2014

Transcriptome-Wide Characterization of APOBEC1-Catalyzed RNA Editing Events in Macrophages

Claire Ellen Hamilton

TRANSCRIPTOME-WIDE CHARACTERIZATION OF APOBEC1-CATALYZED RNA EDITING EVENTS IN MACROPHAGES

A Thesis Oresented to the Faculty of

The Rockefeller University

in Partial Fulfillment of the Requirements for

the degree of Doctor of Philosophy

by

Claire Ellen Hamilton

June 2013

TRANSCRIPTOME-WIDE CHARACTERIZATION OF APOBEC1-CATALYZED RNA EDITING EVENTS IN MACROPHAGES

Claire Ellen Hamilton, Ph.D.
The Rockefeller University 2014

RNA editing refers to the process by which the sequence of RNA is altered through the insertion, deletion or modification of specific nucleotides. Editing of mRNA transcripts can increase the informational complexity encoded by the genome by producing alternative protein isoforms through specific post-transcriptional RNA editing events. Additionally, RNA editing in non-coding regions of mRNA transcripts has been shown to influence gene expression in a tissue-specific manner. In mammals, mRNA editing serves a diverse set of biological roles in neuronal function, host defense and lipid metabolism. The major mRNA editors acting in mammals include the adenosine deaminases acting on RNA (ADARs) and Apolipoprotein B mRNA Editing Catalytic polypeptide-1 (APOBEC1).

The ADARs and APOBEC1 were originally characterized as catalysts for previously characterized biologically important RNA-editing events that resulted in specific coding changes; study of additional editing activity was limited by standard sequencing techniques. APOBEC1 in particular was characterized in the small intestine as mediating a specific editing event in the coding region of *Apolipoprotein B* (*Apob*). APOBEC1-dependent RNA editing in *Apob* mediates the tissue-specific differential expression of *Apob* isoforms, a process important for intestinal lipid metabolism and transport. The development of next-generation sequencing has allowed for transcriptome-wide discovery of RNA editing

activity and has resulted in the identification of more than 10,000 RNA editing events, pointing to more biological functions for RNA editing than had been previously appreciated.

To search for additional APOBEC1 editing events, our lab developed a comparative RNA-Seq screen for the transcriptome-wide identification of enzyme-specific RNA editing events. Applying this technique to small intestine enterocytes, the site of known APOBEC1 activity, we identified over 30 novel APOBEC1 editing events in transcript 3'UTRs, which represents the first example of physiological APOBEC1 editing outside of the *Apob* transcript. These newly identified editing events were located in evolutionarily conserved regions of transcript 3'UTRs, suggesting that this editing activity may have functional relevance. The discovery of additional editing activity for APOBEC1, as well as the fact that it is expressed in a number of immune cell types, suggests that APOBEC1, like other members of the AID/APOBEC family, may contribute to cellular immune processes.

The focus of the work presented in this thesis is the identification and characterization of physiological APOBEC1 editing activity in bone marrow-derived macrophages (BMDMs). Using a comparative RNA-Seq screen, I identified more than 100 novel APOBEC1 editing events in BMDMs. This APOBEC1 activity occurred in two distinct editing patterns and fell within evolutionarily conserved regions of transcript 3'UTRs. Luciferase reporter assays were utilized to assess the consequences of APOBEC1 3'UTR editing on protein expression and identified a number of combinations of editing events that affect translational outcomes. To determine if APOBEC1 editing could modulate protein expression by altering miRNA targeting, high-throughput sequencing of

RNA isolated by cross-linking immunoprecipitation (HITS-CLIP) of the Argonaute (Ago) proteins was performed on wild-type and APOBEC1-deficient cells. HITS-CLIP yielded transcriptome-wide maps of Ago binding and potential miRNA seed regions. While there was considerable overlap between loci targeted by both Ago and APOBEC1, little evidence was found for APOBEC1 disruption or creation of miRNA seed targets. Overall, this work characterizes abundant APOBEC1 activity in BMDMs that can modulate protein expression levels by a miRNA-independent mechanism. These results point to broader functions for APOBEC1 in transcript regulation and host defense.

For Grannie

ACKNOWLEDGEMENTS

My time at Rockefeller has been wonderful, made so by the mentorship and support of a great number of people. First of all I'd like to thank my PI, Nina. She runs a simultaneously nurturing and challenging lab that produces thoughtful, introspective and enthusiastic scientists. I feel honored to have been a member of that group and to have been molded into a productive scientist by someone I consider an incredible role model. And, I promise her that she has rubbed off on me, and I will leave the lab more of an optimist than when I started!

I'd also like to thank my committee: Dr. Robert Darnell, Dr. Sabine Ehrt and Dr. Daniel Mucida. I enjoyed four years of instructive, productive and illuminating committee meetings and ended up collaborating with each of the members of my committee. I am incredibly grateful for their time, support and guidance.

The Papavasiliou lab has grown substantially over my four years but has remained an encouraging, collaborative environment full of great scientists that I'm indebted to. I'd like to thank my bay-mate, Monica Mugnier without whom I couldn't have made it through my PhD! I am also indebted to Eric Fritz for hours of bioinformatic help, scientific advice and experimental support. And, to Dewi Harjanto for bioformatic assistance with miRNA project. Additionally, I would like to thank Rebecca Delker, Danae Schulz and Galadriel Hovel-Miner for their friendship and lots and lots of experimental advice. Another thank-you to Dimitris Garyfallos for keeping me company in the lab late at night and always running my sequencing down to the courier on time! I'd also like to thank Dewi Harjanto for her collaboration with bioinformatics for this project. And finally, I

need to acknowledge Violeta Rayon Estrada for all her experimental help and tons and tons of luciferase assays and macrophage preps. And thank-you to everyone else in the lab for their friendship and advice.

I'd also like to thank Scott Dewell and the rest of the Genomics core. Another thank-you to Emily Conn Gantman in the Darnell lab who did the HITS-CLIP library preps and was an amazing source of advice during all the CLIP experiments. Nikki Goodsmith and Paul Dossa helped me do some infection experiments, and I thank them for their time and assistance. Also thank you to Kate Jeffrey for teaching me how to make macrophages at the beginning of my PhD.

A very big thank-you goes out to Brad Rosenberg who has remained my role model and mentor for last 6 years in the Tri-Institutional program. He has been my friend, my constant supporter, and an invaluable source of advice. I can't thank him enough for everything he's done.

The Tri-Institutional program has been wonderful during my last 6 years. I'd especially like to that Olaf Andersen and Ruth Gotian for all their support.

Finally, I'd like to thank all my friends and family, especially my parents who have always been a source of constant loving support. Finally, I need to extend a huge thank-you to my wonderful boyfriend. Without his support, advice, encouragement, and love I don't think I'd be here today.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 1

CHAPTER 2

CHAPTER 3

LIST OF TABLES

Chapter 1: Introduction

RNA editing describes the multiple processes by which an RNA sequence is altered from that encoded by its DNA, through the insertion, deletion or modification of specific nucleotides (*reviewed in* Gott and Emeson, 2000). The term RNA editing was coined in the 1980s specifically to describe the insertion and deletion of uridine in the mitochondrial RNA of kinetoplastids (Benne et al., 1986). This type of RNA editing is termed insertion/deletion editing (*reviewed in* Gott and Emeson, 2000).

While the historical definition of RNA editing involves the insertion or deletion of mitochondrial nucleotides, the RNA editing field is currently predominantly focused on base-modification editing. Unlike uridine insertion/deletion editing, base modification occurs in all kingdoms of life and functions in a diverse set of biological processes. In base-modification editing, an adenosine or cytidine nucleotide is deaminated, resulting in its conversion to a substitute base. This reaction is catalyzed by two families of deaminase enzymes: adenosine deaminases convert adenosine to inosine (A to I) and cytidine deaminases convert cytidine to uridine (C to U). Adenosine and cytidine deaminases can act on a diverse set of RNA substrates including tRNAs, mRNAs and miRNAs.

1.1 Base-modification editing of tRNAs

One particularly important form of base modification editing contributes to the post-transcriptional modification of tRNAs. Inosine, the product of adenosine deamination, can base pair with U, C or A nucleotides. Therefore, the introduction of inosine into the first position of the tRNA anticodon (I_{34})

increases the base-pairing flexibility of the tRNA. Along with G:U base-pairing, this flexibility is the basis for the "wobble" hypothesis, by which a single tRNA can recognize multiple mRNA codons. Editing of tRNA anticodons is a process catalyzed by the adenosine deaminases acting on tRNAs (ADATs). In bacteria, the adenosine deaminase, TadA/ecADAT2 catalyzes formation of I_{34} in $tRNA^{Arg}$ (Wolf et al., 2002). The homologous eukaryotic enzymes, which function as a heterodimer of Tad2p/ADAT2 and Tad3p/ADAT3, have a broader specificity and act on 7 or 8 tRNA substrates (Gerber and Keller, 1999). These enzymes are essential for cell viability in bacteria and yeast, underscoring the importance of this specific post-transcriptional modification of tRNA (Gerber and Keller, 1999; Wolf et al., 2002).

The ADATs, although acting on adenosine, have considerable homology to cytidine deaminases, particularly in the zinc-dependent deaminase motif (Gerber and Keller, 1999). This motif is shared by an entire super-family of deaminases that include enzymes that act on free cytosine, cytidine and dCMP as well adenosine and cytidine in the context of a polynucleotide. The fact that the ZDD is shared by enzymes that can deaminate both C and A, underscores the biochemical similarities in the deamination reactions that converts A to I and C to U (*reviewed in* Conticello et al., 2007). Indeed, in trypanosomes, the ADAT2/3 heterodimer was shown to be capable of catalyzing A-to-I editing in tRNA as well as C-to-U editing in a ssDNA substrate (Rubio et al., 2007). The ADATs have been proposed to be the evolutionary precursors of the polynucleotide deaminases, the ADARs and the AID/APOBEC family (Conticello, 2008; Conticello et al., 2007).

Although RNA editing also occurs in mitochondrial RNA, tRNA and ribosomal RNA, the focus of this work is mRNA editing. Editing events in mRNA serve diverse biological functions. Editing of mRNAs can increase genetic diversity by altering the genome-encoded transcript sequences, generating alternative protein isoforms. Additionally, editing in untranslated regions has been proposed to lead to altered gene expression. In higher eukaryotes, mRNA editing events are mediated by the adenosine deaminases acting on RNA (ADARs) that catalyze adenosine-to-inosine (A-to-I) editing (Figure 1.1), and apolipoprotein B-editing enzyme, catalytic polypeptide-1 (APOBEC1), a member of the AID/APOBEC family of polynucleotide cytodine deaminases that is responsible for cytidine-to-uridine (C-to-U) editing (Figure 1.2).

1.2. Adenosine deaminases acting on RNA: ADARs

The adenosine deaminases acting on RNA (ADARs) bind to dsRNA substrates and catalyze the conversion of adenosine to inosine (*reviewed in* Bass, 2002; Keegan et al., 2004; Savva et al., 2012; Valente and Nishikura, 2005). Inosine is read as guanosine by translational and reverse transcriptional machinery. A-to-I editing is the most common editing event in higher eukaryotes, occurring predominantly in RNA duplexes formed from inverted Alu or LINE repeats in the untranslated regions (UTRs) of primary mRNA transcripts (Athanasiadis et al., 2004; Kim et al., 2004; Levanon et al., 2004). The ADARs were first identified as dsRNA-unwinding enzymes in Xenopus laevis (Bass and Weintraub, 1987; Rebagliati and Melton, 1987) but were redefined shortly thereafter as dsRNA-editing enzymes (Bass and Weintraub, 1988; Wagner et al., 1989). In mammals, three ADARs (ADAR1-3) have been identified, based on conservation in the C-

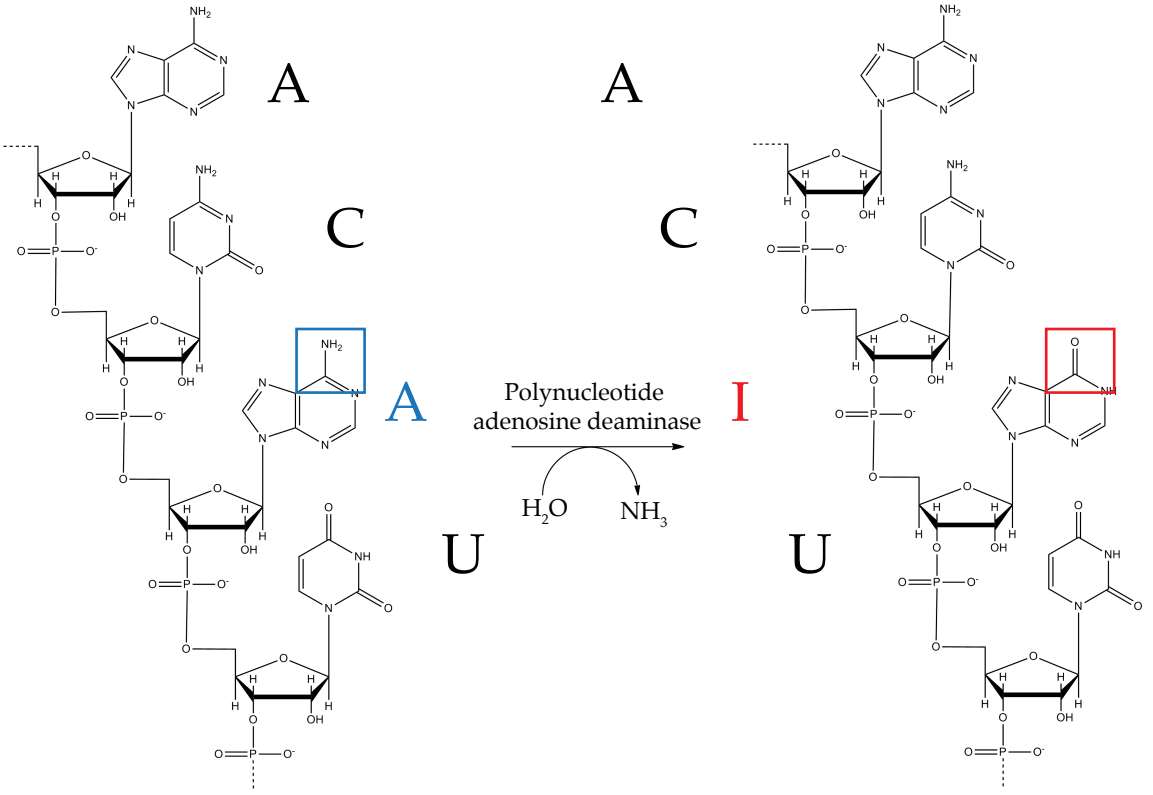


Figure 1.1. The conversion of A-to-I catalyzed by polynucleotide adenosine deaminases.

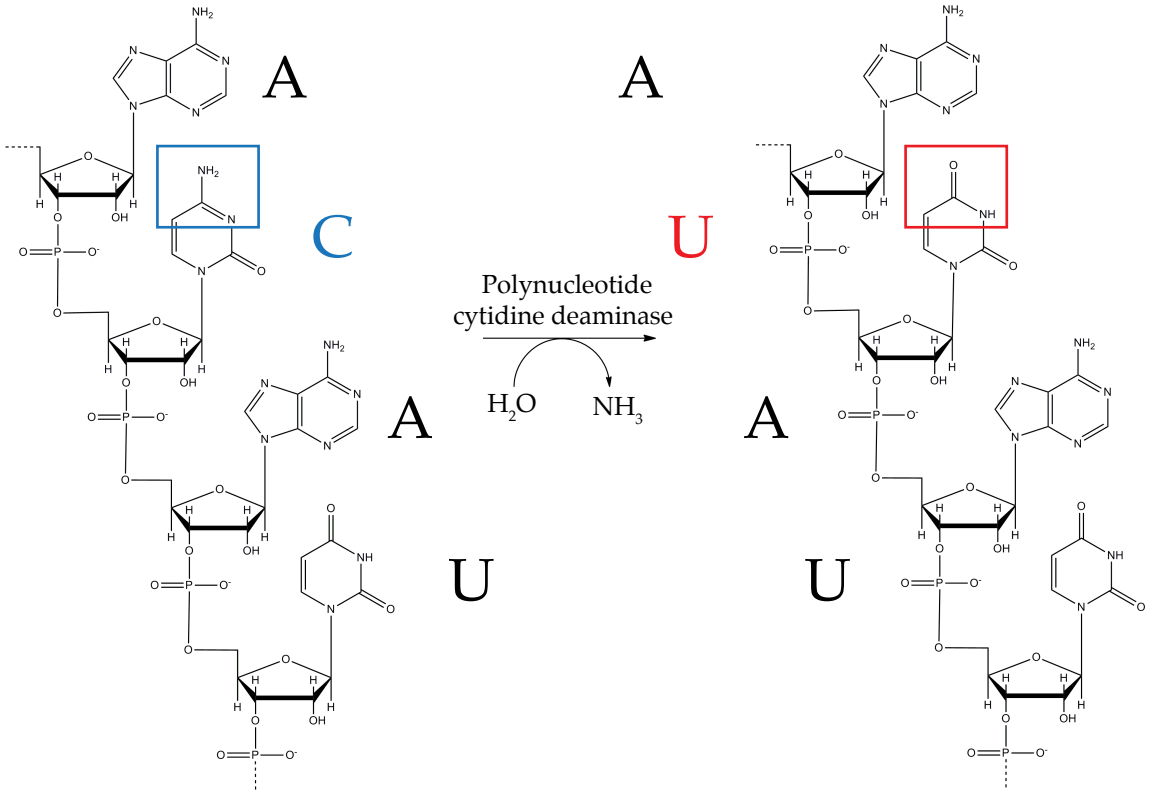


Figure 1.2. The conversion of C-to-U catalyzed by polynucleotide cytidine deaminases.

terminal deaminase and double-stranded RNA-binding domains. ADAR1 and ADAR2 demonstrate deaminase activity (Gerber et al., 1997; Kim et al., 1994; Lai et al., 1997; Melcher et al., 1996b), but the function of the catalytically inactive ADAR3 remains elusive (Chen et al., 2000; Melcher et al., 1996a). ADAR1 is the best-characterized member of this gene family. Two differentially localized isoforms of ADAR1 have been identified. The larger form, ADAR1-L or ADAR p150, is present in both the cytosol and nucleus and is regulated by an interferon-inducible promoter (George and Samuel, 1999b; 1999a; Patterson and Samuel, 1995). The smaller form, ADAR1-S or ADAR1 p110, is exclusive to the nucleus and its transcription is driven by two constitutively active promoters (Kawakubo and Samuel, 2000). The ADAR proteins are highly conserved in metazoa; two ADAR proteins are expressed in *C. elegans* (Tonkin et al., 2002) while one ADAR functions in *Drosophila* (Palladino et al., 2000). However, while mammalian ADARs are essential for life (Higuchi et al., 2000; Wang et al., 2000), ADAR-deficient *Drosophila* or *C. elegans* exhibit significant but non-fatal phenotypes (Palladino et al., 2000; Tonkin et al., 2002).

The domain organization of mammalian ADARs consists of two or three amino-terminal dsRNA binding domains (dsRBDs) and a C-terminal catalytic zinc-dependent deaminase domain (ZDD). ADAR editing exclusively occurs in dsRNA structures and can occur as "selective" editing or "non-selective" editing determined largely by the size of the targeted dsRNA duplex. Smaller (15-40 base-pairs) duplexes promote the deamination of a few "select" adenosines. These smaller duplexes, which typically contain additional structural features such as bulges, mismatched base-pairs and internal loops (Dawson et al., 2004; Lehmann and Bass, 1999), are characteristic of ADAR editing within the coding

regions of mRNA targets (Burns et al., 1997; Higuchi et al., 1993; Lomeli et al., 1994).

Larger structures (>50 base-pairs) support non-selective deamination (Nishikura et al., 1991; Polson and Bass, 1994). This non-selective editing predominantly occurs in the untranslated regions of mRNA transcripts, which have considerable secondary structure and can form long dsRNA duplexes via the pairing of inverted repetitive elements(Athanasiadis et al., 2004; Kawahara and Nishikura, 2006; Levanon et al., 2004). Other dsRNA duplexes found in ADAR-expressing cells, including viral double-stranded replication intermediates and pre-miRNA transcripts (Blow et al., 2006; Kawahara et al., 2007b), are also well-characterized ADAR targets. Overall, ADARs can act on a diverse array of RNA targets and are implicated in many essential biological functions.

The ADAR proteins and APOBEC1 are the only known mammalian mRNA editors. Over the last 20 years ADARs have been extensively characterized as important for neurological function, hematopoeisis, viral immunity and regulation of the interferon response (*reviewed in* Bass, 2002; Hamilton et al., 2010). Furthermore, since the advent of high-throughput sequencing technologies, the transcriptomes of ADAR-expressing organisms have been extensively deep sequenced, producing extensive sequence data and an ever-expanding list of novel RNA editing events in humans (Li et al., 2009b; Peng et al., 2012), mice (Cattenoz et al., 2013; Gu et al., 2012) and *Drosophilia* (Hughes et al., 2012; Rodriguez et al., 2012). In contrast, work on APOBEC1 has been predominantly focused on intestinal editing. The transcriptome-wide editing data has been limited to the work produced by our lab (Rosenberg et al.,

2011b), identifying 32 novel RNA editing events catalyzed by APOBEC1 in intestinal enterocytes.

ADAR and APOBEC1 editing activities closely resemble one another. Both ADAR and APOBEC1 were identified as catalysts of functionally important mRNA editing events within the transcript coding regions. Subsequent transcriptome-wide sequencing of both enzymes has revealed abundant RNA editing within transcript 3'UTRs, with mostly uncertain biological relevance. The following sections provide a comprehensive review of ADAR editing in mammals including the well-characterized functions of ADAR editing in coding regions as well as the more functionally elusive, but extensive, ADAR editing in mRNA UTRs.

1.2.1 ADAR editing in the brain

ADARs are highly expressed in brain tissue and are important for neurological function. Site-specific ADAR editing in the brain leads to codon changes and thereby protein products with altered physiological functions (*reviewed in* Bass, 2002; Keegan et al., 2004; Valente and Nishikura, 2005). ADAR2 extensively edits the mRNA of multiple subunits of the glutamate receptor (GluR). One A-to-I event, termed the Q/R site, in the *GluR-B* subunit is essential (Higuchi et al., 2000). Upon translation, inosine at this position results in a protein with functional differences relative to the non-edited isoform, including decreased calcium permeability and altered channel kinetics (Higuchi et al., 1993; Lomeli et al., 1994). ADAR2 knockout mice exhibit seizures and premature mortality, a phenotype that is rescued by a single nucleotide change at the Q/R site of the *GluR-B* subunit (Higuchi et al., 2000), underscoring the importance of

this editing event. ADAR2 editing in the other *GluR* subunits is less well-characterized and seems to impart more subtle effects on neurological function (*reviewed in* Bass, 2002).

In mammals, ADAR editing also modulates the function of the serotonin receptor, 5-HT_{2C}R. ADAR-catalyzed editing within the mRNA encoding 5-HT_{2C}R, occurs at 5 different sites of the G-protein coupling domain (Burns et al., 1997; Niswender et al., 1998). These editing events allow for the expression of multiple isoforms of 5-HT_{2C}R with different ligand binding and downstream signaling properties. Overall, the ADAR-dependent editing of the serotonin and glutamate receptors in the mammalian brain leads the production of multiple protein isoforms from one mRNA transcript, resembling APOBEC1 editing of *Apob*. These coding-region editing events underscore how RNA editing can increase biological diversity and functional plasticity, a process especially important in the brain.

1.2.2. ADAR1 is essential for immune cell development

Multiple observations demonstrate the importance of A-to-I mRNA editing in the immune system, starting with a requirement for ADAR1 in immune cell development. More specifically, knockout studies have defined a role for ADAR1 in fetal hematopoiesis. ADAR1-deficient mice are known to suffer from liver and bone marrow defects and typically die at embryonic day 11.5–12.5 (Hartner et al., 2004). Using ADAR1 conditional knock-out models, Orkin and colleagues have further demonstrated that ADAR1 is necessary for the maintenance of both fetal liver-derived and adult bone marrow hematopoietic stem cells (HSCs) (Hartner et al., 2009). The authors propose that ADAR1

maintains HSC populations by protecting cells from early apoptotic events, possibly by regulating interferon signaling pathways (see below). Other recent work, however, suggests that ADAR1 activity is essential in the differentiation of hematopoietic progenitor cells (HPCs), rather than HSCs (XuFeng et al., 2009). Although the question remains as to which stage of hematopoiesis requires ADAR1, it is clear ADAR1-deficient precursors do not develop into mature immune cells.

1.2.3. ADAR can restrict or enhance viral infection

As an interferon-inducible gene, it is not surprising that ADAR1 has been implicated in antiviral defense. As described above, it appears to be a critical regulator of the interferon response, which implies activity on host RNA targets. However, as an RNA editing enzyme, ADAR1 could presumably edit viral RNAs in a process not unlike APOBEC3-family hypermutation of retroviral cDNA. Indeed, like HIV antagonism of APOBEC3G by the Vif protein, vaccinia virus and adenovirus have evolved ADAR1 inhibitors that specifically impair ADAR1 deaminase activity (Lei et al., 1998; Liu et al., 2001), suggesting an anti-viral function for ADAR1. ADAR1 does edit a broad spectrum of viral targets but seems to play both pro- and anti-viral roles in infection. A-to-I editing has been observed in diverse viral RNAs, including influenza virus (Tenoever et al., 2007), parainfluenza virus (Murphy et al., 1991), lymphocytic choriomeningitis virus (LCMV) (Zahn et al., 2007), vesicular stomatitis virus (VSV) (O'Hara et al., 1984), measles virus (Li et al., 2009c; Toth et al., 2009; Wang et al., 2008), polyomavirus (Kumar and Carmichael, 1997), hepatitis D virus (HDV) (Luo et al., 1990) and hepatitis C virus (HCV) (Taylor et al., 2005). Despite an early recognition of A-to-

I hyper-editing in viral transcripts during persistent and lytic infections (Cattaneo, 1994), the consequences of many of these editing events is still under investigation. A clear example of direct ADAR antiviral editing has been observed in LCMV RNA transcripts (Zahn et al., 2007). *In vitro* and *in vivo* studies of LCMV infection demonstrated high rates of ADAR1-specific A-to-I editing events, leading to dysfunctional glycoproteins and impaired viral infectivity. In addition, recent work on HCV infection identified an ADAR1 editing-dependent loss of HCV replicons (Taylor et al., 2005). Replicon loss was thought to be attributable to an inosine-specific RNase (Scadden and Smith, 1997; 2001) or to viral genome instability introduced by weakly base-pairing inosine nucleotides (Taylor et al., 2005). ADAR family members can also directly restrict viral replication independent of its editing function. ADAR1 associates with and activates transcription factors involved in anti-viral gene expression, including nuclear factor 90 (NF90) (Nie et al., 2005). ADAR1 interacts with NF90 via an undefined dsRNA bridge and leads to the upregulation of NF90-regulated genes, including IFNβ.

In contrast to these well-characterized anti-viral functions, ADAR proteins can also promote viral infection and replication. A recent screen for effectors of the type I interferon response identified ADAR as a pan-viral stimulatory factor; ADAR enhanced the replication of numerous viruses including HCV, yellow-fever virus, West Nile virus and HIV-1 (Schoggins et al., 2011). While p110 ADAR1 is the predominant isoform during embryogenesis (George et al., 2005), the p150 interferon-inducible form is more prevalent in hematopoietic stem cells of the adult (Hartner et al., 2004). Expanding on their work with the conditional knockout mice, Orkin and colleagues showed that ADAR1 acts as novel

suppressor of the type I interferon response. Specifically, gene signatures of uninfected ADAR1-deficient HSCs and erythroid precursors are highly similar to those of virus-infected or interferon-treated cells. Additionally, ADAR1 knockout embryos were found to have significantly higher levels of type 1 interferon in extracellular fluid. It remains to be explained how ADAR1 dampens the interferon response in the absence of viral infection. ADAR1 could be editing a microRNA molecule or target, neutralizing an unidentified immunostimulatory dsRNA or functioning in the regulation of interferon-induced gene expression. An intriguing hypothesis is that ADAR1 functions as a cytosolic dsRNA binding protein competing for substrates with the DNA-dependent activator of interferon regulatory factors (DAI). In sequestering immunostimulatory dsRNA (such as viral replication intermediates) from DAI, ADAR1 may impair downstream innate immune signaling (Wang et al., 2008). The absence of ADAR1 dsRNA binding activity could lead to aberrant activation of the innate immune response and a corresponding induction of interferon production.

Perhaps related to its role in regulation of the interferon response, ADAR1 has been observed to bind and impair host antiviral response elements. Work on measles viral infection has shown ADAR1 inhibition of protein kinase regulated by RNA (PKR) and interferon regulatory transcription factor-3 (IRF-3) (Toth et al., 2009). Other groups have observed similar ADAR1 regulation of PKR during infections by VSV (Li et al., 2009c; Nie et al., 2007) and HIV (see below) as well. Inhibition of these proteins compromises the host cell's ability to respond to viral signals, potentially promoting persistent infection. Interestingly, hyper-editing of viral transcripts has been observed in patients suffering from a complication of persistent measles infection. This paradoxical situation, in which there is

evidence of ADAR1 activity both supporting and antagonizing viral infection suggests a nuanced role for ADAR1 in host-virus interaction. It appears that ADAR1 can dampen type I interferon signaling by several mechanisms. This function may have evolved to protect the host from an inappropriate (i.e., in uninfected cells) or overactive (i.e., disproportionate response to infection) interferon response. This regulation might serve to counterbalance the effector functions of ADAR1 as an antiviral enzyme. However, it seems that in certain infections, ADAR1 anti-viral activity is not only impaired but the enzyme itself may be co-opted by viruses to support infection. In this context, the significance of ADAR1 hyper-editing viral RNA remains unclear.

Like its deaminase cousin APOBEC3G, ADAR1 has recently been found to target and edit HIV-1 sequences. However, unlike APOBEC3G, which edits retrotranscript cDNA and restricts viral infection, ADAR1 targets viral RNA and enhances HIV protein expression, replication and infectivity (Doria et al., 2009). These effects are mediated by both editing-dependent and editing-independent mechanisms. Over-expression of ADAR1 in HIV-1 producer cells dramatically enhances expression of several viral proteins irrespective of RNA editing, possibly due to ADAR inhibition of PKR (Clerzius et al., 2009). Active RNA editing may also regulate virus production as ADAR1, but not experimentally-engineered catalytic mutants, increased the release of progeny virions 2-fold and enhanced HIV-1 infectivity 2.5-fold (Doria et al., 2009). Further investigation is needed to understand the regulation, impact and mechanism of ADAR1 editing in HIV-1 RNA.

The best-characterized ADAR-mediated viral editing event occurs during HDV infection. HDV is a subviral pathogen that is dependent on a concurrent

infection with HBV; it requires HBV surface antigen to infect hepatocytes (Lai, 1995; Taylor, 2003). The genome of HDV is an ideal substrate for ADARs, as its single-stranded negative sense circular RNA forms secondary structures with frequent duplex regions (Wang et al., 1986). In addition to co-opting the hepatitis B surface antigen, the HDV genome encodes its own HDV-specific surface antigen (HDVAg), the antigenome transcript of which is edited by ADAR1 in a site-specific manner (Casey and Gerin, 1995; Luo et al., 1990; Polson et al., 1996). The HDVAg occurs in two forms, both essential for the viral life cycle. The short form, HDVAg-S is required for viral RNA replication (Kuo et al., 1989) while the long form (HDVAg-L) directs viral genome assembly and packaging (Chang et al., 1991; Ryu et al., 1992). ADAR1 targets the HDV antigenome at a specific "amber/W site," thereby converting a stop codon (UAG) to a tryptophan (UIG). This allows for the translation of the long form of the HDV surface antigen (Casey and Gerin, 1995; Luo et al., 1990; Polson et al., 1996). The HDVAg-L then restricts viral replication in a trans-dominant fashion by binding HDAg-S and interrupting HDAg-S homodimers (Chao et al., 1990; Glenn and White, 1991). In unstimulated cells, amber/W site editing occurs via the ADAR1-S isoform (Jayan and Casey, 2002; Wong and Lazinski, 2002) and serves to support viral assembly; HDVAg-L halts replication and mediates viral packaging by binding clathrin heavy chain (Huang et al., 2009). However, more recent studies have shown that ADAR1 can serve an anti-viral role when editing in a more promiscuous fashion. Over-expression of ADAR1 or ADAR2 (though ADAR2 is not induced in natural infection) leads to hyper-editing at non-amber/W sites, producing higher levels of HDAg-L as well as other HDAg mutants that can also bind HDAg-S and inhibit replication. In an interferon-stimulated system, which more closely

mimics both early natural infection (Hartwig et al., 2004) and IFN-treated infection, the large form of ADAR1 is highly expressed, is the predominant editor, and increases editing 2-fold (Hartwig et al., 2006). It is unclear, however, whether IFN stimulation would induce high enough levels of ADAR1-L to edit promiscuously or to significantly impair viral replication. In fact, while a replication-competent mutant virus with enhanced editing at the amber/W site displays increased levels of HDAg-L early in infection and impaired replication at a later time point, editing and replication seem to be coupled (Sato et al., 2004). Amber/W site editing ceases with replication, indicating a natural feedback mechanism controlling aberrant editing by elevated ADAR1 activity.

1.2.4 ADAR editing and oncogenesis

As ADARs directly alter nucleotide sequences, it has been speculated that ADAR-catalyzed RNA editing could promote oncogenesis. There is some indirect evidence to support this, including altered editing profiles in tumor specimens, especially editing found in transcripts encoding tumor suppressors (*reviewed in* Skarda et al., 2009). Recently ADAR-catalyzed A-to-I editing was directly implicated in hepatocellular carcinoma (HCC), centered on increased A-to-I editing in the antizyme inhibitor 1 (*Azin1)* transcript observed in a substantial set of human HCC specimens (Chen et al., 2013). The elevated editing in *Azin1* led to a gain-of-function phenotype in the encoded protein, increasing its stability and allowing it to bind with greater affinity to antizyme. Antizyme typically functions to target growth-promoting proteins for degradation. Therefore by restricting the inhibitory activity of antizyme, edited AZIN1 promoted cell proliferation and tumor progression. This work is the first well-

characterized direct link between ADAR editing and oncogenesis and points to further functions for RNA editing in promoting cellular transformation.

1.2.5 ADAR editing in untranslated regions

In addition to the ADAR coding targets essential to neurological function, a few additional editing events within the translated regions of mRNA have been characterized (Levanon et al., 2005; Riedmann et al., 2008), but the overwhelming majority occur in untranslated regions (UTRs) of transcripts, introns or miRNA processing intermediates. Most of this untranslated editing occurs within *Alu* or LINE repeats, which through intramolecular base-pairing of inverted repeats, form long RNA duplexes ideal for extensive ADAR editing (Athanasiadis et al., 2004; Blow, 2004; Kim et al., 2004; Li et al., 2009b). *Alu* elements are ~300nt repetitive transposable elements which make up ~10% of the human genome, occurring primarily in the UTRs or introns of transcribed mRNAs (Lander et al., 2001). *Alu* elements can contain cryptic splice sites and can occasionally aberrantly incorporate into mature transcripts, a process termed exonization, posing a threat to transcriptome integrity. In one example, intronic ADAR editing in an *Alu* element prevented its exonization into the mature mRNA (Sakurai et al., 2010). However, the functional consequences for most of these A-to-I editing events within *Alu* regions remain elusive, although they have been proposed to modulate gene expression. Specifically, examples of ADAR editing have been found to induce nuclear retention of the edited transcript, target the transcript for cleavage, and abolish or create miRNA target sites.

Nuclear retention of inosine-containing transcripts occurs via association with a nuclear inosine-specific RNA binding protein $p54^{nrb}$ that sequesters these

mRNAs in nuclear paraspeckles (Chen and Carmichael, 2009; Prasanth et al., 2005; Zhang and Carmichael, 2001). In an experimental system, introduction of inverted *Alu* elements into a GFP gene resulted in dramatic nuclear retention via ADAR editing and binding to $p54^{nrb}$, underscoring the potential importance of this method of *Alu* and ADAR-mediated gene silencing (Chen et al., 2008). However, highly edited endogenous cytoplasmic mRNA transcripts have been identified, suggesting that nuclear retention may not be the fate of all ADAR-targets (Chen and Carmichael, 2009; Chen et al., 2008; Hundley et al., 2008). Furthermore, $p54^{nrb}$ binding does not require inosines, suggesting that an additional feature may be targeting identified inosine-containing transcripts to the nucleus.

Additional inosine-containing mRNA transcripts are targeted for mRNA cleavage by a specific nuclease. *In vitro* work suggests that Tudor Staphylococcal nuclease (Tudor-SN) binds stretches of I:U base-pairs and promotes the cleavage of the hyper-edited transcripts (Scadden, 2005), either through its own nuclease activity (Yang et al., 2006) or by another unidentified factor (Scadden, 2005). Although other endogenous inosine-containing mRNAs are shown to be cleaved upon stress conditions (Prasanth et al., 2005), the Tudor-SN cleavage mechanism has not been established *in vivo*.

ADAR editing events have been shown to modulate miRNA silencing by editing of miRNA targets and double-stranded nuclear precursors. As ADAR editing predominantly occurs in conserved regions of 3'UTRs, it was postulated that ADAR editing could create or disrupt miRNA target sequences (Liang and Landweber, 2007). A few examples of ADAR editing creating miRNA seed targets have been characterized, but few disruptions could be identified

(Borchert et al., 2009). Other work additionally demonstrates that miRNA targeting is rare in *Alu* repeats, suggesting that overlap between ADAR editing and miRNA targeting machinery may be limited (Hoffman et al., 2013).

Precursor miRNAs (pri-miRNAs and pre-miRNAs) can also be edited by ADARs, where introduction of I:U mismatches disrupts miRNA processing (Kawahara et al., 2008; 2007a; Yang et al., 2006) or changes the targeting specificity of the mature miRNA (Kawahara et al., 2007b). In one well-characterized example, ADAR editing in the seed region of pri-miR-376 endowed the mature miRNA product with distinct binding properties, allowing the edited form to target an alternative transcript (Kawahara et al., 2007b). High-throughput miRNA sequencing has shown that this ADAR-catalyzed pri-miRNA editing does occur somewhat frequently (Alon et al., 2012; Vesely et al., 2012), and that some edited miRNAs are differentially expressed in ADAR-deficient embryos (Vesely et al., 2012). Intriguingly, very recent work demonstrated that ADAR promoted miRNA processing by a protein-protein association with DICER (Ota et al., 2013). The potential importance of this process is supported by *Adar*$^{-/-}$ embryos, which exhibit a global inhibition of miRNA expression and subsequent upregulation of targeted transcripts (Ota et al., 2013), suggesting that ADAR editing-independent modulation of DICER has as profound effect on global miRNA expression.

Other *in vitro* experiments have additionally demonstrated that ADAR can compete with DICER for dsRNA substrates as part of siRNA gene silencing (Knight and Bass, 2002; Scadden and Smith, 2001). This work has primarily been observed in *C. elegans* and *Drosophila* knock-down experiments, but taken together with the catastrophic phenotype for ADAR1-null mice discussed above,

could point to an essential function for ADAR in suppressing potentially immunostimulatory dsRNA. All together, this work firmly suggests a role for ADAR editing in small RNAs, primarily in miRNA processing through both editing and editing-independent mechanisms.

1.3. Polynucleotide cytidine deaminases: the AID/APOBEC family

The AID/APOBEC family of zinc-dependent polynucleotide cytidine deaminases catalyze the conversion of cytidine to uridine in a strand of DNA or RNA, effectively altering the sequence of the targeted polynucleotide. The family was named for its founding member Apolipoprotein B mRNA Editing Catalytic polypeptide-1 (APOBEC1), identified as the catalyst for a well-characterized C-to-U editing event in the mRNA transcript of apolipoprotein B. The other family members include Activation-induced cytidine deaminase (AID), APOBEC2, the subfamily of APOBEC3 (A-H) enzymes and the computationally predicted APOBEC4. AID/APOBEC deaminases all share a characteristic zinc dependent deaminase domain (ZDD). In the ZDD three conserved residues, two cysteines and a histdine, coordinate a zinc atom to activate a water molecule for hydrolytic cytidine deamination. Additional conserved glutamic acid and proline residues in the active site also play an essential role in the reaction, transferring a proton from the water molecule to the imino in the pyrimidine ring and ensuring the conformational integrity of the catalytic pocket, respectively (*reviewed in* Smith, 2009). APOBEC1 was originally identified as a cytidine deaminase based on homology to zinc-dependent cytidine deaminases in yeast and *E. coli*, which act on a monomeric substrates as a part of the pyrimidine salvage pathway (Navaratnam et al., 1995; 1993). While the members of the AID/APOBEC family

share significant homology with these enzymes, particularly in the ZDD, they appear to act only on single-stranded polynucleotide substrates.

In contrast to the ancient family of cytidine deaminases acting on free cytidine found in all kingdoms of life, the AID/APOBEC family is a later evolutionary development and is restricted to the vertebrate lineage (Conticello et al., 2005). AID is thought to be the ancestral member of this protein family, which arose concurrently with vertebrates and the development of adaptive immunity. AID, APOBEC2, and APOBEC4 are all present in jawed vertebrates (Conticello et al., 2005). Phylogenetic analyses indicate that APOBEC4 may have evolved independently from AID but APOBEC2 is likely to have arisen by duplication of the AID locus (Conticello, 2008). APOBEC1 and APOBEC3 are derived from more recent AID gene duplication events and are restricted to mammals and placental mammals, respectively (Conticello et al., 2005). APOBEC3, a single locus in the mouse, has undergone a dramatic expansion process in primates into an array of 8 *Apobec3* genes encoding APOBEC3A-H. The emergence of this APOBEC3 subfamily is most likely due to extensive selective pressure by the rapid evolution of its target retroviruses and retrotransposons (Sawyer et al., 2004).

Throughout primate evolution, the majority of the AID/APOBEC enzymes have evolved rapidly, displaying some of the strongest signals of positive selection in the human genome (Sawyer et al., 2004), an evolutionary pattern associated with host defense. Indeed, while members of AID/APOBEC family are implicated in a diverse set of biological processes, the majority function in immunity. AID drives antibody diversification in B cells through the processes of somatic hypermutation and class switch recombination. The

APOBEC3 subfamily acts directly on endogenous viral retrogenomes and restricts exogenous retroviral production. APOBEC2 and APOBEC4 are "orphan" deaminases with no well-established targets. The well-characterized function for APOBEC1, the subject of this discussion, is in lipid absorption and transport in the small intestine. However, the evolutionary history of the AID/APOBEC family and APOBEC1's wide expression in immune cells suggest that this enzyme may have previously unappreciated roles in the immune system. The following sections will give a general overview of the biological functions of the members of the AID/APOBEC family of cytidine deaminases, concluding with a detailed discussion of the known activity of APOBEC1 and the growing evidence for its function beyond the intestine.

