2017

Apobec1-Mediated RNA Editing in Monocytes Contributes to Genetic Heterogeneity and Modulates Monocyte Activity and Development

Violeta Rayón Estrada

APOBEC1-MEDIATED RNA EDITING IN MONOCYTES CONTRIBUTES TO GENETIC HETEROGENEITY AND MODULATES MONOCYTE ACTIVITY AND DEVELOPMENT

A Thesis Presented to the Faculty of

The Rockefeller University

in Partial Fulfillment of the Requirements for

the degree of Doctor of Philosophy

by

Violeta Rayón Estrada

June 2017

APOBEC1-MEDIATED RNA EDITING IN MONOCYTES CONTRIBUTES TO GENETIC HETEROGENEITY AND MODULATES MONOCYTE ACTIVITY AND DEVELOPMENT

Violeta Rayón Estrada, Ph.D.

The Rockefeller University 2017

Post-transcriptional modifications such as alternative splicing have been shown to add to the complexity needed to compensate for the relatively low number of genes found in higher organisms. Many other modifications recently found in mRNA, which cannot be deduced from what is coded in the genome, such a cytidine to uridine and adenosine to inosine editing, reveal that this complexity is ever expanding. Therefore, the current challenge is to understand what is the function of these modifications. In this thesis, I focus on APOBEC1-mediated RNA editing.

In higher eukaryotes RNA editing consists of C to U and A to I transitions mediated by the proteins of the APOBEC1 and ADAR families, respectively. APOBEC1 has been fully characterized in its role of editing the Apolipoprotein B (ApoB) transcript in the intestine, where the C to U modification changes a glutamate codon to a stop codon, creating a smaller version of the ApoB protein. Editing of ApoB is essential for the formation of the chylomicron, a lipid transport protein, making APOBEC1 a crucial enzyme for lipid metabolism.

In recent years, our laboratory developed a comparative approach that uses an Apobec1-deficient mouse in order to find true Apobec1-mediated editing events in *wildtype* mRNA. This method allowed for the discovery of additional sites in a transcriptome-wide manner. Using this, we were able to identify hundreds of additional edited sites in murine intestine and macrophages. This opens up the possibility of alternative APOBEC1 functions. In this thesis, I focus on the edited sites within macrophages, where Apolipoprotein B is not expressed and therefore, where APOBEC1 may play an alternative role to lipid metabolism. Specifically, these consist of 410 high-confidence C to U editing events contained in 275 transcripts, the large majority of which are within the 3'UTR.

First, finding that there are no transcriptional differences between macrophages derived from Apobec1-deficient and *wildtype* mice, I characterized the fate of these transcripts at the molecular level. Previously, ADAR editing has been shown to potentially regulate transcript abundance by nuclear retention and stabilization. Here I demonstrated that this is not the case for APOBEC1. However, even though at the RNA level, there seem to be no alterations due to editing, I showed that editing does regulate translation of protein products, some of which are miRNA mediated.

The changes observed at the protein level are nevertheless quite small, which is to be expected from the low frequency of editing per transcript. However, a very long-standing question in the field is whether this is the reflection of a few cells within the population that possess 100% editing per site or whether each cell has low frequency editing. In order to test for this, we created a statistical model that tested the variability of editing at each site among many cells of the same population. Using this model, we

observed that cells are indeed quite variable in terms of editing. Then I validated the results of the model using barcodes that identify individual RNA molecules to amplify a region surrounding the edited sites in single cells. Altogether, these experiments demonstrated that within the population, cells that are seemingly transcriptionally identical are indeed heterogeneous.

Next, I tested whether the loss of this variability might affect the activity of macrophages. To do this, I designed *in vitro* and *in vivo* assays. I demonstrated that Apobec1-deficient macrophages have altered migration and phagocytosis phenotypes *in vitro*. This predicts that the physiology of monocytes in the Apobec1-deficient mouse would be altered. While setting up a competitive reconstitution *in vivo* assay, where I would be able to test *wildtype* and Apobec1-deficient monocytes side by side, I discovered that the development of monocytes is not equivalent between the two genotypes. Surprisingly, Apobec1-deficient monocytic progenitors tend to outcompete their *wildtype* counterparts and monocytes have an increased preference to form a pro-inflammatory Ly6C positive phenotype. Finally, I present a possible link between loss of Apobec1 and brain disorders.

To an amazing husband and father,

it's a pleasure to walk the path

of life alongside you

ACKNOWLEDGEMENTS

I would first like to thank my advisor Nina for taking a chance on me and for making the enterprise of science a very exciting one. I have learned from you that to persevere in science one has to risk, explore, and take a chance in what sometimes seem to be “crazy” ideas. To embrace frustration, because it can be the fuel needed for success. And very importantly, to stand tall and to never give up. These are precious gifts that I am forever thankful for.

I would also like to thank Karen Bulloch and Bruce McEwen for taking under your wing so we could explore possibilities beyond my ability. Charlie Rice for inviting me to participate in very stimulating meetings with his group. Tom Sankmar for his comments on my paper. My thesis committee members Luciano Maraffini, Charlie Rice and Carl Nathan who have over the years given me great feedback and guidance. Jin Billy Li for kindly agreeing to transverse the county to serve as the external examiner of my thesis defense.

I am immensely grateful to my collaborators Dewi Harjanto, Claire Hamilton, Yamina Berchiche, Khatuna Gagnidze, Dan Cole, Youngcheul and Kate for the exciting science discussions, experiments and data that were crucial to piece together this story. Anastasia Papavasiliou, Theodore Papamarkou and Chris Oates in the UK for describing the wonderful world of editing in numbers. Diego Mourao-Sa for his extreme generosity investing time to contribute to this story.

I would also like to thank my partners in crime Dewi Harjanto, Linda Molla, Jason Pinger, Maryam Zaringhalam and Anita Ramnarain with whom I collaborated, discussed, celebrated, argued and ranted. Dewi, thank you for always being willing and ready to engage in crazy experiments, it's a lot of fun doing science with you and I am thankful that we were able to do so for so many years together. Jason, you are my savior and I will forever be thankful for the many hours of scientific discussion, mice killed and the fantastic company at the bench. Linda, Maryam and Anita thank you for having my back, your friendship and support have been crucial to my success.

I am grateful to my labmates Galadriel Hover-Miner, Monica Mugnier, Danae Schulz, Eric Fritz, Rebecca Delker, Pete Stravopolous, Dimitri Garyfallos, Gianna Triller, Shanin Chowdhury and Nick Economos which in one way or another have marked my life and my science; and my student Shivani Aggarwal for showing me true untampered enjoyment for science.

I am also immensely thankful for the people in the genome resource center, who more than once agreed to provide me with virtually instant services and in the flow cytometry resource center, who have always helped me with a smile.

I would also like to heart fully thank my family and my friends, who are my life support and happiness. Gaby for her immense love. Fran, Eduardo, Laura, Matt and Ariella for making me one of their own. My aunts and uncles in Mexico who have supported me and cared for me and my sister. My parents, who from heaven are cheering me on.

