giorgi, E., Liu, M., Gaschen, B., Daniels, M., Ferrari, G., Haynes, B. F., McMichael, A.*, et al.* (2009). HIV evolution in early infection: selection pressures, patterns of insertion and deletion, and the impact of APOBEC. PLOS Pathog *5*, e1000414.

Xiang, S., Short, S. A., Wolfenden, R., and Carter, C. W., Jr. (1997). The structure of the cytidine deaminase-product complex provides evidence for efficient proton transfer and ground-state destabilization. Biochemistry *36*, 4768-4774.

Xie, K., Sowden, M. P., Dance, G. S., Torelli, A. T., Smith, H. C., and Wedekind, J. E. (2004). The structure of a yeast RNA-editing deaminase provides insight into the fold and function of activation-induced deaminase and APOBEC-1. Proc Natl Acad Sci U S A *101*, 8114-8119.

Xu, H., Svarovskaia, E. S., Barr, R., Zhang, Y., Khan, M. A., Strebel, K., and Pathak, V. K. (2004). A single amino acid substitution in human APOBEC3G antiretroviral enzyme confers resistance to HIV-1 virion infectivity factor-induced depletion. Proc Natl Acad Sci U S A *101*, 5652-5657.

Yamanaka, S., Balestra, M. E., Ferrell, L. D., Fan, J., Arnold, K. S., Taylor, S., Taylor, J. M., and Innerarity, T. L. (1995). Apolipoprotein B mRNA-editing protein induces hepatocellular carcinoma and dysplasia in transgenic animals. Proc Natl Acad Sci U S A *92*, 8483-8487.

Yang, B., Chen, K., Zhang, C., Huang, S., and Zhang, H. (2007). Virion-associated uracil DNA glycosylase-2 and apurinic/apyrimidinic endonuclease are involved in the degradation of APOBEC3G-edited nascent HIV-1 DNA. J Biol Chem *282*, 11667-11675.

Yeung, M. L., Houzet, L., Yedavalli, V. S., and Jeang, K. T. (2009). A genome-wide short hairpin RNA screening of jurkat T-cells for human proteins contributing to productive HIV-1 replication. J Biol Chem *284*, 19463-19473.

Yu, Q., Chen, D., König, R., Mariani, R., Unutmaz, D., and Landau, N. R. (2004a). APOBEC3B and APOBEC3C are potent inhibitors of simian immunodeficiency virus replication. J Biol Chem *279*, 53379-53386.

Yu, Q., Konig, R., Pillai, S., Chiles, K., Kearney, M., Palmer, S., Richman, D., Coffin, J. M., and Landau, N. R. (2004b). Single-strand specificity of APOBEC3G accounts for minus-strand deamination of the HIV genome. Nat Struct Mol Biol *11*, 435-442.

Yu, X., Yu, Y., Liu, B., Luo, K., Kong, W., Mao, P., and Yu, X. F. (2003). Induction of APOBEC3G ubiquitination and degradation by an HIV-1 Vif-Cul5-SCF complex. Science *302*, 1056-1060.

Zhang, H., Yang, B., Pomerantz, R. J., Zhang, C., Arunachalam, S. C., and Gao, L. (2003). The cytidine deaminase CEM15 induces hypermutation in newly synthesized HIV-1 DNA. Nature *424*, 94-98.

Zhang, J., and Webb, D. M. (2004). Rapid evolution of primate antiviral enzyme APOBEC3G. Hum Mol Genet *13*, 1785-1791.

Zhang, W., Du, J., Evans, S. L., Yu, Y., and Yu, X. F. (2012). T-cell differentiation factor CBF-beta regulates HIV-1 Vif-mediated evasion of host restriction. Nature *481*, 376-379.

Zhen, A., Wang, T., Zhao, K., Xiong, Y., and Yu, X. F. (2010). A single amino acid difference in human APOBEC3H variants determines HIV-1 Vif sensitivity. J Virol *84*, 1902-1911.

Zheng, Y. H., Irwin, D., Kurosu, T., Tokunaga, K., Sata, T., and Peterlin, B. M. (2004). Human APOBEC3F is another host factor that blocks human immunodeficiency virus type 1 replication. J Virol *78*, 6073-6076.

Zhou, H., Xu, M., Huang, Q., Gates, A. T., Zhang, X. D., Castle, J. C., Stec, E., Ferrer, M., Strulovici, B., Hazuda, D. J., and Espeseth, A. S. (2008). Genome-scale RNAi screen for host factors required for HIV replication. Cell Host Microbe *4*, 495-504.

Zielonka, J., Bravo, I. G., Marino, D., Conrad, E., Perkovic, M., Battenberg, M., Cichutek, K., and Munk, C. (2009). Restriction of equine infectious anemia virus by equine APOBEC3 cytidine deaminases. J Virol *83*, 7547-7559.

Zielonka, J., Marino, D., Hofmann, H., Yuhki, N., Lochelt, M., and Munk, C. (2010). Vif of feline immunodeficiency virus from domestic cats protects against APOBEC3 restriction factors from many felids. J Virol *84*, 7312-7324.

Zrenner, R., Stitt, M., Sonnewald, U., and Boldt, R. (2006). Pyrimidine and purine biosynthesis and degradation in plants. Annual Rev Plant Biol *57*, 805-836.