1.3.1. Activation-induced cytidine deaminase: AID

AID is a key player in antibody-mediated adaptive immunity, mediating secondary antibody diversification through somatic hypermutation (SHM) and class-switch recombination (CSR) (Figure 1.3). In the decade since AID was identified, AID's role in SHM and CSR has been extensively characterized (*reviewed in* Delker et al., 2009). In SHM, AID introduces point mutations in the recombined variable region of the immunoglobulin locus in germinal center B cells. These mutations are removed by uracil DNA glycosylase (UNG) and then repaired through the activities of error-prone base excision repair (BER) and mismatch repair (MMR) enzymes (Figure 1.3). The combined high editing rate of AID and error rate of these enzymes leads to the rapid introduction of genomic mutations in immunoglobulin (Ig) loci, some of which alter the affinity of the encoded antibody. B cells with improved antigen binding affinity are positivity

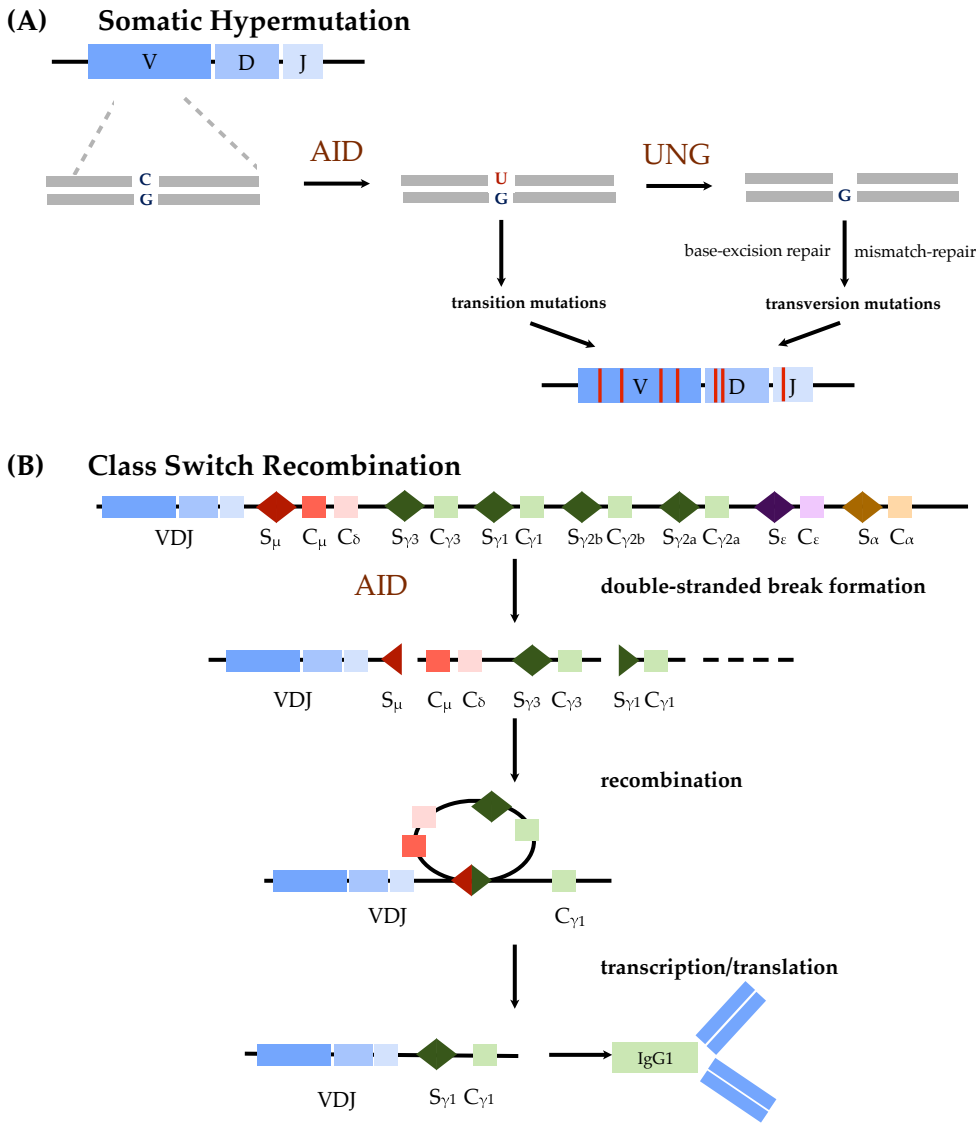


Figure 1.3. AID drives antibody diversity through two distinct mechanisms. (A) In SHM, AID deaminates cytidines within the variable region of Ig loci. Genomic deoxyuridine residues are then resolved by two pathways. Uridine is read as thymidine by replication machinery, leading to transition mutations. Alternatively, UNG excises the edited base, which is followed by abasic site repair via base-excision repair and mismatch-repair enzymes, leading to transversion mutations. (B) In CSR, AID deaminates cytidines within Ig switch (S) regions, leading to double strand breaks. Recombination replaces the primary constant switch region (C_μ) with one of several downstream constant regions (C_γ, C_ε or C_α), altering the effector properties of the encoded antibody. In this representation, the primary C_μ region is replaced with a $C_\gamma 1$ region, thereby causing a switch from the IgM to the IgG1 isotype.

selected for, and cycles of mutation and selection lead to an optimized pool of B cells capable of initiating a highly specific immune response against a particular invading pathogen. In CSR, AID editing in the switch regions of the *Igh* locus mediates the formation of double-stranded breaks that lead to a recombination event, replacing the primary constant region with an alternative downstream constant region. These constant regions encode the Fc region of the antibody, which determines the isotype. Prior to class switch, all B cells express IgM; recombination events lead to the production of secondary isotypes (IgA, IgE, and IgGs in mammals). The processes of SHM and CSR are entirely dependent on AID; *Aicda*$^{-/-}$ B cells express no secondary antibody isotypes and have no apparent somatic mutation in the variable region of the Ig loci (Muramatsu et al., 2000). In humans, loss-of-function mutations in the gene encoding AID results in a comparable deficiency, termed hyper-IgM syndrome 2 (Revy et al., 2000).

At its discovery, AID was proposed to function as an RNA editing enzyme because of its homology to APOBEC1 (Muramatsu et al., 1999). The subsequently developed "RNA-editing hypothesis" of AID function in SHM and CSR suggests that to mediate both processes AID edits the mRNAs of auxiliary factors. This proposal also reflected the APOBEC1 editing model, as it required the assistance of additional co-factors, resembling the relationship between APOBEC1 and its "editosome." The alternative "DNA-editing hypothesis" proposed that AID itself was directly mutating the Ig loci, triggering SHM and CSR. The processes of SHM and CSR described above reflect the predominantly accepted AID DNA-editing model, but the RNA editing hypothesis has not been entirely discarded (Shivarov et al., 2009).

A large body of work has demonstrated the direct DNA editing capacity of AID in support of the DNA editing hypothesis. Endogenous AID is found to be directly bound to the Ig locus in activated splenic B cells (Vuong et al., 2009; Yamane et al., 2011) and ectopically expressed AID is capable of inducing DNA mutations in a diverse set of cell types, including both prokaryotic and eukaryotic cells (Martin and Scharff, 2002; Mayorov et al., 2005; Petersen-Mahrt et al., 2002; Poltoratsky et al., 2004; Yoshikawa et al., 2002). Furthermore, while AID can bind both RNA and DNA *in vitro* (Dickerson et al., 2003; Nonaka et al., 2009), editing assays demonstrated that AID is only capable of deaminase activity on a DNA substrate (Besmer et al., 2006; Bransteitter et al., 2003; Chaudhuri et al., 2003). These *in vitro* studies additionally demonstrated that AID prefers to edit WRC ([A or T][A or G]C) motifs (Pham et al., 2003; Yu et al., 2004a), a pattern also observed in SHM editing of the Ig locus (Rogozin and Diaz, 2004). Finally, deep sequencing of the B cell transcriptome failed to identify any *bona fide* AID-mediated RNA editing events (Fritz et al., 2013). Additional support for the DNA-editing hypothesis was provided by the analysis of mice deficient in UNG, the glycosylase that cleaves mismatched uridines introduced by C-to-U editing. These animals exhibited an altered array of mutations in the Ig locus as compared to their wild-type littermates, with a bias toward C-to-T and G-to-A mutations (Rada et al., 2002). This suggests that the initiating event in SHM is a C-to-U conversion catalyzed by a cytidine deaminase. Error-prone repair of these mismatches, catalyzed by UNG and the base-excision and mismatch repair pathways, leads to accumulation of other transition and transversion mutations, pointing to the importance of both AID and the DNA repair enzymes in SHM. Taken together, this work strongly suggests that the

"RNA-editing hypothesis" is false and demonstrates that the predominant substrate for AID is ssDNA.

As AID activity can mutate genomic DNA and introduce double-stranded breaks, aberrant AID activity has great oncogenic potential. To combat deleterious off-target mutations, while preserving essential immunological function, the transcription, translation, cellular localization and targeting of AID are tightly regulated (*reviewed in* Delker et al., 2009). The many modes of AID regulation are still under intensive investigation and are largely beyond the scope of this discussion. Below, I will briefly expand on the regulation of AID by miRNAs, post-translational modifications, and the targeting of AID to Ig loci. *Aicda* transcripts are targeted by two miRNAs, miR-155 and miR-181, which regulate *Aicda* transcript levels (de Yebenes et al., 2008; Dorsett et al., 2008; Teng et al., 2008). The importance of miRNA-regulation is demonstrated by mouse models. Mice lacking a miR-155 target sequence in the AID 3'UTR exhibit substantially elevated AID expression that is inappropriately sustained after B cell exit from the germinal center (Dorsett et al., 2008; Teng et al., 2008). In addition to the predictably elevated CSR, the disruption of miR-155 regulation leads to hypermutation in non-Ig loci (Teng et al., 2008) and chromosomal translocations (Dorsett et al., 2008), pointing to the importance of miRNAs in preventing promiscuous AID activity.

Post-translational modifications of the AID protein also seem to be essential for physiological AID function. In particular, phosphorylation of the Ser38 residue by the cAMP-dependent protein kinase A (PKA) is required to recruit replication protein A (RPA), an important AID co-factor; disruption of this phosphorylation event decreases SHM and CSR by 70% (Cheng et al., 2009;

McBride et al., 2008). Furthermore, complexes of phosphorylated AID, PKA and RPA are localized to switch regions of the *Igh* locus. When PKA is catalytically inactivated, preventing AID phosphorylation, RPA is not recruited to the switch region and CSR is impaired, highlighting the importance of the post-translational modification and additional targeting factors on the localization of AID activity (Vuong et al., 2009).

The precise mechanism which targets AID activity to the Ig loci is largely unresolved. As highlighted above, PKA and AID are targeted to switch regions, and the phosphorylation of AID by PKA leads to the recruitment of RPA and the initiation of CSR. This finding was further underscored in a genome-wide analysis of RPA occupancy, in which RPA was associated mainly with the Ig loci, an interaction that was dependent on AID phosphorylation (Yamane et al., 2011). Extensive work has pinpointed AID activity to regions of active transcription, where the transcriptional machinery has exposed a ssDNA substrate for AID (*reviewed in* Di Noia and Neuberger, 2007; Nussenzweig and Nussenzweig, 2010; Peled et al., 2008). In a current model, AID is targeted to a stalled RNA polymerase II (Pol II) by the co-factor Suppressor of Ty 5 homolog (Spt5), a stalling factor (Pavri et al., 2010). While this work exemplifies the importance of certain co-factors and post-translational modification in targeting AID to stalled Pol II and actively transcribed genes in the Ig loci, there are many unanswered questions. This is particularly complex, as AID needs to act specifically at the V(D)J region during SHM and the switch region during CSR. How this intricate targeting occurs is especially important in context of the genome-wide non-Ig activity of AID.

Despite many modes of tight regulation, it is clear that AID has significant activity beyond the Ig loci and can contribute to oncogenesis. AID-catalyzed mutations can be found throughout the genome, and in a variety of oncogenes and tumor-suppressor genes (Gordon et al., 2003; Pasqualucci et al., 2001; Pavri et al., 2010; Robbiani et al., 2009; Shen et al., 1998; Yamane et al., 2011). In total, an estimated 25% of the genes expressed in germinal center B cells are mutated by AID, albeit at levels considerably lower than in the Ig loci (Liu et al., 2008). Additionally, deep sequencing identification of the genomic locations of AID places it at the promoters of over 5000 genes in association with its co-factor Spt5 and a stalled RNA Pol II (Pavri et al., 2010; Yamane et al., 2011). While AID targeting seems to be extensive, much of the genome seems to be protected from deleterious AID-mediated mutation and chromosomal translocation. Mice deficient in mismatch and base-excision repair enzymes exhibit strikingly higher mutation rates at non-Ig loci than wild-type littermates, suggesting that physiological error-free base-excision and mismatch repair resolves the majority of aberrant deaminase activity (Liu et al., 2008). AID-mediated tumorigenesis, therefore, occurs more frequently when in concert with an additional pro-oncogenic factor, such as deficient DNA repair or heightened AID activity. Indeed, mice exhibiting uncontrolled AID expression have increased mutation rates and develop tumors as a result of AID mutations in oncogenes and tumor suppressors (Okazaki et al., 2003).

In humans, AID is expressed in some B cell lymphomas (Lenz et al., 2007; Wright et al., 2003), and double-stranded breaks introduced in CSR are proposed to lead to oncogenic chromosomal translocations, especially the c-myc/IgH translocation associated with Burkitt's Lymphoma (Pasqualucci et al., 2001; 2008;

Shaffer et al., 2002). In mice, AID can induce double-stranded breaks throughout the genome (Hasham et al., 2010) and was shown to be essential for c-myc/IgH translocations (Kovalchuk et al., 2007; Ramiro et al., 2006; 2004; Takizawa et al., 2008) as well as other translocations not involving the Ig loci or c-myc (Lin et al., 2009; Robbiani et al., 2009). Interestingly, the dysregulation of AID alone is not sufficient to cause such chromosomal instability or B cell malignancy (Muto et al., 2006; Okazaki et al., 2003); tumors observed in AID transgenic mice were predominantly of epithelial and T cell-origin and not associated with a chromosomal translocation (Okazaki et al., 2003). However, compound p53-deficient/AID transgenic mice rapidly succumb to a set of phenotypically diverse B cell lymphomas, harboring Ig and non-Ig chromosomal translocations (Robbiani et al., 2009). Additionally, exogenous introduction of a DSB at known AID targets, *Myc* or *IgH,* leads to AID-dependent translocations at these loci (Robbiani et al., 2008). Genome-wide profiling of these genetically manipulated translocations reveals a wide spectrum of possible AID-mediated translocations with particular abundance in transcribed genes and at transcription start sites (Chiarle et al., 2011; Klein et al., 2011). Overall, AID-mediated mutations and chromosomal translocations can occur throughout the genome, but physiologically, these unwanted AID-mediated events occur at an extremely low frequency.

AID activity outside of the Ig loci may not always be pathological. AID is expressed in primordial germ cells and embryonic stem cells (Morgan et al., 2004) and reports have implicated AID in vertebrate DNA demethylation (Bhutani et al., 2010; Popp et al., 2010; Rai et al., 2008). However, this result is not corroborated in B cells where genome-wide methylation profiles are unaffected

by AID-deficiency or over-expression (Fritz et al., 2013). Furthermore, specific AID-dependent demethylase activity has not been firmly established and may be only occurring in lower vertebrates. While the proposed model is intriguing, further studies are needed to characterize any *bona fide* role for AID and/or other cytidine deaminases in developmental reprogramming.

1.3.2. The APOBEC3 family of anti-retroviral enzymes

The APOBEC3 subfamily predominantly function in the restriction of retrovirus and endogenous retro-elements. As described above, the APOBEC3 cytidine deaminases have undergone a massive gene expansion in the primate lineage, presumably in response to their rapidly evolving pathogen substrates. The subfamily members, APOBEC3A-H, act on a diverse set of retroviral substrates and vary in their abilities to suppress certain viruses. The best-characterized example is APOBEC3G restriction of HIV infection. The function of APOBEC3G was discovered during the characterization of a specific HIV protein, virion infectivity factor (Vif) and the infectious profile of Vif-deficient (Δ*Vif*) HIV. Δ*Vif* HIV produced in primary human T cells or macrophages (and certain other cell-lines) exhibited a significantly diminished infectivity that was absent from a wild-type HIV (Gabuzda et al., 1992; Schwedler et al., 1993). These cells were termed "non-permissive cells" and were subsequently determined to express a specific anti-viral factor that was normally inhibited by Vif (Madani and Kabat, 1998; Simon et al., 1998). A subtractive cDNA screen comparing the transcript expression profiles of closely related "permissive" and "non-permissive" cell-lines identified this restriction factor as the cytidine deaminase, APOBEC3G (Sheehy et al., 2002).

The potent viral restrictive activity of APOBEC3G in a Δ*Vif* HIV infection occurs partially through its cytidine deaminase activity, specifically in its ability to massively hyper-edit viral cDNAs (Harris et al., 2003; Lecossier et al., 2003; Mangeat et al., 2003). During a Δ*Vif* HIV infection of non-permissive cells, APOBEC3G is incorporated into the budding virus core through an interaction with viral RNA (Khan et al., 2005; Svarovskaia et al., 2004) and the viral Gag nucleocapsid protein (Schäfer et al., 2004). In newly infected cells, APOBEC3G remains associated with reverse transcription machinery and actively deaminates the nascent retroviral (-) strand cDNA (Harris et al., 2003; Mangeat et al., 2003; Yu et al., 2004b; Zhang et al., 2003). Most of these modified retrotranscripts are subsequently degraded. The rest of the edited cDNAs are integrated as devastatingly mutated and therefore defective proviruses (Kieffer et al., 2005). Additionally, APOBEC3G association with viral genomic RNA is thought to inhibit the tRNA priming of reverse-transcription, pointing to both deaminase-mediated as well as editing-independent mechanisms for the APOBEC3G inhibition of Δ*Vif* HIV virion production (Bishop et al., 2006; Guo et al., 2006; 2007).

During wild-type HIV infection, Vif protects the virus from APOBEC3G activity by targeting the deaminase for degradation and preventing its incorporation into packaging virions (Conticello et al., 2003; Marin et al., 2003; Mehle et al., 2004; Sheehy et al., 2003). To accomplish this, Vif binds APOBEC3G and recruits a specific ubiquitin ligase complex that poly-ubquitinates APOBEC3G and targets it for proteosomal degradation (Kobayashi et al., 2005; Yu et al., 2003). The Vif:APOBEC3G interaction is quite specific-specific; human

Vif is unable to inhibit a simian APOBEC3G and simian Vif has no effect on human APOBEC3G (Mariani et al., 2003). Overall, the fact that a retrovirus developed a distinct factor to combat a specific host anti-viral enzyme points to the incredible importance of the APOBEC3s in viral restriction.

Aside from HIV, APOBEC3 family members have been shown to suppress the activity of a diverse set of viruses including simian immunodeficiency virus, equine infectious anemia virus, murine leukemia virus (MLV), foamy virus adeno-associated virus (AAV) and HBV (*reviewed in* Rosenberg and Papavasiliou, 2007; Smith, 2011; Smith et al., 2012 and many others). Overall, the APOBEC3s act predominantly on retroviruses and as described for APOBEC3G, deaminate the single-stranded DNA substrate initially reverse transcribed from the viral RNA genome. While HBV is not a true retrovirus it requires reverse-transcription to replicate its double-stranded DNA, providing the appropriate substrate for APOBEC3s. AAV is a small single-stranded DNA parvovirus that is restricted by APOBEC3A (Chen et al., 2006), suggesting that certain APOBEC3s may exhibit more broad antiviral activity and act on non-retroviral pathogens.

The APOBEC3G gene has exhibited high rates of positive selection throughout primate evolution, but this selection seems to have mostly occurred before the emergence modern primate lentiviruses (Sawyer et al., 2004). As such, the ancestral function of the APOBEC3 family was proposed to also involve the restriction of endogenous retroelements, which include the long terminal repeat (LTR)-containing endogenous retroviruses (ERVs), and non-LTR sequences such as the long interspersed nuclear element-1 (LINE-1). Indeed, human ERVs carry footprints of ancient APOBEC3G deamination (Armitage et al., 2008), and the expansion of the APOBEC3 gene into the eight-member subfamily seems to have

coincided with a marked reduction in active retroelements in humans. In contrast, murine genomes from which only one copy of APOBEC3 is expressed, contain abundant retroelements some of which are still mobile *in vivo* (*reviewed in* Schumann, 2007).

The mechanism by which APOBEC3 cytidine deaminases would restrict these endogenous retroviruses is predominantly thought to mimic APOBEC3G restriction of HIV, but other editing-independent mechanisms have also been proposed. Regardless of mechanism, the majority of the human APOBEC3s have been shown to inhibit retrotransposition of ERVs (Esnault et al., 2005) and LINE-1 elements (Muckenfuss et al., 2006). The APOBEC3s have varied specificities and combat retroelements on two subcellular fronts, suggesting that the expansion of the APOBEC3 lineage into a diverse arsenal of immune enzymes may have been required to restrict the threat of endogeous retroelements.

As has been described for AID, the APOBEC3s have the potential to promote oncogenesis through DNA editing activities. Specifically the nuclear-localized deaminases APOBEC3A and B pose an increased threat to genomic integrity (Bogerd et al., 2006; Landry et al., 2011). As such, APOBEC3B was recently identified as a source of oncogenic mutation in breast cancer (Burns et al., 2013). Presumably for both APOBEC3 and AID, their key host defense functions outweigh the risk they pose for cellular transformation.

1.3.3. APOBEC2 and APOBEC4: orphan deaminases

APOBEC2 was originally identified through an expressed sequence tags (EST) database search for the APOBEC1 ZDD and was proposed to be an RNA-editing enzyme based on its close homology to APOBEC1 (Anant and Davidson,

2000; Liao et al., 1999). However, mutator assays have revealed no deaminase activity on cytidine, either as a free nucleotide or as a part of a DNA strand (Mikl et al., 2005; Nabel et al., 2012). Despite its apparent catalytic inactivity, there is some evidence for a further function for APOBEC2. APOBEC2 is exclusively expressed in cardiac and skeletal muscle (Liao et al., 1999; Mikl et al., 2005) and APOBEC2-deficient mice exhibit a distinct skeletal muscle phenotype. APOBEC2-deficiency leads to a fast-to-slow twitch muscle switch in the soleus muscle and eventual myopathy (Sato et al., 2010). Similarly, in zebrafish, morpholino knockdown of the two APOBEC2 proteins leads to a dystrophic phenotype and diminished heart function (Etard et al., 2010). These studies point to a conserved function for APOBEC2 in muscle development.

Further work in zebrafish has implicated both AID and APOBEC2 in DNA demethylation (Rai et al., 2008). And, APOBEC2 has been shown to be essential for left-right axis determination in *Xenopus* and zebrafish (Vonica et al., 2011). While it is clear that APOBEC2 acts in skeletal muscle function in mice and may have additional functions in lower vertebrates, the particular targets and mechanisms of these roles remain uncertain. Furthermore, it has never been established that APOBEC2 is catalytically active, suggesting that the observed phenotypes are due to an editing-independent function or that APOBEC2 editing requires the assistance of a co-factor or auxiliary protein complex missing from *in vitro* conditions.

APOBEC4 is a computationally predicted member of the AID/APOBEC family and contains the characteristic ZDD domain, suggesting that it is likely a cytidine deaminase of RNA or DNA (Rogozin et al., 2005). APOBEC4 expression

is primarily restricted to the testis but no editing targets or functions for this enzyme have been defined in that tissue.

1.4. APOBEC1

1.4.1. APOBEC1 editing of the *Apob* transcript

The first polynucleotide cytidine deaminase, APOBEC1 was identified as the catalyst of a well-characterized RNA-editing event in the transcript of Apolipoprotein B (*Apob*). ApoB exists in two distinct, tissue-specific protein isoforms, translated from a single mRNA transcript (*reviewed in* Kane, 1983). The full-length isoform, apoB-100, is synthesized in the liver and is an essential component of very low-density lipoproteins (VLDL), intermediate-density lipoproteins (IDL) and low-density (LDL) particles. ApoB-100 is required for the generation of hepatic VLDL particles and for transport of endogenously produced triglycerides in the blood. Additionally, it remains associated with the lipoprotein during the conversion of VLDL to IDL and then LDL particles, and serves as the ligand that mediates the clearance of LDL cholesterol by the LDL receptor pathway. The truncated isoform, apoB-48 is named as such because it is comprised of the N-terminal 48 amino acids of apoB-100. It is solely expressed in the intestine, where it is incorporated into chylomicrons and functions in the absorption and transport of dietary lipid (*reviewed in* Chan, 1992). The translation of apoB-48 is the result of a site-specific deamination of the cytidine to uridine at nucleotide 6666 (C6666) of the *Apob* transcript, the first identified mRNA editing event in mammals. The C-to-U modification creates a pre-mature stop codon (UAA) from a glutamate codon (CAA) and leads to the synthesis of the truncated

apoB-48 isoform from a full-length *Apob* transcript (Chen et al., 1987; Powell et al., 1987) (Figure 1.4).

As the catalytic component of a multi-protein editing complex, APOBEC1 mediates the site-specific deamination of C6666 (Teng et al., 1993). APOBEC1 is a zinc-dependent polynucleotide cytidine deaminase and the founding member of the AID/APOBEC polynucleotide cytidine deaminase family (Barnes and Smith, 1993; Navaratnam et al., 1993; 1995). In humans, APOBEC1 expression is restricted to the small intestine, correlating with the intestine-specific generation of the apoB-48 protein isoform (Lau et al., 1994). Notably, APOBEC1 is not expressed in the human liver, resulting in the exclusive production of apoB-100. As such, APOBEC1 mediates the tissue-specific differential expression of apoB isoforms. In mice, APOBEC1 is expressed in a variety of tissues including both the liver and the small intestine (Hirano et al., 1997; Nakamuta et al., 1995), and apoB-48 is produced in both cell tissues (Greeve et al., 1993). Both *Apob* editing and formation of the apoB-48 isoform are completely dependent on APOBEC1; APOBEC1-deficient mice display no C-to-U editing at C6666 and produce only apoB-100-containing lipoproteins (Hirano et al., 1996; Morrison et al., 1996; Nakamuta et al., 1996), a phenotype rescued by transgenic expression of APOBEC1 in the intestine (Blanc et al., 2012).

1.4.2. Mechanism of APOBEC1 editing

Like the other members of the AID/APOBEC family, APOBEC1 is a zinc-dependent cytidine deaminase and shares the common homologous cytidine deaminase domain (MacGinnitie et al., 1995; Navaratnam et al., 1995). However, it also has a distinct RNA-binding domain and is the only member of the

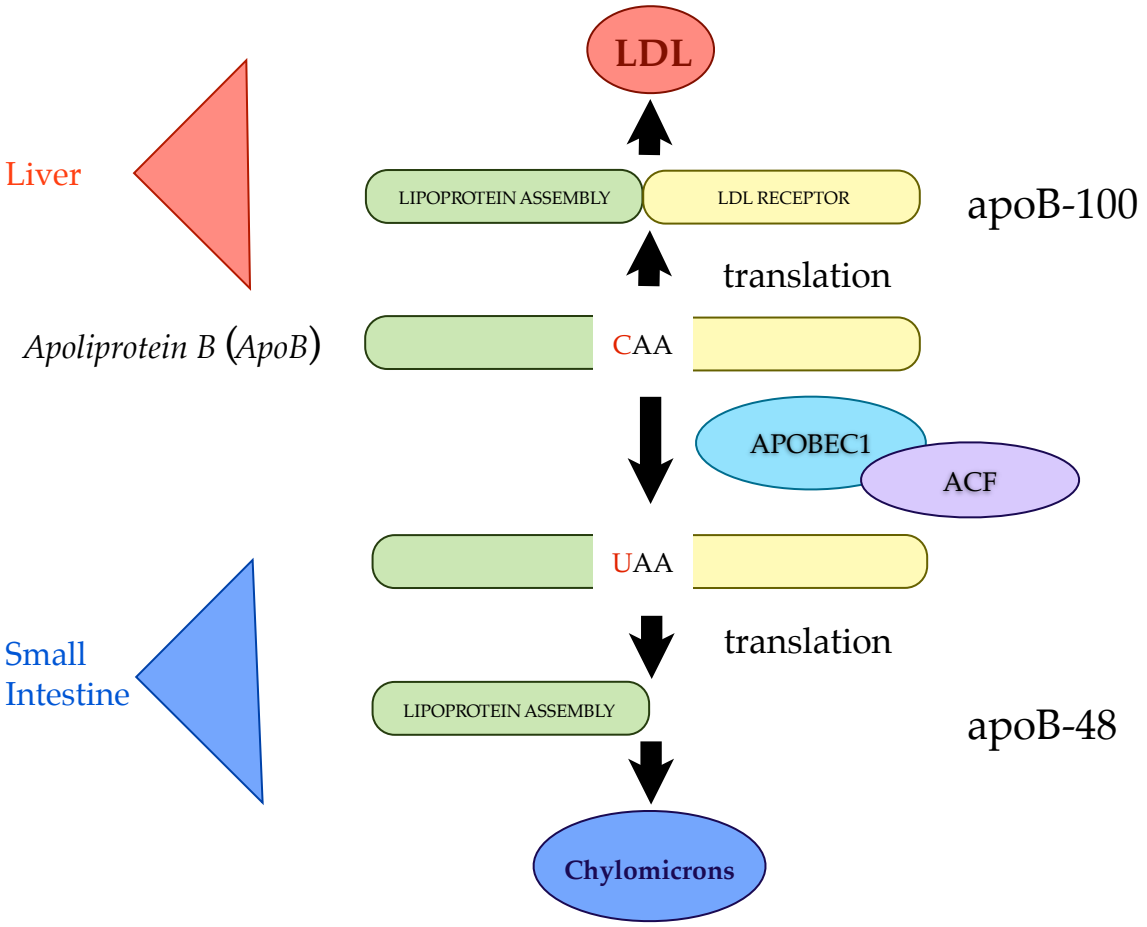


Figure 1.4. APOBEC1 editing of *Apob* transcript. APOBEC1 mediates the tissue-specific distribution of apoB isoforms. In the liver, *Apob* is translated into its full-length isoform, apoB-100, which contains two domains: lipoprotein assembly domain and LDL receptor binding domain. ApoB-100 is an essential component of hepatic LDL particles. In the small intestine, APOBEC1 deaminates cytidine 6666 in the *Apob* transcript, converting a glutamine codon (CAA) into a stop codon (UAA). This leads to the translation of a truncated isoform of apoB, designated apoB-48, which is incorporated into chylomicrons along with dietary lipid absorbed from the small intestine.

deaminase family with a known RNA substrate (Anant et al., 1995; MacGinnitie et al., 1995; Navaratnam et al., 1995). The mechanism and localization of APOBEC1-mediated editing of the *Apob* mRNA transcript is well characterized. *Apob* editing is an intra-nuclear event; APOBEC1 shuttles from the cytoplasm to the nucleus (Chester et al., 2003; Yang et al., 1997) and edits *Apob* in a nuclear process coinciding with, or immediately following, splicing and polyadenylation (Lau et al., 1991). APOBEC1 functions as a homodimer to edit the *Apob* mRNA transcript (Lau et al., 1994; Oka et al., 1997); molecular modeling based on the *E. coli* cytidine deaminase ECCDA, suggests that one active site binds a downstream U, positioning the second functionally active deaminase at C6666 (Navaratnam et al., 1998).

The predominant APOBEC1 editing event in the *Apob* transcript occurs at C6666. At 14kb, *Apob* is exceptionally large mRNA transcript and the precise targeting of the APOBEC1 "editosome" to this particular site is specified by the surrounding sequence elements. In general, APOBEC1 preferentially binds AU-rich regions (Anant et al., 1995; Navaratnam et al., 1995) and is thought to target the conserved AU-rich sequences up and downstream of C6666. Specifically, a downstream 11nt "mooring sequence" (Shah et al., 1991) separated from the edited cytidine by a 4-6nt spacer element is absolutely required for *Apob* editing *in vitro* (Backus and Smith, 1992; Chen et al., 1990); even small deviations in the sequence or position of this "mooring" motif dramatically reduces or eliminates editing at C6666. In the context of the APOBEC1 editosome, introduction of a downstream mooring sequence is also sufficient to induce editing in a heterologous mRNA (Driscoll et al., 1993). Additional sequence features of *Apob*-editing include 5′ and 3′ enhancer sequences which increase the efficiency of

editing (Backus and Smith, 1991; Driscoll et al., 1993; Hersberger and Innerarity, 1998; Nakamuta et al., 1999). The importance of these sequence features was partially explained by secondary structure analysis of the *Apob* transcript. The mooring sequence and 5' enhancer element comprise a conserved but imperfect stem, with the spacer element and edited cytosine forming an octa-loop. This stem-loop structure is essential for APOBEC1 editing; the loop positions C6666 at the APOBEC1 active site and the stem provides binding regions for the APOBEC1 editosome (Maris et al., 2005; Richardson et al., 1998). However, mutations to the mooring sequence which preserved the stem-loop also diminished editing efficiencies, pointing to both structural and sequence requirements for *Apob* editing (Richardson et al., 1998).

APOBEC1 edits *Apob* as a member of an incompletely characterized multi-protein complex, the obligate components of which are APOBEC1 and its cofactor, APOBEC1 complementation factor (ACF) (Lellek et al., 2000; Mehta et al., 2000). ACF is a 64 kDa widely-expressed protein originally isolated from a baboon kidney extract found to "complement" APOBEC1 editing (Mehta et al., 1996). The ACF protein exists in 4 major isoforms generated by alternative splicing events; each isoform has slightly differing abilities to mediate *Apob* editing (Dance et al., 2002; Sowden et al., 2004). The ACF protein contains three N-terminal RNA recognition motifs (RRMs), a C-terminal double-stranded RNA binding domain (RBD), and a nuclear localization sequence (Lellek et al., 2000; Mehta et al., 2000). During *Apob* editing, ACF binds the *Apob* transcript through its RRMs and forms a functional homodimer with an *Apob* mRNA bridge (Galloway et al., 2010).

As a pair, recombinant ACF and APOBEC1 are necessary and sufficient to induce *Apob* editing *in vitro* (Lellek et al., 2000; Mehta et al., 2000). The established mechanism of APOBEC1 editing of C6666 in *Apob* is entirely dependent on ACF at multiple points. First, ACF binding of the mooring sequence targets APOBEC1 to C6666 and mediates the relative fidelity of this editing event (Mehta et al., 2000). As ACF binds to the mooring sequence it melts the stem loop secondary structure and exposes the target cytosine to the APOBEC1 (Maris et al., 2005). The newly edited transcript is transported from the nucleus as a part of the APOBEC1:ACF editosome, a complex that also protects the transcript from nonsense-mediated decay. Association with APOBEC1 alone is not sufficient to suppress degradation and also requires ACF binding to the *Apob* transcript (Chester et al., 2003). Finally, as the atypical NLS sequence in APOBEC1 is not sufficient to target a heterologous protein to the nucleus (Chester et al., 2003; Yang et al., 1997), there is some evidence that ACF may help drive the nuclear localization of APOBEC1, due to its strong NLS sequence (Blanc et al., 2001a; 2003). As such, ACF mediates many aspects of APOBEC1 editing in *Apob* save cytidine deamination: nuclear localization, position of the edited cytosine, substrate access and protection of the final edited product. Interestingly, the function of ACF seems to extend beyond this complementation of APOBEC1 editing; unlike APOBEC1, ACF-deficiency results in embryonic lethality in mice (Blanc et al., 2005).

Other than ACF, the distinct components of the APOBEC1 editosome remain poorly characterized. Some putative editosome factors have been identified based on their ability to bind APOBEC1, ACF or apoB (Harris et al., 1993; Sowden et al., 2002). While they are not essential for editing, the majority

seem to function as regulatory factors. APOBEC1 binding proteins 1 and 2 (ABBP1, ABBP2) have been shown to dramatically enhance editing at C6666 but, the majority of the identified editosome seem to function as negative regulators of C-to-U editing. Heterogeneous nuclear ribonucleoprotein C1 (hnRNP-C1), an mRNA splicing factor, was shown to be stringently bind to the *Apob* transcript around the edited C and function as an inhibitor of the APOBEC1 editing complex (Greeve et al., 1998). Two RNA-binding proteins with distinct homology to ACF also function to inhibit C-to-U editing events in *Apob*. CUG RNA-binding binding protein (CUGBP2) contains three RRMs and associates with APOBEC1 in the cytoplasm and ACF and *Apob* in the nucleus. Anti-sense mediated knockdown of CUGBP2 dramatically increased the efficiency of *Apob* editing (Anant et al., 2001). Glycine-arginine-tyrosine rich RNA binding protein, (GRY-RBP) can also bind to ACF, APOBEC1 and apoB and seems to inhibit C-to-U editing by sequestering ACF from the *Apob* transcript (Blanc et al., 2001b). However, immunodepletion of GRY-RBP abolished C-to-U editing in an APOBEC1-expressing human hepatoma cell-line, suggesting that this protein may function as both a positive and negative regulator of the APOBEC1 editosome (Lau et al., 2001). The exact composition and action of the editosome remains unclear but the literature supports a model in which the assembled editosome works in concert to regulate APOBEC1 editing through inhibitory and stimulatory mechanisms.

1.4.3. Additional APOBEC1 editing in the *Apob* transcript

Despite multiple modes of regulation, APOBEC1 editing of *Apob* is not entirely restricted to C6666. Sequencing of human *Apob* transcripts revealed a

secondary low-frequency APOBEC1 editing event at C6802. This editing event has the potential to cause a Thr to Ile substitution but, as it occurs concurrently with the upstream C6666 editing event that introduces a stop codon, has no consequence for the final protein composition (Navaratnam et al., 1991). Recently, a number of low frequency (~10%) promiscuous hyper-editing events were observed downstream of C6666 in mouse *Apob* transcripts (Blanc et al., 2012). None of these editing events introduced a stop codon and the functional relevance of these additional edited sites is unclear. Likely, either through processive editing or low-affinity binding to AU-rich sequences, APOBEC1's standard behavior includes some low-frequency "background" editing with little biological consequence.

More abundant hyper-editing of *Apob* was observed *in vitro* and *in vivo* during over-expression of APOBEC1 (Sowden et al., 1996a; Yamanaka et al., 1996). This hyper-editing activity is proposed to be due to the altered stoichiometry between APOBEC1 and the editosome complex. In this model, abundant levels of APOBEC1 overwhelm the endogenous repertoire of the regulatory editosome and aberrantly edit as an independent enzyme in AU-rich regions.

However, APOBEC1 has only been shown be capable of editing activity without ACF in elevated temperature environments, where the stem-loop structure of the *Apob* transcript is naturally disrupted (Chester et al., 2004; Maris et al., 2005). Furthermore, there is a dispute in the literature as to whether this hyper-editing is associated with mooring sequences, calling into question whether ACF mediates hyper-editing events (Hersberger and Innerarity, 1998; Yamanaka et al., 1996). Further investigation is needed to determine the

mechanism of APOBEC1-mediated hyper-editing and to elucidate the importance of ACF complementation of APOBEC1 editing in targets beyond the canonical C6666 event.

1.4.4 Editing-independent roles for APOBEC1

Independent of its catalytic activity, APOBEC1 has been demonstrated to bind AU-rich 3'UTRs, altering the transcript stability and modulating protein expression. APOBEC1 regulates the transcript 3'UTRs of *Cox-2* (Anant et al., 2004), *Myc* (Anant et al., 2000) and *Cyp7a1* (Xie et al., 2009) in this manner with varying biological consequences. APOBEC1 stabilization of *Cox-2* transcript serves a protective role after radiation injury (Anant et al., 2004) and may contribute to tumor formation in $Apc^{min/+}$ mice (Blanc et al., 2007). Aberrant regulation of *Cyp7a1* in $Apobec1^{-/-}$ mice is proposed to lead to the gallstone susceptibility phenotype observed in these animals. However, the majority of work on APOBEC1 binding activity has come out from one specific group; alternative views in the field suggest that APOBEC1 has relatively poor RNA-binding capabilities (Smith et al., 2012). Overall, these editing-independent functions of APOBEC1 suggest a broader role for this enzyme than had been previously appreciated but this work remains to be firmly established.

1.4.5 Oncogenic APOBEC1 mRNA targets

Transgenic over-expression mouse models have revealed potentially oncogenic APOBEC1 editing events that do not represent physiological targets. Mice with transgenic hepatic over-expression of APOBEC1 developed hepatocellular carcinomas, associated with the hyper-editing of the novel

APOBEC1 target-1 (*Nat1*) mRNA transcript (Yamanaka et al., 1997). An additional target, the mouse protein tyrosine kinase *Tec*, was also edited in a mooring sequence-dependent fashion. However, this editing resulted in a silent codon change that did not contribute to APOBEC1-mediated oncogenesis (Yamanaka et al., 1995). *Nat1* editing occurs upstream of an imperfect mooring sequence, suggesting that the mechanism of this edited event was comparable to that established for *Apob,* albeit in a non-physiological system.

Analysis of peripheral nerve-sheath tumors from patients with Neurofibromatosis type I (NF1) has revealed a C-to-U RNA-editing event attributable to APOBEC1 in the transcript of the tumor suppressor (also called NF1) whose loss-of-function is associated with development of the disease. This editing event introduces a premature stop codon, leading to the synthesis of an inactive protein. APOBEC1 is expressed in these NF1 tumors and the editing event occurs upstream of a mooring sequence, indicating that it is likely a true APOBEC1-catalyzed event (Mukhopadhyay et al., 2002; Skuse et al., 1996). However, editing was only apparent in a subset of patient samples and at a low editing frequency, suggesting that while this editing may contribute to NF1 pathogenesis in some cases, it is not a physiologically significant APOBEC1 target.