Last but not least, I immensely thank my husband and son, one for his blind faith in my abilities, boundless support, encouragement and love, and the other for all the bliss that having him in our lives brings me. Diego, si alguna vez lees esto, quiero que sepas que para mi, obtener este título ha sido una meta muy preciada que me ha costado mucho trabajo alcanzar; y si yo lo he logrado, tu también podras lograr todas tus metas y tus sueños.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 5

CHAPTER 6

LIST OF TABLES

CHAPTER 5

CHAPTER 1. INTRODUCTION

The processing of messenger RNA leads to the diversification of the genes in the genome by creating alternative proteins though processes such as alternative splicing, alternative initiation and termination, frame shifts etc. However, all of these modifications are variations of what is encoded in the genome. In this chapter I will survey other types of post-transcriptional processing that cannot be deduced from what is coded in the genome. Then I will focus on RNA editing, which is a process that directly recodes the message through adenosine to inosine and cytidine to uracil nucleoside transitions. I will show that these types of modifications have been demonstrated to have big biological implications. Finally, I will survey of the state of apolipoprotein B mRNA editing catalytic polypeptide 1 (APOBEC1) research at the moment I began my studies and will present some preliminary studies performed in my laboratory.

1.1. Post-transcriptional modifications

The post-transcriptional modification of RNA is a naturally occurring process that occurs during the maturation of RNAs. Today, more than 100 distinct ribonucleoside modifications are known and have been shown to be present in all three phylogenetic domains: archaea, bacteria and eukarya[1]. These modifications are most abundant in non-coding RNA (ncRNA), where they are crucial for properly aiding in translation and splicing[1].

Transfer RNA (tRNA) modification is a good example where modifications play a very important role in its structure and function. tRNA is both the most highly modified

(with 25% of its nucleosides altered) and the one with the most complex chemical variety[1]. For example, different combinations of modifications provide plasticity so that tRNA can adopt subtle features like increased rigidity or flexibility[2]. Rigidity can be enhanced for example, by the presence of ubiquitous pseudouridines partly by coordinating stabilizing water molecules[3]. On the other hand dihydrouridine provides flexibility[4]. Maintenance of the optimal tRNA structure however, seems to require both modifications, suggesting that coordination of a network of modifications is needed for proper stability[5].

Modification of mRNA with N^6-mathyladenosine (m^6A), 5-methylcytidine (m^5C) and ribose methylations have been known for decades[6–8]. Besides these, four additional base modifications have been discovered so far in mRNA: inosine (I), pseudouridine (Ψ), 5-hydroxymethylcytidine (hm^5C) and N^1-methyladenosine (m^1A)[9]. A few of these modifications have been well characterized, for example inosine, pseudouridine and m^6A have been shown to influence the metabolism, function[1], localization, or stability of transcripts to rapidly adjust the transcriptome in response to developmental and environmental cues[10–12]. Moreover m^6A is found in the 5' untranslated region and marks the beginning of transcripts in mRNA[13]. It is now clear that all these modifications in mRNA are not passive marks. However, research in this area is still emerging, and new techniques that incorporate high throughput RNA sequencing are allowing for the identification of hundred of additional marks in mRNA. Therefore little is known about their role and which are the writers (enzymes that put the marks on) and readers of these emerging marks.

1.2. RNA editing

The term RNA editing was originally coined to describe the insertion of 4 uridines in the coding region of the mitochondrial cytochrome oxidase subunit II gene in *Trypanosoma brucei*, which restores the frame shift encoded in the DNA[14]. This nucleotide insertion, which cannot be deduced from genomic sequence, allows for the formation of a functional protein, essential for generating functional mitochondrial proteins in trypanosomes[15]. The discovery of other changes in the sequence of transcripts led to the re-definition of RNA editing to include all type of modifications that result in sequence changes in the RNA from what is encoded in the genome. These modifications now include insertion, deletion and modification of nucleotides[16].

Modification of nucleotides is performed by the cytidine deaminase (CDA) superfamily of enzymes that includes both the adenosine deaminases acting on RNA (ADAR) and the apolipoprotein B mRNA editing catalytic polypeptide 1 (APOBEC1) family of proteins, as well as the adenosine deaminase acting on the tRNA (ADAT) family[17]. In higher eukaryotes, base modifications are the major type of RNA editing[18]. RNA editing is restricted to base modifications of two types: cytidines to uridines (C to U) and adenosines to inosines (A to I). These modifications are the result of enzymatic deamination catalyzed by enzymes of the APOBEC1 and ADAR families, respectively (Figure 1.1).

RNA editing, as well as other RNA post-transcriptional modifications like alternative splicing, serves to diversify the genome, expanding the limited number of genes given the complexity of higher organisms. Using RNA modification strategies,

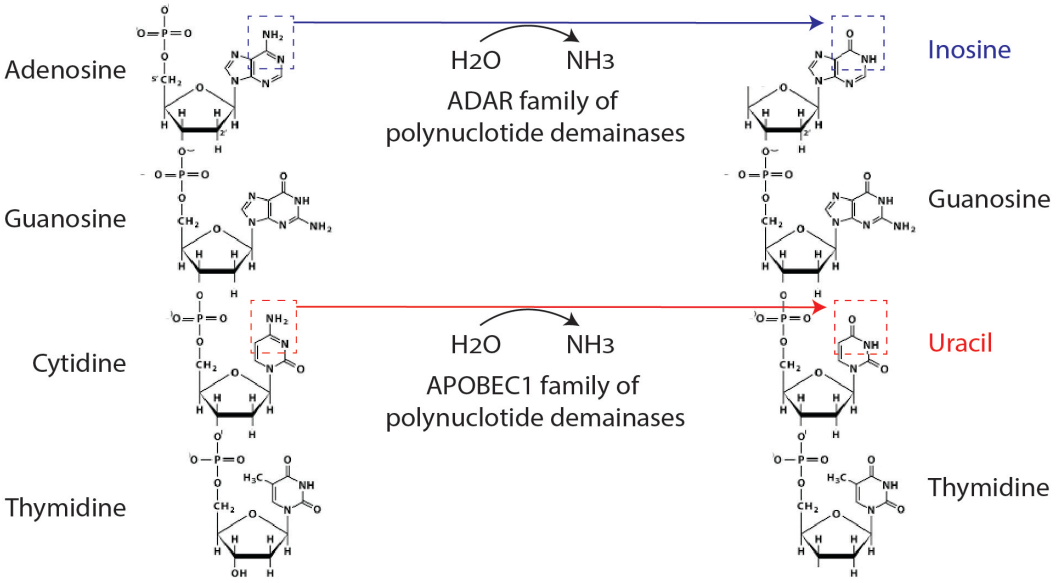

Figure 1.1 | Deamination by the ADAR and APOBEC1 families. Deamination of adenosine to inosine (A to I) in an RNA polynucleotide is catalyzed by the ADAR family of adenosine deaminases (top). APOBEC1 family members catalyze the deamination of cytidine to uridine (C to U; bottom).

many different gene products can be created that fill in the requirements of complex structural, enzymatic and regulatory functions.

1.2.1. Adenosine deaminases

The ADAR family of CDAs posses a conserved modular domain constitution consisting of a variable N-terminal domain, one to three repeats of a double stranded RNA-binding domain (dsRBD) and a catalytic deaminase domain in the c terminal[19]. The catalytic domain consists of amino acid residues conserved with cytidine deaminases, including Apobec1[20]. Mammals have 3 ADAR proteins[21], which are highly conserved in vertebrates. In contrast, lower metazoan organisms such as *Caenorhabditis elegans*[22] and *Drosophila melanogaster*[23] only posses one and in a few organisms, like protozoa, yeast and plants, ADARs are absent[24]. ADAR1 and ADAR2 arose early in metazoan evolution, proven by their presence in in sea urchin and sea anemones[24]. ADAR3 probably arose later in evolution, as a duplication of ADAR2 in vertebrates. Interestingly, ADAR was lost in some species such as insects and squid during subsequent evolution[24].