A link between APOBEC1 and oncogenesis was further supported by studies implicating APOBEC1 in susceptibility to testicular germ cell tumors (TGCTs) (Nelson et al., 2012) and adenocarcinoma of the small intestine (Blanc et al., 2007) in tumor-susceptible mouse models. Most interestingly, APOBEC1-deficiency reduced the small intestine tumor burden in compound $Apc^{min/+}$;$Apobec1^{-/-}$ mice. Adenomas isolated from these mice had a marked reduction in

Cox-2 mRNA abundance that was not associated with any C-to-U editing (Blanc et al., 2007). While this points to an APOBEC1-mediated editing-independent mechanism of oncogenesis, there was no comprehensive search for APOBEC1-mediated editing in these tumors. Transcriptome-wide sequencing of $Apc^{min/+}$ mice may reveal additional novel oncogenic APOBEC1 targets.

1.4.6 Viral APOBEC1 targets

There is some evidence supporting a role for APOBEC1 hyper-editing in the inhibition of viral transcripts. However, although viral restriction by other adenosine and cytidine deaminases (ADARs, APOBEC3s) is an abundant and widely investigated phenomenon, APOBEC1 editing of viral RNA and DNA is still poorly characterized. Mouse APOBEC1 has been shown to be capable of hyper-editing viral genomes and RNA during MLV infection. This hyper-editing activity was observed for *in vitro* and *in vivo* infections but occurred at an exceptionally low frequency and was not linked to viral restriction (Petit et al., 2009). Similarly, *in vitro* infection assays suggested that both human and mouse APOBEC1 could hyper-edit HBV DNA, but patient samples revealed only negligible editing in a pattern attributable to APOBEC1 (Gonzalez et al., 2009). Although these studies suggest some viral editing by APOBEC1, the experimental techniques involved raise questions about the strength of the conclusions. First, the editing is detected using PCR techniques that selectively amplify edited products that naturally occur at an exceptionally low frequency. Also, the enzyme responsible for *in vivo* C-to-U editing in this work is determined by the enzyme preference for nucleotide 5' to the edited cytosine, suggested to be TpC for APOBEC1. In *Apob,* C6666 is preceded by an A and

3'UTR targets of APOBEC1 are predominantly preceded by either an A or T(U), calling into question the validity of any exclusive preference for a 5' T. However, more convincing studies have implicated APOBEC1 hyper-editing in the restriction of HSV-1 (Gee et al., 2011) and HIV (Ikeda et al., 2008) viruses. Overall, it is clear that APOBEC1 can edit viral ssDNA and RNA *in vitro* and some work points to a broader anti-viral function for APOBEC1 in certain viral infections.

1.5. Transcriptome-wide discovery of RNA editing events

The field of RNA editing has been dramatically changed with the advent of high-throughput sequencing technology and the ability to search transcriptome-wide for novel RNA editing events. Much of the early work on RNA editing started with the identification of an important RNA modification, followed by the subsequent characterization of the enzyme responsible. With transcriptome-wide sequencing data, the number of known RNA editing events has dramatically expanded, but the characterization of the functional relevance of the majority of these novel sites has lagged behind. This is especially true for ADAR-catalyzed RNA editing events in mRNAs and miRNAs, which have been extensively profiled in humans, mice, *C. elegans* and *Drosophila*. Our lab has previously demonstrated the utility of a screen for cytidine deaminase-specific RNA editing in mice, which compares wild-type sequence to a deaminase-deficient control to filter single-nucleotide variants (SNVs) for *bona fide* RNA editing (Rosenberg et al., 2011b). Here, I will focus on the many strategies employed in transcriptome-wide identification of A-to-I editing and C-to-U editing and the advantages and limitations of these techniques.

1.5.1 Transcriptome-wide strategies for identifying A-to-I editing

Initial transcriptome-wide discovery of RNA editing events involved mining publically available cDNA, EST, and DNA sequences for A-to-G mutations (Athanasiadis et al., 2004; Kim et al., 2004; Levanon et al., 2004). Another comparable study generated and sequenced a cDNA library, with similar results (Blow, 2004). As these publically available sequences have high error rates, the authors limited the search to likely ADAR targets, i.e. apparent ADAR clustering events or areas containing predicted RNA duplexes. These studies identified thousands of previously uncharacterized A-to-G(I) editing events, predominantly occurring in untranslated *Alu* repeats. This early work is additionally important as it set the stage for later high-throughput sequencing analyses and used many of the same bioinformatic techniques and strategies for sequence alignment and subsequent filtering of putative RNA editing events.

Next-generation sequencing technology allows for the massively parallel sequencing of whole genomes and transcriptomes. The majority of work searching for novel RNA editing has utilized transcriptome-wide sequencing (RNA-Seq). In classic mRNA-Seq, whole RNA isolated from a cell, tissue or organism of interest is subjected to a poly(A) selection, isolating polyadenylated mRNA transcripts, which are then fragmented, reverse transcribed and massively sequenced, yielding millions of reads that can be aligned to the reference genome or assembled *de novo*. Another permutation of RNA-Seq involving ribosomal RNA depletion in place of a poly(A)-selection, yields larger sequencing libraries as it includes non-poly(A) transcripts, such as many long non-coding RNAs (lncRNAs). Typically, to identify RNA editing events, single-nucleotide variants (SNVs) between the RNA-Seq reads and the reference

genome are identified and massively filtered to remove false positives introduced by errors in DNA amplification, library sequencing, and read mapping. Additionally, many strategies involve the removal of single-nucleotide polymorphisms (SNPs) encoded in the genome.

The first example of the identification of RNA editing by high-throughput sequencing used a "target-capture" technique followed by massively parallel DNA sequencing of cDNA and gDNA derived from a single human individual (Li et al., 2009b). A set of ~35,000 padlock probes were designed based on previously identified RNA editing sites, excluding repetitive *Alu* elements. These padlocks probes are single-stranded primers containing two regions of complementarity designed to flank a specific region of interest in the target DNA strand. Once the complementary regions are hybridized to the target molecules, the gap between them is filled in by a polymerase and the subsequent pool of thousands of padlocks can be amplified and massively sequenced with an Illumina sequencer. This represented a landmark for the field of RNA editing as it demonstrated the utility of deep sequencing in the transcriptome-wide identification of RNA editing and identified over 200 novel RNA editing events. However, this technique has considerable limitations. First, it required the laborious generation of thousands of padlock probes. As the cost of oligo synthesis has subsequently fallen more slowly than that of high-throughput sequencing, it is now significantly less efficient than traditional RNA-Seq. Second, the padlock probe strategy was inherently biased as it relies on patterns established from previously identified editing events.

A subsequent technique was developed based on inosine cyanoethylation, termed inosine chemical erasing (ICE) (Sakurai et al., 2010). Cyanoethylated

inosines form N^1-cyanoethylinosine (ce^1I), which cannot base pair with C and therefore stalls reverse transcription machinery. In ICE, both cyanoethylated (Ce+) and un-cyanoethylated (Ce-) RNA samples from the same tissue were reverse transcribed, amplified and subjected to Sanger sequencing. The sequences of true RNA editing events would exhibit an A-to-G conversion in the Ce- sample that is absent from the Ce+ sample, due to inhibition of the reverse transcription machinery that occurred during cDNA conversion. False positives, such as A-to-G sequencing errors or SNPs, would be equally converted in both Ce- and Ce+ samples, and easily removed from the screen. The ICE technique was applied to hundreds of regions predicted to contain ADAR editing and led to the discovery of ~2500 novel editing events. ICE sequencing has been highly successful in identifying novel ADAR editing events and its strategy greater lowers false positive rates. However, like the padlock method, it is biased toward "likely" ADAR targets and can only identify A-to-I events, not the particular enzyme responsible.

As the price of next generation sequencing has dropped, more groups were able to use RNA-Seq to search for RNA-editing events in an unbiased fashion across the transcriptome (Bahn et al., 2012). The major barrier for these techniques is differentiating *bona fide* RNA editing from SNPs or additional errors introduced during sequencing and mapping. A highly controversial study that identified widespread RNA and DNA differences (RDDs) across the transcriptome highlighted the challenge of this process and lead to follow-up work that defined more optimal practice for the identification of RNA editing. In their much discussed paper, Cheung and colleagues presented a transcriptome-wide comparison of RNA and DNA sequences from human B cells, in which

they reported over 10,000 RDDs across the transcriptome, introducing tremendous informational complexity (Li et al., 2011). The most surprising part about this study was that many of these events were transversions incapable of being catalyzed by adenosine or cytidine deaminases, the only known enzymes to introduce RDDs. However, technical comments demonstrated that the majority of these events (>90%) were due to technical artifacts and genetic variation (Lin et al., 2012; Pickrell et al., 2012; Schrider et al., 2011). First, the Cheung group aligned reads to an incomplete transcriptome; although reads were aligned uniquely (i.e., discarding reads which aligned to more than one position), reads originating from gene paralogs missing from the incomplete reference aligned incorrectly, introducing apparent RDDs. Second, many of the identified RDDs displayed positional and strand biases, indicating that they were technical artifacts. A common place for these biases to be introduced was during first-strand cDNA synthesis, where 5′ mismatches can be introduced with the random hexamer, resulting in a 5′ end bias restricted to the negative strand (Lin et al., 2012). Finally, a substantial number were identified as genomic SNPs.

The limitations of this report indicate that transcriptome-wide identification of RNA editing events must be employed with stringent alignment and filtering parameters. Reads must be uniquely and stringently mapped to the genome. Additional pre-alignment read trimming can remove end bias, but these can also be filtered later. After read alignment, a series of filters can reduce a false positives by eliminating strand-bias, low coverage sites, sites edited at 100%, positional bias, repetitive regions, and known SNPs. A number of subsequent studies applying these more stringent alignment and filtering parameters to varying degrees have established a higher-confidence and ever-

expanding inosome in introns, mRNAs, and miRNAs (Alon et al., 2011; Bahn et al., 2012; Danecek et al., 2012; Gu et al., 2012; Peng et al., 2012; Ramaswami et al., 2012).

1.5.2. Transcriptome-wide identification of APOBEC1 editing

As discussed above, a major challenge in the identification of RNA-editing events is how to filter *bona fide* RNA editing events from genomically encoded SNPs. In 2009, I contributed to a manuscript from our lab describing a "comparative" RNA-Seq screen, which uses a deaminase-deficient control to partially overcome this barrier and identify *bona fide* deaminase-specific RNA editing events (Rosenberg et al., 2011b). However, this strategy is less effective in the identification of ADAR-specific events as *Adar*$^{-/-}$ mice are not viable. Some studies have been successful in siRNA-mediated knock-down strategies to identify ADAR editing events in certain cells lines, but this can be limited by the stringency of the knock-down (Bahn et al.). Applying the comparative RNA-Seq strategy to the transcriptome of small-intestinal enterocytes, we identified 32 additional APOBEC1 editing targets, dramatically expanding the physiological editing repertoire of an editing enzyme previously reported edit one target.

The comparative RNA-Seq strategy exploits the complete lack of APOBEC1-catalyzed *Apob* editing observed in *Apobec1*$^{-/-}$ mice. The transcriptomes of small intestine enterocytes derived from wild-type and APOBEC1-deficient mice were subjected to next-generation sequencing, yielding millions of 36nt reads. These reads were stringently aligned to the genome; after a series of quality filters, loci which contained a C-to-T mismatch in the wild-type sample that was absent from the APOBEC1-deficient sample were defined as APOBEC1-

catalyzed RNA editing events. In this point, our technique exhibits a considerable advantage over many previously described A-to-I identification strategies in mice. The deaminase-deficient control greatly reduces the number of false positives introduced by genetic variation and sequencing errors and defines only deaminase-specific events. Furthermore, our screen was performed with highly stringent read quality, alignment and filtering parameters. Subsequently, further validation of the APOBEC1 editing events yielded a low false positive rate of ~15%.

These newly identified APOBEC1 editing events were validated by standard Sanger sequencing of cDNA and gDNA from an additional pair of wild-type and *Apobec1*$^{-/-}$mice. 33 of the 39 candidate sites were validated as *bona fide* APOBEC1 editing events. The edited frequencies of these sites were calculated from the RNA-Seq reads and ranged from 0.92 (in *Apob*) to 0.18. Additionally, analysis of Sanger sequences revealed some additional low-frequency hyper-editing events surrounding certain identified editing sites. These resemble the low-frequency promiscuous editing observed in *Apob* (Blanc et al., 2012) and are of unknown relevance.

Further analysis of these validated editing events revealed characteristic sequence features of APOBEC1 editing in *Apob*. Most prominently, a downstream mooring motif comparable to the 11-nucleotide mooring sequence was observed in most of the editing targets. This mooring motif, defined as WRAUYANUAU, is more flexible than the previously established mooring sequence (UGAUCAGUAU) essential for *Apob* editing and predominantly occurs 4-6nt downstream of the targeted cytidine. A transcriptome-wide search for this mooring motif revealed that it is present in numerous transcripts, both in coding

and untranslated regions. We sequenced these transcripts with standard Sanger sequencing to look for APOBEC1 editing upstream of the putative mooring sequences. Although we identified 9 additional APOBEC1 3'UTR targets that had been missed by the original RNA-Seq screen, no editing was observed in coding regions. These results underscore the importance of the mooring sequence in directing APOBEC1 editing activity while also suggesting that APOBEC1 editing in coding regions may be a rare event regulated by auxiliary factors.

Earlier analyses of APOBEC1 editing were essentially restricted to one target, so very little has been defined about the targeting and editing preferences of this enzyme. It has been demonstrated previously that APOBEC1 binds highly AU-rich regions and that the region flanking C6666 is AU-rich. To establish whether this was true for 3'UTR editing, we determined the nucleotide composition of the 100nt regions surrounding the validated APOBEC1 editing in 3'UTRs. As a set, these were found to be substantially more AU-rich than random sets of 101nt. Additionally, the 101nt region within each 3'UTR that was edited by APOBEC1 was significantly more enriched in AU content than the surrounding sequence. Together, this suggests that, as in *Apob,* APOBEC1 binds AU-rich regions to promote efficient editing. The members of the AID/APOBEC family exhibit local preferences for the nucleotides flanking the targeted cytidine. To define similar preferences for APOBEC1, we assessed the nucleotide composition of the 4 nucleotides up and downstream of the edited cytidines. We found that there was a significant preference for A or U nucleotides at the positions immediately flanking the edited C.

Finally, we assessed whether the regions with observed APOBEC1 editing in 3'UTRs was conserved over evolution. We determined the phastCon scores for the 101nt windows centered upon edited cytidines and compared it to random 101nt windows in the same 3'UTRs. Together, the APOBEC1 targeted regions in 3'UTRs were significantly more conserved, suggesting that these regions may be of functional relevance.

Overall, this work established a highly-specific RNA-Seq screen for RNA editing and identified 32 novel APOBEC1 editing events in evolutionarily conserved regions of transcript 3'UTRs. While the functional importance of this untranslated RNA editing remains elusive, these results point to additional functions for APOBEC1 beyond its well-characterized role in lipid metabolism.

1.6. Statement of problem

Polynucleotide RNA and DNA editing function in a diverse set of biological processes in mammals. In particular, the activities of some well-characterized cytidine and adenosine deaminases centers upon host defense, where editing events contribute to antibody diversification, restriction of retroviruses and endogenous retroelements, and suppression of the interferon response. Although much of this editing activity has also been linked to oncogensis, the essential immune functions imparted by this editing presumably outweigh the risk for cellular transformation. RNA editing is often additionally touted as a driver of transcriptome diversity, where post-transcriptional modifications increase the complexity of the genome (Bass, 2002). ADAR, the best characterized polynucletoide RNA editing enzyme, does seem to introduce some sequence diversity in targeted transcripts, especially in neurological tissue.

The majority of ADAR editing events, however, occurs within untranslated regions of targeted transcripts and may function to modulate gene expression rather than protein isoforms *per se*.

Our lab has developed a comparative RNA-Seq screen to identify APOBEC1-specific RNA editing events transcriptome-wide. Application of this screen in small intestine enterocytes identified 31 novel APOBEC1-catalyzed editing events within transcript 3'UTRs. These newly identified targets were predominantly located in areas of high phylogenic conservation, implying functional relevance. Additionally, a subset overlapped with miRNA seed target regions, suggesting a role for APOBEC1 editing in modulating miRNA targeting. Overall, our work increased the known editing repertoire for APOBEC1 and points a broader function for APOBEC1 than previously established.

Despite its expression in a variety of tissues (Hirano et al., 1996; Nakamuta et al., 1995), including secondary lymphoid organs and a number of immune cell types (Rosenberg and Papavasiliou, unpublished data), the study of APOBEC1 has remained focused on its roles in the digestive system. However, there is some evidence that APOBEC1 could also function in immune cells. Throughout primate evolution, members of the AID/APOBEC family have evolved rapidly, displaying some of the strongest signals of positive selection in the human genome (Sawyer et al., 2004), a pattern associated with host defense. Indeed, the majority of this family has well-conserved functions in the immune system. Furthermore, a role for RNA editing in immune function has been established for ADARs, who exhibit well-characterized functions in the interferon response, viral infection and the suppression of exogenous dsRNAs.

Here, I have used an adapted comparative RNA-Seq strategy to identify over 100 novel APOBEC1 editing events in transcript 3′UTRs of bone marrow-derived macrophages (BMDMs), a cell type that express APOBEC1 without ACF or ApoB. Unlike APOBEC1 editing identified in the intestine, BMDM editing is only weakly associated with a mooring motif. These editing events occur predominantly in conserved regions of 3′UTRs and can be grouped into two distinct editing patterns: site-specific editing and hyper-editing. I further analyzed the downstream consequences of the newly identified APOBEC1 editing events and found editing events that repressed protein expression in an experimental system. Finally, I assessed the interaction between APOBEC1 editing and miRNA targeting and found little evidence that APOBEC1 was affecting miRNA binding. My results demonstrate dramatic physiological ACF-independent APOBEC1 editing outside of the intestine and point to a role for these editing events in miRNA-independent transcript regulation in BMDMs.

Chapter 2: APOBEC1 mRNA editing in bone marrow-derived macrophages

The development of high-throughput sequencing technology has led to the rapid identification of thousands of previously unknown RNA editing events, the majority of which fall within the untranslated regions of target transcripts. Although the functional relevance of most of these newly discovered editing events remain elusive, they point to a broader function for RNA editing than has been previously appreciated. In particular for APOBEC1, an enzyme whose activity was thought to be restricted to a lone target and a single role in lipid metabolism, the recognition that APOBEC1 can edit additional transcripts in enterocytes raises the possibility that its physiological activity may also not be constrained to the intestine. The hypothesis that APOBEC1 may edit transcripts in additional cell types was tested using an adapted comparative RNA-Seq strategy in bone marrow-derived macrophages (BMDMs), a cell-type which expresses APOBEC1 but lacks ACF or ApoB. Abundant APOBEC1 editing was identified in BMDM transcript 3′UTRs, occurring in two editing patterns: site-specific editing and hyper-editing. These editing events represent the first example of physiological APOBEC1 editing outside of the intestine and point to further functions for APOBEC1 in the immune system.

2.1 Identification of APOBEC1 editing events in BMDMs

2.1.1. APOBEC1 expression in BMDMs

While APOBEC1 RNA editing activity has previously been identified only in the small intestine and liver, it is expressed in a diverse set of tissues, including a number of immune cell types (Rosenberg and Papavasiliou, unpublished data). Notably, microarray data has shown that APOBEC1 is expressed in macrophages and that in this cell type expression levels are regulated by LPS stimulation (Mabbott et al., 2010). As these results point to a function for APOBEC1 in macrophages, I reasoned that bone marrow-derived macrophages (BMDMs) were an ideal experimental system with which to explore APOBEC1 editing in the immune system. Bone-marrow precursors were incubated with macrophage-specific cytokines (M-CSF) to generate mature macrophage cells, confirmed by two typical macrophage-specific cell surface markers, F4/80 and CD11b. Notably, *Apobec1*$^{-/-}$ BMDMs have no maturation defects and final cultures are >97% mature macrophages in both wild-type and APOBEC1-deficent littermates (Figure 2.1).

APOBEC1 expression was confirmed by RT-PCR of mRNA derived from wild-type BMDMs. Surprisingly, in contrast to small intestine enterocytes, BMDMs lack APOBEC1's known cofactor, *A1cf*, and its canonical editing target, *ApoB* (Figure 2.2A). This finding was further supported by RNA-Seq performed on wild-type BMDMs; both *A1cf* and *ApoB* transcript expression (represented as FPKM; fragments per kilobase per million reads mapped) were calculated to be zero, while APOBEC1 was expressed with an FPKM of approximately 261 (Figure 2.2B). Additionally, *Apobec1* was confirmed to be differentially expressed

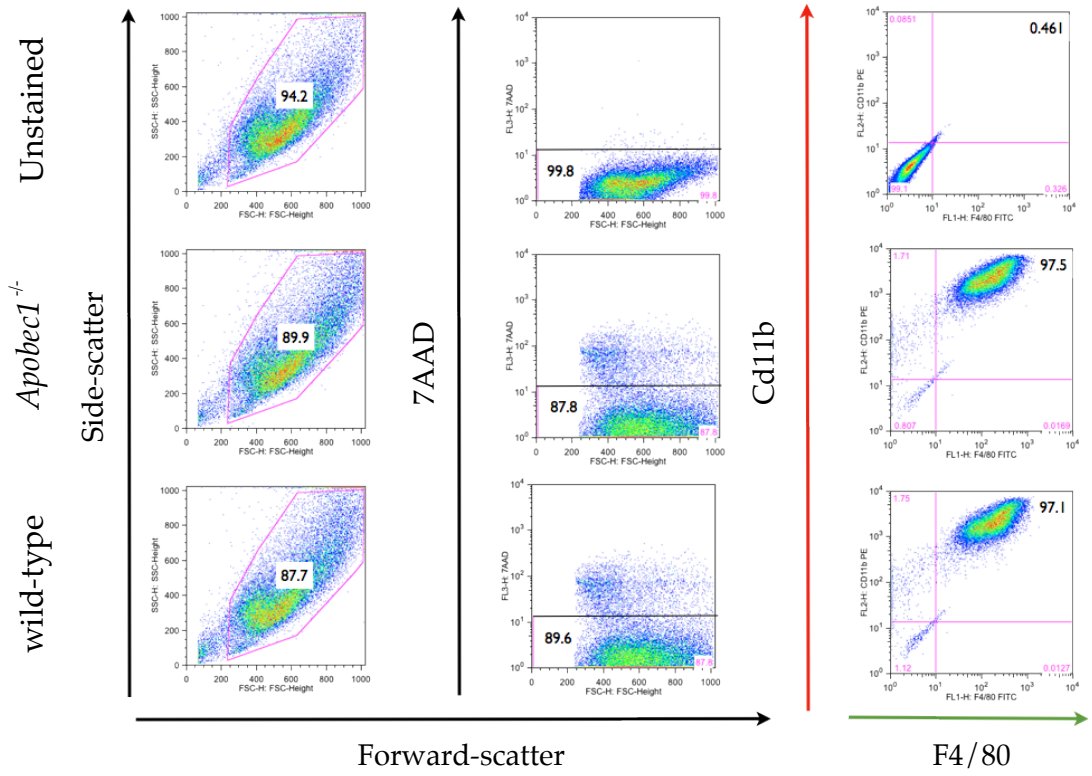
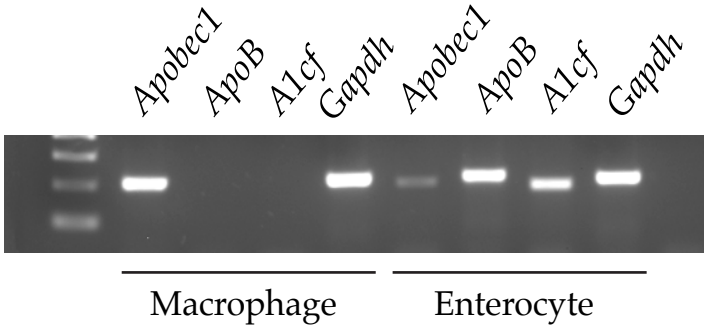


Figure 2.1. Flow cytometry analysis of BMDM maturation. Bone marrow precursors were derived from wild-type and APOBEC1-deficient mice and matured into bone marrow-derived macrophags (BMDMs). The final composition of the "mature" BMDM cultures was confirmed via flow cytometry analysis. Live cells were gated based on forward and side-scatter and lack of 7AAD staining. Mature macrophages were identified with two cell surface markers, Cd11b and F4/80.

A



B

	BMDM wildtype FPKM
Apobec1	261.006
ApoB	0
A1cf	0

C

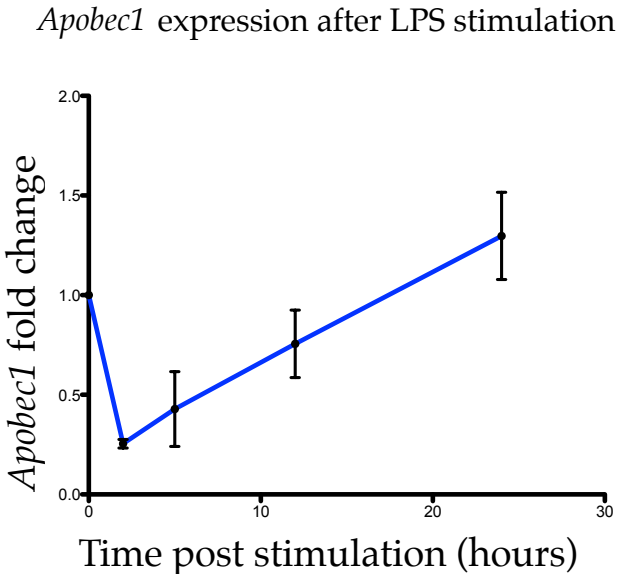


Figure 2.2. APOBEC1 expression in BMDMs. (A) PCR amplification of *Apobec1, A1cf and Apob* cDNA from macrophage and enterocyte mRNA. (B) Transcript expression data for *Apobec1, A1cf and Apob* from polyA$^+$ RNA-Seq as calculated by cuffdiff, representated as fragments per kiolobase per million reads mapped (FPKM). (C) APOBEC1 expression after LPS stimulation as assessed by qRTPCR.

in BMDMs stimulated by LPS (100ng/mL). *Apobec1* transcript levels drop to 20% of baseline after 2 hours of LPS stimulation, and subsequently recover at 24 hours. Together these results further point to a role for APOBEC1 in BMDMs.

2.1.2. RNA-Seq reveals abundant APOBEC1 editing in BMDMs.

The Papavasiliou lab has previously established a comparative RNA-Seq strategy with high specificity for the identification of RNA-editing events, which utilizes an APOBEC1-deficient control to filter single-nucleotide variants (SNVs) for APOBEC1-dependent RNA editing events (Rosenberg et al., 2011b; 2011a). As described in section 1.5.2, this screen revealed 32 additional APOBEC1 editing events in enterocyte transcript 3'UTRs, pointing to further functions for APOBEC1. To investigate the extent of APOBEC1 deaminase activity in murine macrophages, we performed 75nt, single-end RNA-Seq on poly-A^+ RNA from wild-type and *Apobec1*$^{-/-}$ BMDMs, yielding approximately 28 and 33 million reads, respectively. These reads were trimmed and then aligned to the reference genome (mm9). The alignment strategy was designed mindful of the fact that the edited population of reads in the wild-type could contain many mismatches to the reference genome, essential to the downstream editing analysis. With this in mind, moderately permissive alignment parameters were used that allowed for up to 6 mismatches within a 75nt read.

To identify putative APOBEC1-catalyzed deamination events, C-to-T SNVs between the reads and the reference genome were identified. To do this a pileup file was generated using the SAMtools program (Li et al., 2009a). The strategy for identifying these APOBEC1-mediated mismatches differed slightly from previously established protocols. Previously, read to reference mismatches

were determined using the SNP-calling algorithm defined by the SAMtools program during the generation of a pile-up file. But, as RNA editing events can occur in different patterns from genomic SNPs, it was suboptimal to use this SNP-calling algorithm to assign a consensus base. Therefore, process was somewhat simplified and SNVs were identified purely by the ratio of nucleotides present at that position in aligned reads.

These identified APOEC1-specific C-to-T SNVs were further filtered, identifying those that fell within known genes, were covered by at least 20 reads, had at least 20% apparent editing, were not located in regions that were non-isogenic between the mice used, were not significantly strand biased, and were a minimum distance from non-C-to-T SNVs. Finally, we retained only those C-to-T mismatches that did not occur in the *Apobec1*$^{-/-}$ sample, isolating true APOBEC1-dependent events (Figure 2.3). To increase both sensitivity and specificity, we lowered parameters for the APOBEC1-deficient sample, identifying a base as "edited" if it was covered by at least 10 reads and had an editing frequency of at least 5%. Based on this filtering strategy, 110 putative APOBEC1 editing events within 72 transcripts were established (Figure 2.4). A comparable analysis was performed in the *Apobec1*$^{-/-}$ sample as compared to wild-type to estimate an implied false positive rate (IFPR) of <1%.

A subset of editing events were validated by standard Sanger sequencing or subclone sequencing of cDNA and genomic DNA (gDNA) derived from an independent littermate pair of wild-type and *Apobec1*$^{-/-}$ BMDMs (Figure 2.6 - 2.8). We confirmed 42 editing events of 48 screened, for a false positive rate of 12.5%. This false positive rate was moderately higher than the predicted IFPR. A number of factors contributed to this slightly higher FDR including the plasticity

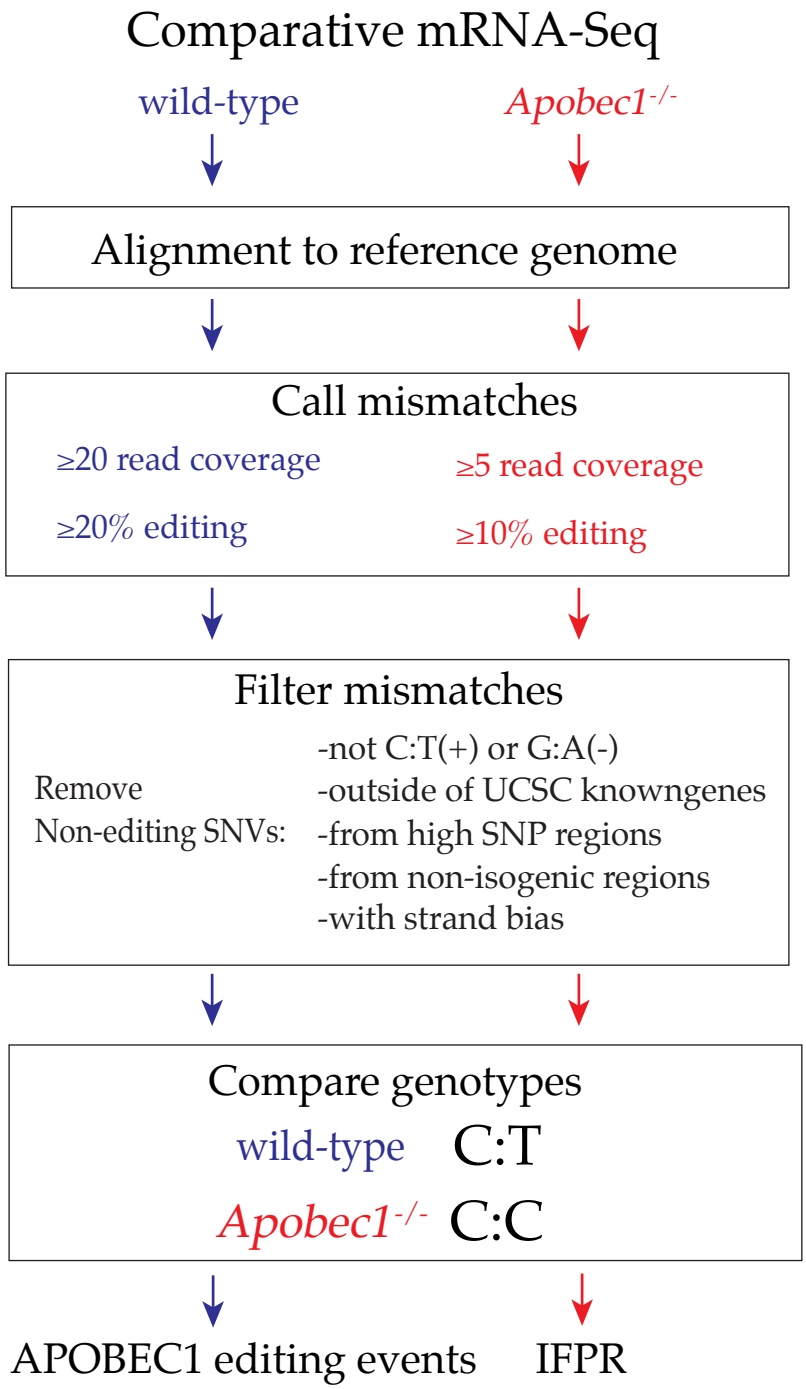


Figure 2.3. Workflow for comparative mRNA-Seq strategy. The strategy for comparative RNA-Seq read alignment and SNV filtering is depicted here. After filtering, C-to-T mismatches identified in the sample wildtype sample that are absent from the APOBEC1-deficient BMDMs are designated as *bona fide* APOBEC1 editing events. The converse, C-to-T mismatches in the APOBEC1-deficient sample that are absent from wild-type yields an implied false positive rate (IFPR).

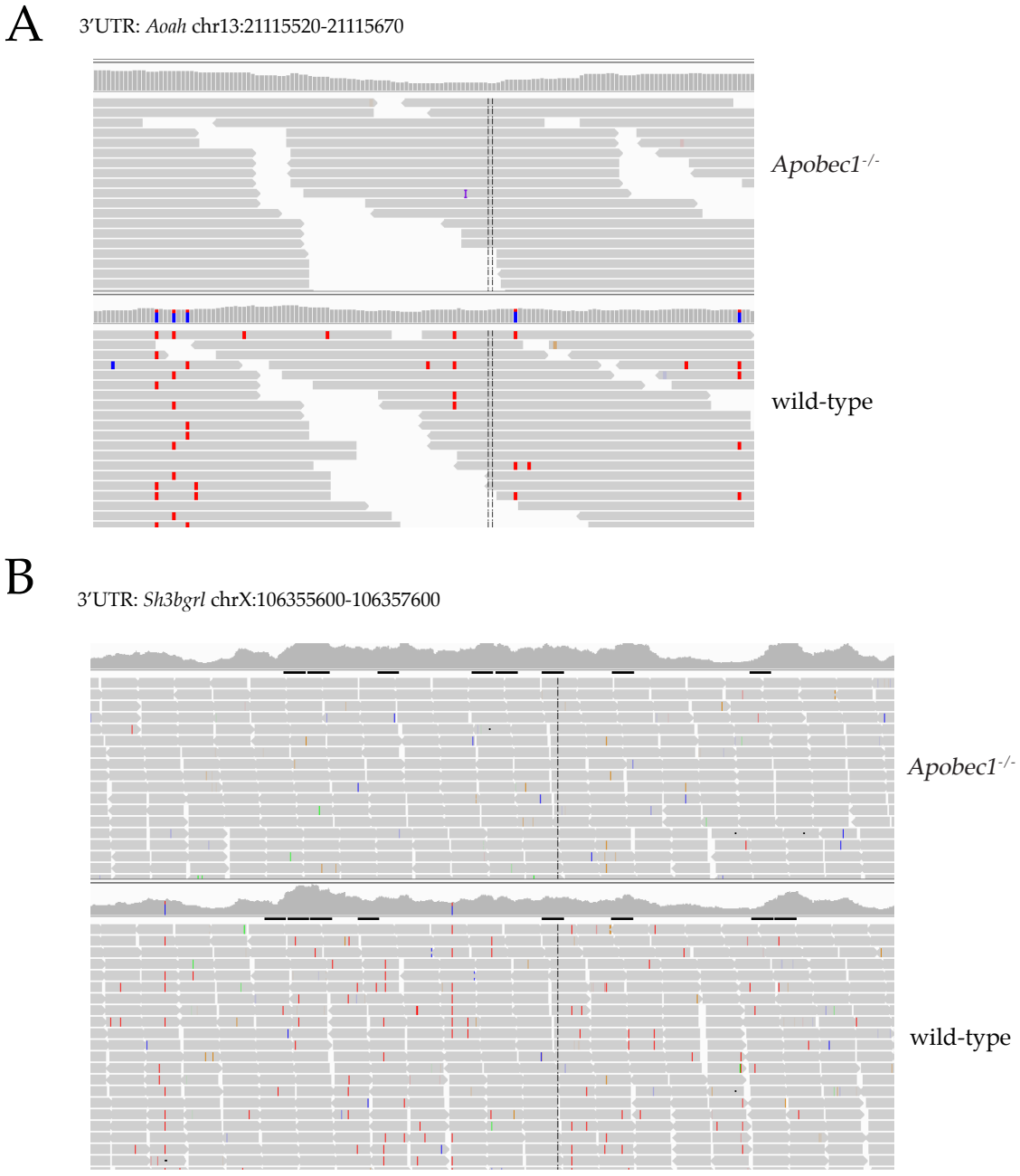


Figure 2.4. Examples of APOBEC1 editing events identified by RNA-Seq. (A/B) RNA-Seq reads from wildtype and *Apobec1*$^{-/-}$ libraries covering the 3′UTRs of *Aoah* (A) and *Sh3bgrl* (B) as visualized by the Integrative Genome Viewer. Nucleotides which match the reference genomic sequence are represented in grey and nucleotide mismatches to the reference genome are represented by a variety of colors with T mismatches in red.

of the APOBEC1 editing repertoire within each transcript, the elimination of a genomic SNP filter and the moderately permissive read depth cutoffs. After validation, we compiled a list of 104 high-confidence APOBEC1 editing events within 68 transcript 3′UTRs (Figure 2.5;Table 2.1), representing the first example of physiological APOBEC1 editing outside of the intestine.

2.1.3 Two patterns of APOBEC1 editing: site-specific and hyper-editing

In contrast to physiological APOBEC1 RNA editing in enterocytes, we observed two distinct APOBEC1 editing patterns in BMDMs (Figure 2.5). In site-specific editing, a single cytosine is edited consistently at high frequency, with negligible frequencies of editing events surrounding the targeted base. Site-specific editing, comparable to previously established APOBEC1 editing in *ApoB*, was observed in 54 transcripts 3′UTRs (containing 60 identified editing events) (Figure 2.5-2.7). In hyper-edited transcripts the 3′UTR contained multiple (2-8) high frequency editing events (editing "hotspots") and many scattered low-frequency editing events surrounding each "hotspot" base. APOBEC1-dependent hyper-editing was observed in 14 APOBEC1-targeted transcripts (containing 44 identified editing events) (Figures 2.4, 2.8, 2.9). Some "promiscuous" APOBEC1 hyper-editing activity has been previously observed in transgenic over-expressing mice or a viral infection model, but never in a physiological system. Therefore, in addition to identifying abundant APOBEC1 editing in a novel cell type, this represents a previously uncharacterized physiological editing modality for APOBEC1.

Table 2.1. APOBEC1 editing events in BMDMs

Gene name	Edited Site	Ref base	Read base	Read depth	Strand	Edit Frequency	Hyperedited
1110002B05Rik	chr12:55747165	G	R	127	-	0.43	no
2010106G01Rik	chr2:126716570	G	R	46	-	0.35	no
2210013O21Rik	chrX:150163433	C	Y	42	+	0.21	no
2610001J05Rik	chr6:13819970	G	R	65	-	0.25	no
4930579G24Rik	chr3:79436318	C	Y	22	+	0.23	no
6330578E17Rik	chr1:37475131	G	R	30	-	0.20	no
Acadl	chr1:66877651	G	R	20	-	0.20	no
ADAM10	chr9:70626825	C	Y	62	+	0.21	no
Anxa5	chr3:36347952	G	R	594	-	0.31	no
Aoah	chr13:21115534	C	Y	24	+	0.25	yes
Aoah	chr13:21115538	C	Y	23	+	0.30	yes
Aoah	chr13:21115541	C	Y	23	+	0.22	yes
Aoah	chr13:21115616	C	Y	24	+	0.21	yes
Aoah	chr13:21115667	C	Y	23	+	0.22	yes
App	chr16:84954725	G	R	50	-	0.20	yes
App	chr16:84954758	G	R	133	-	0.46	yes
App	chr16:84955024	G	R	304	-	0.24	yes
App	chr16:84955039	G	R	257	-	0.20	yes
App	chr16:84955062	G	R	241	-	0.21	yes
App	chr16:84955086	G	R	254	-	0.23	yes
App	chr16:84955113	G	R	311	-	0.32	yes
App	chr16:84955194	G	R	265	-	0.28	yes
Arih1	chr9:59239901	G	R	22	-	0.23	no
Atp6ap2	chrX:12193513	C	Y	157	+	0.27	yes
Atp6ap2	chrX:12193524	C	Y	149	+	0.23	yes
Atp6ap2	chrX:12193607	C	Y	126	+	0.40	yes
Atp6v1a	chr16:44087436	G	R	143	-	0.23	no
B2m	chr2:121978476	C	Y	2350	+	0.51	yes
B2m	chr2:121978523	C	Y	1863	+	0.28	yes
B2m	chr2:121978638	C	Y	1770	+	0.54	yes
BC013529	chr10:7487994	G	R	20	-	0.25	no
Bcap31	chrX:70931693	G	R	112	-	0.21	yes
Bcap31	chrX:70931742	G	R	127	-	0.20	yes
Bcap31	chrX:70931744	G	R	126	-	0.31	yes
Casp6	chr3:129616676	C	Y	28	+	0.29	no
Ccni	chr5:93611225	G	R	51	-	0.27	no
Cd36	chr5:17288955	G	R	47	-	0.77	no
Cybb	chrX:9012852	G	R	196	-	0.29	yes
Cybb	chrX:9013719	G	R	260	-	0.22	yes
Dpp8	chr9:64929147	C	Y	36	+	0.25	no
Dpp8	chr9:64930384	C	Y	29	+	0.29	no
Dynlt3	chrX:9232850	G	R	23	-	0.35	yes

Table 2.1. Continued. APOBEC1 editing events in BMDMs

Gene name	Edited Site	Ref base	Read base	Read depth	Strand	Edit Frequency	Hyperedited
Dynlt3	chrX:9233008	G	R	31	-	0.26	yes
Eif4a2	chr16:23113791	C	Y	20	+	0.20	no
Entpd5	chr12:85716971	G	R	20	-	0.20	no
Epsti1	chr14:78402303	C	Y	25	+	0.24	no
Fkbp1a	chr2:151386579	C	Y	34	+	0.21	yes
Fkbp1a	chr2:151387062	C	Y	22	+	0.32	yes
Fuca2	chr10:13235355	C	Y	20	+	0.25	no
Gdbd1	chr11:86847804	G	R	23	-	0.26	no
Hibadh	chr6:52496349	G	R	38	-	0.24	no
Hmgn3	chr9:83003686	G	R	22	-	0.23	no
Impa1	chr3:10314195	G	R	22	-	0.23	no
Itgb2	chr10:77028321	C	Y	200	+	0.23	yes
Itgb2	chr10:77028356	C	Y	144	+	0.38	yes
Lamp1	chr8:13174686	C	Y	368	+	0.21	yes
Lamp1	chr8:13174689	C	Y	363	+	0.29	yes
Lamp1	chr8:13174696	C	Y	418	+	0.32	yes
Lamp1	chr8:13174720	C	Y	383	+	0.29	yes
Lamp1	chr8:13174732	C	Y	399	+	0.39	yes
Lamp2	chrX:35774512	G	R	67	-	0.27	no
Lamp2	chrX:35774742	G	R	81	-	0.30	no
Lypla1	chr1:4836242	C	Y	37	+	0.30	no
Mbnl1	chr3:60432599	C	Y	147	+	0.20	no
mcmbp	chr7:135841366	G	R	25	-	0.32	no
Mmd	chr11:90139728	C	Y	31	+	0.29	no
Mospd2	chrX:161374490	G	R	23	-	0.43	no
Nptn	chr9:58500000	C	Y	150	+	0.36	yes
Nptn	chr9:58500149	C	Y	162	+	0.20	yes
Ola1	chr2:72931513	G	R	27	-	0.22	no
Ola1	chr2:72931552	G	R	22	-	0.23	no
Papss1	chr3:131306342	C	Y	24	+	0.21	no
Papss1	chr3:131306346	C	Y	23	+	0.22	no
Pla2g7	chr17:43749067	C	Y	86	+	0.24	no
Prkacb	chr3:146395143	G	R	36	-	0.22	no
Ptma	chr1:88427075	C	Y	43	+	0.23	no
Rab18	chr18:6789843	C	Y	25	+	0.28	no
Rac1	chr5:144266732	G	R	207	-	0.29	no
Reep5	chr18:34506411	G	R	44	-	0.21	no
Rpl15	chr14:19100950	G	R	32	-	0.25	no
Rpl15	chr14:19101348	G	R	21	-	0.29	no
Sdcbp	chr4:6322479	C	Y	45	+	0.56	yes
Sdcbp	chr4:6322512	C	Y	85	+	0.64	yes
Sep15	chr3:144259976	C	Y	286	+	0.37	yes

Table 2.1. Continued. APOBEC1 editing events in BMDMs

Gene name	Edited Site	Ref base	Read base	Read depth	Strand	Edit Frequency	Hyperedited
Sep15	chr3:144260406	C	Y	244	+	0.21	yes
Serinc1	chr10:57235791	G	R	62	-	0.81	no
Serinc3	chr2:163450893	G	R	109	-	0.23	no
Sgk3	chr1:9889105	C	Y	30	+	0.37	no
Sgk3	chr1:9889140	C	Y	21	+	0.29	no
Sh3bgrl	chrX:106355759	C	Y	112	+	0.28	yes
Sh3bgrl	chrX:106356391	C	Y	101	+	0.28	yes
Sh3bgrl	chrX:106357513	C	Y	182	+	0.25	yes
Spcs2	chr7:106987604	G	R	31	-	0.26	no
Spp1	chr5:104869859	C	Y	3284	+	0.22	no
Srgn	chr10:61957357	G	R	47	-	0.21	no
St8sia4	chr1:97484976	G	R	20	-	0.25	no
Syap1	chrX:159295116	G	R	20	-	0.20	no
Sypl	chr12:33661093	C	Y	45	+	0.20	no
Tes	chr6:17055467	C	Y	40	+	0.30	no
Tmed7	chr18:46747592	G	R	26	-	0.31	no
Tmem30a	chr9:79617629	G	R	28	-	0.61	no
Tmem55a	chr4:14841457	C	Y	20	+	0.35	no
Tspan3	chr9:55983987	G	R	34	-	0.26	no
Zfp871	chr17:32906699	G	R	20	-	0.21	no

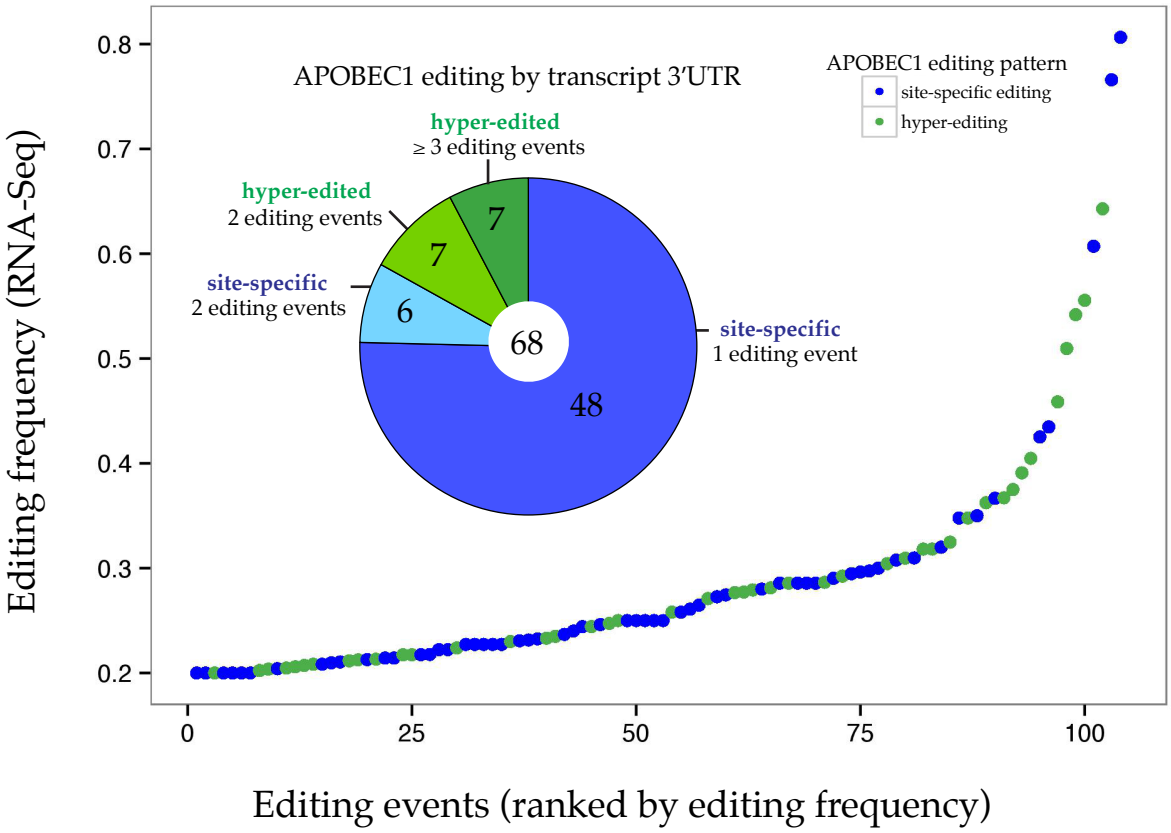


Figure 2.5. APOBEC1 site-specific editing and hyper-editing. Editing frequency of single-site edited (blue) and hyper-edited (green) sites. Inset: APOBEC1-targeted transcripts represented by editing pattern and number of events per transcript.

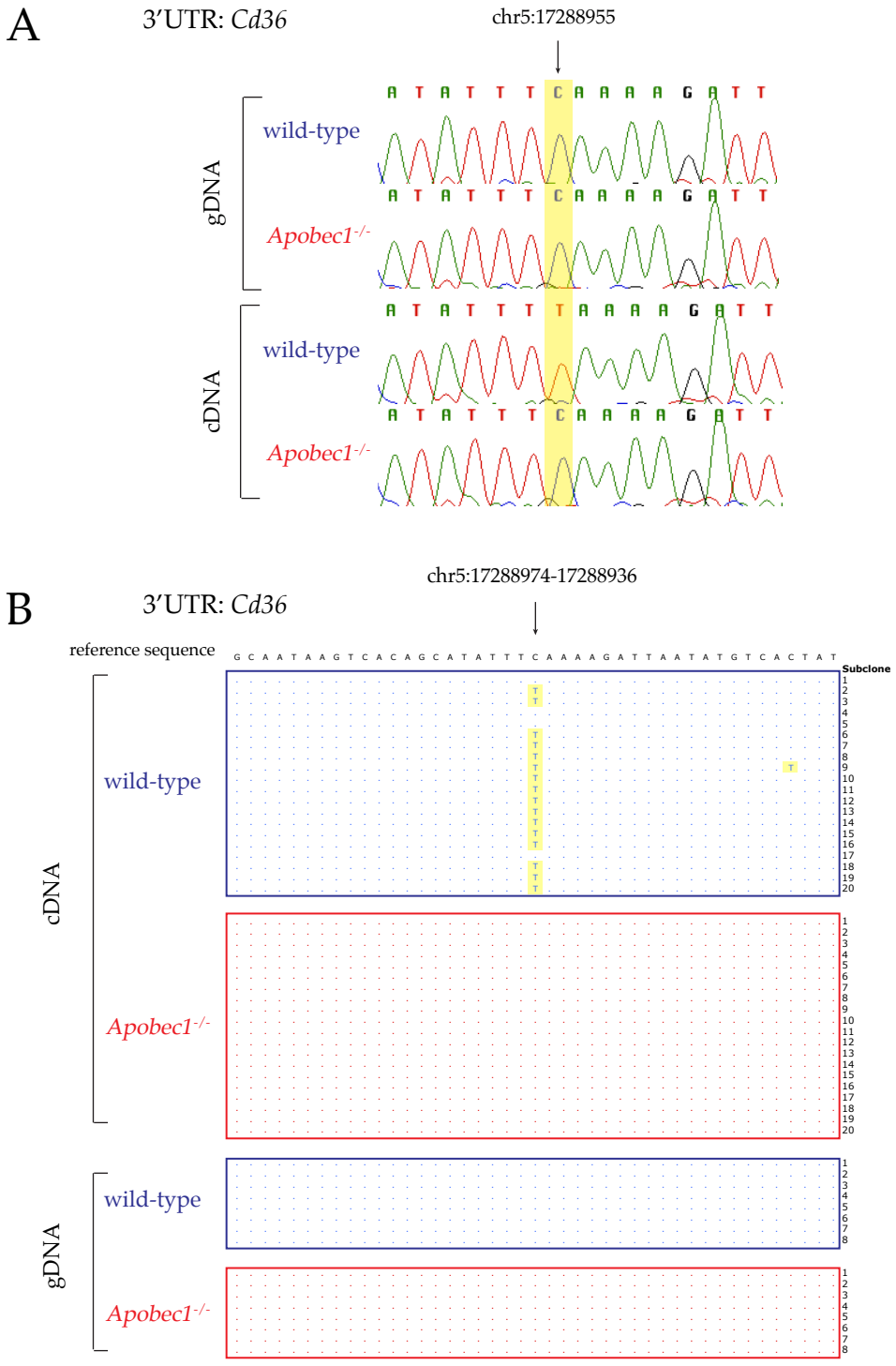


Figure 2.6 Representative example of APOBEC1 site-specific editing. (A) Sanger sequencing and (B) subclone sequencing of the *Cd36* transcript 3′UTR in BMDM wild-type and *Apobec1*$^{-/-}$ genomic DNA (gDNA) and cDNA

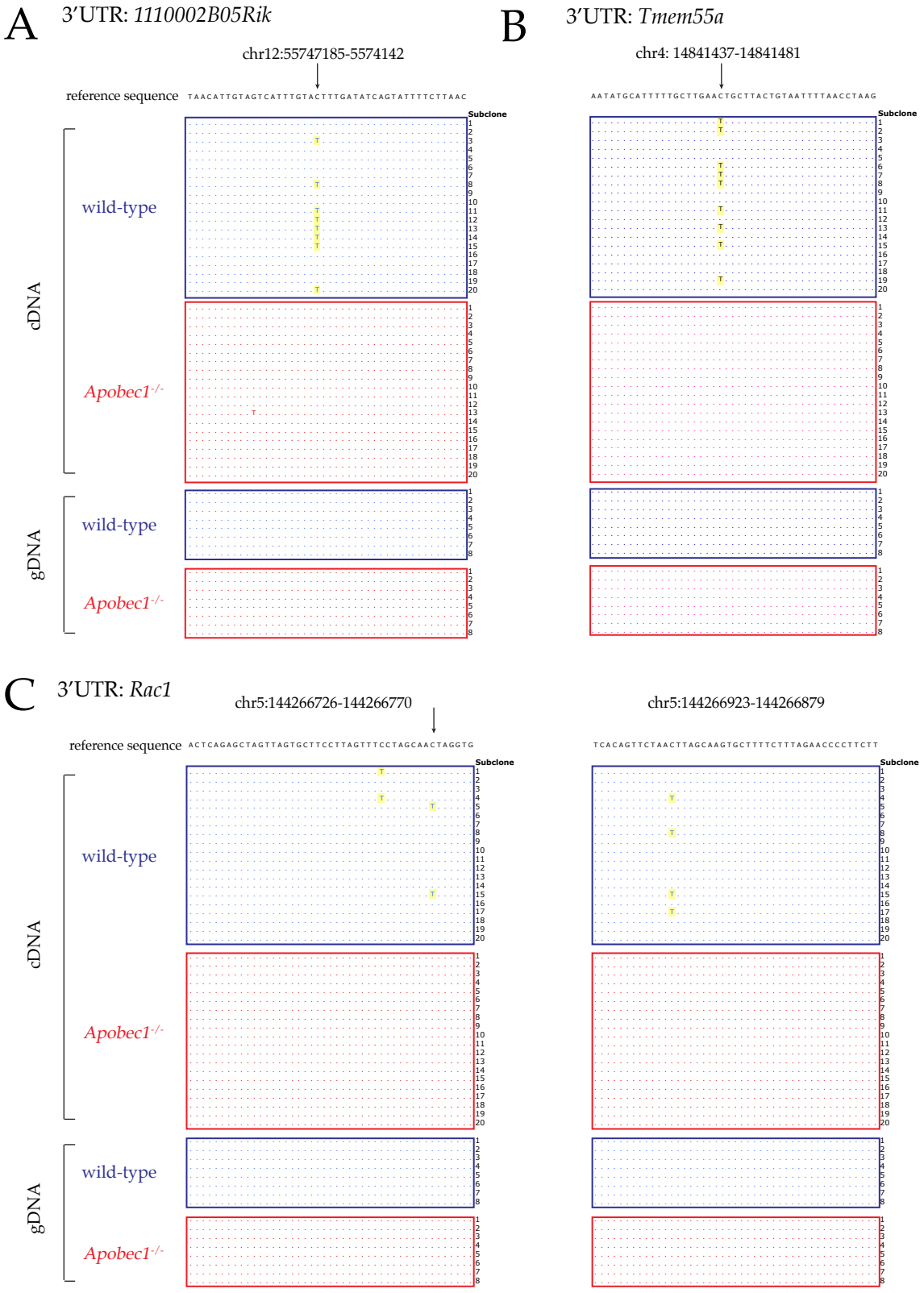


Figure 2.7. Additional examples of APOBEC1-mediated site-specific editing. Subclone sequencing of (C) *1110002B05Rik* (D) *Tmem55a* (E) *Rac1* 3′UTRs.

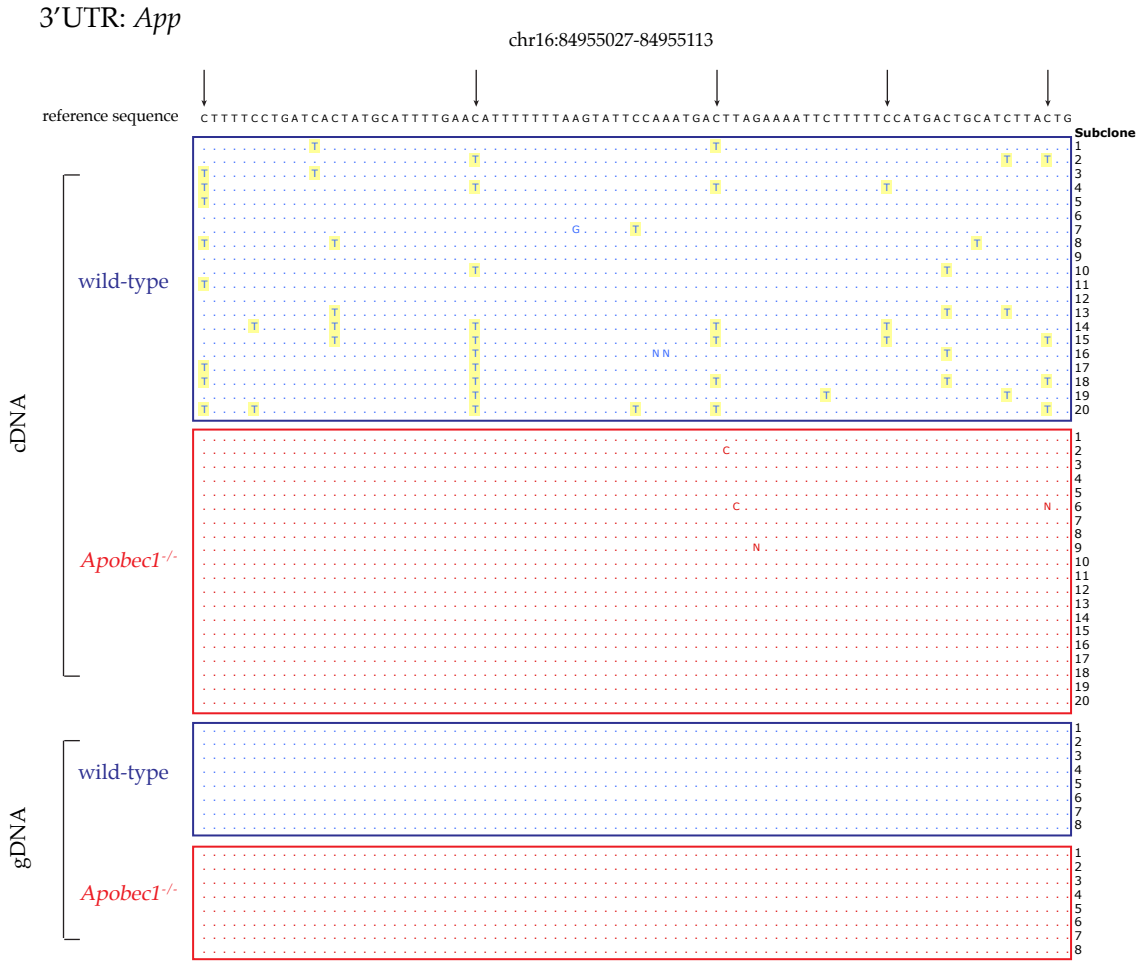


Figure 2.8. Representative example of APOBEC1 hyper-editing. Subclone sequences from the cDNA of wildtype and *Apobec1*$^{-/-}$*App* transcripts aligned to the reference genomic sequence. C-to-T mismatches to reference are highlighted in yellow and editing events identified by RNA-Seq screen are indicated by arrows.

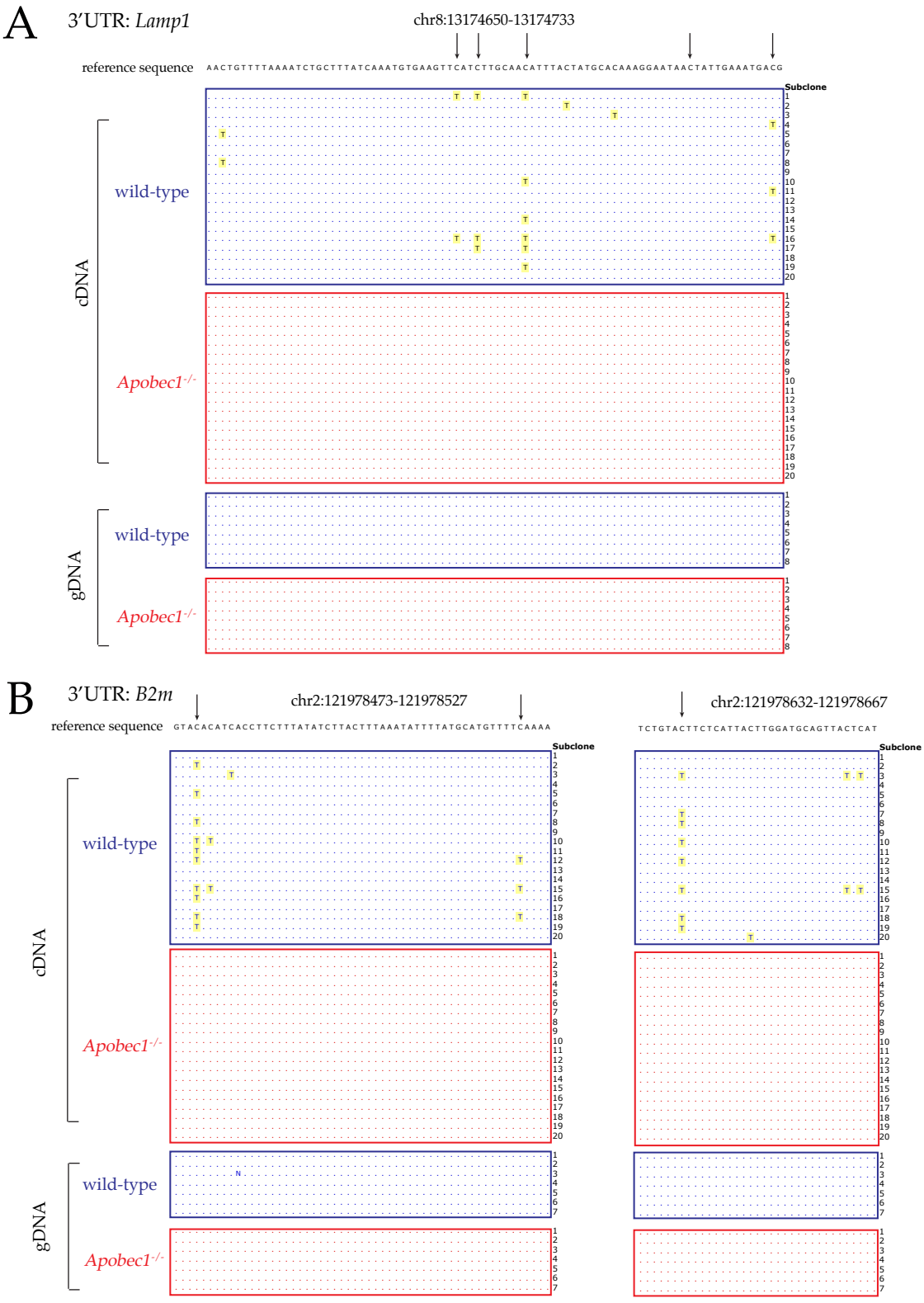


Figure 2.9. Additional examples of APOBEC1-mediated hyper-editing. Subclone sequencing of (A) *Lamp1* and (B) *B2m* transcript 3'UTRs.

2.2 Characteristics of APOBEC1 editing events in BMDMs

High-frequency APOBEC1 editing events had been identified in a cell type that lacks the co-factor ACF, suggesting that APOBEC1 could be editing in a previously uncharacterized mechanism. Analysis of APOBEC1 intestinal editing events has revealed sequence characteristics for editing that include A or U flanking nucleotides (Rosenberg et al., 2011b) and a downstream mooring motif (Backus and Smith, 1991; Rosenberg et al., 2011b; Shah et al., 1991). In *ApoB* editing, ACF binds to the mooring sequence and directs APOBEC1 to C6666 with high fidelity. Intestinal 3′UTR editing events are largely associated with a similar mooring motif, suggesting that ACF binding to the target transcript may also be required. To gain insight into mechanistic difference for APOBEC1 editing without ACF, I performed analyses to determine whether macrophage-specific APOBEC1 editing events shared the features of 3′UTR editing in the intestine.

2.2.1 Adjacent nucleotide analysis

Other members of the AID/APOBEC family have been shown to have preferences for the nucleotides neighboring the targeted cytosine. In small intestine enterocytes, APOBEC1 was found to exhibit a strong preference for A or U nucleotides immediately flanking the edited C (Rosenberg et al., 2011b). Analysis of the flanking nucleotides in BMDM edited transcripts revealed that again the edited cytosine was flanked by A or U more often than would be expected by chance (Figure 2.10). This result suggests that factors leading to a flanking nucleotide preference remain consistent in ACF-independent APOBEC1 editing in BMDMs.

Figure 2.10. APOBEC1 adjacent nucleotide preferences. Sequence log depicting the frequency of nucleotides in positions flanking the APOBEC1-targeted cytosine. The height of the nucleotide depicts the relative frequency of each base at the specified position.

2.2.2 Mooring sequence analysis

A downstream 11-nucleotide APOBEC1 mooring sequence was established in the intestine to be essential for *ApoB* editing (Shah et al., 1991) and was found to be strongly associated with 3′UTR editing (Rosenberg et al., 2011b). The sequence of the mooring sequence and the further refined "mooring motif" are depicted in Figure 2.12A. To further investigate features of ACF-independent APOBEC1 editing, the relationship of macrophage-specific APOBEC1 editing events with this mooring sequence was investigated. In contrast to enterocyte editing, only 33 APOBEC1-targeted transcripts (43% of hyper-edited transcripts and 54% of site-specific edited transcripts) were associated with a high-quality mooring sequence, defined as ≤ 2 deviations from the established mooring motif (Figure 2.11).

To determine the correlation between mooring sequence strength and editing frequency, a mooring motif scoring system was defined based on sequence quality and position of the mooring motif in relation to the targeted cytosine (Figure 2.12A). For all APOBEC1 editing events identified, mooring sequence score was only weakly correlated (Pearson's correlation coefficient r = 0.39) with editing frequency (Figure 2.12B). However, hyper-editing events could skew these results, as only one mooring sequence may be associated with many editing events in one transcript 3′UTR. A comparable analysis limited to site-specific editing events yielded only a slightly higher correlation (r = 0.55) (Figure 2.12B). In hyper-edited transcripts, a very poor correlation was found between number of high-frequency edits (≥ 20%) in each transcript and mooring sequence stringency of the best mooring motif (r = 0.17) (Figure 2.12C). These data

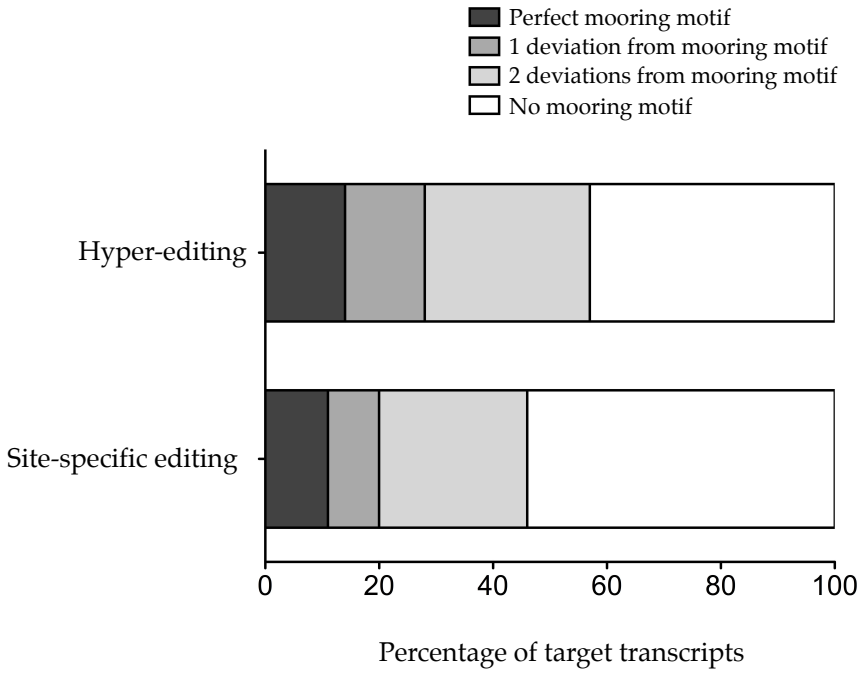


Figure 2.11. Presence of mooring motif in APOBEC1-targeted transcripts. Quality of the mooring motif in hyper-edited and single-site edited transcripts. Mooring motifs are grouped by the number of mismatches (0, 1, 2, > 2) to the established motif and represented as a percentage of the total number of hyper-edited or site-specific editing transcripts. In transcripts with more than one edited C and more than one mooring sequence, the best motif was chosen.

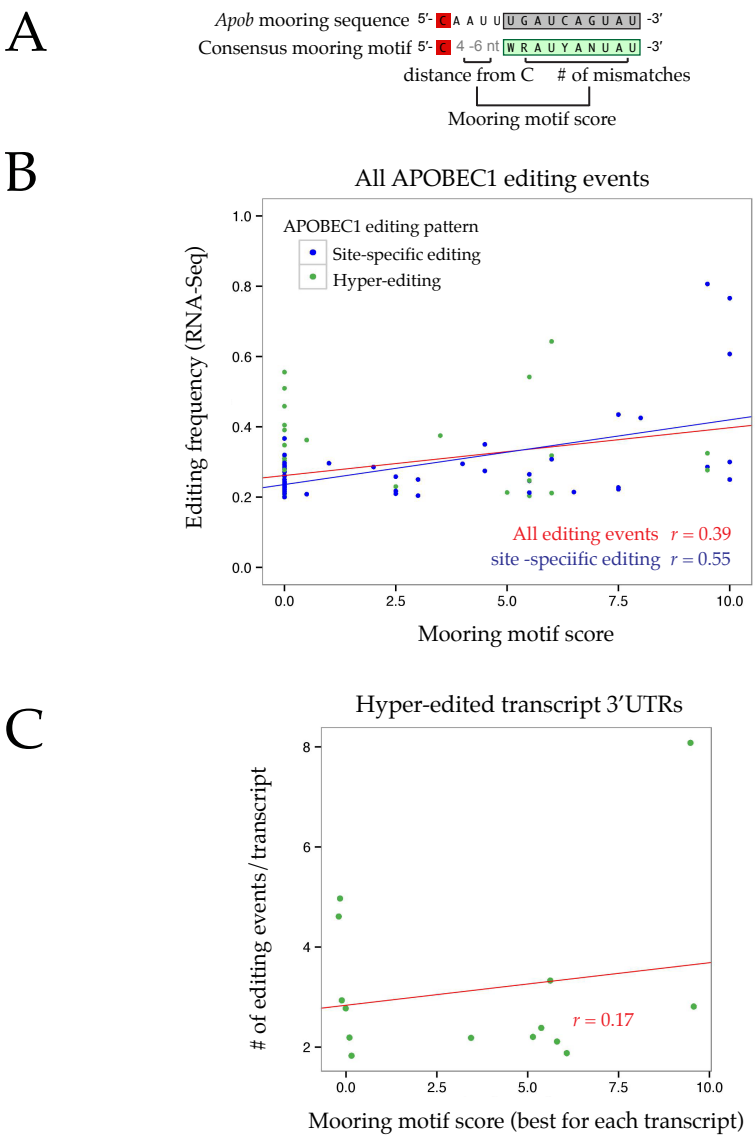


Figure 2.12. APOBEC1 editing is weakly associated with mooring motifs. (A) Schematic of APOBEC1 mooring sequence essential for ApoB editing of C6666 (grey) and consensus mooring motif established for APOBEC1 3'UTR editing in enterocytes (green). Edited cytidine is represented in red. The mooring sequence stringency score was based on the distance of the mooring sequence and the quality of the mooring motif. (B) APOBEC1 editing events (site-specific, blue; hyper-edited, green) plotted as a factor of the mooring sequence score and editing frequency. Best-fit lines were plotted for all editing events (red) and only site-specific events (blue) (r = Pearson's correlation coefficient). (C) The mooring motif score for hyper-edited transcripts is plotted against the number of APOBEC1 editing events/transcript. The "best" mooring motif score was used for transcripts with multiple motifs.

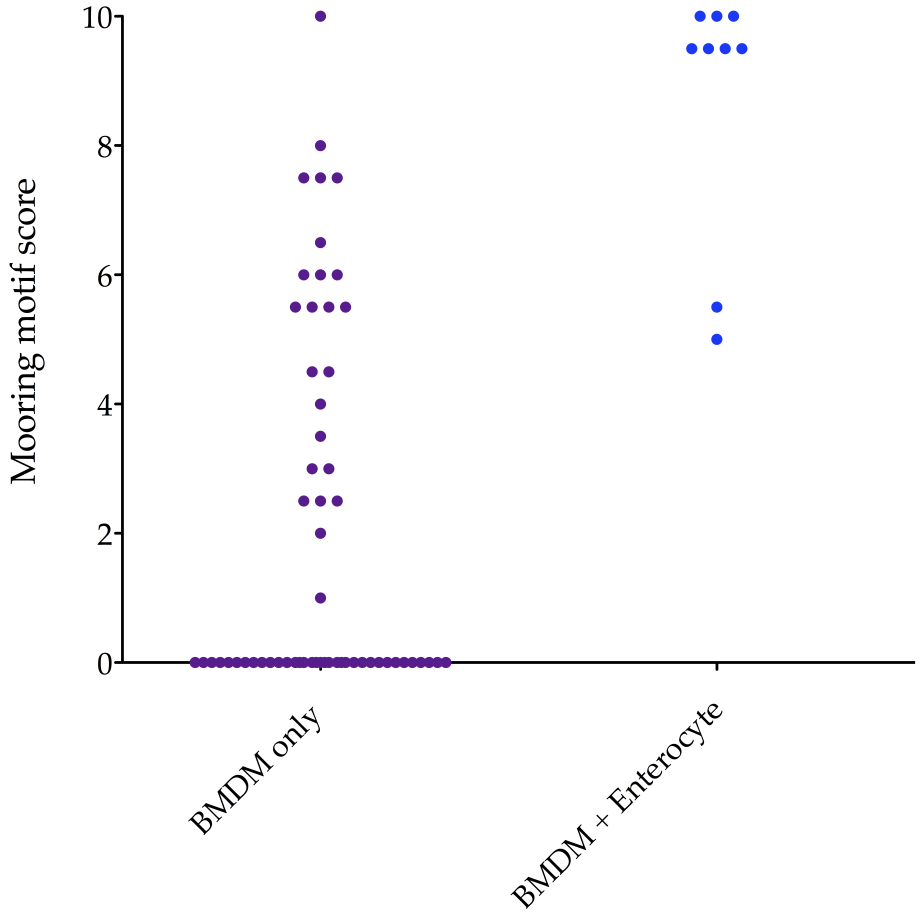


Figure 2.13. Mooring sequence scores in BMDMs vs enterocyte editing targets. Edited transcripts were split into those only edited in BMDMs (purple) and those with editing events identified by RNA-Seq in both small-intestine enterocytes and BMDMs (blue). Each transcript is plotted for the score of its mooring sequence. A score of 0 denotes the absence of a high-quality mooring sequence. A score of 10 indicates a mooring sequence with no deviations from the established motif that occurs 4-6nt downstream of the edited cytosine.

demonstrate that ACF-independent APOBEC1 editing in BMDMs is occasionally associated with, but not absolutely dependent on, a downstream mooring sequence. Furthermore, this suggests that the requirements for APOBEC1-dependent RNA editing in BMDMs may be substantially different than previously established for *ApoB* and points to an alternative editing requirements for these ACF-independent editing events.

2.2.3 Comparison of APOBEC1 intestine and BMDM editing

The editing of C6666 in the *ApoB* transcript remains the only example of true APOBEC1 editing in a coding region, but some APOBEC1 transcript 3′UTR substrates displayed similar editing in both enterocytes and BMDMs. In isolation, this result is not surprising; *cis*-acting sequence requirements for APOBEC1 editing in preferred substrates could "mark" the transcript for editing. However, one of these sequence requirements is the binding motif for a co-factor not expressed in BMDMs. As discussed, while half of the transcripts edited in BMDMs lack a downstream mooring sequence, some transcripts are associated with high quality mooring motifs. However, when the transcripts are grouped into those edited only in BMDMs and those also edited in the intestine, transcripts edited in both cell types tend to have higher mooring sequence scores (Figure 2.13). This finding points to additional uncharacterized sequence requirements or an alternative co-factor involved in APOBEC1-mediated transcript 3′UTR editing.

Chapter 3: Consequences of APOBEC1 editing events in 3'UTRs

Transcriptome-wide high-throughput sequencing has revealed thousands of previously uncharacterized ADAR and APOBEC1 RNA editing events in transcript 3'UTRs. The functional relevance of this untranslated region editing has remained largely unknown. Specific examples of ADAR editing in 3'UTRs have been shown to modulate genes expression through nuclear retention of ADAR target transcripts (Chen et al., 2008), mRNA cleavage (Osenberg et al., 2009) or potential modification of miRNA target regions (Borchert et al., 2009; Liang and Landweber, 2007). For APOBEC1, the identification of additional editing events in enterocyte transcript 3'UTRs pointed to a broader role for the enzyme than had been previously appreciated. These additional editing events intestine enterocytes were shown to occur predominantly in evolutionarily conserved regions, suggesting that untranslated APOBEC1 may have functional importance. Furthermore, the majority of APOBEC1 editing events were shown to occur in sequences matching miRNA seed targets, suggesting that APOBEC1 editing could alter gene expression by influencing miRNA targeting. However, as functional testing of APOBEC1 editing in enterocytes was complicated by APOBEC1's role in lipid metabolism, the specific consequences of APOBEC1 editing in 3'UTRs remained elusive.

Chapter 2 described the identification of abundant APOBEC1 editing events in BMDMs, the first example of physiological APOBEC1 editing outside of the digestive system. BMDMs are a preferable experimental system to test functional relevance of APOBEC1 editing in transcript 3'UTRs as:

1- BMDMs are a robust primary cell type, providing both a physiological milieu and abundant, easily culturable material.

2- APOBEC1-deficient BMDMs are unaffected by the lipid defects that afflict *Apobec1*$^{-/-}$ small intestine enterocytes due to the loss of apoB-48. These apoB-mediated effects complicate analysis of small-intestinal enterocyte editing where it is impossible to differentiate apoB-mediated defects from those caused by alternative APOBEC1 editing events.

3- As discussed, there is significant evidence that APOBEC1 may function in the immune system. As BMDMs have abundant APOBEC1 enzymatic activity, this cell type represents an appropriate system to assess roles for APOBEC1 in immunity.

Therefore, BMDM APOBEC1 editing events in 3'UTRs were assessed for consequences for transcript expression, gene expression and cell function. As a set, APOBEC1 editing events were found in evolutionarily conserved regions. Although transcripts found to be edited by APOBEC1 had no significant differences in overall protein expression between wild-type and APOBEC1-deficient mice, certain editing events were shown to repress protein expression in a luciferase reporter assay. To test whether this protein repression was due to changes in miRNA processing, HITS-CLIP for the Argonaute (Ago) proteins was performed on wild-type and APOBEC1-deficient BMDMs. Although Ago and APOBEC1 seemed to target similar transcript substrates, there was little evidence for APOBEC1 altering miRNA binding. Overall, in BMDMs APOBEC1 editing

events in 3′UTRs can modulate protein expression in a miRNA-independent mechanism.

3.1. BMDM APOBEC1 editing occurs in evolutionarily conserved regions.

Intestinal APOBEC1-dependent 3′UTR editing, although located in a non-coding region of the transcript, was identified be largely located in regions of substantial phylogenic conservation, implying functional importance to the cell (Rosenberg et al., 2011b). In BMDMs, transcript 3′UTRs have been identified which contain such a large number of C-to-T editing events; the composite edited sequence represents a dramatic change from the genomically-encoded DNA sequence. Such abundant editing if located in functional regulatory regions could have dramatic downstream consequences. To assess the conservation of BMDM APOBEC1 editing, the PhastCon scores of 101nt windows surrounding APOBEC1-edited cytosines were compared to random sets of 101nt windows located in 3′UTRs. On average, edited windows were significantly ($p < 0.01$, student's t-test) more conserved than the random windows (Figure 3.1), suggesting that sequences edited by APOBEC1 may be functionally relevant.

3.2. Consequences of APOBEC1 editing: luciferase reporter assay.

As APOBEC1 editing events occur in such conserved regions of transcript 3′UTRs, I sought to determine whether APOBEC1 editing events in non-coding 3′UTRs can modulate transcript and protein expression. To assess the effects of APOBEC1 on the transcriptome, transcript expression profiles generated from RNA-Seq were analyzed, and transcript expression was highly correlated between the two genotypes ($r > 0.99$) (Figure 3.2). A small subset of unedited

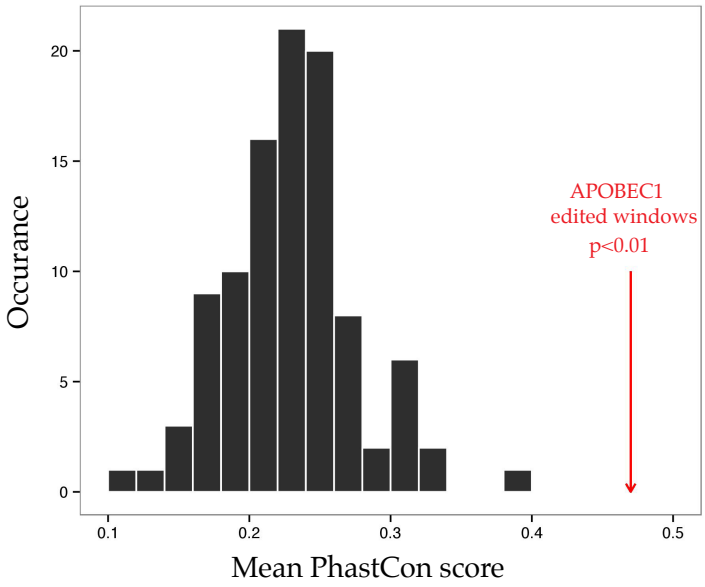


Figure 3.1. APOBEC1 editing events occur in conserved regions of 3'UTRs. Mean PhastCon score of 100 sets of 68 random 101nt windows in 3'UTRs as compared to the mean PhastCon score of the 68 101nt windows surrounding APOBEC1 editing sites (red arrow).