ADAR1 and 2 have confirmed enzymatic ability to convert A to I in double stranded RNA substrates[20,25–27] and are present in many cell types, with highest amounts in the brain[28]. ADAR3 is brain specific and can bind to single as well as double stranded RNA, but it is thought to be catalytically inactive[27]. The editing activity of ADARs requires homodimerization, being that ADAR3 can bind to ADAR 1 and 2, it is thought to act as an inhibitor[27]. ADARs have a diverse set of substrates: ADAR 1 and 2 will edit almost any double stranded structure, made up of inter- or intra-molecular interactions that make at least 2 helical turns (~20nt)[29]. ADARs are also quite promiscuous, with adenosines

contained in long stretches of dsRNA edited ~50%. In contrast to this, short dsRNAs or long RNAs with imperfect mismatches (loops, bulges, etc.) are edited selectively, indicating that the selectivity is dictated by the secondary structure of the dsRNA[30]. In addition to the double stranded structure, selective editing requires a downstream complementary sequence (ECS) that is crucial for exonic editing, for example in editing of the Ampa receptor subunit GluR-B[31].

1.2.2. Role of ADARs in development and protein diversification

ADAR 1 and 2 are expressed in many tissues, whereas ADAR3 is only expressed in the brain[20,25–27]. In mouse development, ADAR 1 and 2 expressions start at around E10 in the heart. Both ADAR1 and its regulator miR-1 are important for the embryonic heart development[32]. Models that contain deletions in ADAR genes give an insight into the importance of ADAR in development. ADAR1 deficient mice die quickly in utero due to generalized apoptosis[32,33]. Conditional activation studies of ADAR 1 also demonstrated that it is important for the maintenance of the hematopoietic stem cell[33]. ADAR2 deficient mice die soon after birth due to neurological deficiencies resulting from de-regulated GluR[34]. These mice can be rescued with genomic editing of the receptor[34].

Originally, ADAR editing was thought to be restricted to ~30 transcripts, within coding regions. Most of these transcripts are ion channels and neurotransmitters, for example *GluR*[31], the serotonin 2c-receptor G-protein subtype *5-HT$_{2C}$R*[35] and the potassium channel *Kv1.1A*[36]. In these cases the recoding results in the generation of protein isoforms and the diversification of their functions. A mutation of a glutamine (C<u>A</u>G) to an arginine (C<u>I</u>G) in the GluR receptor at a site known as the Q/R site changes

the proteins tertiary structure such that the whole channel becomes impermeable[31,37] to Ca^{2+}. But the complexity does not end there. The glutamate receptor unit GluR-6 harbors 3 exonic editing sites (known as I/V, Y/C and Q/R) that all together give rise to 8 different sequence combinations. Because the extent of editing at each site varies between 10 and 80%, all possible theoretical combinations of the protein products can be found in the nervous system[38]. Just as complex is editing of five sites within 3 codons of the 5-HT_{2C}R (isoleucine AUA, asparagine AAU and isoleucine AUU), which results in up to 6 amino acid changes and in combination, 24 receptor isoforms with variable potency and ligand binding[35,39]. Finally, 14 edited sites identified in the potassium channel Kv1.1 lead to many kinds of outcomes such as inhibition of tetramerization and changes in the rate of deactivation[40].

Editing has been implicated in certain diseases of the brain. For example, underediting of the Q/R site of the GluR has been observed in the death of motor neurons of sporadic amyotrophic lateral sclerosis (ALS) patients[41]. Also, the editing pattern of 5-HT_{2C}R mRNA is significantly altered in the prefrontal cortex of suicide victims[42] .

1.2.3. Functional roles of ADAR editing

The protein coding ADAR-mediated modifications mentioned above are just a small fraction of the transcripts targeted by ADAR. Using new bioinformatic approaches, a wide number of sites have been discovered, mostly within inversely oriented repetitive elements in non-coding regions, especially within Alu regions (short interspersed elements that are unique to primates) and Long interspersed nuclear elements (LINE)[43,44].

Editing of non-coding regions involves ADAR into to several processes: alternative splicing, nuclear retention and interaction with miRNA pathways. Alternative splicing can come about when ADAR targets 2 nucleotides to create a novel splice site: the highly conserved canonical 5' GU (from AU) recognition sequence or the 3' splice acceptor AG (form AA) site[43]. In the same way, it can destroy the 3' splice site AG (from GG)[44].

Nuclear retention has been identified for the mouse cationic amino acid transporter 2 (CAT2). The mechanism is believed to involve A to I editing on SINE elements present within its 3'UTR, which mediate binding to the nuclear retention protein $p54^{ntb}$. This interaction traps the RNA of CAT2 in nuclear speckles known as interchromatin granule clusters[45].

It has been shown that ADARS modulate different pathways of miRNA synthesis, leading to suppression or enhancement of miRNA processing steps. In hematopoietic cells for example, ADAR1 and 2 edit pri-miRNA-142, which results in degraded by the nuclease Tudor-SN[46]. Another example is the processing of pri-miRNA-151, which is also a substrate of both ADAR1 and 2. Editing of this pri-miRNA inhibits its cleavage by Dicer and its accumulation suppresses the expression of mature miRNA-151[46].

ADAR editing can also antagonize miRNA-mediated gene silencing, through interference with the miRNA-induced silencing complex (miRISC) assembly. One example where this occurs with the Epstein Barr virus encoded pri-miRNA BART6. Upon editing of this pri-miRNA the loading efficiency of miR-BART6 to the miRISC reduces, lowering the silencing of it's target mRNA[47]. Additionally, mature miRNAs have been found to be edited and are predicted to repress a set of genes that differ from

those targeted by the unedited miRNAs. This is the case for miRNA-376, whose seed sequence is edited by ADAR2 in the +4site, essential for its binding to the complementary sequence in mRNA[48].

Finally, a role for miRNA regulation has been proposed for ADAR 1 and 2, independently from their RNA editing role. A catalytically inactive ADAR2 has been shown to inhibit Drosha processing of pri-miRNA-376a2 through RNA binding. Furthermore, there is evidence that ADAR1 can form a heterodimer complex with Dicer to promote cleavage of miRNAs and facilitates loading of miRNA into other silencing complexes[49].

1.2.4 ADARs in the innate immune response

In its role as an adenosine deaminase, ADAR participates in the response to viral infection, for example in brain infections with the measles virus, samples form patient brains show that 2% of the nucleotides were edited, resulting in alteration of reading frames and fused genes[50]. This reaction has been characterized to involve the IFN-inducible ADAR p150 isoform[51]. Another example is the hepatitis C virus (HCV) sensitivity to interferon-α (IFNα). Many clinical HCV infections are resistant to IFN-α therapy. However, subgenomic in vitro self-replicating HCV RNAs (HCV replicon) have a marked IFNα sensitivity[52]. In this case IFNα increases the expression of ADAR1 resulting in a decrease of viral replication that leads to genomic instability.

ADAR can also establish a proviral response, depending on the type of virus. The hepatitis delta virus (HDV) requires a larger version of its short DNA encoded HDV antigen for packaging its genome into virions. In order to do so it takes advantage of

ADAR1 editing to provide the change necessary for the formation of the larger HDV antigen by changing a UAG stop codon to a UIG tryptophan codon[53].