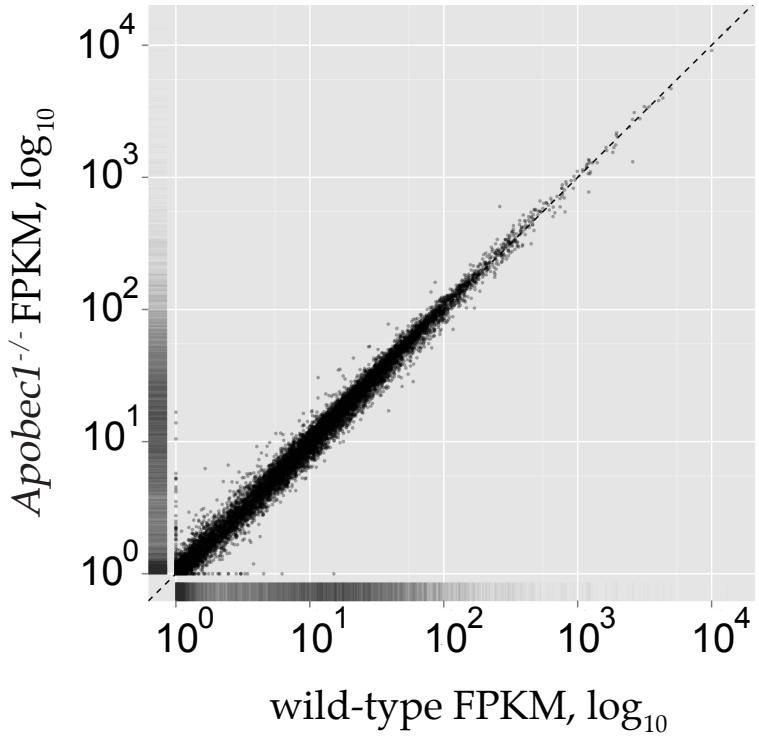


Figure 3.2. Gene expression profiles for wild-type and *Apobec1*$^{-/-}$ BMDMs. Comparison of transcript expression levels as calculated by cuffdiff (represented as FPKM, fragment per kilobase per million reads mapped) from RNA-Seq for wild-type and *Apobec1*$^{-/-}$ BMDMs.

transcripts were differentially expressed (Table 3.1), indicating that while APOBEC1 editing could be influencing transcript levels of a few downstream targets, there were no quantifiable changes in expression of edited transcripts.

However, the effects of transcript regulatory processes may only be appreciable at the protein level. For example, translational repression has been shown to be the principal mechanism of miRNA-mediated gene expression changes and required for subsequent mRNA degradation (Meijer et al., 2013). Therefore, to directly assess the consequences of APOBEC1 editing events on protein expression, a standard luciferase reporter assay was utilized. APOBEC1-targeted transcript 3′UTRs of interest were amplified from wild-type BMDM cDNA and a selection of clones with representative levels of C-to-U alterations (either single-site edited or with increasing amounts of hyper-editing) were cloned into a dual-luciferase reporter construct. "Edited" (with C-to-U change) and "unedited" (with no C-to-U change) luciferase constructs were transfected directly into *Apobec1*$^{-/-}$ BMDMs and the relative change in luciferase expression between "edited" and "unedited" constructs was determined (Figure 3.3A).

A number of these constructs containing C-to-U changes showed significant repression of luciferase levels (Figure 3.3B, 3.4), suggesting that APOBEC1 editing events can occur in regions important to transcript regulation. Of the single-site "edited" constructs tested, only the C-to-U change in *Cd36* modulated luciferase levels (Figure 3.3B). However, the "hyper-edited" constructs tested (*App*, *Lamp1*, *B2m)* had a least one clone with a specific repertoire of C-to-U changes that reduces protein expression levels in a luciferase reporter assay, suggesting that hyper-edited events were more likely to alter translational outcomes of targeted transcripts (Figure 3.4). Importantly, these

Table 3.1. Genes differentially expressed between wild-type and *Apobec1*$^{-/-}$ BMDMs. Transcriptome-wide mRNA expression data (FPKM; fragments per kilobase per million reads mapped) was obtained from RNA-Seq of wild-type and *Apobec1*$^{-/-}$ BMDMs. Genes listed have significantly lower (top) or higher (bottom) expression in APOBEC1-deficient BMDMs as compared to the wild-type sample. No significantly differentially expressed genes are edited by APOBEC1.

Gene	wild-type FPKM	*Apobec1* FPKM	fold change, log	p value
Clk1	37.7471	16.1321	-1.22643	0
Clec7a	58.1322	27.4597	-1.08202	0
Apoe	2576.81	1311.63	-0.974225	4.88E-11
S100a9	47.8244	25.1486	-0.927271	3.51E-08
Atp6v0d2	10.1526	5.40898	-0.908424	2.02E-07
Il7r	16.8209	9.01861	-0.899281	1.99E-10
Kcnj2	15.7805	8.83474	-0.836885	1.90E-10
Lyrm5	9.42068	5.49518	-0.777665	0.00028054
Cysltr1	9.18537	5.37057	-0.774262	1.76E-05
Plxdc2	18.4787	10.8933	-0.762425	1.59E-08
Hpgds	53.2339	31.4061	-0.761298	5.16E-11
Eif2s3y	19.8506	11.7219	-0.759981	1.27E-06
Mrps28	9.30101	15.7603	0.760832	0.000302543
Emp1	79.3218	134.575	0.762617	1.54E-11
Rhof	5.38609	9.20734	0.773546	1.68E-05
Clec4e	9.3831	16.1609	0.78437	5.24E-07
Gm5424	12.1444	21.9633	0.854808	2.22E-08
Smpdl3b	7.35261	13.3654	0.862171	3.18E-07
Hist1h2bc	16.7332	30.4343	0.862989	7.71E-07
Plxnd1	85.3102	159.429	0.902127	7.04E-14
Klf2	27.0208	51.6073	0.933504	5.24E-13
Hist1h1c	26.4887	51.1862	0.950378	2.68E-13
Rsc1a1	7.94624	15.3601	0.950842	2.17E-07
Osgin1	10.5094	20.4214	0.958395	2.07E-10
Fos	37.5034	76.4846	1.02815	0
Gdf3	20.1619	41.9364	1.05657	2.22E-15
Tnf	5.05998	10.5908	1.06561	1.55E-07
Clec4n	3.34772	7.20979	1.10678	3.32E-06
Scd1	3.80326	8.23143	1.1139	5.47E-12
Dusp1	15.8032	34.2969	1.11786	1.33E-15
Ccl2	4.01672	8.85682	1.14077	3.69E-05
Ier3	15.3149	33.8081	1.14243	4.25E-12
Asns	4.5777	10.5256	1.20121	1.39E-09
Apobec1	261.006	601.018	1.20333	0
Irg1	3.01941	8.19314	1.44015	2.10E-13
Dusp6	10.5318	29.5225	1.48707	0

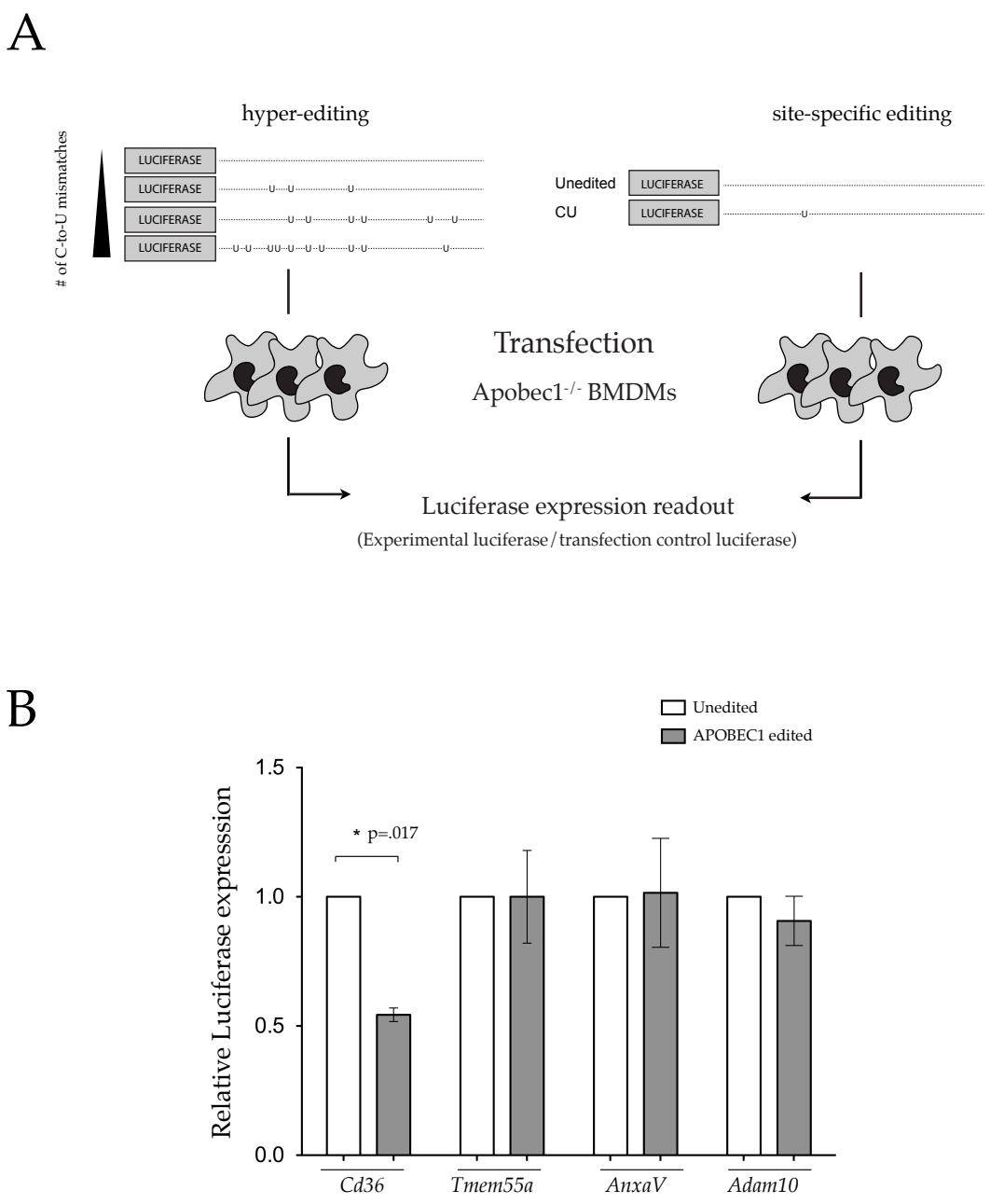


Figure 3.3. Consequences of APOBEC1 editing: luciferase reporter assay. (A) Strategy for testing effects of APOBEC1 3'UTR editing on protein expression. (B) Luciferase levels for representative cDNA clones of site-specifc edited transcript 3'UTRs, normalized to unedited.

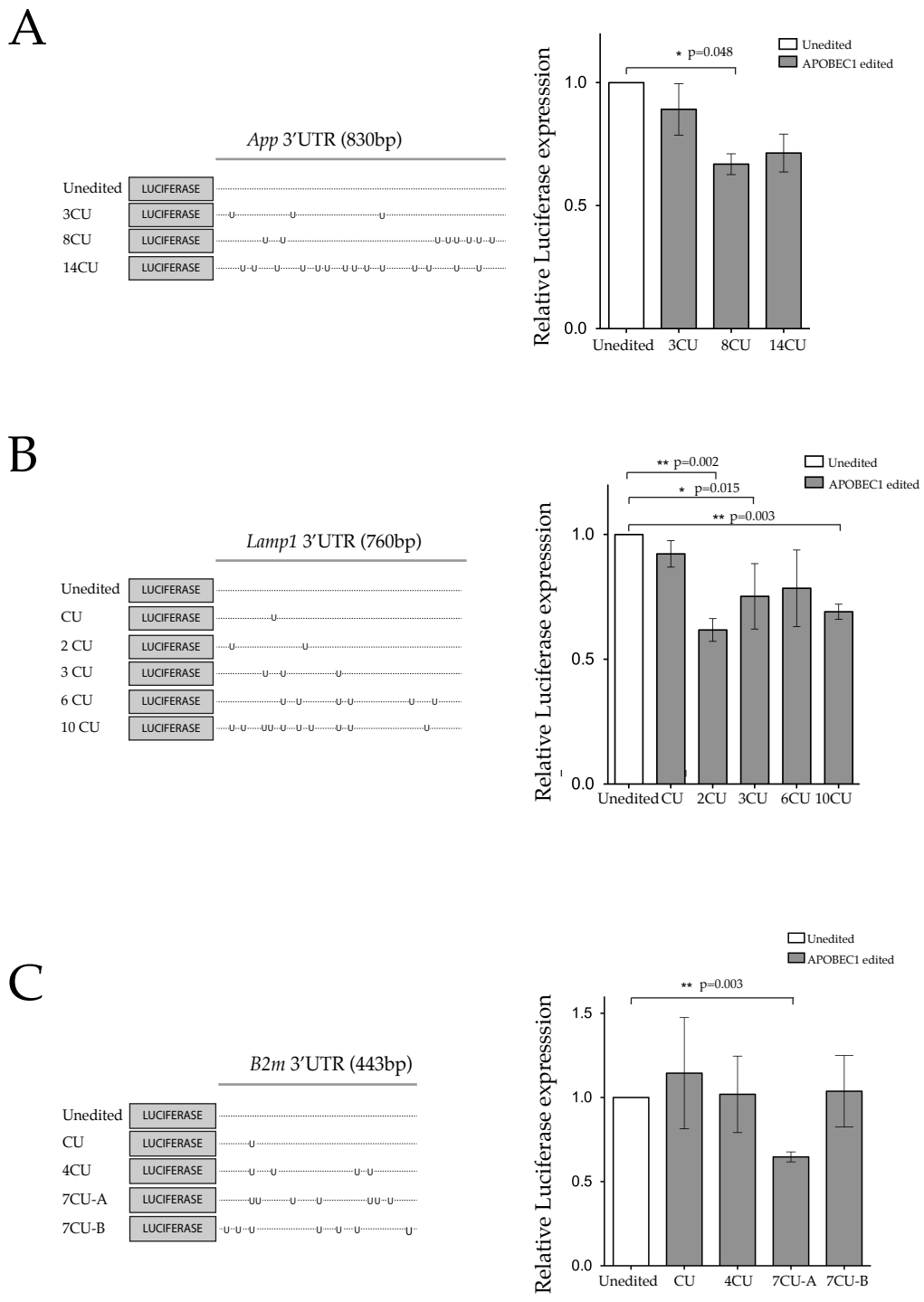


Figure 3.4. Consequences ofAPOBEC1 hyper-editing: luciferase reporter assay. Luciferase levels for representative cDNA clones of hyper-edited 3′UTRs (A) *App* (B) *Lamp1* and (C) *B2m*, normalized to unedited. Maps of hyper-edited constructs indicate the level and distribution of editing events.

hyper-edited transcripts tested were generated from cDNA cloning rather than targeted mutagenesis and so were reflective of the diversity of transcript sequences found in the cell. These results point to a role for APOBEC1 editing events, particularly hyper-editing in transcript regulation. However, some hyper-edited clones had no effect on protein expression and one singly edited clone significantly reduced luciferase levels, indicating that certain APOBEC1-dependent C-to-U editing events or combinations of C-to-U changes can modulate mRNA stability or transcript regulatory factors, but not all C-to-U changes alter protein expression.

3.3. Consequences of APOBEC1 editing on endogenous protein levels

As APOBEC1 was demonstrated to modulate protein levels in a simplified experimental system, I sought to verify this *in vivo*, looking for endogenous protein expression changes in APOBEC1-deficient BMDMs. First, protein levels were assessed by Western blot for a selection of targets with publically available antibodies. No apparent differences in protein levels could be observed between wild-type and APOBEC1-deficient samples. Representative immunoblots for APP and B2m proteins are shown in Figure 3.5. Protein level was then evaluated via flow cytometry for both cell surface (Figure 3.6A) and intracellular (Figure 3.6B) APOBEC1 targets. Again, no significant changes in protein levels were observed for *Apobec1*$^{-/-}$ BMDMs. Representative flow cytometry plots are depicted in Figure 3.6. Finally, to attempt to detect any subtle changes within a specific population of cells that may be mediated by APOBEC1 editing, fluorescence-activated cell sorting (FACS) was used to isolate wild-type BMDMs that expressed the highest cell-surface levels (~top 10% of the population) or

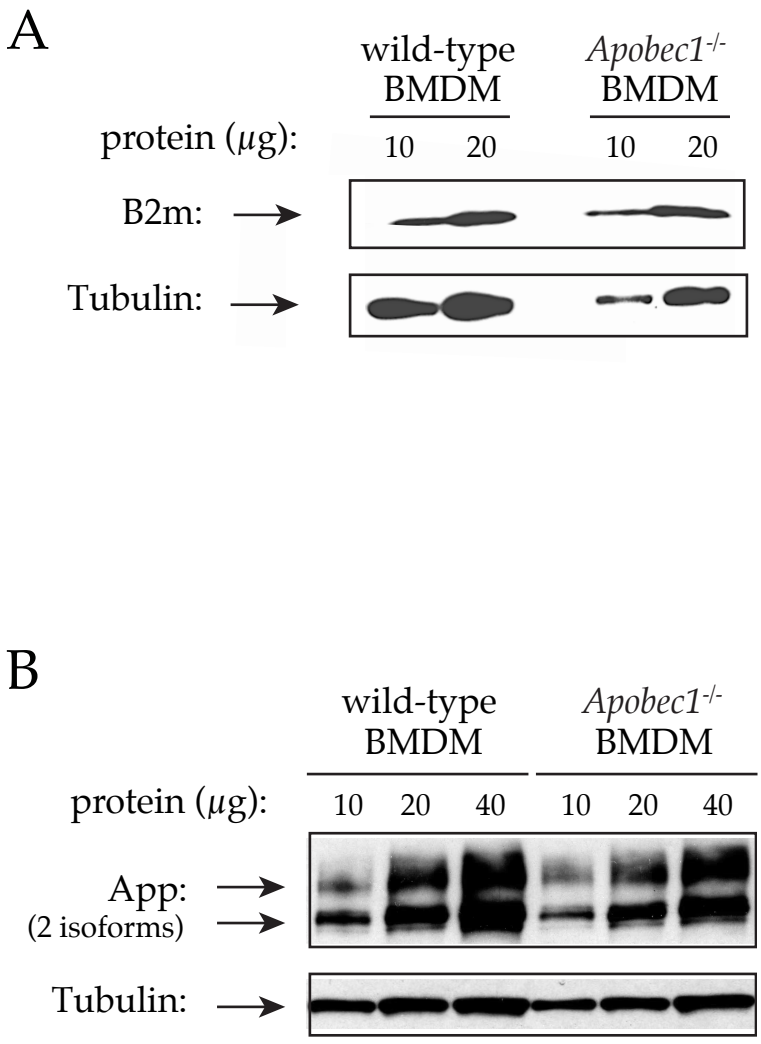


Figure 3.5 Consequences of APOBEC1 editing: Western blotting. Immunoblot analysis of B2m (A) and App (B) protein expression in BMDM lysates reveals no appreciable differences between wild-type and APOBEC1-deficient samples.

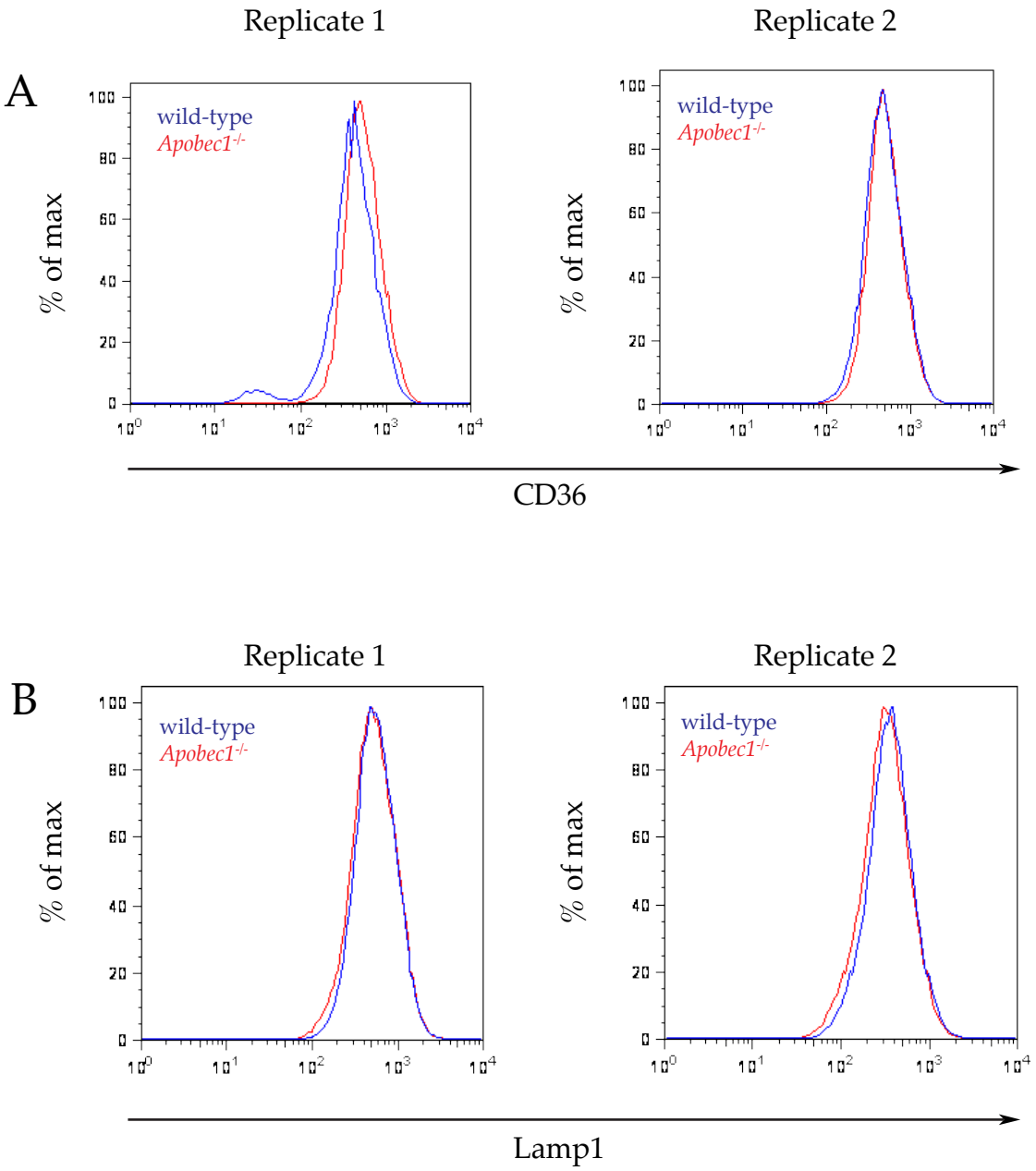


Figure 3.6. Consequences of APOBEC1 editing: flow-cytometry. Flow-cytometry was used to measure the surface expression of CD36 (A) and the internal expression of LAMP1 (B) in wild-type (blue) and APOBEC1-deficient (red) BMDMs. Two representative histograms are shown, depiciting negiglable differences appreciable between the two genotypes.

lowest levels (~bottom 10%) of the CD36 receptor. Editing frequencies of the *Cd36* transcript were then quantified for CD36-high and CD36-low populations. However, both populations were approximately 100% edited (data not shown), suggesting that transcript editing abundance was not appreciably correlated to CD36 expression across a population of wild-type cells. Overall, no changes in endogenous protein expression in APOBEC1-deficient BMDMs could be observed by Western blot, flow cytometry or FACS analysis. These results suggest that APOBEC1-mediated expression changes observed by luciferase reporter assay are too subtle to be appreciated endogenously across a population of cells.

3.4. Interaction between APOBEC1 editing and miRNA targeting.

As APOBEC1 editing events in BMDMs occur within highly conserved regions of transcript 3'UTRs and can modulate protein expression levels, it is possible that APOBEC1 editing could be influencing miRNA targeting. Of note, both miRNA binding and APOBEC1 editing have been found to occur preferentially in regions of high AU nucleotide composition (Grimson et al., 2007; Rosenberg et al., 2011b), suggesting that miRNAs and APOBEC1 target similar 3'UTR substrates and could be interacting on the same transcripts. To investigate the possible consequences of APOBEC1 editing for miRNA targeting, high-throughput sequencing of RNA isolated by cross-linking immunoprecipitation (HITS-CLIP) of the Argonaute (Ago) proteins (Chi et al., 2009) was performed in BMDMs derived from wild-type and *Apobec1*$^{-/-}$ littermates (Figure 3.7). The HITS-CLIP protocol was performed in collaboration with Emily Conn Gantman of the Darnell lab. BMDMs were UV crosslinked,

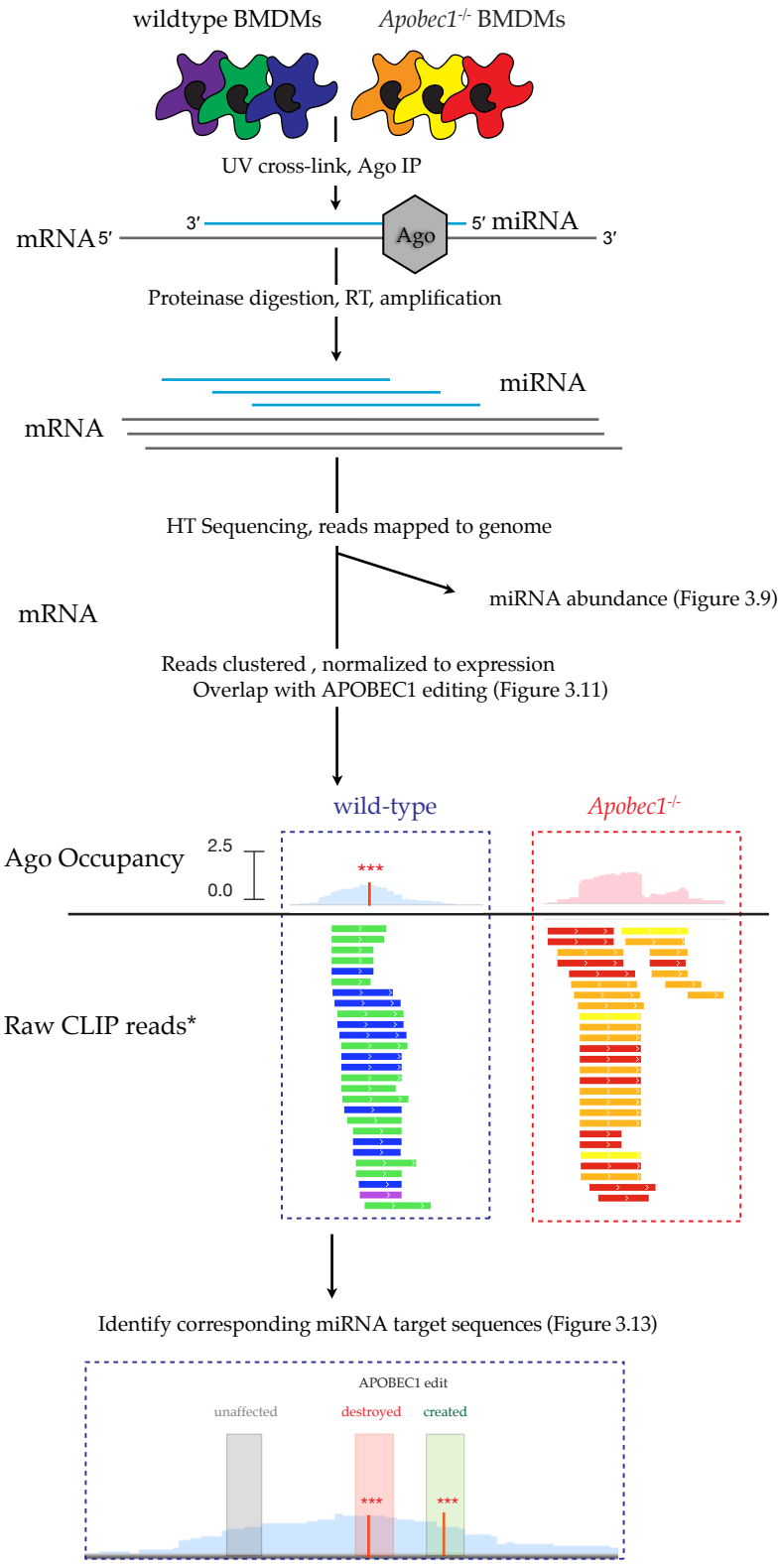


Figure 3.7. HITS-CLIP workflow. A schematic of HITS-CLIP library preparation and bioinformatic analysis is depicted. *CLIP reads are representative not quantitative. ***APOBEC1 editing events.

RNAse digested and subjected to Ago immunoprecipitation. The RNA in the Ago complexes was radiolabeled, purified by gel electrophoresis and then visualized by radiography, where two complex sizes were apparent at ~110kDa and ~130kDa, corresponding to miRNA and mRNA complexes respectively (Figure 3.8A). After proteinase digestion to remove Ago, previously Ago-bound mRNA and miRNA pools were reverse transcribed, PCR amplified and subjected to ultra high-throughput sequencing. A representative PCR amplification of mRNA and miRNA RNA pools is depicted in Figure 3.8B.

Resultant high-throughput sequencing reads were separated by read length into mRNA (≥25nt) and miRNA (≤24nt) pools and aligned to the genome (mm9). Expression of miRNAs was highly correlated between wild-type and *Apobec1*$^{-/-}$ ($r > 0.99$) samples and a list of bound miRNAs was generated from the miRNA alignment (Figure 3.9). Ago clusters, or loci with abundant Ago binding, were defined as regions with ≥ 5 nucleotide overlap and a total of ≥ 8 mRNA reads. The Ago footprint (-30,+32; or more narrowly -24,+22) was extracted from the peak of the cluster, and these footprints were normalized to RNA-Seq transcript expression, defining the relative "Ago occupancy" of the Ago-bound region. High-confidence footprints were defined as occurring in at least 2 replicates of one genotype (biological complexity of 2 for either wild-type or *Apobec1*$^{-/-}$). These high-confidence footprints were predominantly located in coding regions of transcripts and transcript 3'UTRs (Figure 3.10), a pattern of Ago binding previously observed in other applications of Ago HITS-CLIP (Chi et al., 2009; 2012; Helwak et al., 2013; Loeb et al., 2012).

High-confidence Ago footprints were then intersected with APOBEC1 editing events to identify regions of Ago-APOBEC1 overlap across the

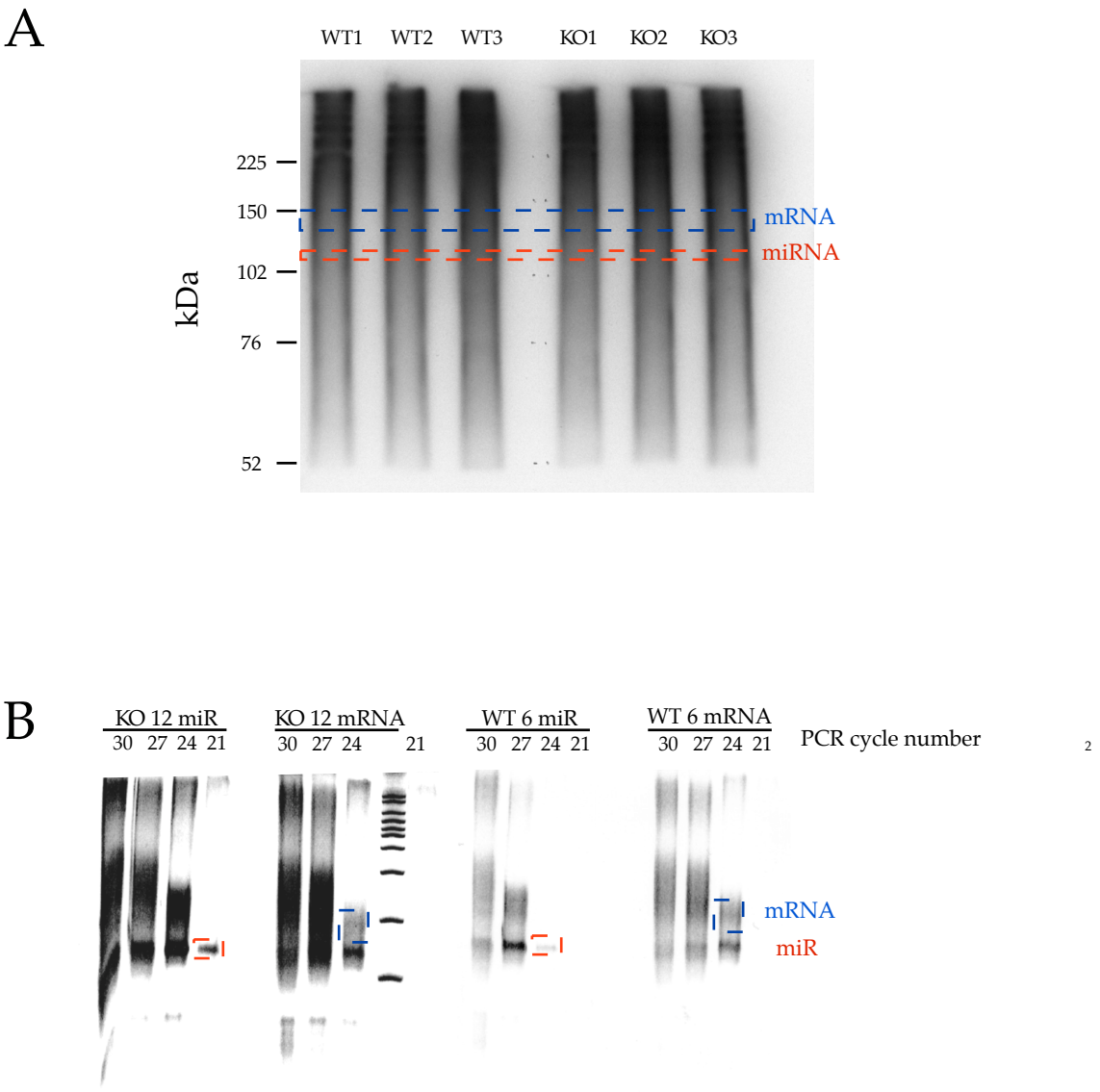


Figure 3.8. Representative HITS-CLIP gels. (A) Autoradiograph of purified radiolabeled Ago:RNA complexes. Regions excised from the nitrocellulose membrane are indicated and represent the separate mRNA (blue; ~130kDa) and miRNA (red;~110kDa) complexes. (B) Representative DNA gel depicting mRNA and miRNA PCR products after linker ligation and PCR amplification. mRNA products cover a range of sizes while miRNAs form a distinct band. Regions excised from the gel are indicated.

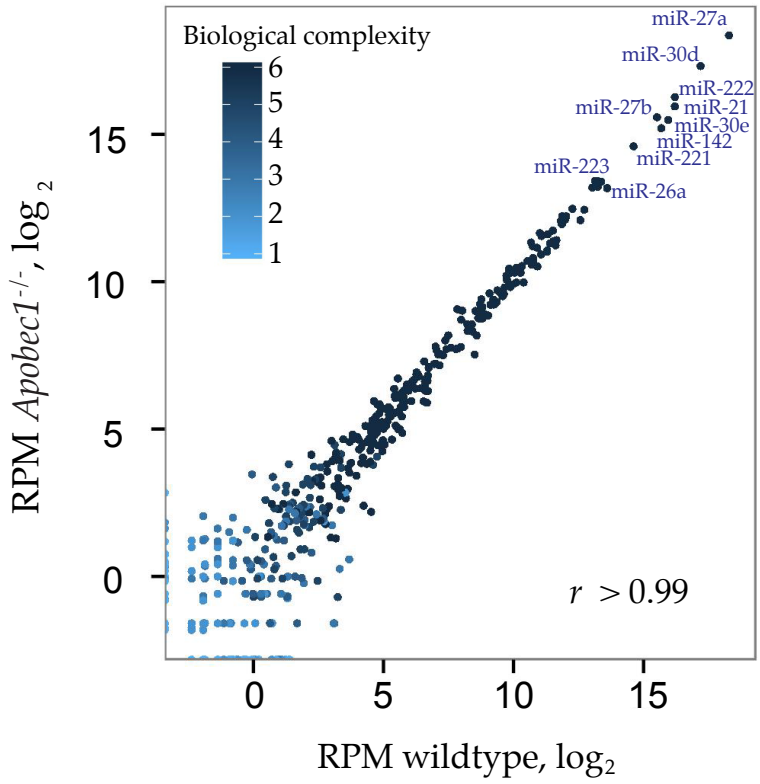


Figure 3.9. Abundant miRNAs bound to Ago in BMDMs. miRNA abundance is plotted as reads per million mapped (RPM) of wild-type and *Apobec1*$^{-/-}$ replicates. The top 10 expressed miRNAs in the wild-type samples are labeled. Total biological complexity of each miRNA is indicated by color.

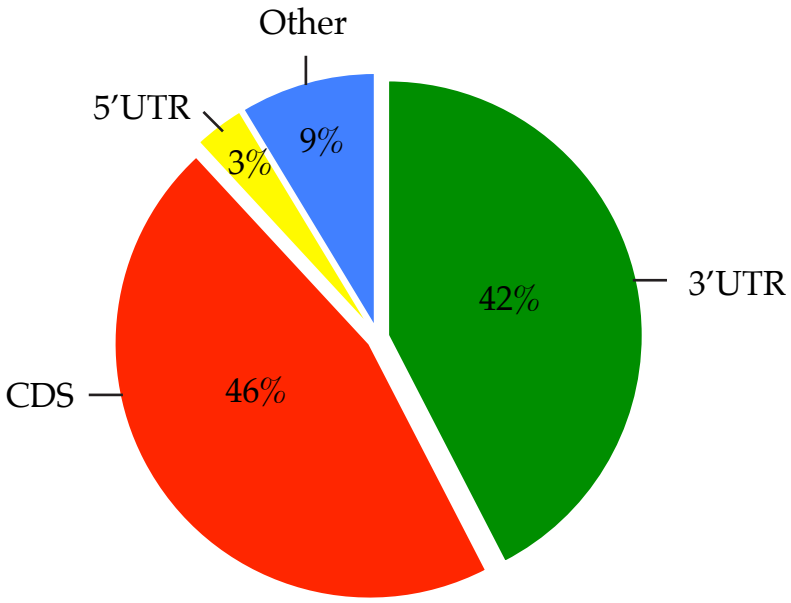


Figure 3.10. Genomic locations of Ago footprints. High-confidence Ago footprints were predominantly located in the coding sequence (CDS) and 3′UTR of targeted transcripts. All annotations were taken from the RefSeq database. A small number of footprints occured in the 5′UTRs or "Other,"which could include: non-coding RNAs, intergenic regions, intronic regions or unannotated transcripts.

transcriptome (Figure 3.11). APOBEC1 editing events were overrepresented in well-covered (20x) Ago footprints by a factor of 5.1 (odds ratio 6.48, 95% confidence interval (3.98, 10.21), $p = 4.52 \times 10^{-12}$). However, Ago occupancy levels were well correlated between wild-type and *Apobec1*$^{-/-}$ samples ($r = 0.82$), with negligible differences in occupancy levels between samples in APOBEC1-targeted footprints, indicating that overall miRNA targeting was not dramatically influenced by APOBEC1 binding or editing.

3.5. Consequences of APOBEC1 editing on miRNA targeting.

To look for more subtle effects on specific miRNAs, I attempted to assign miRNAs to APOBEC1-edited footprints and identify regions where APOBEC1 could create or destroy miRNA seed targets. As has been previously characterized in comparable HITS-CLIP experiments, the top-expressed miRNAs represented the majority of the miRNA pool (Figure 3.12A). However, when I tried to assign only these highly expressed miRNAs to high-confidence Ago footprints by canonical miRNA binding rules, many remained "orphan" (Figure 3.12B). Therefore, to definitively assess APOBEC1:Ago interaction, it was necessary to include potential involvement of both lower-expressed miRNAs and non-canonical miRNA binding. Based on recent reports characterizing modes of miRNA non-canonical binding (Chi et al., 2012; Helwak et al., 2013; Loeb et al., 2012), the search was expanded to include canonical miRNA seed matches (position 2-7 of the mature sequence), as well as non-canonical binding with an exact match to the target sequence (binding positions 1-8 or 3-9) and non-canonical binding with mismatches to the target sequence (one nucleotide

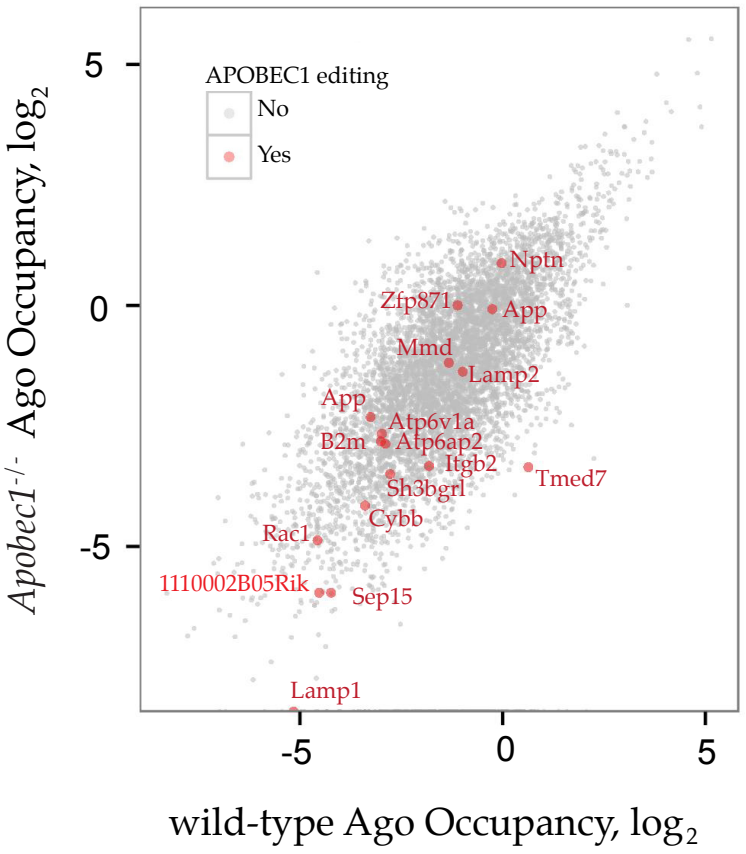
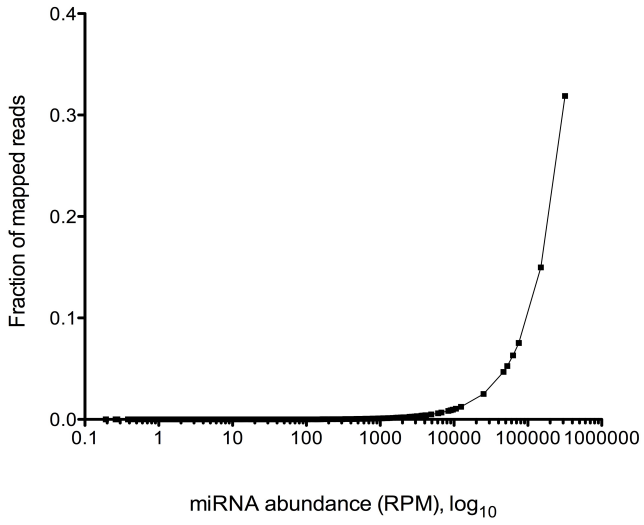


Figure 3.11. Overlap between APOBEC1 editing events and Ago footprints. Ago occupancy values (defined as CLIP read coverage/RNA-Seq read coverage) for Ago footprints in wild-type and *Apobec1*$^{-/-}$ samples. Footprints that contain APOBEC1 editing events are represented in red.

A



B

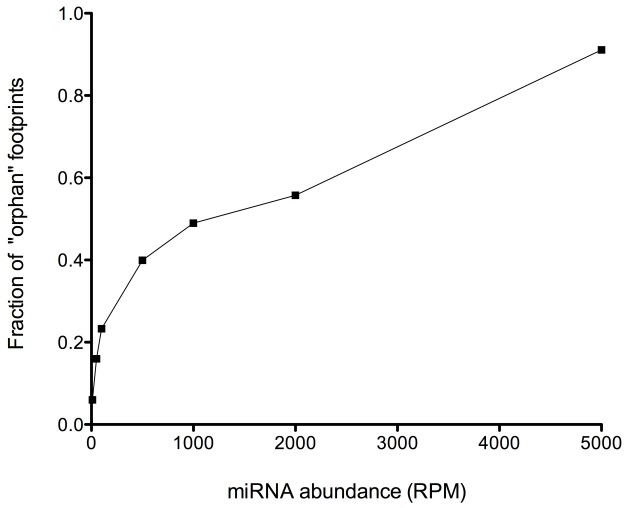


Figure 3.12. Assigning miRNA targets to Ago footprints. (A) Abundance of miRNAs plotted as the fraction of the total miRNA reads mapped. The few highest "expressed" miRNAs (>1000RPM) make up ~98% of the total reads mapped. (B) Fraction of Ago footprints that were "orphan" (no canonical miRNA seed match) when searched for miRNA at various abundance (RPM) cut-offs. miRNA abundance levels were calculated as number of CLIP reads that mapped to the mature miRNA sequence per million mapped. A series of lists of miRNAs were created based on abundance with cut-offs at 10, 50, 100, 500, 1000, 2000 and 5000 RPMs, with decreasing numbers of miRNAs at each cut-off. The footprints were scanned for the canonical miRNA target regions represented in each list. At 1000RPM ~45% of the footprints had no canonical seed match.

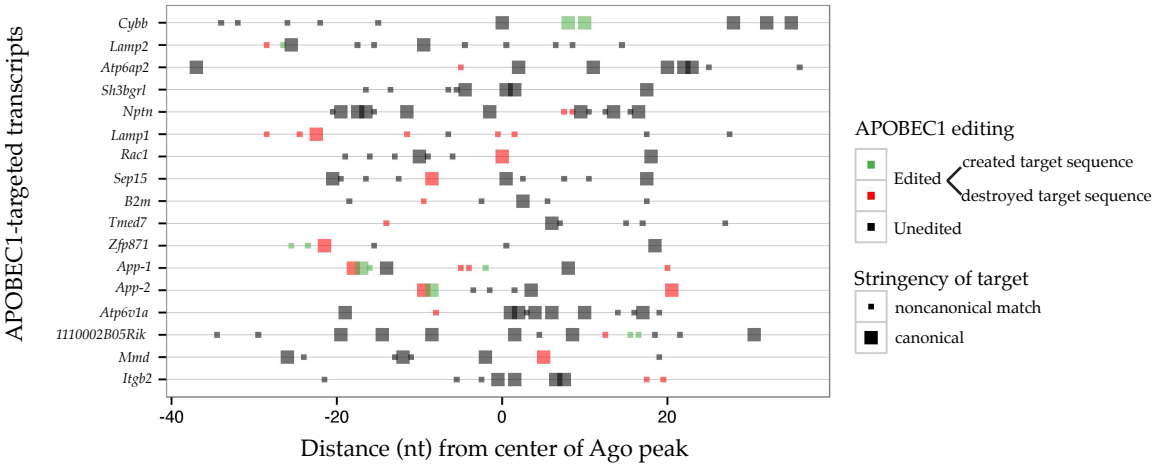


Figure 3.13. Putative miRNA seed targets in APOBEC1-edited regions. "Edited" (with APOBEC1-dependant C-to-T change) and "Unedited" (without editing events) footprint sequences were scanned for expressed miRNA targets regions ("canonical" = match to position 2-7 of the mature miRNA, "non-canonical match" = match to positions 1-6 or 3-8). miRNA targets which would be created (green) or disrupted (red) by an APOBEC1 editing event are depicted.

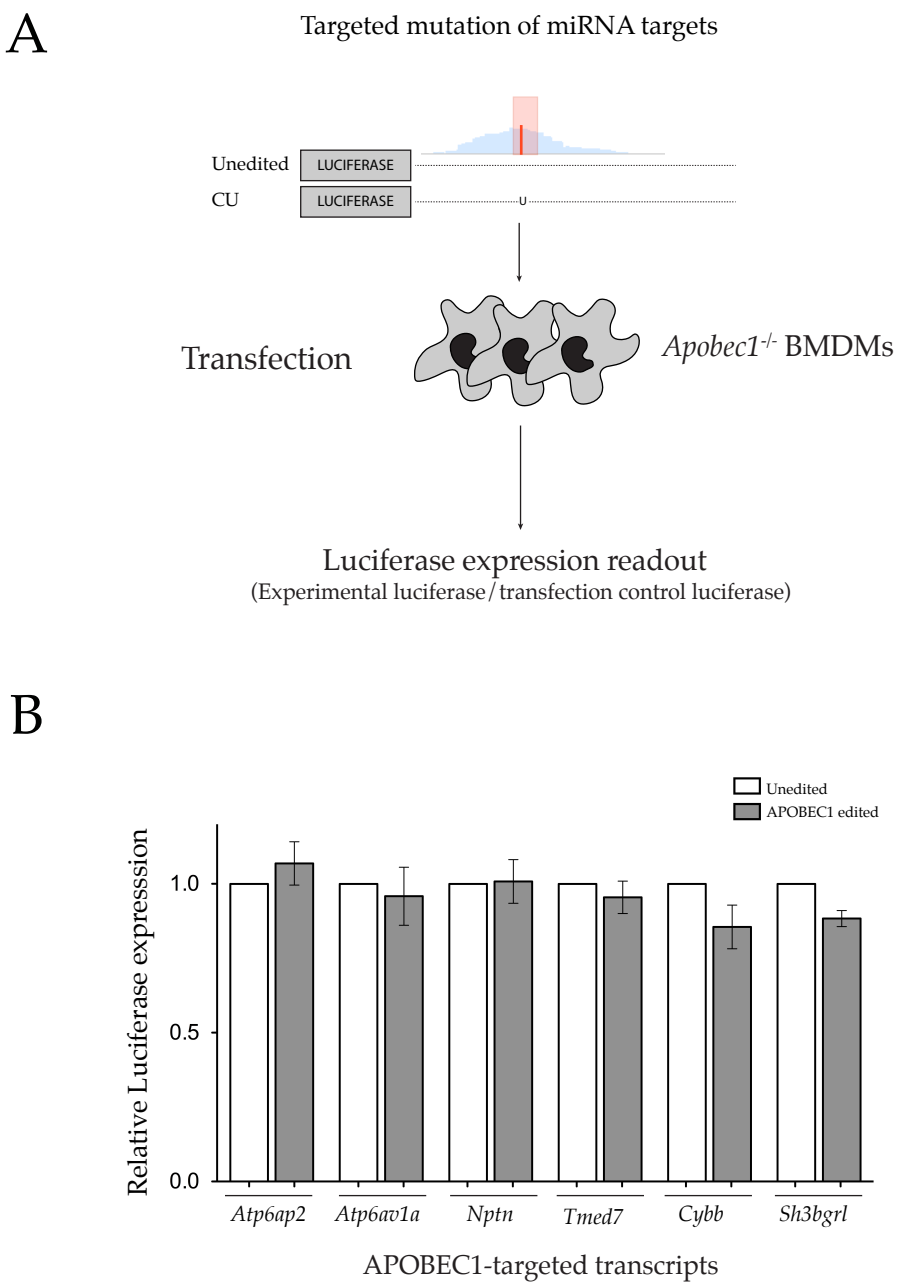


Figure 3.14. Testing APOBEC1-edited miRNA seed targets by luciferase assay. (A) Luciferase reporter strategy. APOBEC1 editing events were introduced via site-directed mutagenesis and tested for effect on miRNA-targeting via luciferase reporter assay. (B) Luciferase reporter assays yielded no significant luciferase expression differences between "edited" and "unedited" constructs.

mismatch or G-bulge insertion). From this analysis, an APOBEC1-targeted footprint map was created, identifying putative miRNA binding sites in each footprint, and identifying those, which overlapped with an APOBEC1 editing event (Figure 3.13). Using the narrowest definition of the Ago footprint (46nt), a set of putative miRNA target regions for each footprint was established after filtering for miRNAs which were expressed in 3 replicates and whose target regions was within a distance of 50 nucleotides from the center of the Ago peak. For each footprint we identified a set of likely miRNA targets, a number of which were disrupted or created by APOBEC1 editing events (Figure 3.13).

I used a luciferase reporter assay to test the functionality of potential APOBEC1-altered miRNA binding sites. Identified 3'UTRs were cloned into a dual-luciferase construct and site-specific mutagenesis was performed at the location of APOBEC1 editing events that had been observed in Ago-CLIP to potentially disrupt or enhance miRNA binding (Figure 3.13, Figure 3.14A). Constructs containing five individual C-to-U mutations identified to disrupt a miRNA seed target were tested with luciferase reporter assay and failed to alter luciferase levels (Figure 3.14B). These results demonstrate that while APOBEC1 editing events can modulate protein expression, this phenomenon is unlikely mediated by changes in miRNA targeting.

Chapter 4: Discussion

The use next-generation sequencing has greatly expanded the editing repertoires of RNA editing enzymes, ADAR and APOBEC1. In particular for APOBEC1, our lab used a comparative RNA-Seq screen to identify 32 additional mRNA targets contained in transcript 3'UTRs, pointing to additional functions for APOBEC1 that had been previously unappreciated. Based on the observation that APOBEC1 was expressed in immune cell types, I decided to use comparative RNA-Seq to identify novel APOBEC1 editing events in BMDMs. The RNA-Seq screen revealed abundant APOBEC1 dependent mRNA editing in BMDMs in highly-conserved regions of transcript 3'UTRs. These newly identified APOBEC1 editing events occurred in two distinct patterns: site-specific editing and hyper-editing. Further analysis revealed that these editing events shared adjacent nucleotide preferences with intestinal APOBEC1 editing events but were only loosely associated with a downstream mooring sequence, previously demonstrated to be essential for APOBEC1 editing of *Apob* and associated with intestinal 3'UTR editing. Although differences in endogenous transcript and protein levels were not observed, certain patterns of editing resulted in altered luciferase activity via standard luciferase reporter assay, suggesting that in some cases editing can alter transcript fate. Finally, transcriptome-wide profiling of miRNA binding sites revealed little functional overlap between miRNA targeting and APOBEC1 editing events in 3'UTRs, pointing to a miRNA-independent mechanism for APOBEC1-mediated transcript regulation. Below I will discuss the questions remaining about the

implications for ACF-independent APOBEC1 site-specific and hyper-editing events in transcript 3'UTRs.

4.1 Overestimation of false-positive rate

Here we used a comparative RNA-Seq screen in BMDMs to identify the first examples of APOBEC1 physiological editing outside of the digestive system. We identified over 100 high-frequency editing events, more than tripling the number of previously established APOBEC1 substrates. A subset of these APOBEC1-dependent C-to-U changes was validated with standard Sanger sequencing with a false positive rate (FDR) of 12.5%. The implied false-positive rate was calculated at less than 1% based on the amount of C-to-U editing identified in the APOBEC1-deficient samples that was absent from the wild-type sample, indicating that this FDR may be an overestimate. This overestimation may have occurred for a few reasons. First, I chose to forgo a genomic SNP filter during the post-screen filtering process. This choice was based on an observation I made during the validation of the Rosenberg data early in my thesis work. One target, *Cd36*, was not identified in the initial comparative RNA-Seq screen. However, *Cd36* was found to have a strong mooring sequence, and its mRNA transcript sequence was later determined to be highly (> 90%) edited by APOBEC1. I subsequently realized that this specific C-to-U edit was included in a SNP database, based on sequencing that had been performed on a cDNA substrate. As highly edited targets are more likely to have been previously identified as SNPs, I decided to abandon the SNP filter and validate a subset of the editing targets with close to 50% or 100% editing and throw out sites with

50% or 100% editing in coding regions. One site in a coding region, which was included in the FDR calculation, was thrown out.

Additionally, analysis of the APOBEC1 editing events has revealed that there may be some plasticity in the editing repertoire, leading to unvalidated editing events. For example, the two identified sites within the *Rhoa* transcript 3′UTR could not be confirmed via subclone sequencing, but further analysis of the transcript sequence revealed a set of additional low frequency APOBEC1-dependent C-to-U editing events (data not shown). This result suggests that identification of the full spectrum of possible APOBEC1 editing events would require the compilation of transcript-wide sequencing of many samples.

4.2 APOBEC1 editing is predominantly restricted to 3′UTRs

Through an established RNA-Seq screen, I have identified over 100 novel APOBEC1 editing events in 3′UTRs, but no additional APOBEC1-catalyzed events within transcript 5′UTRs or coding regions. To date, the only physiological APOBEC1 editing event occurring in a coding region is the editing of C6666 in *Apob* transcript, where the majority of editing serves an important role for lipid metabolism and transport. Previously we established that while the APOBEC1 "mooring sequence" could be found in the coding region of a set of other expressed transcripts, but no APOBEC1 editing was observed at these loci. Taken together, these data suggest that *Apob* is the only known APOBEC1 editing event that alters a codon, and the remainder of APOBEC1 editing is restricted to transcript 3′UTRs. The reasons for this remain elusive. Possibly, APOBEC1 editing in 3′UTRs is "sacrificial," absorbing aberrant potentially harmful editing activity in a safe region of the transcript where editing will have

a minimal effect on final protein composition. This hypothesis is somewhat logical for intestinal APOBEC1 activity, in which there is one highly important editing event in a coding region. Theoretically other coding regions could be protected from APOBEC1-mediated editing by allowing 3'UTRs to be edited and limiting the amounts of free APOBEC1 not bound to *Apob*. However, in immune cells where there is no established role for APOBEC1 this is less plausible, making it more likely that 3'UTR editing fills an important role in this cell type.

Prior to the identification of APOBEC1 editing in 3'UTRs, it was postulated that access to most coding exons for the large multi-protein editosome was blocked by RNA-splicing machinery. APOBEC1 editing in *Apob* occurs in the middle of an exceptionally large exon (> 7kb), perhaps avoiding the RNA splicing machinery localized to distant exon-intron boundaries (Sowden et al., 1996b). This mechanism could explain the abundant APOBEC1 editing found in untranslated regions. Additionally, APOBEC1 could be recruited by an alternative RNA-binding factor, which is exclusively targeted to 3'UTRs. As 3'UTR editing seems to occur most abundantly in a cell type that lacks ACF, a model in which APOBEC1 3'UTR editing occurs in the absence of ACF but under the influence of another auxiliary factor is appealing. Further analysis of specific targeting factors and the influence of exogenously introduced ACF on BMDM editing could provide insight into some of these questions.

Additional mechanisms mediating APOBEC1 3'UTR editing might involve a specific cellular factor that protects coding exons from APOBEC1 editing, potentially by targeting edited transcripts for degradation. Indeed, APOBEC1 editing of C6666 in *Apob* introduces a premature stop codon, making it a target for the nonsense-mediated decay (NMD) pathway. The edited

transcript escapes NMD by remaining in complex with the APOBEC1 editosome as it is exported from the nucleus. Interestingly, this process requires ACF involvement; edited *Apob* transcripts associated APOBEC1 alone are subject to NMD but transcripts bound to an APOBEC1:ACF complex are protected (Chester et al., 2003). As such, in ACF-deficient BMDMs, some catastrophic editing events occurring in coding regions may be degraded via NMD, and elude detection via RNA-Seq. However, silent or non-synonymous mutations that alter codon specificity but do not introduce a stop codon should be unaffected by NMD. As none of these types of mutations have been observed in BMDMs, NMD cannot be the only protective factor against APOBEC1 editing in coding exons.

Lastly, APOBEC1 editing of the *Apob* transcript is associated with a specific transcript secondary structure, a conserved stem-loop that serves to introduce the edited cytidine into the APOBEC1 catalytic pocket. 3'UTRs are associated with abundant secondary structure, indicating that they might naturally serve as APOBEC1 substrates. This is reminiscent of ADAR editing where inverted repeats in 3'UTRs provide the optimal dsRNA substrate for ADAR editing. Duplexes in coding regions require more complex binding with a complementary intronic region and occur at much lower frequency, partially explaining why ADAR editing is overrepresented in transcript 3'UTRs. A similar mechanism could be influencing APOBEC1 editing, shifting the known repertoire toward highly structured untranslated regions of transcripts.

4.3. Implications for APOBEC1 hyper-editing

Identification of BMDM-specific APOBEC1 RNA editing revealed two distinct editing patterns that I have termed single-site editing and hyper-editing. Single-site editing events are reminiscent of the canonical physiological *ApoB* editing and most APOBEC1-dependent editing events identified in the intestine transcript 3'UTRs: one major editing site, edited at a high frequency, occasionally associated with nearby low frequency C-to-U editing events. We were surprised to discover dramatic APOBEC1 hyper-editing activity in BMDMs, an unusual physiological behavior for APOBEC1. As discussed, there is some background activity for APOBEC1 observed in *Apob* (Blanc et al., 2012) and 3'UTR targets in enterocytes (Rosenberg et al., 2011b). This editing was typically very low frequency (< 10%) and occurred in association with one highly edited site. In contrast, there are a few previous examples of more high-frequency APOBEC1-dependent promiscuous or hyper-editing activity observed in the *ApoB* transcript (Sowden et al., 1996b; 1998; Yamanaka et al., 1996) and in viral transcripts(Petit et al., 2009), but this editing behavior occurs in the context of APOBEC1 over-expression or MLV infection and does not represent the physiological editing behavior of a steady-state cell. Is there a difference between this higher frequency " hyper-editing" we observe and background editing found in *Apob?* Why in some transcripts is there one dominant site-specific event and in other transcripts targeting seems seem to be promiscuous and plastic? The mechanisms of APOBEC1 hyper-editing in BMDMs and how certain transcripts are targeted for site-specific editing versus hyper-editing remain elusive. The possible contribution of an additional co-factor *in lieu* of ACF is discussed below. Regardless of mechanism, APOBEC1 hyper-editing results in a dramatic shift in

the RNA sequence, increasing the likelihood that editing could be affecting important regulatory regions of the UTR.

4.4. ACF-independent APOBEC1 editing

APOBEC1 editing activity in BMDMs is especially intriguing because it occurs in the absence of ACF and is only loosely associated with the downstream 11-nucleotide mooring sequence. These results suggest that the mechanism of APOBEC1 editing in BMDMs may differ considerably to that established for *ApoB*. Interestingly, while as a unit recombinant APOBEC1 and ACF were shown *in vitro* to be necessary and sufficient to induce C-to-U editing at C6666 in *ApoB* (Lellek et al., 2000; Mehta et al., 2000), no study has shown *in vivo* that ACF is essential for APOBEC1 editing. In the intestine, these studies have been complicated by the difficulties in generating an ACF-deficient mouse, as disruption of the ACF locus leads to embryonic lethality (Blanc et al., 2005). Our data definitively show that physiological APOBEC1 enzymatic editing activity occurs in 3'UTRs in the absence of ACF.

Although we have shown that APOBEC1 can edit mRNA transcripts without ACF, the mechanism by which this occurs is still an outstanding question. Some *in vitro* studies suggest that under specific elevated temperature conditions APOBEC1 may be capable of ACF-independent editing (Chester et al., 2004), indicating that in certain environments it could be the sole mediator of both RNA targeting and editing. Indeed, in transgenic APOBEC1-overexpressing mice, a proposed mechanism for the observed hyper-editing activity is that the ACF:APOBEC1 stoichiometry is shifted, and the abundant levels of free APOBEC1 aberrantly edits the *Apob* transcript (Blanc and Davidson, 2010;

Chester et al., 2004). Presumably, any alternative co-factor would also be overwhelmed by the over-expression of APOBEC1, pointing to independent APOBEC1 activity in this non-physiological system. Alternatively, APOBEC1 editing in BMDMs could be happening in conjunction with another auxiliary factor, potentially one or more candidate members of the multimeric editosome, many of which are expressed in BMDMs based on RNA-Seq transcript expression, including ABBP-1, ABBP-2, CUGBP2, and GRY-RBP (data not shown). Furthermore, it is also a point of debate whether APOBEC1 is capable of nuclear import when not in complex with ACF (Yang et al., 2001), suggesting that APOBEC1 cytoplasmic-nuclear shuttling may need to occur in complex with an associated protein, underscoring the importance of some sort of associated editosome.

Regardless of mechanism, the fact that APOBEC1 editing occurs without ACF in 3'UTRs brings up questions about the necessity of ACF involvement of other APOBEC1 targets. As described above, APOBEC1:ACF complexes were only shown to be essential for editing *in vitro*, and only for C6666 in the *Apob* transcript. Whether ACF is necessary for editing in alternative sites in apoB or intestinal 3'UTR targets is unclear. BMDMs and intestinal enterocytes share a number of APOBEC1 targets, indicating that at least some examples of 3'UTR editing in small-intestinal enterocytes might occur without ACF involvement.

A popular hypothesis in the field, based largely on the over-expression phenomenon observed in mouse models, is that ACF primarily acts as a "chaperone," limiting aberrant APOBEC1 editing. This is somewhat supported by an increase in APOBEC1 editing frequency observed in heterozygous *A1cf*$^{+/-}$ mice (Blanc et al., 2005). The abundant APOBEC1-mediated hyper-editing in

BMDMs also lends support for the idea that APOBEC1 might edit more promiscuously without ACF. However, additional examples of site-specific, seemingly strictly targeted editing are also prevalent in BMDMs, complicate the idea that APOBEC1 would require ACF for targeted mutation. Overall, the discovery of high frequency APOBEC1 RNA editing in BMDMs dramatically expands the role for APOBEC1 beyond the digestive system and brings up many questions about precise mechanisms of APOBEC1 editing in 3′UTRs. As they express APOBEC1 and lack ACF, BMDMs provide an ideal experimental system to further investigate the influence of cofactors on APOBEC1-dependent RNA editing.

4.5. APOBEC1 editing can modulate protein expression

In BMDMs, APOBEC1 editing events occurred more frequently than expected by chance in regions conserved by evolution, suggesting functional relevance. When the specific translational outcomes of APOBEC1-dependent C-to-U editing events in 3′UTRs were assessed via luciferase reporter assays, examples of C-to-U changes were observed that led to significant reduction in luciferase activity. As hyper-edited transcripts tested were generated from directly from cloned cDNA they reflected the variation in transcript sequences found in the cell. In general, the majority of clones with one or two C-to-U changes had no consequence for luciferase expression whereas hyper-edited transcript 3′UTRs were more likely to result in differential translation outcomes. However, some hyper-edited clones had no effect on protein expression and one singly edited clone, *Cd36*, significantly repressed luciferase levels. These results suggest that certain combinations APOBEC1-dependent C-to-U editing events or

combinations of C-to-U changes can affect protein expression but random C-to-U changes alone are not sufficient to yield altered translational outcomes. Importantly, these experiments were performed in APOBEC1-deficient cells and therefore only tested the effects of C-to-U changes on the transcript sequence itself, not other modes of APOBEC1-mediated regulation that could occur via RNA-binding.

An important aspect of these luciferase reporter assays is that they were performed in the primary cell type of interest, BMDMs. As BMDMs are resistant to lipofectamine and similar transfection modalities, I used Amaxa "nucleofection" technology to introduce reporter constructs into the nucleus through electroporation. This efficiently (~20%) introduced luciferase constructs into BMDMs although led to substantial cell activation. Therefore, a caveat of these experiments is that they occur in an activated setting rather than a steady-state cell where editing events were originally identified.

Interestingly, a large part of my initial work in the lab involved luciferase reporter assays designed to test the translational outcomes of APOBEC1 intestinal 3'UTR targets. These experiments were performed in 293Ts and yielded no significant luciferase expression differences in edited constructs as compared to unedited controls (data not shown). Of note, in establishing a BMDM system to look for luciferase changes, I used one of these previously cloned constructs, *Cd36*, a highly edited target in both the intestine and BMDMs. In contrast to the assays performed in 293Ts, the *Cd36* edited construct when transfected into BMDMs, exhibited significant luciferase repression. This result highlights the limitations of luciferase reporter assays performed in unassociated cell lines, a commonly used experimental method. Cell-specific expression of

miRNAs, RNA-binding proteins, and other auxiliary factors that substantially affect transcript regulation and the specific expression profiles may vary widely between cell-types, especially for transformed cells with uncertain karyotypes.

Despite the ability of "edited" 3'UTRs to alter protein expression in a luciferase reporter assay, no transcript level differences were observed for targeted transcripts and no protein expression could be appreciated by standard assays. This is not completely surprising, as changes to translational efficiency or subtle changes in transcript stability may not be appreciable at the transcript level. Additionally, the predominantly two-fold protein expression differences we observe are likely below the threshold of Western blotting and flow-cytometric analysis, the two techniques utilized.

4.6. HITS-CLIP reveals little interaction between miRNA targeting and APOBEC1 editing.

MiRNAs preferentially target phylogenically conserved regions of 3'UTRs, which are also the principle sites of APOBEC1 editing in BMDMs. To comprehensively assess the interaction of APOBEC1 editing with miRNA targeting, we used HITS-CLIP to generate both a list of miRNAs expressed in BMDMs and a transcriptome-wide map of Ago binding in wild-type and *Apobec1*$^{-/-}$ cells. As described above, we exhaustively searched for miRNA seed targets within regions of Ago-binding that might be affected by APOBEC1 editing events. Even with such a comprehensive search for miRNA targets, Ago HITS-CLIP analysis revealed significant overlap between regions of Ago targeting and APOBEC1 editing events, but little evidence for miRNA target creation or disruption by APOBEC1 editing. Some editing events shown to

repress luciferase expression, such as the high frequency event in the *Cd36* transcript, did not overlap with a region of Ago binding, suggesting that APOBEC1 editing was affecting other mechanisms of transcript regulation. Other hyper-edited transcripts that repressed protein expression levels contained editing events that overlapped with sites of Ago binding, but no miRNA sites generated by APOBEC1 editing events could be identified. All other potential sites of APOBEC1:Ago interaction, including potential miRNA target generation in *App*, could not be validated by luciferase reporter assays, suggesting that the identified miRNA site was non-functional, or that the C-to-T change had negligible effect on the efficiency of miRNA repression. Despite this, we cannot definitively rule out a role for miRNAs in APOBEC1-mediated transcript regulation. Lower frequency combinations of events could lead to more subtle changes below the resolution of HITS-CLIP or unappreciated modes of miRNA binding could be mediating some of the protein expression differences. In particular, a recent study used an adapted HITS-CLIP technique to link and then sequence interacting mRNAs and miRNAs (Helwak et al., 2013). Although this process was quite inefficient, the authors are able to construct a profile of the many forms of predicted miRNA binding, the majority of which requires additional binding events beyond the canonical seed (positions 2-7 of the mature miRNA). This work has substantial limitations as it was performed using an ectopically expressed tagged Ago protein in a human cell line, but it points to the signficiant complexity in miRNA targeting.

4.7. **Alternative mechanisms for APOBEC1-mediated transcript regulation**

While the majority of APOBEC1 editing events in BMDMs do not seem to affect miRNA targeting, the specific mechanism for APOBEC1 editing-mediated protein repression remains unclear. As has been discussed above with RNA editing, next-generation sequencing has greatly expanded the breadth of knowledge regarding the complexity of RNA modification and regulation. Most importantly for this discussion, *cis*-acting elements in transcript 3'UTRs and *trans*-acting factors that bind to them have been shown to contribute to transcript regulation through a variety of mechanisms, a number of which could be influenced by RNA editing.

Specific ADAR editing events in 3'UTRs have been implicated in targeting transcripts for cleavage via a specific nuclease (Scadden, 2005) and inducing nuclear retention by promoting binding to a dedicated nuclear factor (Chen and Carmichael, 2009; Prasanth et al., 2005; Zhang and Carmichael, 2001). The caveats for these hypotheses have been discussed in Section 1.2.5, but they remain popular models for the regulation certain ADAR-targeted transcripts. Similarly, APOBEC1 editing in 3'UTRs could target the transcript for degradation or result in its retention in the nucleus. However, there seems to be negligible transcript expression differences between wild-type and APOBEC1-deficient samples, and no edited transcripts exhibit differential expression between the two genotypes. This result detracts from the degradation hypothesis, in which expression changes should be appreciable at the transcript level. Recent work in our lab also points against the nuclear retention of APOBEC1 edited transcripts. Transcripts isolated specifically from the nucleus tended to have lower editing frequencies than those extracted from the cytosol and there is no differences in nuclear of

cytosolic transcript expression between wild-type and APOBEC1-deficient BMDMs. This work is still in progress, but points to the efficient nuclear export of APOBEC1 edited transcripts. Together, these data suggest that unlike ADAR, APOBEC1-mediated transcript regulation is not occurring via nuclear retention or cleavage. Furthermore, as we have observed that edited transcripts are efficiently exported from the nucleus, the regulation of these transcripts most likely occurs in the cytoplasm and is not associated with changes in splicing, 5'-capping and alternative polyadenylation. Indeed, analysis of RNA-Seq reads reveals no differentially expressed isoforms or alternative polyadenylation sites in edited transcripts between wild-type and APOBEC1-deficient samples, further underscoring this point.

Alternatively, APOBEC1 editing could be altering transcript stability, potentially by modulating the binding properties of stabilizing or de-stabilizing RNA-binding proteins. AU Rich Elements (AREs), which consist of sets of AUUUA pentamers, are bound by a variety of RNA-binding proteins and the number of AREs in a transcript is inversely correlated with transcript stability (Hao and Baltimore, 2009). As APOBEC1 introduces C-to-U changes within AU-rich regions of 3'UTRs, it seems likely that APOBEC1 editing could introduce AREs. However, none of the targets shown to be differentially expressed via luciferase reporter assay have AREs introduced by C-to-U editing events. This indicates that, while APOBEC1 editing could be influencing stability via another mechanism, AREs are not likely to mediate this control. Alternatively, APOBEC1 itself has been shown to mediate transcript stability through its RNA-binding capabilities (Anant and Davidson, 2000; Anant et al., 2004), pointing to potential

editing and RNA-binding mechanisms involved in APOBEC1 regulation of transcript stability.

There are a variety of other mechanisms in which APOBEC1 could be influencing transcript regulation but as in the above these mainly involve modifying 3′ UTR *cis*-elements and leading the altered functions of RNA-binding proteins (RBPs). In this way, structural elements within transcript 3′UTRs can also influence protein expression. For example, in response to environmental cues an RNA secondary structure in the *Vegfa* transcript 3′UTR undergoes a conformational change that regulates a RBP-mediated change in VEGFA protein expression (Ray et al., 2009). As *Apob* editing requires a stem-loop structure, it seems likely that 3′UTR editing occurs in the vicinity of secondary structure and that base-pairing within structural elements could be disrupted by editing. Additionally, mRNA subcellular localization can be mediated by specific "zip codes" within transcripts 3′UTRs; these motifs are bound by RBPs that mediate the targeting of the mRNA to a precise cytoplasmic environment. This asymmetric localization of mRNAs generates cell polarity by controlling the sites of translation. A zip code mechanism of subcellular localization has been specifically characterized in neurons and fibroblasts, where β-*actin* mRNA transcripts are localized to the fibroblast leading edge (Kislauskis et al., 1994) and neuronal growth cones (Bassell et al., 1998), promoting accumulation of β-actin protein and subsequent forward movement. Intriguingly, this subcellular localization is partially mediated by KSRP, a known inhibitor of APOBEC1 (Gu et al., 2002), but no further link between RNA editing and subcellular localization of this manner has been established.

While there are many modes of potential APOBEC1-mediated regulation of gene expression, the lack of observed transcript expression differences in APOBEC1-deficient cells points to a mechanism in which APOBEC1 editing more directly affects translational efficiency, potentially by modulating the binding properties of RBPs and repressing translation through a variety of potential mechanisms including de-adenylation, ribosome stalling or other modes of inhibition of translational machinery. To test this hypothesis, we plan to conduct ribosomal profiling of APOBEC1-deficient BMDMs.

4.8 Closing remarks

This thesis presents a body of work that establishes abundant physiological APOBEC1 editing activity beyond the digestive system and demonstrates that these untranslated editing events can modulate protein expression. These findings are highly significant to the fields of RNA editing and transcriptional regulation. First, we demonstrate APOBEC1 catalytic activity in an immune cell type, the first time physiological APOBEC1 editing activity has been characterized outside of the digestive system. This editing also occurs in the absence of the "essential" APOBEC1 co-factor, calling into question the dogma regarding the regulation and editing mechanism of this highly characterized RNA-editing enzyme. Additionally, APOBEC1 hyper-editing, observed in a number of transcripts in BMDMs, represents a unique editing modality for APOBEC1. Finally, we establish a role for APOBEC1 editing in BMDM transcript 3'UTRs in the regulation of transcript expression through a non-miRNA-mediated mechanism.

Although we have definitively shown that C-to-T changes at the sites of APOBEC1 editing can led to protein expression differences in a simplified experimental system, analysis of the downstream consequences of APOBEC1 editing *in vivo* presents a significant challenge. Targeted transcripts contain editing events with editing frequencies that range from 20-80% and hyper-edited transcripts can contain thousands of combinations of editing events and editing frequencies. The editing repertoire of APOBEC1-targeted transcript 3′UTRs present in a cell at any given moment could be vast, making the resolution of any specific event difficult to assess over a cell population. Therefore although it seems that APOBEC1 editing can have effect on individual transcript fate, further study is required to definitively assess the specific mechanisms behind APOBEC1 editing and its direct consequences for cellular function and host defense.

Chapter 5: Materials and Methods

5.1. Materials and methods for the identification of mRNA editing in BMDMs

5.1.1 Mice and isolation of BMDMs

C57BL/6 littermate or age-matched mice were used at 6-12 weeks of age. *Apobec1*$^{-/-}$ mice were generated as previously described (Hirano et al., 1996) and provided by N. Davidson (Washington University School of Medicine, St. Louis, MO). *Apobec1*$^{-/-}$ mice are healthy, viable and need no specific husbandry. To isolate BMDMs, mice were euthanized by cervical dislocation and the hind leg removed and cleaned of hair, skin and muscle. The bare bone was cleaned with a 70% EtOH. The two ends of the femurs were cut and the marrow was flushed with 22G needle with cold PBS onto a cell strainer. The cells were pelleted and resuspended in macrophage media (DMEM, 10% FBS, 1% Non-essential amino acids (Invitrogen), 0.1% BME, 20ng/mL M-CSF (Peprotech)) and plated onto one untreated 10cm and incubated in a Precursor cells were plated onto untreated 10cm dishes and incubated in humidified 37°C/5% CO_2 incubator overnight. On day 2, cells were counted and re-plated at a concentration of 2 million cells per 10mL macrophage media onto untreated 10cm dishes. Cells were matured for 7 days in, replacing half of the media (with macrophage media supplemented with 40ng/mL M-CSF) every 3 days. Macrophage surface markers: F4/80 (Invitrogen) and Cd11b (BD biosciences) were confirmed via flow cytometry on a FACS Calibur flow cytometer after the 7-day maturation (described in detail in 5.2.6)

5.1.2. RNA extraction and amplification of *Apobec1*, *A1cf* and *Apob*.

BMDMs were washed with cold PBS and RNA was extracted with either Trizol (Invitrogen) or an RNAeasy kit (Qiagen). RNA was normalized and then subjected to DNAse treatment with RQ1 RNAse-free DNAase (Promega). Reverse transcription was performed with SuperScript III (Invitrogen) and either oligo-dT primers, transcript-specific primers or random hexamers, depending on the experiment in question (described below).

For RTPCR analysis of APOBEC1 expression: cDNA was generated using oligo-dT primers. Target specific primers were designed to amplify a 200bp region of APOBEC1, ACF, ApoB and GAPDH. PCR amplification was performed with a Hot-Start Taq Polymerase (Qiagen) and amplicons were run on a 1% agarose gel and visualized.

For qRTPCR analysis of APOBEC1 expression: cDNA was generated using random hexamers. Target-specific primers were designed to amplify a 100bp exon-spanning region of APOBEC1 and a 100bp exon-spanning region of an endogenous control, Rpl32 or GAPDH. qRTPCR was performed using the Sybr Green Master Mix (Life Technologies) and run on a Roche Life cycler 480 system.

5.1.3. LPS stimulation.

BMDMs were derived as described above and then plated onto untreated 6-well plates. Cells were stimulated with 100ng/uL of LPS (Sigma) for 0, 2, 6, 12, and 24 hours. At each time-point, RNA was extracted with the RNAeasy

(Qiagen) kit and cDNA was derived as described above with random hexamer primers. qRTPCR was performed as described above.

5.1.4. mRNA-Seq library preparation

mRNA-Seq library preparation was adapted from standard Illumina protocols. RNA was extracted from macrophage cultures using the Ribopure kit (Ambien). DNAse treatment was performed with Turbo DNAse (Ambien). RNA quality was determined by Bioanalyzer analysis (Agilent Bioanalyzer) and only high quality RNA preps were used for sequencing. RNA was diluted to 10ug in 50uL and poly-A^+ selection was performed with Sera Mag oligo-dT magnetic beads (Thermo), RNA was eluted off the beads with 10mM Tris and analyzed for quality with a Bioanalyzer (Figure 5.1B). All eluted mRNA was fragmented with fragmentation buffer (final composition: 40 mM Tris acetate, pH 8.2, 100 mM potassium acetate, 30 mM magnesium acetate) in a PCR thermocycler at 94°C for 4 min 45s. RNA was washed and concentrated by ethanol precipitation (performed with 5M sodium acetate, pH 5.2 and 100% ethanol) and was analyzed for quality on a Bioanalyzer (Figure 5.1C).

First strand synthesis was performed using a SuperScript III first-strand synthesis system kit (Invitrogen) and was primed with random primers. Second strand synthesis was performed with the Superscript double-stranded cDNA synthesis kit (Invitrogen), which utilizes *E. coli* DNA ligase (10U/uL), *E. coli* DNA polymerase (10U/uL) and *E. coli* RNAse H (2U/uL). After double-stranded synthesis, ends were repaired using dNTPs, T4 DNA polymerase, Klenow DNA polymerase and T4 PNK (All provided by Illlumina). End-repaired cDNA was purified in a PCR purification kit (Qiagen) and eluted in provided EB buffer.

To facilitate PCR adaptor ligation, additional adenosines were added to the 3′ ends of the double-stranded cDNA (ds-cDNA) using Klenow exo (3′ to 5′ exo$^-$) in the presence of dATP (both provided by Illumina). Illumina PCR adaptors were ligated to the ds-cDNA using T4 DNA ligase (Illumina) and cDNA templates were purified by gel electrophoresis on a 2% agarose gel and gel extraction. Fragments that ran from 275-325bp were excised from the gel and DNA was extracted using a gel extraction kit (Qiagen) and eluted in provided EB buffer.

Purified cDNA templates were enriched with 15 cycles of PCR amplification using Illumina PE 1.0 and 2.0 primers and amplified with Phusion DNA polymerase (Qiagen). The concentration and quality of final amplified libraries was determined by Nanodrop spectrophotometer and Bioanalyzer analysis.

5.1.5. RNA-Seq: sequencing, read processing and alignment

Single-end 75nt sequencing was performed on Illumina Genome Analyzer IIx (GAIIx) yielding 28-33 million reads. Initial read quality and trimming was performed with the Fastx toolkit software package (available at http://hannonlab.cshl.edu/fastx_toolkit/index.html). Read quality was analyzed with the Fastx quality stats tool, which revealed a C/T bias at the first base of the reads. Subsequently, the first base was trimmed using the Fastx trimmer tool.

Trimmed reads were mapped to the C57LBL/6 mouse reference genome (NCBI37/mm9) using Tophat (v1.3.3) and Bowtie(v0.12.8) (Trapnell et al., 2009) with the parameters "--solexa1.3-quals -g 1 --coverage-search." The details of the

alignment parameters are as follows: -g 1 suppresses alignments for reads that map to more than 1 location in the reference genome; --coverage-search enables a coverage-based search for junctions and it recommended for read 75bp or higher. The default number of mismatches to the reference genome was used, 2/25b segment or up to 6/ 75bp read.

5.1.6. Identification of APOBEC1 dependent C-to-U mismatches from mRNA-Seq

Pileups were assembled using SAMtools (v0.1.7a). Filters for non-editing SNVs, as described in the text, were implemented with a custom Python script and editing events were visualized with IGV (Robinson et al., 2011; Thorvaldsdóttir et al., 2013). C-to-T (in positive transcripts) or G-to-A (in negative transcripts) mutations occurring at a frequency of 20% in at least 20 reads in the wild-type sample or occurring at a frequency of 10% in 5 reads for the APOBEC1-deficient sample were first identified from SAMtools pileups. Then, these putative sites were put through a series of filters to remove false positives. Specifically, sites were retained if they occurred in the UCSC known gene database, did not occur exclusively in one strand, were more than 50nt from another non C-to-T mutation and were not in the non-isogenic region. The non-isogenic region was determined as the range of putative "editing events" that occurred in the KO sample in chromosome 6. For this experiment is was defined as chr6:56829100-137411200. Real editing events occurring in this relatively large region would be removed from any screen, but as this region is highly variable it is impossible to discern true APOBEC1 editing events from SNPs without considerable standard Sanger sequencing validation. Finally, the SNVs identified

in both the wild-type and APOBEC1-deficient samples were compared and those editing events that also occurred in the knock-out sample were removed.