Another function of ADAR in the immune response is in the distinction of "self" dsRNA from that generated through viral infections. Viral dsRNA comes from many sources including dsRNA genomes, hairpin secondary structures in mRNA and replication of ssRNA viruses. If the dsRNA is found in the cytoplasm, it will encounter the retinoic acid-inducible gene I (RIG-I) and melanoma differentiation-associated gene 5 (MDA5) receptors, which will induce a type I IFN response, unless dsRNA possesses inosine in its structure[54]. In this scenario, ADAR1 is a negative regulator of the type I IFN response by editing viral dsRNA, which is then confused as "self", through specific binding to RIG-I and MDA5[55]. The elucidation of this process was pieced together though studies of ADAR deletion. First, mouse embryonic fibroblasts deficient in ADAR1 p150 are highly susceptible to infection and have enhanced viral induces cytotoxic effects[51]. This study implicated ADAR1 in the repression of type I IFN response. Also, in the search for the mechanism by which ADAR1 deletion causes embryonic lethality, the generation of a double mutant Adar1+/- and Mavs-/- led to the discovery that embryos that were homozygotes for both survived to birth[54]. The surviving embryos had a much less heightened IFN response, placing ADAR1 upstream of RIG-I and MDA5 in the IFN activation pathway[54].

1.2.4. Cytidine deaminases

The AID/Apobec family of cytidine deaminases consists of several members that catalyze cytidine deamination in both DNA and RNA. APOBEC1 was the first member

of its family to be discovered and thus bears the name of its first discovered substrate Apolipoprotein B. Other family members include AID (Activation-induced cytidine deaminase), Apobec2, Apobec3 (A-H) and Apobec4. All of these share a conserved catalytic domain, which is zinc dependent (reviewed in Smith 2009), which has a big resemblance to the ancestral domain found in cytidine deaminases acting on free cytidine (Conticello, 2005). However, the AID/Apobec family is a later evolutionary development restricted to the vertebrate lineage. The first member is thought to have been AID, which arose concurrently with the development of adaptive immunity in vertebrates (Conticello, 2005).

Throughout evolution, AID/Apobec enzymes evolved rapidly, displaying the strongest signals of positive selection in the human genome, which is a pattern associated with host defense (Sawyer, 2004). For example, Apobec3, has undergone a dramatic expansion in primates into 8 subfamily members, in comparison to mouse, which possesses only one. Indeed, most family members function in immunity: AID is crucial for antibody diversification through somatic hypermutation and class switch recombination; and Apobec3's act on both endogenous and exogenous retroviral genomes (reviewed in Hamilton et al., 2010).

Until recently, the only known member of the AID/APOBEC family to act on RNA was Apobec1: AID is restricted to double stranded RNA, Apobec3's to single stranded DNA and Apobec2 and 4 have no identified substrates. However, recent studies published during the course of my thesis work, have demonstrated that a member of the Apobec3 subfamily, Apobec3A can also edit RNA substrates[56] (REF: Sharma 2015). The canonical role of Apobec3A is related to its DNA editing ability in the inhibition of the

replication of the adeno-associated virus (AAV), a small single stranded DNA parvovirus[57]. Sharma and colleagues showed that in human monocytes, Apobec3A is capable of editing a large number of endogenous mRNA transcripts and that editing increased under different physiological conditions such as hypoxia and interferon gamma induced inflammation[56].

1.2.5. APOBEC1 editing

Apobec1 has a well-characterized function in the intestine, where it targets cytidine 6666 in the transcript of ApoB, changing a glutamine codon (**C**AA) to a termination codon (**U**AA)[58,59] (Figure 1.2). The resulting truncated protein (ApoB48) is functionally different from its full-length counterpart. This is very important for lipoprotein assembly and transport because ApoB is a structural protein of VLDL particles, whereas ApoB48 is crucial for the formation of chylomicrons. As expected, the loss of APOBEC1 in mice results in elevated LDL levels and cholesterol[60] and is associated with susceptibility to atherosclerosis[61].

Editing of ApoB is regulated in a tissue specific manner through restricted expression of Apobec1. In adult human intestine, where Apobec1 is expressed, the efficiency of editing is >90%, in contrast to the liver, where ApoB remains unedited (in mice ApoB is highly edited in both liver and intestine). Furthermore, editing of the specific cytidine within the ApoB transcript seems to be exquisitely regulated and is correlated to Apobec1 expression levels. In order to achieve its specificity, APOBEC1 requires the interaction with a multi-protein editing complex. The principal component of this complex is the RNA binding subunit Apobec-1 complementation factor (ACF),

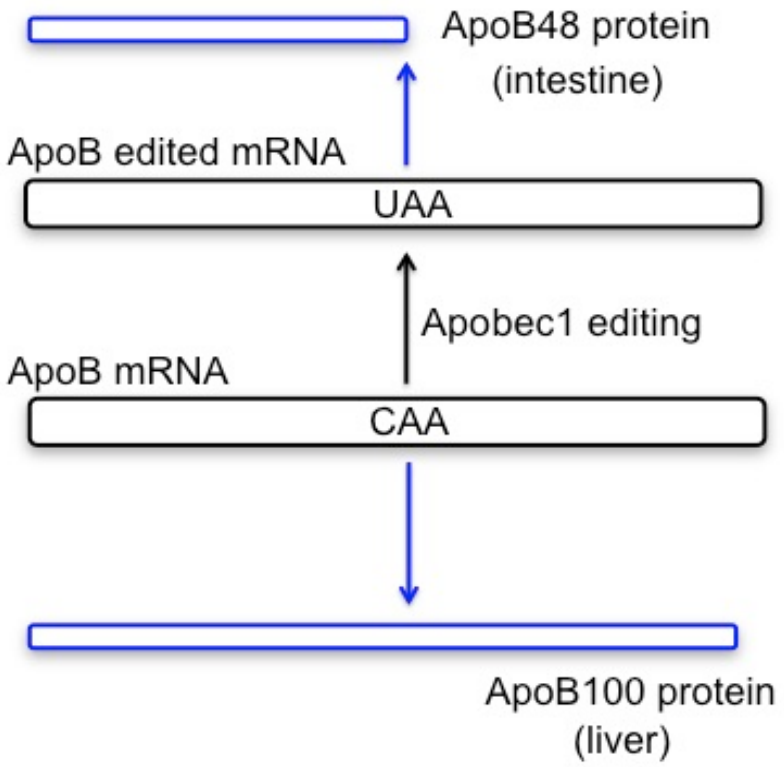

Figure 1.2 | APOBEC1 editing of the Apolipoprotein B transcript results in the creation of a smaller protein. APOBEC1 mediated deamination of cytosine to uridine changes a glutamate (CAA) to a stop codon (UAA), which results in the creation of a truncated protein, ApoB48. This is an organ-speciLic mechanism that allows for the creation of two different lipid particles, VLDL and chylomicrons, composed of ApoB100 and ApoB48 respectively.

which together with Apobec1 forms the minimal functional editing enzyme complex[62,63]. In contrast to Apobec1[58], ACF is an essential gene and mice lacking it die in the pre-implantation stage of development[64]. ACF binds the transcript of ApoB through recognition of an 11 nucleotide mooring sequence surrounded by an AU rich environment. ACF has also been shown to protect the ApoB transcript with a premature stop codon from nonsense-mediated decay[65].