5.1.7. Designation of single-site editing vs. hyper-editing

Initially some transcripts were identified to have numerous C-to-T changes at high frequency throughout the transcript 3'UTRs. However, some single high-frequency events were also associated with numerous low frequency events, comparable to the " hyper-editing" noticed in enterocyte 3'UTR editing. From this, it was concluded that APOBEC1 editing is always associated with a few up- and down-stream editing events. To differentiate between true hyper-edited and this associated editing, a set of hyper-editing rules were established. Transcripts with 3 or more high-confidence, high-frequency C-to-T changes (editing frequency ≥ 20%) or 2 high-frequency editing events and 3 or more moderate frequency events (editing fraction ≥ 0.09 and ≤ 0.19) were established as "hyper-edited". All the remaining events were characterized as single-site edited, even those with 2 editing events. These rules resulting in a few transcripts with 2 editing events being designated as hyper-edited and a few as single-site edited based on the nature of the surrounding editing events.

5.1.8. Validation of Editing targets

Putative APOBEC1 editing events were validated with standard Sanger sequencing and subclone sequencing of cDNA and gDNA. cDNA was prepared from total RNA using Superscript III (Invitrogen) and oligo-dT priming. For Sanger sequencing, 3'UTRs were amplified from cDNA and genomic DNA using

Pfu Turbo high-fidelity polymerase (Stratagene). Sequencing was performed by GENEWIZ, Inc.. For subclone sequencing, amplicons were cloned into a Strataclone Blunt cloning vector and transformation colonies were selected by blue/white screening with Xgal. Individual colonies were picked and screened by GENEWIZ, Inc. One editing event occurring in a coding region with the frequency of approximately 50%, was thrown out under the presumption it was a genomic SNP. This event is included in the false positive calculation.

5.1.9. Analysis of additional features of APOBEC1 editing

Mooring sequences was identified using a custom python script, using tools in BioPython. In this script, genomic sequences were extracted for the 25bp downstream of the edited cytosine. Sequences of negative transcripts were reverse transcribed. These 25bp sequences were scanned for a perfect mooring motif (WRAUYANUAU) or a motif with 1 deviation (nucleotide mismatch or deletion) or a motif with 2 deviations. Scoring system for the mooring motif was based on mooring sequence distance from the edited cytosine and the sequence fidelity of the motif itself. Scoring system was a scale of 0 (no mooring motif) to 10 (perfect mooring motif): -1pt for 2 bases beyond 4-6nt, -2 points for every mismatch.

Flanking nucleotide analysis was performed as follows: the genomic sequence composition of the 5 nucleotides up and downstream of each edited cytosine was determined and submitted to the logo generating program found at http://weblogo.berkeley.edu/logo.cgi (Kohli et al., 2010).

5.2. Materials and methods specific to the characterization of the consequences of mRNA transcript 3′UTR editing in BMDMs

5.2.1. Conservation analysis

The conservation of APOBEC1 edited regions was calculated using a custom python script and the shuffleBed operation of BEDtools (Quinlan and Hall, 2010). 68 101nt windows surrounding APOBEC1 editing events were defined. For transcripts with more than one editing event, the highest frequency event (as defined from RNA-Seq) was used to eliminate bias. ShuffleBed was used to create sets of 68 101nt windows at random locations in 3′UTR regions of the UCSC known genes. PhastCons scores were obtained from the multialignment of mouse and 19 other placental mammals. The mean phastCon score for APOBEC1-edited windows and 100 random sets of 101nt windows was computed. If the window had no assigned PhastCon scores, it was thrown out. Based on this strategy, one edited window was eliminated from analysis. The mean phastCon score for 67 edit-containing windows was 0.47. The mean score for the random sets was never > 0.47, therefore we report a p value of < 0.01. Conservation analysis was performed with Python coding assistance from Eric Fritz, Papavasiliou lab.

5.2.2. RNA-Seq transcript expression profiling

Using aligned RNA-Seq data, transcript expression levels were calculated using the cuffdiff tool in the Cufflinks (v1.2.1) software package (Trapnell et al., 2013) based on the Ensembl gene set.

5.2.3. Cloning dual-luciferase vectors

For random cDNA cloning analysis: a series of edited and unedited 3′UTRs were amplified using a high-fidelity Pfu Turbo polymerase (Invitrogen) from wild-type ("edited") and *Apobec1*$^{-/-}$ ("unedited") macrophage cDNA (cDNA was generated as descried in 5.1.2) and sub-cloned (Strataclone Blunt PCR cloning kit). Clones were sequenced (GENEWIZ) and representative clones for each degree of editing (single-site, number of hyper-edited events etc.) were inserted downstream of Firefly luciferase in a dual-luciferase vector (Promega pmirGLO dual-expression luciferase vector).

For targeted miRNA analysis: "Un-edited" target 3′UTRs were amplified and cloned as described above. Site-specific mutagenesis was performed on unedited target 3′UTRs. Mutagenesis primers were designed using online tools provided by Agilent technologies (www.genomics.agilent.com). Mutated constructs were amplified with Pfu Turbo (Invitrogen) polymerase in the presence of mutagenesis primers followed by digestion with DpnI (New England Biolabs).

5.2.4. Luciferase reporter assays

Apobec1$^{-/-}$ BMDMs were transfected with dual-luciferase constructs and pmaxGFP transfection control vector using the Amaxa Mouse Macrophage Nucleofactor kit (Lonza). Renilla luciferase served as an internal control. Cells were incubated for 24 hours, lysed with passive lysis buffer by shaking at RT for 15min and subjected to one freeze-thaw cycle at -80°C. Firefly and Renilla luciferase expression were measured using the dual-luciferase reporter system (Promega) and a FLUOstar Omega plate reader. Background luciferase levels

(pmaxGFP) were subtracted from experimental samples. Firefly expression was normalized to Renilla for each construct. Then, luciferase values measured for each "edited" construct was normalized to its "unedited" counterpart for graphical visualization. Significance for the difference between each "edited" and "unedited" pair was determined through a Student's t-test (Excel).

5.2.5. Immunoblotting

BMDMs were lysed in RIPA buffer (50mM Tris pH 8, 150mM NaCl, 0.5% deoxycholate, 0.1% SDS, 1% NP-40, 1mM DTT, protease inhibitor cocktail, 0.5mM PMSF) on ice for 20 min. Protein concentrations were determined via Lowry assay (BioRad) and normalized lysates were run on pre-cast protein gels (BioRad Criterion, Tris-HCl) in running buffer (25mM Tris, 192mM glycine, 0.1% SDS). Proteins were transferred onto a nitrocellulose membrane by semi-dry transfer in transfer buffer (25mM Tris, 192mM glycine, 0.1% SDS, 20% methanol). Membranes were blocked in 5% milk in TBS-T and incubated with primary antibodies: App (clone 22C11, Millipore), B2m (Abcam), tubulin (Sigma, clone DM1A).

5.2.6. Flow cytometry

BMDMs were removed from 10cm dishes by scraping or vigorous pipetting and pelleted by centrifugation (1000xg, 3 min). Cells were incubated with an Fc-receptor blocking antibody (BD Biosciences) for 10min at 4°C.

For the staining of cell-surface markers: BMDMs were then incubated with the following fluorophore-conjugated primary antibodies:

1) Macrophage-specific cell surface markers: F4/80 (conjugated to AF-488, Invitrogen), Cd11b (conjugated to PE, BD biosciences).
2) Live:dead stain: 7AAD (BD Biosciences)
3) CD36 antibody conjugated to APC; isotype control conjugated to APC (BD biosciences)

For the staining of intracellular proteins: BMDMs were fixed and permeabilized according to parameters defined in the Cytofix/Cytoperm kit (BD biosciences). Lamp1 antibodies were conjugated to FITC and compared to a FITC isotype control (BD biosciences). No live/dead staining was used and the same Cd11b-PE antibody from above and a comparable F4/80 conjugated to APC were used to isolate macrophages.

5.2.7. FACS

BMDMs were stained for F4/80, Cd11b and CD36 as described above. Cells were sorted using a BD-FACS Aria machine for the top ~10% and bottom ~10% cells expressing CD36. RNA was extracted, reverse transcribed, PCR amplified with CD36 specific primers and subjected to subclone sequencing as previously described.

5.2.8. HTS-CLIP protocol

HTS-CLIP analysis was performed as previously described (Chi et al., 2009) with a few alterations. HITS-CLIP sample preparation was performed by Dr. Emily Conn Gantman in the Darnell lab. A brief summary of the CLIP protocol adapted from materials provided by Dr. Conn Gantman is described below.

BMDMs were prepared from 3 wild-type and APOBEC1$^{-/-}$ littermate pairs. BMDMs were matured as described above and crosslinked 3x at 200mJ/cm^2 on the original maturation plates in 3mL of 1xPBS on a bed of ice. BMDMs were scrapped off the plates, flash-frozen and stored at -80C. Frozen cells were thawed and lysed in 1mL PXL lysis buffer (1x PBS, 0.1% SDS, 0.5% Na-DOC, 0.5% NP-40) with complete protease inhibitor. After lysing on ice with occasional vortexing, the cells were subjected to DNAse treatment with RQ1 RNAse-free DNAse (Promega) for 5 minutes at 37°C shaking in a Thermomixer. RNAse treatment performed with high (1:100) or low (1:10,000) RNAse A solutions and incubated for 5 minutes shaking at 37°C. Lysates were spun down (14,000RPM, 30minutes at 4°C) and the supernatant transferred to a new tube.

Protein A beads were pre-loaded with rabbit anti-mouse IgG bridging antibody (Jackson ImmunoResearch, at 2.3 mg/mL), incubated for 35 minutes at room temperature. Beads were washed (in 0.02% Tween) and then loaded with 3 uL Ago antibody (2A8 ascites provided by Dr. Zissimos Mourelatos- (Nelson et al., 2007)) per 400 uL of beads and rotated for 3 hours at 4°C. Cleared lysates were incubated with primary antibody-loaded beads for 2 hours at 4°C and washed in a series of washes: 1) 2x 1x PXL buffer; 2) 2x 5x PXL (5xPBS 0.1 % SDS, 0.5% Na-DOC, 0.5% NP-40); 3) 2x with high stringency buffer (15mM Tris-HCl, pH 7.5, 5mM EDTA, 2.5mM EGTA, 1% Triton X-100, 1% Na-DOC, 0.1% SDS, 120mM NaCl, 25mM KCl); 4) 1x with high salt buffer (15mM Tris-HCl, pH 7.5, 5mM EDTA, 2.5mM EGTA, 1% Triton X-100, 1% Na-DOC, 0.1% SDS, 1M NaCl); 5) 2x with low salt buffer (15mM Tris-HCl, pH 7.5, 5mM EDTA); 6) 2x NT-2

buffer (50mM Tris-HCl pH 7.4, 150mM NaCl, 1mM $MgCl_2$, 0.05% NP-40); 7) 2x PNK buffer (50mM Tris-HCl, pH 7.5, 10mM $MgCl_2$, 0.5% NP-40).

Dephosphorylation of the 5′ phosphate was performed with calf intestinal phosphatase (Roche) with RNAsin Plus RNAse Inhibitor (Promega) at 37°C for 20 minutes, shaking at 1000rpm for 15s every 2 min. Beads were then washed 1) 1x PNK buffer 2) 2x PNK buffer supplemented with 20mM EGTA 3) 2x PNK buffer.

The puromycin blocked linker was radio-labeled with T4 PNK (New England Biolabs) and ^{32}P-γ-ATP for 30 min at 37°C. Radiolabeled linker was spun through a G-25 column to remove free-ATP. Labeled 3′RNA linker was ligated to the 3′ end of the RNA with T4 RNA ligase (Fermentas) incubated at 16°C for 1 hour, shaking at 1000rpm for 15s every 4 min. After 1 hour, 80 pmole of cold L32 RNA linker with 5′ phosophate was added. Sampels were incubated overnight and washed 3x with PNK buffer. 5′ ends were re-phosphorylated with with T4 PNK for 20 min at 37°C shaking at 1000rpm for 15s every 4 min.

Protein:RNA complexes were eluted off the beads in NuPAGE loading buffer (Invitrogen) at 70°C for 10 min shaking at 1000rpm. Supernatants were run on Novex NuPAGE 8% Bis-Tris gels (Invitrogen) in MOPS running buffer (Invitrogen) for 2hr at 175V and transferred onto Protran BA85 nitrocellulose (Whatman) and exposed to Biomax MR film (Kodak).

Regions that corresponded to Ago:mRNA and Ago:miRNA complexes were excised from the membrane (Figure 3.8A), diced and treated with proteinase K (4mg/mL Roche) for 20 min at 37°C shaking at 1100rpm. RNA was then extracted via phenol-chloroform extraction and ethanol precipitation. 5′

linkers with a degenerate nucleotide end were ligated to the extracted RNA with T4 RNA ligase (Fermentas) at 16°C for 5 hours. The ligated reaction was then subjected to DNAse treatement with RQ1 DNAse (Promega) for 20min at 37°C and extracted with phenol-chloroform and subjected to ethanol precipitation.

Precipitated RNA was reverse transcribed with Superscript III (Invitrogen) and PCR amplified with Accuprime Pfx Supermix (Invitrogen) for 20-35 cycles. PCR products were run on a 10% denaturing polacrylamide gel and visualized with SYBR Gold (Molecular Probes) staining. 60-100nt products were excised from the lowest cycle number with visual product and gel extracted. Additional PCR amplification followed by gel extraction was performed as described with fusion primers to provide the platform for Illumina sequencing. 10-30uL of 10nM DNA was submitted for sequencing on Illumina HiSeq.

5.2.9. Processing and alignment of HITS-CLIP reads

HITS-CLIP reads were filtered by quality (the first 5 nucleotides had a minimum quality score of 15 and the next 45 had a minimum mean score of 15) and exact sequences were collapsed. The 5′ linker was stripped off and Illumina adapter sequences were clipped from the 3′ end (Fastx Toolkit). Reads were then parsed by size into mRNA (≥25nt) and miRNA(≥ 17nt and ≤ 24nt) fractions using a custom python script. miRNA reads were aligned to mm9 using bowtie (v0.12.8) with the following specifications: "-l 17 -v 2 --best --strata -m 12". To determine the best alignment strategy for his highly duplicated dataset we determined that the mmu-miRNAs mapped to a maximum of 12 separate genomic positions, so therefore we allowed up to 12 alignments per read. Read

counts were quantified with SeqMonk (http://www.bioinformatics.babraham.ac.uk/projects/seqmonk/) using miRNA intervals defined by miRBase (v18). miRNAs mapping to multiple positions in the genome were then collapsed.

mRNA reads were uniquely aligned to mm9 using bowtie (v0.12.8) with the following options: "-v 2 --best --strata -m 1". A second step of PCR duplicate removal was performed as previously described (Chi et al., 2012) in which reads with the same 5′ 5nt degenerate linker and the same coordinates were removed. This step eliminates true PCR duplicates, in which sequencing errors were also introduced, and would therefore be missed by an exact sequence collapser.

5.2.10. Identifying Ago footprints

Clusters were defined as regions with ≥ 8 reads that overlapped by at least 5 nucleotides. The peaks of the clusters were identified as previously described (Chi et al., 2009). The "Ago footprint" around the cluster has been previously broadly defined as the region 32nt upstream and 30nt downstream from the peak (previously characterized in (Chi et al., 2009) as the region in which Ago in bound 95% of the time) or narrowly defined as the region 22nt upstream and 24nt downstream (region in which Ago is bound 100% of the time). For this analysis, we identified APOBEC1:Ago overlap by looking at the broad definition of the footprint but identified potential miRNA target regions within the narrow footprint. The read depth of each footprint from HITS-CLIP and RNA-Seq was calculated using SeqMonk and CLIP depth was normalized to transcript expression (RNA-Seq read depth) to define the "Ago occupancy" (reads per million mapped CLIP/reads per million mapped RNA-Seq). This method should

provide accurate normalization, as it takes into account regions within each transcript that could be differentially expressed or differentially mapped, a phenomenon neglected when normalizing to total transcript expression. Ago footprints were filtered to 17,477 that were contained within "expressed" regions, defined as an RPM of greater than or equal to 1, as this provided reasonable coverage of these regions, and would eliminate the any problems with artifactually increasing Ago footprint occupancy values by normalizing to a value less than 1. The biological complexity (the number of replicates contributing to each Ago Footprint) was calculated based on the replicate contributing ≥2 reads to the footprint. "High-confidence" footprints (14,781) were defined as having a biological complexity of ≥ 2. As all APOBEC1 editing and most miRNA targeting happening within the 3′UTR of a given transcript, we narrowed our search to only 3′UTRs. We generated a permissive 3′UTR database from RefSeq by merging (Bedtools, mergeBed) overlapping 3′UTR regions, thereby defining a region as a 3′UTR if it was catalogued as such in any transcript isoform. 6,270 high-confidence footprints were contained within these merged 3′UTRs. To determine the additional genomic locations of footprints not located in 3′UTRs, a similar "merged" database of 5′UTRs and CDS were created. Footprints that do not occur in either 5′UTRs or CDS are listed as "other."

5.2.11 Identifying overlap between Ago and APOBEC1 targeting

Overlap between Ago footprints and APOBEC1 editing was determined with the intersectBed function of Bedtools. The over-representation of APOBEC1 editing events in CLIP footprints was calculated as follows. The fraction of the

total 3′UTR in which APOBEC1 editing could be identified (or "accessible 3′UTRs") was defined as regions of well-expressed 3′UTRs that were covered by ≥ 20 reads. The number of nucleotides in the accessible 3′UTRs was calculated (A). The number of nucleotides in footprints that overlapped with accessible 3′UTRs was also calculated (B). If we consider the number of editing events identified as C and the number of editing events within Ago footprints and D. Then the over-representation ratio of APOBEC1 editing events in footprints was calculated as (D/B)/(C/A). Odds ratio was calculated (D/B)/((C-D)/(A-B)) with significance and confidence intervals calculated in R.

5.2.12 miRNA seed target search

After the identification of high-confidence Ago footprints, we performed an exhaustive search to assign miRNA targets to those footprints and identify regions where APOBEC1 editing could create or destroy a miRNA target region. The search for miRNA target regions was performed with coding assistance from Dr. Dewi Harjanto, Papavasiliou lab. Using miRNA alignment data, we generated a list of bound miRNAs, defined as those that had a biological complexity of 3 in one of the two genotypes. Using a custom Python script, we scanned the footprint sequences for "canonical" matched miRNA 6mer seed regions (positions 2-7 of the mature miRNA sequence), as well as other non-canonical matched 6mers from the 5′ end of the mature sequence (positions 1-6 or 3-8). We also scanned the footprint sequences for "fuzzy" 6mers and 7mers (1 nucleotide mismatch) and G-bulge seed regions (1 G insertion). Footprints were analyzed for the "best" miRNA target region fit, based on the sequence proximity to the footprint peak and the amount of the miRNA bound to Ago. We

identified a number of target regions that were either created or destroyed by APOBEC1 editing events and tested these with standard luciferase reporter assay (described above).

REFERENCES

Alon, S., Mor, E., Vigneault, F., Church, G.M., Locatelli, F., Galeano, F., Gallo, A., Shomron, N., and Eisenberg, E. (2012). Systematic identification of edited microRNAs in the human brain. Genome Res *22*, 1533–1540.

Alon, S., Vigneault, F., Eminaga, S., Christodoulou, D.C., Seidman, J.G., Church, G.M., and Eisenberg, E. (2011). Barcoding bias in high-throughput multiplex sequencing of miRNA. Genome Res *21*, 1506–1511.

Anant, S., and Davidson, N.O. (2000). An AU-rich sequence element (UUUN[A/U]U) downstream of the edited C in apolipoprotein B mRNA is a high-affinity binding site for Apobec-1: binding of Apobec-1 to this motif in the 3' untranslated region of c-myc increases mRNA stability. Mol Cell Biol *20*, 1982–1992.

Anant, S., Henderson, J.O., Mukhopadhyay, D., Navaratnam, N., Kennedy, S., Min, J., and Davidson, N.O. (2001). Novel role for RNA-binding protein CUGBP2 in mammalian RNA editing. CUGBP2 modulates C to U editing of apolipoprotein B mRNA by interacting with apobec-1 and ACF, the apobec-1 complementation factor. J Biol Chem *276*, 47338–47351.

Anant, S., MacGinnitie, A.J., and Davidson, N.O. (1995). apobec-1, the catalytic subunit of the mammalian apolipoprotein B mRNA editing enzyme, is a novel RNA-binding protein. J Biol Chem *270*, 14762–14767.

Anant, S., Murmu, N., Houchen, C.W., Mukhopadhyay, D., Riehl, T.E., Young, S.G., Morrison, A.R., Stenson, W.F., and Davidson, N.O. (2004). Apobec-1 protects intestine from radiation injury through posttranscriptional regulation of cyclooxygenase-2 expression. Gastroenterology *127*, 1139–1149.

Armitage, A.E., Katzourakis, A., de Oliveira, T., Welch, J.J., Belshaw, R., Bishop, K.N., Kramer, B., McMichael, A.J., Rambaut, A., and Iversen, A.K.N. (2008). Conserved footprints of APOBEC3G on Hypermutated human immunodeficiency virus type 1 and human endogenous retrovirus HERV-K(HML2) sequences. J Virol *82*, 8743–8761.

Athanasiadis, A., Rich, A., and Maas, S. (2004). Widespread A-to-I RNA editing of Alu-containing mRNAs in the human transcriptome. PLoS Biol *2*, e391.

Backus, J.W., and Smith, H.C. (1991). Apolipoprotein B mRNA sequences 3' of the editing site are necessary and sufficient for editing and editosome assembly. Nucleic Acids Res *19*, 6781–6786.

Backus, J.W., and Smith, H.C. (1992). Three distinct RNA sequence elements are required for efficient apolipoprotein B (apoB) RNA editing in vitro. Nucleic Acids Res *20*, 6007–6014.

Bahn, J.H., Lee, J.-H., Li, G., and al, E. Accurate identification of A-to-I RNA

editing in human by transcriptome sequencing.

Bahn, J.H., Lee, J.-H., Li, G., Greer, C., Peng, G., and Xiao, X. (2012). Accurate identification of A-to-I RNA editing in human by transcriptome sequencing. Genome Res *22*, 142–150.

Barnes, C., and Smith, H.C. (1993). Apolipoprotein B mRNA editing in vitro is a zinc-dependent process. Biochem Biophys Res Commun *197*, 1410–1414.

Bass, B.L., and Weintraub, H. (1987). A developmentally regulated activity that unwinds RNA duplexes. Cell *48*, 607–613.

Bass, B.L., and Weintraub, H. (1988). An unwinding activity that covalently modifies its double-stranded RNA substrate. Cell *55*, 1089–1098.

Bass, B.L. (2002). RNA editing by adenosine deaminases that act on RNA. Annu Rev Biochem *71*, 817–846.

Bassell, G.J., Zhang, H., Byrd, A.L., Femino, A.M., Singer, R.H., Taneja, K.L., Lifshitz, L.M., Herman, I.M., and Kosik, K.S. (1998). Sorting of beta-actin mRNA and protein to neurites and growth cones in culture. J Neurosci *18*, 251–265.

Benne, R., Van den Burg, J., Brakenhoff, J.P., Sloof, P., Van Boom, J.H., and Tromp, M.C. (1986). Major transcript of the frameshifted coxII gene from trypanosome mitochondria contains four nucleotides that are not encoded in the DNA. Cell *46*, 819–826.

Besmer, E., Market, E., and Papavasiliou, F.N. (2006). The transcription elongation complex directs activation-induced cytidine deaminase-mediated DNA deamination. Mol Cell Biol *26*, 4378–4385.

Bhutani, N., Brady, J.J., Damian, M., Sacco, A., Corbel, S.Y., and Blau, H.M. (2010). Reprogramming towards pluripotency requires AID-dependent DNA demethylation. Nature *463*, 1042–1047.

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.

Blanc, V., Henderson, J.O., Kennedy, S., and Davidson, N.O. (2001a). Mutagenesis of apobec-1 complementation factor reveals distinct domains that modulate RNA binding, protein-protein interaction with apobec-1, and complementation of C to U RNA-editing activity. J Biol Chem *276*, 46386–46393.

Blanc, V., Navaratnam, N., Henderson, J.O., Anant, S., Kennedy, S., Jarmuz, A., Scott, J., and Davidson, N.O. (2001b). Identification of GRY-RBP as an apolipoprotein B RNA-binding protein that interacts with both apobec-1 and apobec-1 complementation factor to modulate C to U editing. J Biol Chem *276*, 10272–10283.

Blanc, V., and Davidson, N.O. (2010). APOBEC-1-mediated RNA editing. Wiley Interdiscip Rev Syst Biol Med *2*, 594–602.

Blanc, V., Henderson, J.O., Newberry, E.P., Kennedy, S., Luo, J., and Davidson, N.O. (2005). Targeted deletion of the murine apobec-1 complementation factor (acf) gene results in embryonic lethality. Mol Cell Biol *25*, 7260–7269.

Blanc, V., Henderson, J.O., Newberry, R.D., Xie, Y., Cho, S.-J., Newberry, E.P., Kennedy, S., Rubin, D.C., Wang, H.L., Luo, J., et al. (2007). Deletion of the AU-rich RNA binding protein Apobec-1 reduces intestinal tumor burden in Apc(min) mice. Cancer Res *67*, 8565–8573.

Blanc, V., Kennedy, S., and Davidson, N.O. (2003). A novel nuclear localization signal in the auxiliary domain of apobec-1 complementation factor regulates nucleocytoplasmic import and shuttling. J Biol Chem *278*, 41198–41204.

Blanc, V., Xie, Y., Luo, J., Kennedy, S., and Davidson, N.O. (2012). Intestine-specific expression of Apobec-1 rescues apolipoprotein B RNA editing and alters chylomicron production in Apobec1$^{-/-}$ mice. J Lipid Res *53*, 2643–2655.

Blow, M. (2004). A survey of RNA editing in human brain. Genome Res *14*, 2379–2387.

Blow, M.J., Grocock, R.J., van Dongen, S., Enright, A.J., Dicks, E., Futreal, P.A., Wooster, R., and Stratton, M.R. (2006). RNA editing of human microRNAs. Genome Biol *7*, R27.

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

Borchert, G.M., Gilmore, B.L., Spengler, R.M., Xing, Y., Lanier, W., Bhattacharya, D., and Davidson, B.L. (2009). Adenosine deamination in human transcripts generates novel microRNA binding sites. Hum Mol Genet *18*, 4801–4807.

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 USA *100*, 4102–4107.

Burns, C.M., Chu, H., Rueter, S.M., Hutchinson, L.K., Canton, H., Sanders-Bush, E., and Emeson, R.B. (1997). Regulation of serotonin-2C receptor G-protein coupling by RNA editing. Nature *387*, 303–308.

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.

Casey, J.L., and Gerin, J.L. (1995). Hepatitis D virus RNA editing: specific modification of adenosine in the antigenomic RNA. J Virol *69*, 7593–7600.

Cattaneo, R. (1994). Biased (A-->I) hypermutation of animal RNA virus genomes. Curr Opin Genet Dev *4*, 895–900.

Cattenoz, P.B., Taft, R.J., Westhof, E., and Mattick, J.S. (2013). Transcriptome-wide identification of A > I RNA editing sites by inosine specific cleavage. Rna *19*, 257–270.

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

Chang, F.L., Chen, P.J., Tu, S.J., Wang, C.J., and Chen, D.S. (1991). The large form of hepatitis delta antigen is crucial for assembly of hepatitis delta virus. Proc Natl Acad Sci USA *88*, 8490–8494.

Chao, M., Hsieh, S.Y., and Taylor, J. (1990). Role of two forms of hepatitis delta virus antigen: evidence for a mechanism of self-limiting genome replication. J Virol *64*, 5066–5069.

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

Chen, C.X., Cho, D.S., Wang, Q., Lai, F., Carter, K.C., and Nishikura, K. (2000). A third member of the RNA-specific adenosine deaminase gene family, ADAR3, contains both single- and double-stranded RNA binding domains. Rna *6*, 755–767.

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, L., Li, Y., Lin, C.H., Chan, T.H.M., Chow, R.K.K., Song, Y., Liu, M., Yuan, Y.-F., Fu, L., Kong, K.L., et al. (2013). Recoding RNA editing of AZIN1 predisposes to hepatocellular carcinoma. Nat Med *19*, 209–216.

Chen, L.-L., and Carmichael, G.G. (2009). Altered nuclear retention of mRNAs containing inverted repeats in human embryonic stem cells: functional role of a nuclear noncoding RNA. Mol Cell *35*, 467–478.

Chen, L.-L., DeCerbo, J.N., and Carmichael, G.G. (2008). Alu element-mediated gene silencing. The EMBO Journal *27*, 1694–1705.

Chen, S.H., Habib, G., Yang, C.Y., Gu, Z.W., Lee, B.R., Weng, S.A., Silberman, S.R., Cai, S.J., Deslypere, J.P., and Rosseneu, M. (1987). Apolipoprotein B-48 is the product of a messenger RNA with an organ-specific in-frame stop codon. Science *238*, 363–366.

Chen, S.H., Li, X.X., Liao, W.S., Wu, J.H., and Chan, L. (1990). RNA editing of apolipoprotein B mRNA. Sequence specificity determined by in vitro coupled transcription editing. J Biol Chem *265*, 6811–6816.

Cheng, H.-L., Vuong, B.Q., Basu, U., Franklin, A., Schwer, B., Astarita, J., Phan, R.T., Datta, A., Manis, J., Alt, F.W., et al. (2009). Integrity of the AID serine-38 phosphorylation site is critical for class switch recombination and somatic hypermutation in mice. Proceedings of the National Academy of Sciences *106*, 2717–2722.

Chester, A., Somasekaram, A., and Tzimina, M. (2003). The apolipoprotein B mRNA editing complex performs a multifunctional cycle and …. The EMBO Journal.

Chester, A., Weinreb, V., Carter, C.W., and Navaratnam, N. (2004). Optimization of apolipoprotein B mRNA editing by APOBEC1 apoenzyme and the role of its auxiliary factor, ACF. Rna *10*, 1399–1411.

Chi, S.W., Hannon, G.J., and Darnell, R.B. (2012). An alternative mode of microRNA target recognition. Nat Struct Mol Biol *19*, 321–327.

Chi, S.W., Zang, J.B., Mele, A., and Darnell, R.B. (2009). Argonaute HITS-CLIP decodes microRNA-mRNA interaction maps. Nature *460*, 479–486.

Chiarle, R., Zhang, Y., Frock, R.L., Lewis, S.M., Molinie, B., Ho, Y.-J., Myers, D.R., Choi, V.W., Compagno, M., Malkin, D.J., et al. (2011). Genome-wide translocation sequencing reveals mechanisms of chromosome breaks and rearrangements in B cells. Cell *147*, 107–119.

Clerzius, G., Gélinas, J.-F., Daher, A., Bonnet, M., Meurs, E.F., and Gatignol, A. (2009). ADAR1 interacts with PKR during human immunodeficiency virus infection of lymphocytes and contributes to viral replication. J Virol *83*, 10119–10128.

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

Conticello, S.G., Thomas, C.J.F., 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.

Dance, G.S.C., Sowden, M.P., Cartegni, L., Cooper, E., Krainer, A.R., and Smith, H.C. (2002). Two proteins essential for apolipoprotein B mRNA editing are expressed from a single gene through alternative splicing. J Biol Chem *277*, 12703–12709.

Danecek, P., Nellåker, C., McIntyre, R.E., Buendia-Buendia, J.E., Bumpstead, S., Ponting, C.P., Flint, J., Durbin, R., Keane, T.M., and Adams, D.J. (2012). High levels of RNA-editing site conservation amongst 15 laboratory mouse strains. Genome Biol *13*, 26.

Dawson, T.R., Sansam, C.L., and Emeson, R.B. (2004). Structure and sequence determinants required for the RNA editing of ADAR2 substrates. J Biol Chem *279*, 4941–4951.

de Yebenes, V.G., Belver, L., Pisano, D.G., González, S., Villasante, A., Croce, C., He, L., and Ramiro, A.R. (2008). miR-181b negatively regulates activation-induced cytidine deaminase in B cells. J Exp Med *205*, 2199–2206.

Delker, R.K., Fugmann, S.D., and Papavasiliou, F.N. (2009). A coming-of-age story: activation-induced cytidine deaminase turns 10. Nat Immunol *10*, 1147–1153.

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

Dickerson, S.K., Market, E., Besmer, E., and Papavasiliou, F.N. (2003). AID mediates hypermutation by deaminating single stranded DNA. J Exp Med *197*, 1291–1296.

Doria, M., Neri, F., Gallo, A., Farace, M.G., and Michienzi, A. (2009). Editing of HIV-1 RNA by the double-stranded RNA deaminase ADAR1 stimulates viral infection. Nucleic Acids Res *37*, 5848–5858.

Dorsett, Y., McBride, K.M., Jankovic, M., Gazumyan, A., Thai, T.-H., Robbiani, D.F., Di Virgilio, M., Reina-San-Martin, B., Heidkamp, G., Schwickert, T.A., et al. (2008). MicroRNA-155 suppresses activation-induced cytidine deaminase-mediated Myc-Igh translocation. Immunity *28*, 630–638.

Driscoll, D.M., Lakhe-Reddy, S., Oleksa, L.M., and Martinez, D. (1993). Induction of RNA editing at heterologous sites by sequences in apolipoprotein B mRNA. Mol Cell Biol *13*, 7288–7294.

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.

Etard, C., Roostalu, U., and Strähle, U. (2010). Lack of Apobec2-related proteins causes a dystrophic muscle phenotype in zebrafish embryos. J Cell Biol *189*, 527–539.

Fritz, E.L., Rosenberg, B.R., Lay, K., Mihailović, A., Tuschl, T., and Papavasiliou, F.N. (2013). A comprehensive analysis of the effects of the deaminase AID on the transcriptome and methylome of activated B cells. Nat Immunol *14*, 749–755.

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.

Galloway, C.A., Kumar, A., Krucinska, J., and Smith, H.C. (2010). APOBEC-1 complementation factor (ACF) forms RNA-dependent multimers. Biochem Biophys Res Commun *398*, 38–43.

Gee, P., Ando, Y., Kitayama, H., Yamamoto, S.P., Kanemura, Y., Ebina, H., Kawaguchi, Y., and Koyanagi, Y. (2011). APOBEC1-mediated editing and attenuation of herpes simplex virus 1 DNA indicate that neurons have an antiviral role during herpes simplex encephalitis. J Virol *85*, 9726–9736.

George, C.X., and Samuel, C.E. (1999a). Human RNA-specific adenosine deaminase ADAR1 transcripts possess alternative exon 1 structures that initiate from different promoters, one constitutively active and the other interferon inducible. Proc Natl Acad Sci USA *96*, 4621–4626.

George, C.X., and Samuel, C.E. (1999b). Characterization of the 5'-flanking region of the human RNA-specific adenosine deaminase ADAR1 gene and identification of an interferon-inducible ADAR1 promoter. Gene *229*, 203–213.

George, C.X., Wagner, M.V., and Samuel, C.E. (2005). Expression of interferon-inducible RNA adenosine deaminase ADAR1 during pathogen infection and mouse embryo development involves tissue-selective promoter utilization and alternative splicing. J Biol Chem *280*, 15020–15028.

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

Gerber, A., O'Connell, M.A., and Keller, W. (1997). Two forms of human double-stranded RNA-specific editase 1 (hRED1) generated by the insertion of an Alu cassette. Rna *3*, 453–463.

Glenn, J.S., and White, J.M. (1991). trans-dominant inhibition of human hepatitis delta virus genome replication. J Virol *65*, 2357–2361.

Gonzalez, M.C., Suspène, R., Henry, M., Guétard, D., Wain-Hobson, S., and Vartanian, J.-P. (2009). Human APOBEC1 cytidine deaminase edits HBV DNA. Retrovirology *6*, 96.

Gordon, M.S., Kanegai, C.M., Doerr, J.R., and Wall, R. (2003). Somatic hypermutation of the B cell receptor genes B29 (Igbeta, CD79b) and mb1 (Igalpha, CD79a). Proc Natl Acad Sci USA *100*, 4126–4131.

Gott, J.M., and Emeson, R.B. (2000). Functions and mechanisms of RNA editing. Annu. Rev. Genet. *34*, 499–531.

Greeve, J., Altkemper, I., Dieterich, J.H., Greten, H., and Windler, E. (1993). Apolipoprotein B mRNA editing in 12 different mammalian species: hepatic expression is reflected in low concentrations of apoB-containing plasma

lipoproteins. J Lipid Res *34*, 1367–1383.

Greeve, J., Lellek, H., Rautenberg, P., and Greten, H. (1998). Inhibition of the apolipoprotein B mRNA editing enzyme-complex by hnRNP C1 protein and 40S hnRNP complexes. Biol. Chem. *379*, 1063–1073.

Grimson, A., Farh, K.K.-H., Johnston, W.K., Garrett-Engele, P., Lim, L.P., and Bartel, D.P. (2007). MicroRNA targeting specificity in mammals: determinants beyond seed pairing. Mol Cell *27*, 91–105.

Gu, T., Buaas, F.W., Simons, A.K., Ackert-Bicknell, C.L., Braun, R.E., and Hibbs, M.A. (2012). Canonical A-to-I and C-to-U RNA editing is enriched at 3'UTRs and microRNA target sites in multiple mouse tissues. PLoS ONE *7*, e33720.

Gu, W., Pan, F., Zhang, H., Bassell, G.J., and Singer, R.H. (2002). A predominantly nuclear protein affecting cytoplasmic localization of beta-actin mRNA in fibroblasts and neurons. J Cell Biol *156*, 41–51.

Guo, F., Cen, S., Niu, M., Saadatmand, J., and Kleiman, L. (2006). Inhibition of formula-primed reverse transcription by human APOBEC3G during human immunodeficiency virus type 1 replication. J Virol *80*, 11710–11722.

Guo, F., Cen, S., Niu, M., Yang, Y., Gorelick, R.J., and Kleiman, L. (2007). The interaction of APOBEC3G with human immunodeficiency virus type 1 nucleocapsid inhibits tRNA3Lys annealing to viral RNA. J Virol *81*, 11322–11331.

Hamilton, C.E., Papavasiliou, F.N., and Rosenberg, B.R. (2010). Diverse functions for DNA and RNA editing in the immune system. RNA Biol *7*, 220–228.

Hao, S., and Baltimore, D. (2009). The stability of mRNA influences the temporal order of the induction of genes encoding inflammatory molecules. Nat Immunol *10*, 281–288.

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, S.G., Sabio, I., Mayer, E., Steinberg, M.F., Backus, J.W., Sparks, J.D., Sparks, C.E., and Smith, H.C. (1993). Extract-specific heterogeneity in high-order complexes containing apolipoprotein B mRNA editing activity and RNA-binding proteins. J Biol Chem *268*, 7382–7392.

Hartner, J.C., Schmittwolf, C., Kispert, A., Müller, A.M., Higuchi, M., and Seeburg, P.H. (2004). Liver disintegration in the mouse embryo caused by deficiency in the RNA-editing enzyme ADAR1. J Biol Chem *279*, 4894–4902.

Hartner, J.C., Walkley, C.R., Lu, J., and Orkin, S.H. (2009). ADAR1 is essential for the maintenance of hematopoiesis and suppression of interferon signaling. Nat Immunol *10*, 109–115.

Hartwig, D., Schütte, C., Warnecke, J., Dorn, I., Hennig, H., Kirchner, H., and Schlenke, P. (2006). The large form of ADAR 1 is responsible for enhanced hepatitis delta virus RNA editing in interferon-alpha-stimulated host cells. J Viral Hepat *13*, 150–157.

Hartwig, D., Schoeneich, L., Greeve, J., Schütte, C., Dorn, I., Kirchner, H., and Hennig, H. (2004). Interferon-alpha stimulation of liver cells enhances hepatitis delta virus RNA editing in early infection. J Hepatol *41*, 667–672.

Hasham, M.G., Donghia, N.M., Coffey, E., Maynard, J., Snow, K.J., Ames, J., Wilpan, R.Y., He, Y., King, B.L., and Mills, K.D. (2010). Widespread genomic breaks generated by activation-induced cytidine deaminase are prevented by homologous recombination. Nat Immunol *11*, 820–826.

Helwak, A., Kudla, G., Dudnakova, T., and Tollervey, D. (2013). Mapping the Human miRNA Interactome by CLASH Reveals Frequent Noncanonical Binding. Cell *153*, 654–665.

Hersberger, M., and Innerarity, T.L. (1998). Two efficiency elements flanking the editing site of cytidine 6666 in the apolipoprotein B mRNA support mooring-dependent editing. J Biol Chem *273*, 9435–9442.

Higuchi, M., Maas, S., Single, F.N., Hartner, J., Rozov, A., Burnashev, N., Feldmeyer, D., Sprengel, R., and Seeburg, P.H. (2000). Point mutation in an AMPA receptor gene rescues lethality in mice deficient in the RNA-editing enzyme ADAR2. Nature *406*, 78–81.

Higuchi, M., Single, F.N., Köhler, M., Sommer, B., Sprengel, R., and Seeburg, P.H. (1993). RNA editing of AMPA receptor subunit GluR-B: a base-paired intron-exon structure determines position and efficiency. Cell *75*, 1361–1370.

Hirano, K., Min, J., Funahashi, T., and Davidson, N.O. (1997). Cloning and characterization of the rat apobec-1 gene: a comparative analysis of gene structure and promoter usage in rat and mouse. J Lipid Res *38*, 1103–1119.

Hirano, K., Young, S.G., Farese, R.V., Ng, J., Sande, E., Warburton, C., Powell-Braxton, L.M., and Davidson, N.O. (1996). Targeted disruption of the mouse apobec-1 gene abolishes apolipoprotein B mRNA editing and eliminates apolipoprotein B48. J Biol Chem *271*, 9887–9890.

Hoffman, Y., Dahary, D., Bublik, D.R., Oren, M., and Pilpel, Y. (2013). The majority of endogenous microRNA targets within Alu elements avoid the microRNA machinery. Bioinformatics *29*, 894–902.

Huang, C., Chang, S.C., Yang, H.-C., Chien, C.-L., and Chang, M.-F. (2009). Clathrin-mediated post-Golgi membrane trafficking in the morphogenesis of hepatitis delta virus. J Virol *83*, 12314–12324.

Hughes, M.E., Grant, G.R., Paquin, C., Qian, J., and Nitabach, M.N. (2012). Deep sequencing the circadian and diurnal transcriptome of Drosophila brain. Genome

Res *22*, 1266–1281.

Hundley, H.A., Krauchuk, A.A., and Bass, B.L. (2008). C. elegans and H. sapiens mRNAs with edited 3' UTRs are present on polysomes. Rna *14*, 2050–2060.

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.

Jayan, G.C., and Casey, J.L. (2002). Inhibition of hepatitis delta virus RNA editing by short inhibitory RNA-mediated knockdown of ADAR1 but not ADAR2 expression. J Virol *76*, 12399–12404.

Kane, J.P. (1983). Apolipoprotein B: structural and metabolic heterogeneity. Annu. Rev. Physiol. *45*, 637–650.

Kawahara, Y., and Nishikura, K. (2006). Extensive adenosine-to-inosine editing detected in Alu repeats of antisense RNAs reveals scarcity of sense-antisense duplex formation. FEBS Lett. *580*, 2301–2305.

Kawahara, Y., Megraw, M., Kreider, E., Iizasa, H., Valente, L., Hatzigeorgiou, A.G., and Nishikura, K. (2008). Frequency and fate of microRNA editing in human brain. Nucleic Acids Res *36*, 5270–5280.

Kawahara, Y., Zinshteyn, B., Chendrimada, T.P., Shiekhattar, R., and Nishikura, K. (2007a). RNA editing of the microRNA-151 precursor blocks cleavage by the Dicer-TRBP complex. EMBO Rep *8*, 763–769.

Kawahara, Y., Zinshteyn, B., Sethupathy, P., Iizasa, H., Hatzigeorgiou, A.G., and Nishikura, K. (2007b). Redirection of silencing targets by adenosine-to-inosine editing of miRNAs. Science *315*, 1137–1140.

Kawakubo, K., and Samuel, C.E. (2000). Human RNA-specific adenosine deaminase (ADAR1) gene specifies transcripts that initiate from a constitutively active alternative promoter. Gene *258*, 165–172.

Keegan, L.P., Leroy, A., Sproul, D., and O'Connell, M.A. (2004). Adenosine deaminases acting on RNA (ADARs): RNA-editing enzymes. Genome Biol *5*, 209.

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.

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.

Kim, D.D.Y., Kim, T.T.Y., Walsh, T., Kobayashi, Y., Matise, T.C., Buyske, S., and

Gabriel, A. (2004). Widespread RNA editing of embedded alu elements in the human transcriptome. Genome Res *14*, 1719–1725.

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 USA *91*, 11457–11461.

Kislauskis, E.H., Zhu, X., and Singer, R.H. (1994). Sequences responsible for intracellular localization of beta-actin messenger RNA also affect cell phenotype. J Cell Biol *127*, 441–451.

Klein, I.A., Resch, W., Jankovic, M., Oliveira, T., Yamane, A., Nakahashi, H., Di Virgilio, M., Bothmer, A., Nussenzweig, A., Robbiani, D.F., et al. (2011). Translocation-capture sequencing reveals the extent and nature of chromosomal rearrangements in B lymphocytes. Cell *147*, 95–106.

Knight, S.W., and Bass, B.L. (2002). The role of RNA editing by ADARs in RNAi. Mol Cell *10*, 809–817.

Kobayashi, M., Takaori-Kondo, A., Miyauchi, Y., Iwai, K., and Uchiyama, T. (2005). Ubiquitination of APOBEC3G by an HIV-1 Vif-Cullin5-Elongin B-Elongin C complex is essential for Vif function. J Biol Chem *280*, 18573–18578.

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.

Kovalchuk, A.L., duBois, W., Mushinski, E., McNeil, N.E., Hirt, C., Qi, C.-F., Li, Z., Janz, S., Honjo, T., Muramatsu, M., et al. (2007). AID-deficient Bcl-xL transgenic mice develop delayed atypical plasma cell tumors with unusual Ig/Myc chromosomal rearrangements. J Exp Med *204*, 2989–3001.

Kumar, M., and Carmichael, G.G. (1997). Nuclear antisense RNA induces extensive adenosine modifications and nuclear retention of target transcripts. Proc Natl Acad Sci USA *94*, 3542–3547.

Kuo, M.Y., Chao, M., and Taylor, J. (1989). Initiation of replication of the human hepatitis delta virus genome from cloned DNA: role of delta antigen. J Virol *63*, 1945–1950.

Lai, F., Chen, C.X., Carter, K.C., and Nishikura, K. (1997). Editing of glutamate receptor B subunit ion channel RNAs by four alternatively spliced DRADA2 double-stranded RNA adenosine deaminases. Mol Cell Biol *17*, 2413–2424.

Lai, M.M. (1995). The molecular biology of hepatitis delta virus. Annu Rev Biochem *64*, 259–286.

Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., et al. (2001). Initial sequencing

and analysis of the human genome. Nature *409*, 860–921.

Landry, S., Narvaiza, I., Linfesty, D.C., and Weitzman, M.D. (2011). APOBEC3A can activate the DNA damage response and cause cell-cycle arrest. EMBO Rep *12*, 444–450.

Lau, P.P., Chang, B.H., and Chan, L. (2001). Two-hybrid cloning identifies an RNA-binding protein, GRY-RBP, as a component of apobec-1 editosome. Biochem Biophys Res Commun *282*, 977–983.

Lau, P.P., Xiong, W.J., Zhu, H.J., Chen, S.H., and Chan, L. (1991). Apolipoprotein B mRNA editing is an intranuclear event that occurs posttranscriptionally coincident with splicing and polyadenylation. J Biol Chem *266*, 20550–20554.

Lau, P.P., Zhu, H.J., Baldini, A., Charnsangavej, C., and Chan, L. (1994). Dimeric structure of a human apolipoprotein B mRNA editing protein and cloning and chromosomal localization of its gene. Proc Natl Acad Sci USA *91*, 8522–8526.

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.

Lehmann, K.A., and Bass, B.L. (1999). The importance of internal loops within RNA substrates of ADAR1. J Mol Biol *291*, 1–13.

Lei, M., Liu, Y., and Samuel, C.E. (1998). Adenovirus VAI RNA antagonizes the RNA-editing activity of the ADAR adenosine deaminase. Virology *245*, 188–196.

Lellek, H., Kirsten, R., Diehl, I., Apostel, F., Buck, F., and Greeve, J. (2000). Purification and molecular cloning of a novel essential component of the apolipoprotein B mRNA editing enzyme-complex. J Biol Chem *275*, 19848–19856.

Lenz, G., Nagel, I., Siebert, R., Roschke, A.V., Sanger, W., Wright, G.W., Dave, S.S., Tan, B., Zhao, H., Rosenwald, A., et al. (2007). Aberrant immunoglobulin class switch recombination and switch translocations in activated B cell-like diffuse large B cell lymphoma. J Exp Med *204*, 633–643.

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 Biotechnol *22*, 1001–1005.

Levanon, E.Y., Hallegger, M., Kinar, Y., Shemesh, R., Djinovic-Carugo, K., Rechavi, G., Jantsch, M.F., and Eisenberg, E. (2005). Evolutionarily conserved human targets of adenosine to inosine RNA editing. Nucleic Acids Res *33*, 1162–1168.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and 1000 Genome Project Data Processing Subgroup (2009a). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078–2079.

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

Li, M., Wang, I.X., Li, Y., Bruzel, A., Richards, A.L., Toung, J.M., and Cheung, V.G. (2011). Widespread RNA and DNA sequence differences in the human transcriptome. Science *333*, 53–58.

Li, Z., Wolff, K., and Samuel, C. (2009c). RNA adenosine deaminase ADAR1 deficiency leads to increased activation of protein kinase PKR and reduced vesicular stomatitis virus growth following interferon treatment. Virology.

Liang, H., and Landweber, L.F. (2007). Hypothesis: RNA editing of microRNA target sites in humans? Rna *13*, 463–467.

Liao, W., Hong, S.H., Chan, B.H., Rudolph, F.B., Clark, S.C., and Chan, L. (1999). APOBEC-2, a cardiac- and skeletal muscle-specific member of the cytidine deaminase supergene family. Biochem Biophys Res Commun *260*, 398–404.

Lin, C., Yang, L., Tanasa, B., Hutt, K., Ju, B.-G., Ohgi, K., Zhang, J., Rose, D.W., Fu, X.-D., Glass, C.K., et al. (2009). Nuclear receptor-induced chromosomal proximity and DNA breaks underlie specific translocations in cancer. Cell *139*, 1069–1083.

Lin, W., Piskol, R., Tan, M.H., and Li, J.B. (2012). Comment on "Widespread RNA and DNA sequence differences in the human transcriptome". Science *335*, 1302–authorreply1302.

Liu, M., Duke, J.L., Richter, D.J., Vinuesa, C.G., Goodnow, C.C., Kleinstein, S.H., and Schatz, D.G. (2008). Two levels of protection for the B cell genome during somatic hypermutation. Nature *451*, 841–845.

Liu, Y., Wolff, K.C., Jacobs, B.L., and Samuel, C.E. (2001). Vaccinia virus E3L interferon resistance protein inhibits the interferon-induced adenosine deaminase A-to-I editing activity. Virology *289*, 378–387.

Loeb, G.B., Khan, A.A., Canner, D., Hiatt, J.B., Shendure, J., Darnell, R.B., Leslie, C.S., and Rudensky, A.Y. (2012). Transcriptome-wide miR-155 Binding Map Reveals Widespread Noncanonical MicroRNA Targeting. Mol Cell *48*, 760–770.

Lomeli, H., Mosbacher, J., Melcher, T., Höger, T., Geiger, J.R., Kuner, T., Monyer, H., Higuchi, M., Bach, A., and Seeburg, P.H. (1994). Control of kinetic properties of AMPA receptor channels by nuclear RNA editing. Science *266*, 1709–1713.

Luo, G.X., Chao, M., Hsieh, S.Y., Sureau, C., Nishikura, K., and Taylor, J. (1990). A specific base transition occurs on replicating hepatitis delta virus RNA. J Virol *64*, 1021–1027.

Mabbott, N.A., Kenneth Baillie, J., Hume, D.A., and Freeman, T.C. (2010). Meta-analysis of lineage-specific gene expression signatures in mouse leukocyte

populations. Immunobiology *215*, 724–736.

MacGinnitie, A.J., Anant, S., and Davidson, N.O. (1995). Mutagenesis of apobec-1, the catalytic subunit of the mammalian apolipoprotein B mRNA editing enzyme, reveals distinct domains that mediate cytosine nucleoside deaminase, RNA binding, and RNA editing activity. J Biol Chem *270*, 14768–14775.

Madani, N., and Kabat, D. (1998). An endogenous inhibitor of human immunodeficiency virus in human lymphocytes is overcome by the viral Vif protein. J Virol *72*, 10251–10255.

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.

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

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.

Maris, C., Masse, J., Chester, A., Navaratnam, N., and Allain, F.H.-T. (2005). NMR structure of the apoB mRNA stem-loop and its interaction with the C to U editing APOBEC1 complementary factor. Rna *11*, 173–186.

Martin, A., and Scharff, M.D. (2002). Somatic hypermutation of the AID transgene in B and non-B cells. Proc Natl Acad Sci USA *99*, 12304–12308.

Mayorov, V.I., Rogozin, I.B., Adkison, L.R., Frahm, C., Kunkel, T.A., and Pavlov, Y.I. (2005). Expression of human AID in yeast induces mutations in context similar to the context of somatic hypermutation at G-C pairs in immunoglobulin genes. BMC Immunol. *6*, 10.

McBride, K.M., Gazumyan, A., Woo, E.M., Schwickert, T.A., Chait, B.T., and Nussenzweig, M.C. (2008). Regulation of class switch recombination and somatic mutation by AID phosphorylation. J Exp Med *205*, 2585–2594.

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.

Mehta, A., Banerjee, S., and Driscoll, D.M. (1996). Apobec-1 interacts with a 65-kDa complementing protein to edit apolipoprotein-B mRNA in vitro. J Biol Chem *271*, 28294–28299.

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

Meijer, H.A., Kong, Y.W., Lu, W.T., Wilczynska, A., Spriggs, R.V., Robinson, S.W., Godfrey, J.D., Willis, A.E., and Bushell, M. (2013). Translational repression and eIF4A2 activity are critical for microRNA-mediated gene regulation. Science *340*, 82–85.

Melcher, T., Maas, S., Herb, A., Sprengel, R., Higuchi, M., and Seeburg, P.H. (1996a). RED2, a brain-specific member of the RNA-specific adenosine deaminase family. J Biol Chem *271*, 31795–31798.

Melcher, T., Maas, S., Herb, A., Sprengel, R., Seeburg, P.H., and Higuchi, M. (1996b). A mammalian RNA editing enzyme. Nature *379*, 460–464.

Mikl, M.C., Watt, I.N., Lu, M., Reik, W., Davies, S.L., Neuberger, M.S., and Rada, C. (2005). Mice deficient in APOBEC2 and APOBEC3. Mol Cell Biol *25*, 7270–7277.

Morgan, H.D., Dean, W., Coker, H.A., Reik, W., and Petersen-Mahrt, S.K. (2004). Activation-induced cytidine deaminase deaminates 5-methylcytosine in DNA and is expressed in pluripotent tissues: implications for epigenetic reprogramming. J Biol Chem *279*, 52353–52360.

Morrison, J.R., Pászty, C., Stevens, M.E., Hughes, S.D., Forte, T., Scott, J., and Rubin, E.M. (1996). Apolipoprotein B RNA editing enzyme-deficient mice are viable despite alterations in lipoprotein metabolism. Proc Natl Acad Sci USA *93*, 7154–7159.

Muckenfuss, H., Hamdorf, M., Held, U., Perkovic, M., Löwer, 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.

Mukhopadhyay, D., Anant, S., Lee, R.M., Kennedy, S., Viskochil, D., and Davidson, N.O. (2002). C-->U editing of neurofibromatosis 1 mRNA occurs in tumors that express both the type II transcript and apobec-1, the catalytic subunit of the apolipoprotein B mRNA-editing enzyme. Am. J. Hum. Genet. *70*, 38–50.

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. Journal of Biological ….

Murphy, D.G., Dimock, K., and Kang, C.Y. (1991). Numerous transitions in human parainfluenza virus 3 RNA recovered from persistently infected cells. Virology *181*, 760–763.

Muto, T., Okazaki, I.-M., Yamada, S., Tanaka, Y., Kinoshita, K., Muramatsu, M.,

Nagaoka, H., and Honjo, T. (2006). Negative regulation of activation-induced cytidine deaminase in B cells. Proc Natl Acad Sci USA *103*, 2752–2757.

Nabel, C.S., Jia, H., Ye, Y., Shen, L., Goldschmidt, H.L., Stivers, J.T., Zhang, Y., and Kohli, R.M. (2012). AID/APOBEC deaminases disfavor modified cytosines implicated in DNA demethylation. Nat. Chem. Biol. *8*, 751–758.

Nakamuta, M., Chang, B.H., Zsigmond, E., Kobayashi, K., Lei, H., Ishida, B.Y., Oka, K., Li, E., and Chan, L. (1996). Complete phenotypic characterization of apobec-1 knockout mice with a wild-type genetic background and a human apolipoprotein B transgenic background, and restoration of apolipoprotein B mRNA editing by somatic gene transfer of Apobec-1. J Biol Chem *271*, 25981–25988.

Nakamuta, M., Oka, K., Krushkal, J., Kobayashi, K., Yamamoto, M., Li, W.H., and Chan, L. (1995). Alternative mRNA splicing and differential promoter utilization determine tissue-specific expression of the apolipoprotein B mRNA-editing protein (Apobec1) gene in mice. Structure and evolution of Apobec1 and related nucleoside/nucleotide deaminases. J Biol Chem *270*, 13042–13056.

Nakamuta, M., Tsai, A., Chan, L., Davidson, N.O., and Teng, B.B. (1999). Sequence elements required for apolipoprotein B mRNA editing enhancement activity from chicken enterocytes. Biochem Biophys Res Commun *254*, 744–750.

Navaratnam, N., Bhattacharya, S., and Fujino, T. (1995). Evolutionary origins of apoB mRNA editing: catalysis by a cytidine deaminase that has …. Cell.

Navaratnam, N., Fujino, T., Bayliss, J., Jarmuz, A., How, A., Richardson, N., Somasekaram, A., Bhattacharya, S., Carter, C., and Scott, J. (1998). Escherichia coli cytidine deaminase provides a molecular model for ApoB RNA editing and a mechanism for RNA substrate recognition. J Mol Biol *275*, 695–714.

Navaratnam, N., Morrison, J.R., Bhattacharya, S., Patel, D., Funahashi, T., Giannoni, F., Teng, B.B., Davidson, N.O., and Scott, J. (1993). The p27 catalytic subunit of the apolipoprotein B mRNA editing enzyme is a cytidine deaminase. J Biol Chem *268*, 20709–20712.

Navaratnam, N., Patel, D., Shah, R.R., Greeve, J.C., Powell, L.M., Knott, T.J., and Scott, J. (1991). An additional editing site is present in apolipoprotein B mRNA. Nucleic Acids Res *19*, 1741–1744.

Nelson, P.T., De Planell-Saguer, M., Lamprinaki, S., Kiriakidou, M., Zhang, P., O'Doherty, U., and Mourelatos, Z. (2007). A novel monoclonal antibody against human Argonaute proteins reveals unexpected characteristics of miRNAs in human blood cells. Rna *13*, 1787–1792.

Nelson, V.R., Heaney, J.D., Tesar, P.J., Davidson, N.O., and Nadeau, J.H. (2012). Transgenerational epigenetic effects of the Apobec1 cytidine deaminase deficiency on testicular germ cell tumor susceptibility and embryonic viability.

Proc Natl Acad Sci USA *109*, E2766–E2773.

Nie, Y., Ding, L., Kao, P.N., Braun, R., and Yang, J.-H. (2005). ADAR1 interacts with NF90 through double-stranded RNA and regulates NF90-mediated gene expression independently of RNA editing. Mol Cell Biol *25*, 6956–6963.

Nie, Y., Hammond, G.L., and Yang, J.-H. (2007). Double-stranded RNA deaminase ADAR1 increases host susceptibility to virus infection. J Virol *81*, 917–923.

Nishikura, K., Yoo, C., Kim, U., Murray, J.M., Estes, P.A., Cash, F.E., and Liebhaber, S.A. (1991). Substrate specificity of the dsRNA unwinding/modifying activity. The EMBO Journal *10*, 3523–3532.

Niswender, C.M., Sanders-Bush, E., and Emeson, R.B. (1998). Identification and characterization of RNA editing events within the 5-HT2C receptor. Ann N Y Acad Sci *861*, 38–48.

Nonaka, T., Doi, T., Toyoshima, T., Muramatsu, M., Honjo, T., and Kinoshita, K. (2009). Carboxy-terminal domain of AID required for its mRNA complex formation in vivo. Proc Natl Acad Sci USA *106*, 2747–2751.

Nussenzweig, A., and Nussenzweig, M.C. (2010). Origin of chromosomal translocations in lymphoid cancer. Cell *141*, 27–38.

O'Hara, P.J., Nichol, S.T., Horodyski, F.M., and Holland, J.J. (1984). Vesicular stomatitis virus defective interfering particles can contain extensive genomic sequence rearrangements and base substitutions. Cell *36*, 915–924.

Oka, K., Kobayashi, K., Sullivan, M., Martinez, J., Teng, B.B., Ishimura-Oka, K., and Chan, L. (1997). Tissue-specific inhibition of apolipoprotein B mRNA editing in the liver by adenovirus-mediated transfer of a dominant negative mutant APOBEC-1 leads to increased low density lipoprotein in mice. J Biol Chem *272*, 1456–1460.

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.

Osenberg, S., Dominissini, D., Rechavi, G., and Eisenberg, E. (2009). Widespread cleavage of A-to-I hyper-editing substrates. Rna *15*, 1632–1639.

Ota, H., Sakurai, M., Gupta, R., Valente, L., Wulff, B.-E., Ariyoshi, K., Iizasa, H., Davuluri, R.V., and Nishikura, K. (2013). ADAR1 Forms a Complex with Dicer to Promote MicroRNA Processing and RNA-Induced Gene Silencing. Cell *153*, 575–589.

Palladino, M.J., Keegan, L.P., O'Connell, M.A., and Reenan, R.A. (2000). A-to-I pre-mRNA editing in Drosophila is primarily involved in adult nervous system function and integrity. Cell *102*, 437–449.

Pasqualucci, L., Neumeister, P., Goossens, T., Nanjangud, G., Chaganti, R.S., Küppers, R., and Dalla-Favera, R. (2001). Hypermutation of multiple proto-oncogenes in B-cell diffuse large-cell lymphomas. Nature *412*, 341–346.

Pasqualucci, L., Bhagat, G., Jankovic, M., Compagno, M., Smith, P., Muramatsu, M., Honjo, T., Morse, H.C., Nussenzweig, M.C., and Dalla-Favera, R. (2008). AID is required for germinal center-derived lymphomagenesis. Nat. Genet. *40*, 108–112.

Patterson, J.B., and Samuel, C.E. (1995). Expression and regulation by interferon of a double-stranded-RNA-specific adenosine deaminase from human cells: evidence for two forms of the deaminase. Mol Cell Biol *15*, 5376–5388.

Pavri, R., Gazumyan, A., Jankovic, M., Di Virgilio, M., Klein, I., Ansarah-Sobrinho, C., Resch, W., Yamane, A., Reina-San-Martin, B., Barreto, V., et al. (2010). Activation-induced cytidine deaminase targets DNA at sites of RNA polymerase II stalling by interaction with Spt5. Cell *143*, 122–133.

Peled, J.U., Kuang, F.L., Iglesias-Ussel, M.D., Roa, S., Kalis, S.L., Goodman, M.F., and Scharff, M.D. (2008). The biochemistry of somatic hypermutation. Annu. Rev. Immunol. *26*, 481–511.

Peng, Z., Cheng, Y., Tan, B.C.-M., Kang, L., Tian, Z., Zhu, Y., Zhang, W., Liang, Y., Hu, X., Tan, X., et al. (2012). Comprehensive analysis of RNA-Seq data reveals extensive RNA editing in a human transcriptome. Nat Biotechnol *30*, 253–260.

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., Guétard, 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.

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

Pickrell, J.K., Gilad, Y., and Pritchard, J.K. (2012). Comment on "Widespread RNA and DNA sequence differences in the human transcriptome". Science *335*, 1302–authorreply1302.

Polson, A.G., and Bass, B.L. (1994). Preferential selection of adenosines for modification by double-stranded RNA adenosine deaminase. The EMBO Journal *13*, 5701–5711.

Polson, A.G., Bass, B.L., and Casey, J.L. (1996). RNA editing of hepatitis delta virus antigenome by dsRNA-adenosine deaminase. Nature *380*, 454–456.

Poltoratsky, V.P., Wilson, S.H., Kunkel, T.A., and Pavlov, Y.I. (2004).

Recombinogenic phenotype of human activation-induced cytosine deaminase. J Immunol *172*, 4308–4313.

Popp, C., Dean, W., Feng, S., Cokus, S.J., Andrews, S., Pellegrini, M., Jacobsen, S.E., and Reik, W. (2010). Genome-wide erasure of DNA methylation in mouse primordial germ cells is affected by AID deficiency. Nature *463*, 1101–1105.

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.

Prasanth, K.V., Prasanth, S.G., Xuan, Z., Hearn, S., Freier, S.M., Bennett, C.F., Zhang, M.Q., and Spector, D.L. (2005). Regulating gene expression through RNA nuclear retention. Cell *123*, 249–263.

Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841–842.

Rada, C., Williams, G.T., Nilsen, H., Barnes, D.E., Lindahl, T., and Neuberger, M.S. (2002). Immunoglobulin isotype switching is inhibited and somatic hypermutation perturbed in UNG-deficient mice. Curr Biol *12*, 1748–1755.

Rai, K., Huggins, I.J., James, S.R., Karpf, A.R., Jones, D.A., and Cairns, B.R. (2008). DNA demethylation in zebrafish involves the coupling of a deaminase, a glycosylase, and gadd45. Cell *135*, 1201–1212.

Ramaswami, G., Lin, W., Piskol, R., Tan, M.H., Davis, C., and Li, J.B. (2012). Accurate identification of human Alu and non-Alu RNA editing sites. Nat Methods *9*, 579–581.

Ramiro, A.R., Jankovic, M., Callen, E., Difilippantonio, S., Chen, H.T., McBride, K.M., Eisenreich, T.R., Chen, J., Dickins, R.A., Lowe, S.W., et al. (2006). Role of genomic instability and p53 in AID-induced c-myc-Igh translocations. Nature *440*, 105–109.

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.

Ray, P.S., Jia, J., Yao, P., Majumder, M., Hatzoglou, M., and Fox, P.L. (2009). A stress-responsive RNA switch regulates VEGFA expression. Nature *457*, 915–919.

Rebagliati, M.R., and Melton, D.A. (1987). Antisense RNA injections in fertilized frog eggs reveal an RNA duplex unwinding activity. Cell *48*, 599–605.

Revy, P., Muto, T., Levy, Y., Geissmann, F., Plebani, A., Sanal, O., Catalan, N., Forveille, M., Dufourcq-Labelouse, R., Gennery, A., et al. (2000). Activation-induced cytidine deaminase (AID) deficiency causes the autosomal recessive form of the Hyper-IgM syndrome (HIGM2). Cell *102*, 565–575.

Richardson, N., Navaratnam, N., and Scott, J. (1998). Secondary structure for the apolipoprotein B mRNA editing site. Au-binding proteins interact with a stem loop. J Biol Chem *273*, 31707–31717.

Riedmann, E.M., Schopoff, S., Hartner, J.C., and Jantsch, M.F. (2008). Specificity of ADAR-mediated RNA editing in newly identified targets. Rna *14*, 1110–1118.

Robbiani, D.F., Bothmer, A., Callen, E., Reina-San-Martin, B., Dorsett, Y., Difilippantonio, S., Bolland, D.J., Chen, H.T., Corcoran, A.E., Nussenzweig, A., et al. (2008). AID is required for the chromosomal breaks in c-myc that lead to c-myc/IgH translocations. Cell *135*, 1028–1038.

Robbiani, D.F., Bunting, S., Feldhahn, N., Bothmer, A., Camps, J., Deroubaix, S., McBride, K.M., Klein, I.A., Stone, G., Eisenreich, T.R., et al. (2009). AID produces DNA double-strand breaks in non-Ig genes and mature B cell lymphomas with reciprocal chromosome translocations. Mol Cell *36*, 631–641.

Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., and Mesirov, J.P. (2011). Integrative genomics viewer. Nat Biotechnol *29*, 24–26.

Rodriguez, J., Menet, J.S., and Rosbash, M. (2012). Nascent-seq indicates widespread cotranscriptional RNA editing in Drosophila. Mol Cell *47*, 27–37.

Rogozin, I.B., and Diaz, M. (2004). Cutting edge: DGYW/WRCH is a better predictor of mutability at G:C bases in Ig hypermutation than the widely accepted RGYW/WRCY motif and probably reflects a two-step activation-induced cytidine deaminase-triggered process. J Immunol *172*, 3382–3384.

Rogozin, I.B., Basu, M.K., Jordan, I.K., Pavlov, Y.I., and Koonin, E.V. (2005). APOBEC4, a new member of the AID/APOBEC family of polynucleotide (deoxy)cytidine deaminases predicted by computational analysis. Cell Cycle *4*, 1281–1285.

Rosenberg, B.R., and Papavasiliou, F.N. (2007). Beyond SHM and CSR: AID and related cytidine deaminases in the host response to viral infection. Adv Immunol *94*, 215–244.

Rosenberg, B.R., Dewell, S., and Papavasiliou, F.N. (2011a). Identifying mRNA editing deaminase targets by RNA-Seq. Methods Mol Biol *718*, 103–119.

Rosenberg, B.R., Hamilton, C.E., Mwangi, M.M., Dewell, S., and Papavasiliou, F.N. (2011b). Transcriptome-wide sequencing reveals numerous APOBEC1 mRNA-editing targets in transcript 3′ UTRs. Nat Struct Mol Biol *18*, 230–236.

Rubio, M.A.T., Pastar, I., Gaston, K.W., Ragone, F.L., Janzen, C.J., Cross, G.A.M., Papavasiliou, F.N., and Alfonzo, J.D. (2007). An adenosine-to-inosine tRNA-editing enzyme that can perform C-to-U deamination of DNA. Proc Natl Acad Sci USA *104*, 7821–7826.

Ryu, W.S., Bayer, M., and Taylor, J. (1992). Assembly of hepatitis delta virus particles. J Virol *66*, 2310–2315.

Sakurai, M., Yano, T., Kawabata, H., Ueda, H., and Suzuki, T. (2010). Inosine cyanoethylation identifies A-to-I RNA editing sites in the human transcriptome. Nat. Chem. Biol. *6*, 733–740.

Sato, S., Cornillez-Ty, C., and Lazinski, D.W. (2004). By inhibiting replication, the large hepatitis delta antigen can indirectly regulate amber/W editing and its own expression. J Virol *78*, 8120–8134.

Sato, Y., Probst, H.C., Tatsumi, R., Ikeuchi, Y., Neuberger, M.S., and Rada, C. (2010). Deficiency in APOBEC2 leads to a shift in muscle fiber type, diminished body mass, and myopathy. J Biol Chem *285*, 7111–7118.

Savva, Y.A., Rieder, L.E., and Reenan, R.A. (2012). The ADAR protein family. Genome Biol *13*, 252.

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.

Scadden, A.D., and Smith, C.W. (1997). A ribonuclease specific for inosine-containing RNA: a potential role in antiviral defence? The EMBO Journal *16*, 2140–2149.

Scadden, A.D., and Smith, C.W. (2001). RNAi is antagonized by A-->I hyper-editing. EMBO Rep *2*, 1107–1111.

Scadden, A.D.J. (2005). The RISC subunit Tudor-SN binds to hyper-edited double-stranded RNA and promotes its cleavage. Nat Struct Mol Biol *12*, 489–496.

Schäfer, 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.

Schoggins, J.W., Wilson, S.J., Panis, M., Murphy, M.Y., Jones, C.T., Bieniasz, P., and Rice, C.M. (2011). A diverse range of gene products are effectors of the type I interferon antiviral response. Nature *472*, 481–485.

Schrider, D.R., Gout, J.-F., and Hahn, M.W. (2011). Very few RNA and DNA sequence differences in the human transcriptome. PLoS ONE *6*, e25842.

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

Schwedler, von, 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.

Shaffer, A.L., Rosenwald, A., and Staudt, L.M. (2002). Lymphoid malignancies: the dark side of B-cell differentiation. Nat Rev Immunol *2*, 920–932.

Shah, R.R., Knott, T.J., Legros, J.E., Navaratnam, N., Greeve, J.C., and Scott, J. (1991). Sequence requirements for the editing of apolipoprotein B mRNA. J Biol Chem *266*, 16301–16304.

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.

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.

Shen, H.M., Peters, A., Baron, B., Zhu, X., and Storb, U. (1998). Mutation of BCL-6 gene in normal B cells by the process of somatic hypermutation of Ig genes. Science *280*, 1750–1752.

Shivarov, V., Shinkura, R., Doi, T., Begum, N.A., Nagaoka, H., Okazaki, I.-M., Ito, S., Nonaka, T., Kinoshita, K., and Honjo, T. (2009). Molecular mechanism for generation of antibody memory. Philos Trans R Soc Lond, B, Biol Sci *364*, 569–575.

Simon, J.H., Gaddis, N.C., Fouchier, R.A., and Malim, M.H. (1998). Evidence for a newly discovered cellular anti-HIV-1 phenotype. Nat Med *4*, 1397–1400.

Skarda, J., Amariglio, N., and Rechavi, G. (2009). RNA editing in human cancer: review. Apmis *117*, 551–557.

Skuse, G.R., Cappione, A.J., Sowden, M., Metheny, L.J., and Smith, H.C. (1996). The neurofibromatosis type I messenger RNA undergoes base-modification RNA editing. Nucleic Acids Res *24*, 478–485.

Smith, H.C. (2009). The APOBEC1 Paradigm for Mammalian Cytidine Deaminases That Edit DNA and RNA. Landes bioScience.

Smith, H.C. (2011). APOBEC3G: a double agent in defense. Trends Biochem Sci *36*, 239–244.

Smith, H.C., Bennett, R.P., Kizilyer, A., McDougall, W.M., and Prohaska, K.M. (2012). Functions and regulation of the APOBEC family of proteins. Semin. Cell Dev. Biol. *23*, 258–268.

Sowden, M.P., Eagleton, M.J., and Smith, H.C. (1998). Apolipoprotein B RNA sequence 3' of the mooring sequence and cellular sources of auxiliary factors determine the location and extent of promiscuous editing. Nucleic Acids Res *26*, 1644–1652.

Sowden, M., Hamm, J.K., and Smith, H.C. (1996a). Over-expression of APOBEC-

1 results in mooring sequence-dependent promiscuous RNA editing. J Biol Chem *271*, 3011–3017.

Sowden, M., Hamm, J.K., Spinelli, S., and Smith, H.C. (1996b). Determinants involved in regulating the proportion of edited apolipoprotein B RNAs. Rna *2*, 274–288.

Sowden, M.P., Ballatori, N., Jensen, K.L. de M., Reed, L.H., and Smith, H.C. (2002). The editosome for cytidine to uridine mRNA editing has a native complexity of 27S: identification of intracellular domains containing active and inactive editing factors. J Cell Sci *115*, 1027–1039.

Sowden, M.P., Lehmann, D.M., Lin, X., Smith, C.O., and Smith, H.C. (2004). Identification of novel alternative splice variants of APOBEC-1 complementation factor with different capacities to support apolipoprotein B mRNA editing. J Biol Chem *279*, 197–206.

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.

Takizawa, M., Tolarová, H., Li, Z., duBois, W., Lim, S., Callen, E., Franco, S., Mosaico, M., Feigenbaum, L., Alt, F.W., et al. (2008). AID expression levels determine the extent of cMyc oncogenic translocations and the incidence of B cell tumor development. J Exp Med *205*, 1949–1957.

Taylor, D.R., Puig, M., Darnell, M.E.R., Mihalik, K., and Feinstone, S.M. (2005). New antiviral pathway that mediates hepatitis C virus replicon interferon sensitivity through ADAR1. J Virol *79*, 6291–6298.

Taylor, J.M. (2003). Replication of human hepatitis delta virus: recent developments. Trends Microbiol *11*, 185–190.

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

Teng, G., Hakimpour, P., Landgraf, P., Rice, A., Tuschl, T., Casellas, R., and Papavasiliou, F.N. (2008). MicroRNA-155 is a negative regulator of activation-induced cytidine deaminase. Immunity *28*, 621–629.

Tenoever, B.R., Ng, S.-L., Chua, M.A., McWhirter, S.M., García-Sastre, A., and Maniatis, T. (2007). Multiple functions of the IKK-related kinase IKKepsilon in interferon-mediated antiviral immunity. Science *315*, 1274–1278.

Thorvaldsdóttir, H., Robinson, J.T., and Mesirov, J.P. (2013). Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Brief. Bioinformatics *14*, 178–192.

Tonkin, L.A., Saccomanno, L., Morse, D.P., Brodigan, T., Krause, M., and Bass, B.L. (2002). RNA editing by ADARs is important for normal behavior in Caenorhabditis elegans. The EMBO Journal *21*, 6025–6035.

Toth, A.M., Li, Z., Cattaneo, R., and Samuel, C.E. (2009). RNA-specific adenosine deaminase ADAR1 suppresses measles virus-induced apoptosis and activation of protein kinase PKR. J Biol Chem *284*, 29350–29356.

Trapnell, C., Hendrickson, D.G., Sauvageau, M., Goff, L., Rinn, J.L., and Pachter, L. (2013). Differential analysis of gene regulation at transcript resolution with RNA-seq. Nat Biotechnol *31*, 46–53.

Trapnell, C., Pachter, L., and Salzberg, S.L. (2009). TopHat: discovering splice junctions with RNA-Seq. Bioinformatics *25*, 1105–1111.

Valente, L., and Nishikura, K. (2005). ADAR gene family and A-to-I RNA editing: diverse roles in posttranscriptional gene regulation. Prog Nucleic Acid Res Mol Biol *79*, 299–338.

Vesely, C., Tauber, S., Sedlazeck, F.J., Haeseler, von, A., and Jantsch, M.F. (2012). Adenosine deaminases that act on RNA induce reproducible changes in abundance and sequence of embryonic miRNAs. Genome Res *22*, 1468–1476.

Vonica, A., Rosa, A., Arduini, B.L., and Brivanlou, A.H. (2011). APOBEC2, a selective inhibitor of TGFβ signaling, regulates left-right axis specification during early embryogenesis. Dev. Biol. *350*, 13–23.

Vuong, B.Q., Lee, M., Kabir, S., Irimia, C., Macchiarulo, S., McKnight, G.S., and Chaudhuri, J. (2009). Specific recruitment of protein kinase A to the immunoglobulin locus regulates class-switch recombination. Nat Immunol *10*, 420–426.

Wagner, R.W., Smith, J.E., Cooperman, B.S., and Nishikura, K. (1989). A double-stranded RNA unwinding activity introduces structural alterations by means of adenosine to inosine conversions in mammalian cells and Xenopus eggs. Proc Natl Acad Sci USA *86*, 2647–2651.

Wang, K.S., Choo, Q.L., Weiner, A.J., Ou, J.H., Najarian, R.C., Thayer, R.M., Mullenbach, G.T., Denniston, K.J., Gerin, J.L., and Houghton, M. (1986). Structure, sequence and expression of the hepatitis delta (delta) viral genome. Nature *323*, 508–514.

Wang, Q., Khillan, J., Gadue, P., and Nishikura, K. (2000). Requirement of the RNA editing deaminase ADAR1 gene for embryonic erythropoiesis. Science *290*, 1765–1768.

Wang, Z., Choi, M.K., Ban, T., Yanai, H., Negishi, H., Lu, Y., Tamura, T., Takaoka, A., Nishikura, K., and Taniguchi, T. (2008). Regulation of innate immune responses by DAI (DLM-1/ZBP1) and other DNA-sensing molecules. Proceedings of the National Academy of Sciences *105*, 5477–5482.

Wolf, J., Gerber, A.P., and Keller, W. (2002). tadA, an essential tRNA-specific adenosine deaminase from Escherichia coli. The EMBO Journal *21*, 3841–3851.

Wong, S.K., and Lazinski, D.W. (2002). Replicating hepatitis delta virus RNA is edited in the nucleus by the small form of ADAR1. Proc Natl Acad Sci USA *99*, 15118–15123.

Wright, G., Tan, B., Rosenwald, A., Hurt, E.H., Wiestner, A., and Staudt, L.M. (2003). A gene expression-based method to diagnose clinically distinct subgroups of diffuse large B cell lymphoma. Proc Natl Acad Sci USA *100*, 9991–9996.

Xie, Y., Blanc, V., Kerr, T.A., Kennedy, S., Luo, J., Newberry, E.P., and Davidson, N.O. (2009). Decreased expression of cholesterol 7alpha-hydroxylase and altered bile acid metabolism in Apobec-1-/- mice lead to increased gallstone susceptibility. J Biol Chem *284*, 16860–16871.

XuFeng, R., Boyer, M.J., Shen, H., Li, Y., Yu, H., Gao, Y., Yang, Q., Wang, Q., and Cheng, T. (2009). ADAR1 is required for hematopoietic progenitor cell survival via RNA editing. Proc Natl Acad Sci USA *106*, 17763–17768.

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 USA *92*, 8483–8487.

Yamanaka, S., Poksay, K.S., Arnold, K.S., and Innerarity, T.L. (1997). A novel translational repressor mRNA is edited extensively in livers containing tumors caused by the transgene expression of the apoB mRNA-editing enzyme. Genes Dev *11*, 321–333.

Yamanaka, S., Poksay, K.S., Driscoll, D.M., and Innerarity, T.L. (1996). Hyper-editing of multiple cytidines of apolipoprotein B mRNA by APOBEC-1 requires auxiliary protein(s) but not a mooring sequence motif. J Biol Chem *271*, 11506–11510.

Yamane, A., Resch, W., Kuo, N., Kuchen, S., Li, Z., Sun, H.-W., Robbiani, D.F., McBride, K., Nussenzweig, M.C., and Casellas, R. (2011). Deep-sequencing identification of the genomic targets of the cytidine deaminase AID and its cofactor RPA in B lymphocytes. Nat Immunol *12*, 62–69.

Yang, W., Chendrimada, T.P., Wang, Q., Higuchi, M., Seeburg, P.H., Shiekhattar, R., and Nishikura, K. (2006). Modulation of microRNA processing and expression through RNA editing by ADAR deaminases. Nat Struct Mol Biol *13*, 13–21.

Yang, Y., Sowden, M.P., Yang, Y., and Smith, H.C. (2001). Intracellular trafficking determinants in APOBEC-1, the catalytic subunit for cytidine to uridine editing of apolipoprotein B mRNA. Exp Cell Res *267*, 153–164.

Yang, Y., Yang, Y., and Smith, H.C. (1997). Multiple protein domains determine

the cell type-specific nuclear distribution of the catalytic subunit required for apolipoprotein B mRNA editing. Proc Natl Acad Sci USA *94*, 13075–13080.

Yoshikawa, K., Okazaki, I.-M., Eto, T., Kinoshita, K., Muramatsu, M., Nagaoka, H., and Honjo, T. (2002). AID enzyme-induced hypermutation in an actively transcribed gene in fibroblasts. Science *296*, 2033–2036.

Yu, K., Huang, F.-T., and Lieber, M.R. (2004a). DNA substrate length and surrounding sequence affect the activation-induced deaminase activity at cytidine. J Biol Chem *279*, 6496–6500.

Yu, Q., König, 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.

Zahn, R.C., Schelp, I., Utermöhlen, O., and Laer, von, D. (2007). A-to-G hypermutation in the genome of lymphocytic choriomeningitis virus. J Virol *81*, 457–464.

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, Z., and Carmichael, G.G. (2001). The fate of dsRNA in the nucleus: a p54(nrb)-containing complex mediates the nuclear retention of promiscuously A-to-I edited RNAs. Cell *106*, 465–475.