Liver –specific overexpression of Apobec1 in mice revealed Apobec1 changes its specificity of editing in ApoB, resulting in extensive editing at multiple cytidines downstream of the canonical one, a phenomenon known as hyperediting[66]. It is presumed that the reason for this loss of specificity is the altered stoichiometry of Apobec1 to ACF, which was confirmed with forced over expression of Apobec1 *in vitro*[66–68]. However, this is not the case when the ratio is altered via reduced ACF in Acf+/- mice, who exhibit increased hepatic editing at the ApoB canonical site but not hyperediting[69]. This indicates that there is still much to be investigated about the stoichiometry of the proteins involved in the regulation of the editing complex.

1.2.6. Alternative APOBEC1 activities

Besides its RNA editing ability, Apobec1 has been shown to have two activities which have been observed *in vitro* but that are yet to be confirmed in physiological conditions: modification of transcript stability and DNA editing.

APOBEC1 has been shown to be an RNA binding protein, with specificity to AU-rich sites, similar to the one found surrounding the ApoB mooring sequence. Specifically, it has been shown to bind the 3' untranslated region (UTR) of c-myc, TNF-α and IL-2

when fused to luciferase reporters and alter their stability[70]. Similarly, it has been shown that APOBEC1 can confer stability to the cyclooxygenase-2 (COX-2) transcript, which contains AU-rich elements that mediate its decay, through binding of its 3'UTR, when cloned into a luciferase vector. Altogether, these observations suggest that Apobec1 may have a role in transcript stability, unrelated to its editing ability.

When tested in a DNA mutation model, the mutation induced rifampicin-resistant *E. coli*, Apobec1 was shown to be a strong DNA mutator, displaying comparable if not better activity as its close relative AID[71]. This was also confirmed biochemically using *in vitro* deamination assays with bacteria-purified Apobec1 and DNA substrates. These assays revealed that Apobec1 possessed weak C to U deamination, which was restricted to single-stranded DNA substrates[72]. However, there is thus far no evidence of APOBEC1-induced mutations in the genome of naturally expressing Apobec1 cells.

1.2.7 APOBEC1 in cancer

The jury is out on whether APOBEC1 is involved in natural tumor formation. Evidence from over-expression models of APOBEC1 has shown that, in its role as a DNA editor, it is a potential oncogene. However, there is still very limited evidence that links APOBEC1 editing to cancer originating from patient samples. Evidence from both instances is presented below.

Until recently, the only other known target of APOBEC1 editing was the neurofibromatosis type 1 RNA (NF1) tumor suppressor, shown to be edited in a subset of the patients affected with peripheral nerve sheath tumors (PNSTs)[73,74]. These patients have a particular alternatively spliced version of the Nf1 transcript, which includes exon

24A[74]. Apobec1 changes an arginine (CGA) to a translational stop codon (UGA) through binding to a mooring motif similar to the one in ApoB. The resulting truncated protein lacks the GTPase-activating protein region that is responsible for tumor suppression[73]. This finding represents the first time that Apobec1 was found to act outside of the intestine and opened up the possibility that Apobec1 may have alternative targets and therefore additional function. This shows that Apobec1 could be involved with cancer formation.

A second line of evidence that APOBEC1 could be involved in cancer formation comes from studies in over-expression mouse models. For example, APOBEC1 caused hepatocellular carcinoma in mice with transgenic hepatic overexpression. Upon examination of edited targets, hyperediting in a novel Apobec1 target 1(*Nat1)* was identified[75]. Nat1 binds the initiation factor 4a (eIF4A), which inhibits both cap-dependent and cap-independent translation. The aberrant editing alters amino acids and creates multiple stop codons, which in turn result in the reduction of the protein. This also shows an oncogenic potential for APOBEC1, however this evidence does not come from physiological levels of APOBEC1 expression.

Further links of APOBEC1 with oncogenesis come from studies of testicular germ cell tumors (TGCTs)[76] and adenocarcinoma of the small intestine in tumor-susceptible mouse models[77]. In the $Apc^{min/+}$ mouse model of small intestinal adenoma formation, loss of APOBEC1 reduced tumor burden, increased apoptosis and reduced proliferation[77]. However, the authors speculated that the mechanism is through the stabilization of transcripts such as COX-2, given that adenomas from $Apc^{min/+}$Apobec-/- mice had a marked reduction in *Cox-2* mRNA. This could point to an editing-independent

oncogenic mechanism, however a genome wide search of alternative APOBEC1 targets was not pursued, and so the possibility remains that editing could be involved.

However, a study that searched for aberrant APOBEC1 editing in tumor specimens of different types, including hepatocellular, bile duct, gastric, colorectal, pancreatic and breast carcinomas, failed to confirm that APOBEC1 editing is involved in tumor genesis of these cancers[78]. This is the result of either low levels of APOBEC1, such as in colorectal and lung carcinoma; or because whether an active version APOBEC1 or the required cofactors are absent from the tumor sample[78].

1.2.8. High throughput RNA seq identification of APOBEC1 editing

After a several year hiatus, research on APOBEC1 re-emerged as new sequencing technologies became available and enabled the identification of new targets. In this section I will describe research in my laboratory that set the basis for the discovery of hundreds of additional sites, as well as preliminary experiments in macrophages.

High throughput detection of RNA modifications relies on the traces that such modifications leave behind in cDNA, as the reverse transcriptase is confronted with non-canonical substrates. For example, large modifications tend to lead to arrest the reverse transcriptase and modifications on the Watson-Crick pairing will lead to misincorporation of non-complementary dNTP. The former results in shorter cDNA or abortive sequences, while in the latter case, the signature results in apparent mutations in the cDNA sequence. For example inosine, which is the result of ADAR-mediated deamination of adenosine, pairs with cytidine leading to the incorporation of dCTP instead of dTTP onto the cDNA.

These mutations can be then identified bioinformatically, through the comparison of the cDNA sequence to its corresponding DNA sequence.

In order to find alternative functions for APOBEC1, we need to uncover the universe of sites that are edited. To do this, our laboratory developed a comparative approach to identify high confidence APOBEC1-mediated edited sites in a genome-wide manner. The method is based on the analysis of mRNA high throughput sequencing data taking advantage of the editing signatures in cDNA, where C to U editing causes the misincorporation of dATP instead of dGTP. The pipeline takes the aligned reads from the genome and quantifies the number of C to T mismatches in comparison to the genome. Then a series of filters is applied: mismatches that map to known single nucleotide variants are discarded. Finally, the list of the mismatches in wildtype BMDMs is compared to their *Apobec1*-/- counterparts. Bona fide editing sites are ones that occur only in the wildtype sample and are absent in the *Apobec1*-/- sample.

This method was used to identify APOBEC1 mediated C to U editing in mouse intestine. As a result, 36 additional sites were identified. Interestingly, the site at the Apolipoprotein B transcript is the only site located at the 3'UTR; all additional sites were found in the 3' untranslated region of transcripts.

1.3. Statement of purpose

It is a very exciting time to do research in the field of epitranscriptomics given that, with the development of new tools, we are now able to map large numbers of modifications in mRNA. We know that many of them are found in tRNA and rRNA and are crucial for the cell. In a similar manner, it is possible that such modifications play

crucial roles not only for transcript processing and fate, but also for the adaptability of cells to changing environments.

One of the modifications that have been thus far overlooked is C to U editing. This is mainly because until recently, it was believed to be restricted to one enzyme (Apobec1), one substrate (ApoB) and one system (intestine). Our group uncovered a bigger universe of transcripts affected by Apobec1 and its physiological activity in immune cells. This opens up the possibility that C to U editing has a bigger role than previously thought. It is becoming clear that C to U editing is a widely used mechanism of genetic diversification that warrants further studies.

Alternative functions of Apobec1 still remain to be established. The methods our group created to globally identify APOBEC1 edited sites allowed us to have a good catalog of APOBEC1 targets and edited sites in macrophages. In this system, ApoB is not expressed, which raises the possibility that Apobec1 editing has a novel role unrelated to lipid metabolism. I set out to provide answers, even if partial, to the following questions: What is the purpose of Apobec1 editing in macrophages? Does editing confer additional transcript regulation? Does editing regulate the activity of macrophages?

With these questions as main motivation of my work, I focused on two main objectives: (1) to examine the functional consequences of editing at a molecular and cell level; and (2) to find out what, if any, is the role of APOBEC1 in tissues other than the intestine and liver.

CHAPTER 2. IDENTIFICATION OF HIGH CONFIDENCE EDITED SITES IN BONE MARROW DERIVED MACROPHAGES

The editing detection method mentioned in section 1.2.8 (above) can be directly applicable to any kind of cell that expresses APOBEC1 or other editing enzymes. However, in an effort to improve the detection method, our group further developed a new method of editing identification that allowed us to confidently call low frequency edited sites. Our group then looked at other systems that expressed APOBEC1 in which it's activity could be characterized. The cell selected was bone marrow derived macrophages, which are cells of the immune system that express very high levels of APOBEC1, are readily available through differentiation of bone marrow progenitors (see methods) and do not express the canonical co-activator ACF nor its canonical target ApoB. Recently, RBM47, an RNA binding protein, was discovered to be a cofactor protein for Apobec1 editing of ApoB mRNA[79]. Rbm47 has homology with ACF and is able to form the core editing enzyme complex. This cofactor is expressed in BMDMs and is a good candidate for aiding Apobec1 editing in these cells, although this remains to be investigated.

The new editing pipeline is also based on a comparative RNA-seq approach[80]. Briefly, RNA from resting wildtype and Apobec1-/- BMDMs was extracted and mRNA libraries were made and sequenced. Once the reads were permissively aligned to the reference genome, each site was scanned and mismatches to the reference genome were quantified. Each type of mismatch was then used to build vectors (C, T, A, G) for each chromosomal coordinate, with a corresponding vector for the Apobec1-/- sample. In

order to determine true edited sites, the angle between the two vectors is compared. A larger angle is a true edited coordinate (Figure 2.1). The method also incorporates filters for artifacts: blat filter that eliminates reads mapping to alternative genomic sites; filters for known variants, strand bias and variants at the end of reads, where a higher rate of sequencing errors reside.

Using this method, we were able to detect 410 high-confidence C to U editing events in 275 transcripts. The large majority of them (97%) were contained in the 3'UTR of the transcripts and most of the sites were edited at low frequency 10-20% (Figure 2.2). We also observed that the sites could be classified into single edited and hyperedited (high frequency editing in one location, with additional lower frequency editing in alternative nearby sites, figure 2.3.B). A great number of sites were validated by Sanger sequencing of bacterial colonies containing fragments amplified from cDNA from wildtype and Apobec1-/- BMDMs (Figure 2.3).

However, we found that the global gene expression of the genes between the two genotypes is highly correlated (not shown), indicating that there are no differences in gene expression between the two.

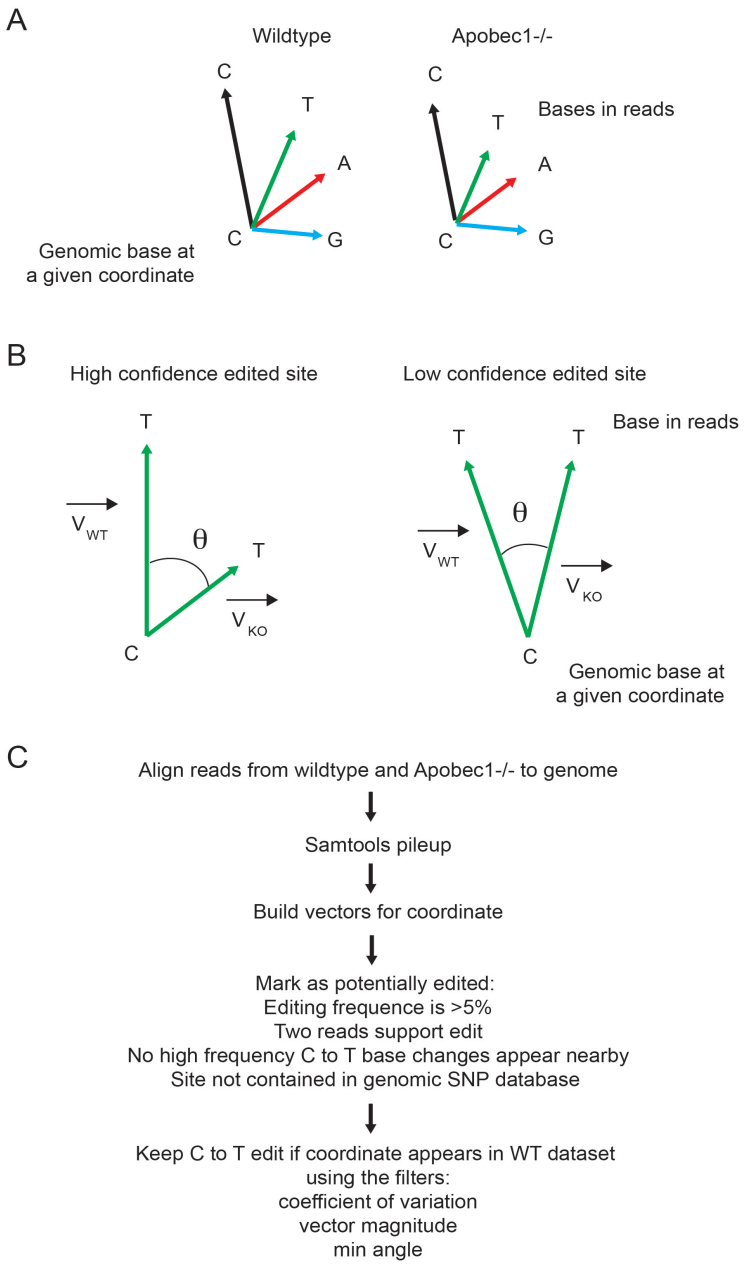

Figure 2.1 | Pipeline for editing detection. (A) Depiction of vectors in a 4D space, built for every substitution at every base in both genotypes using data from the pileup. (B) Vectors for both wildtype (WT) and Apobec1-/- (KO) are compared and the magnitude, angle and variation between them is quantiLied. A high conLidence edited site possesses a large magnitude angle (left), whereas a low conLidence site has a low angle (right). (C) Pipeline used for bioinformatic detection of edited sites.

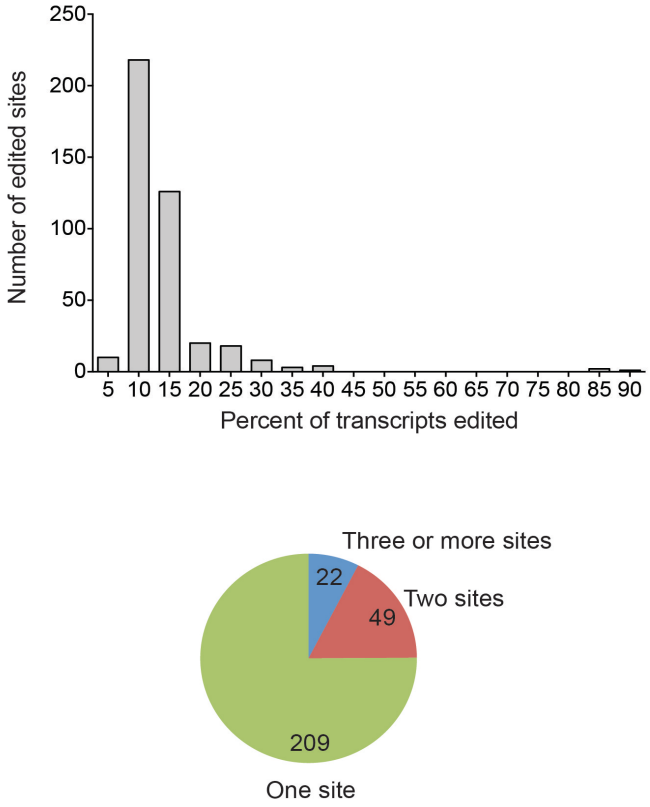

Figure 2.2 | Distribution of APOBEC1 editing in BMDMs. Most transcripts found were edited at a low rate. Transcripts with more than one edited site were also found.

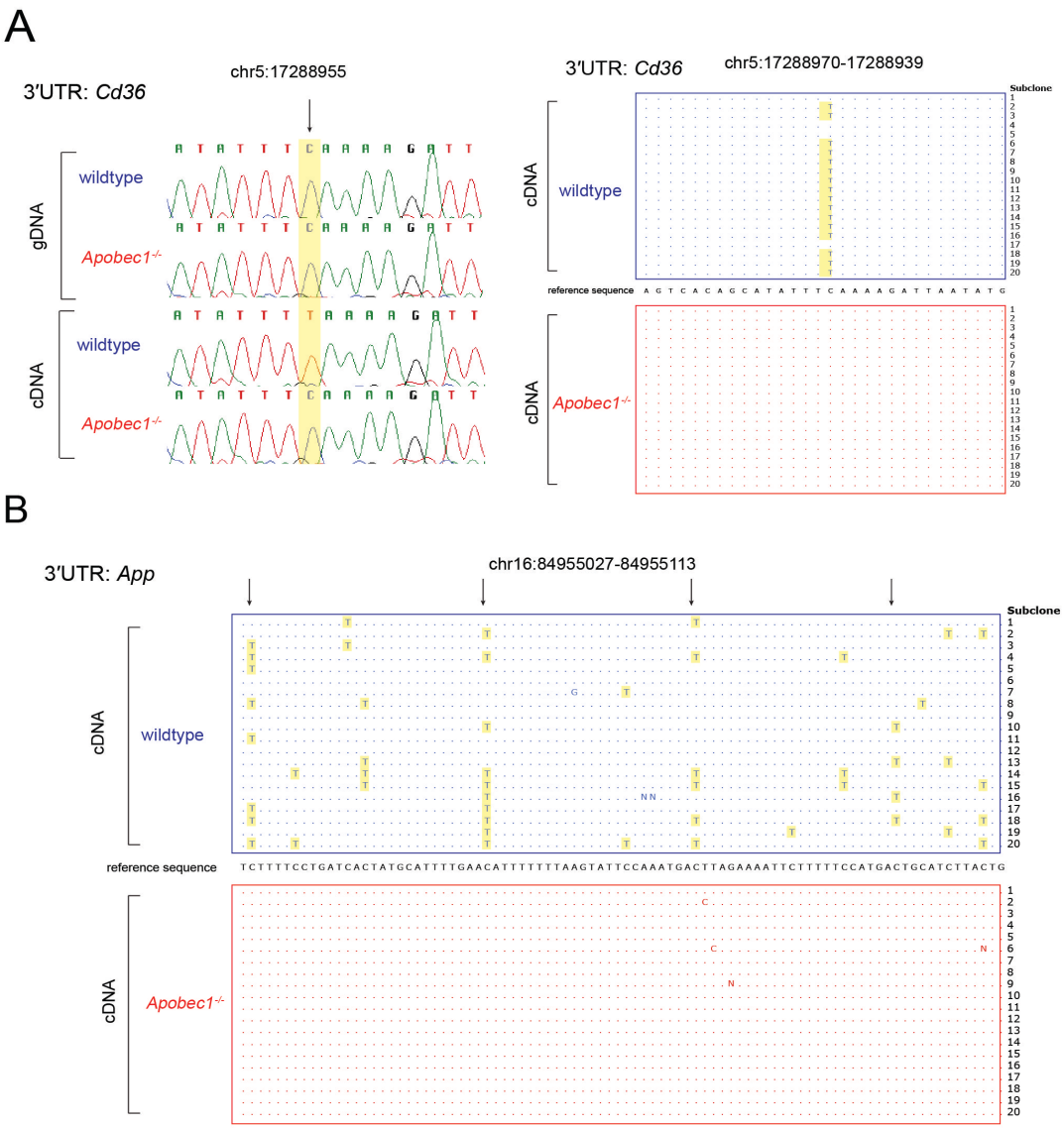

Figure 2.3 | Validation of edited targets via Sanger sequencing. Examples of Sanger sequencing validation of bioinformatically identified Apobec1 C to U editing. (A) Comparison of a fragment of cDNA to gDNA of wildtype and Apobec1-/- Cd36. The edited site identified is highlighted in yellow (left). Sequences of individual colonies, aligned to the reference genome (right). Dots indicate matches to the reference sequence whereas mismatches are indicated with letters. C to T editing events are highlighted in yellow. (B) Fragment of the transcript App revealing sites of hyperediting. Produced by Claire Hamilton.

CHAPTER 3. MOLECULAR CONSEQUENCES OF APOBEC1 EDITING

The 3' UTR region of transcripts is a regulatory hub that controls many processes, such as nuclear export, stability, degradation and ribosome loading. Therefore, alterations in this region have the potential to result in regulation of transcript fates. This has been previously shown to be a possible consequence of ADAR editing[81,45]; therefore it is possible that Apobec1 editing has similar consequences.

3.1. APOBEC1 editing in transcript fate

Given that Apobec1 editing in BMDMs did not seem to alter gene expression globally, I first characterized other potential functions of 3'UTR alterations at the transcript level. The specificity of APOBEC1 to the 3' untranslated region of the UTR suggests that editing plays a regulatory mechanism for transcripts. This has been previously observed for editing enzymes of the ADAR family. ADAR editing has been shown to have the potential to alter transcript stability[81] and localization[45], causing nuclear retention of the edited transcript. To test whether APOBEC1 editing resulted in altered localization of transcripts, I designed a cell fractionation experiment to compare the levels of editing between the cytoplasmic and nuclear fractions. I selected various transcripts that had different levels of editing at a single site, as well as hyperediting. After cell fractionation using cell lysis buffers with different composition of surfactants, I was able to isolate cytoplasm, nucleus and chromatin. Then, I extracted RNA form each fraction, made cDNA and tested for abundance of each transcript using qPCR.

In order to compare the frequency of editing at each site in the cytoplasm and nucleus, I amplified a region surrounding the edited site and submitted individual bacterial clones for Sanger sequencing. Western blot analysis of the different fractions using the antibodies for the histone H3 and tubulin confirmed the purity of the nuclear and cytoplasmic fractions respectively (not shown).

As expected, I found that all edited transcripts in the *wildtype* sample were edited in the nucleus. This is consistent with the previously reported mode of action of APOBEC1[10]. Figure 3.1.1 shows two examples of editing in the different fractions of *wildtype* samples. Individual sequences derived from the colony sequencing, amplified from each of the fraction's cDNA, were aligned to the reference genome. Using this data, I quantified the fraction of edited transcripts, counting as edited any transcript with at least one C-T change in the whole length of the amplified fragment, even if the previously predicted site did not show editing. As can be seen in Figure 3.1.2.A, the fraction of edited transcripts in the nucleus is very similar to the cytoplasm, which indicates that editing does not cause selective retention of edited transcripts in the nucleus. This is confirmed by the qPCR data in Figure 3.1.2.B, which shows that, even though the amount of transcripts within the nucleus is higher than in the cytoplasm, there is no difference between the transcripts of *wildtype* and *Apobec1-/-* BMDMs.

An alternative mechanism of transcript regulation that could be affected by editing is the stability of transcripts. To test whether editing would increase or decrease transcript stability, I quantified several edited transcripts at different time points upon addition of actinomycin D to the cell culture, which inhibits transcription. This allows for evaluation of the decay of transcripts, because no new ones are added to the pool. In

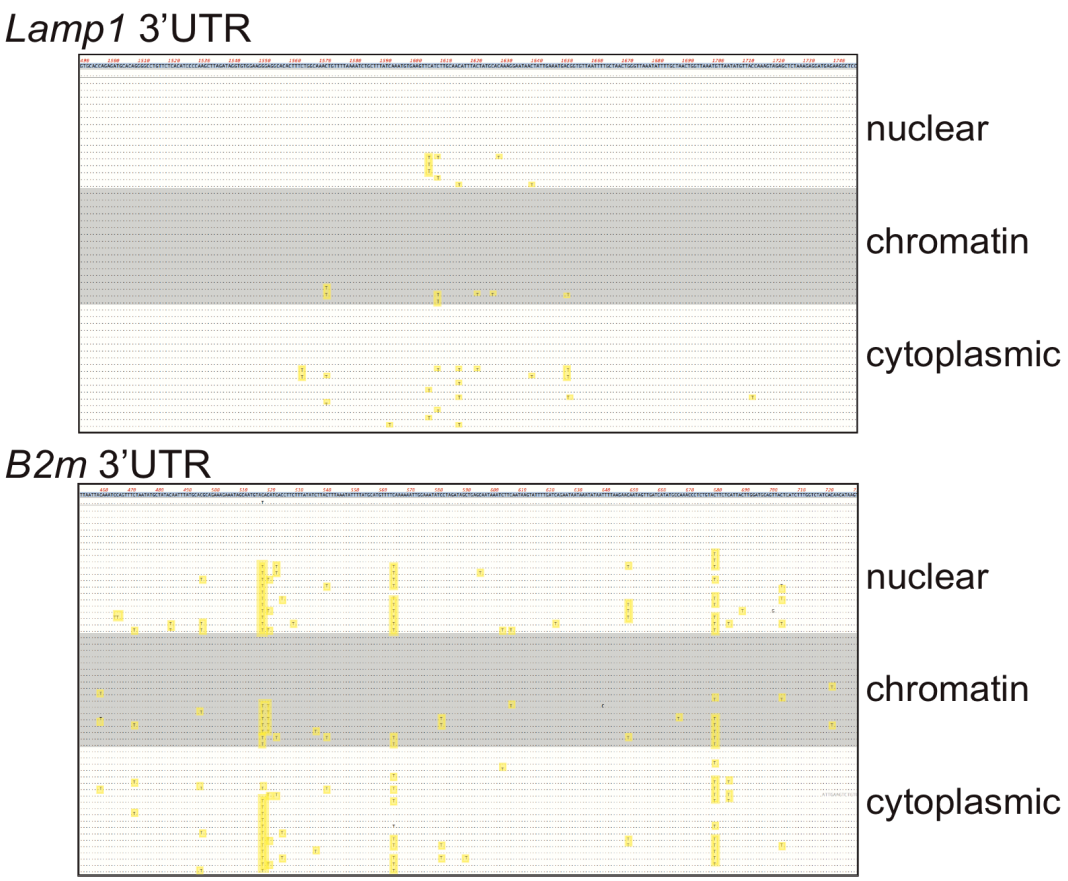

Figure 3.1.1 | Effect of editing in transcript localization. Aligned sequences of the PCR fragments amplified around the edited sites of the *Lamp1* and *B2m* 3'UTRs. The C to T changes, with respect to the genomic sequence, are highlighted in yellow.

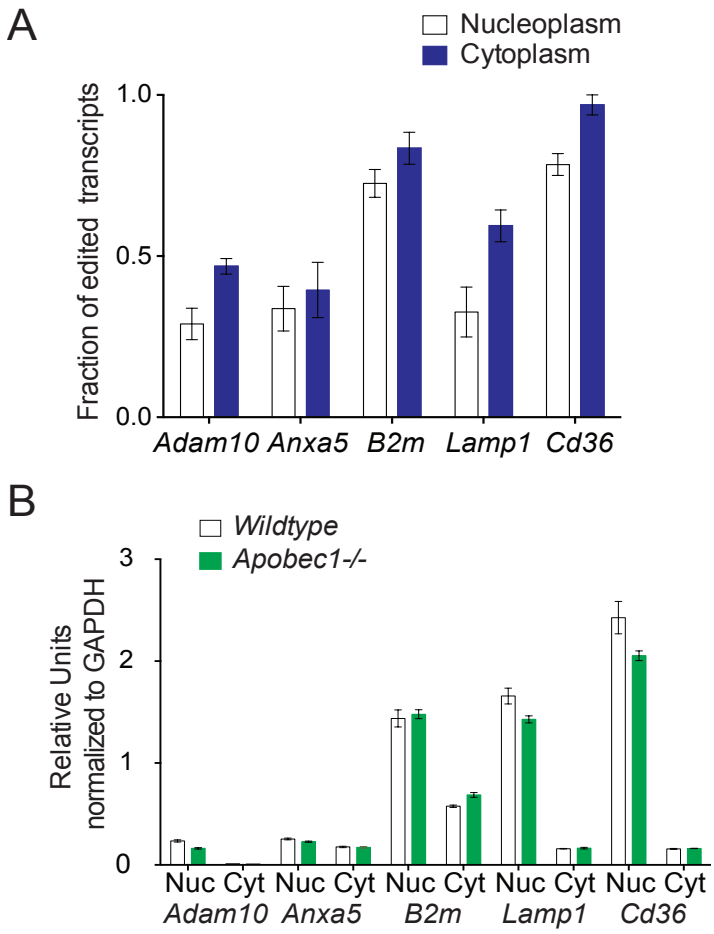

Figure 3.1.2 | Effect of editing in transcript localization. (A) Quantification of the fraction of edited transcripts in the nuclear and cytoplasmic cell fractions (n=3). (B) Relative levels of transcripts in the nuclear and cytoplasmic fractions of *wildtype* and *Apobec1*$^{-/-}$ BMDMs. Transcript levels were normalized to *Gapdh*. All error bars represent the standard error of the mean.

addition to the edited transcripts, I also quantified known stable and unstable transcripts for comparison. The decay of edited transcripts in both *wildtype* and *Apobec1-/-* BMDMs was quantified via qPCR and plotted. Figure 3.1.3 shows that, as expected, there remains above 50% of the known stable transcript *Atp5a1* after 4 hours after Actinomycin D addition, whereas there is almost no trace of the unstable transcript *Jun* at the same time. When the edited transcripts were examined, I observed that they are stable, such that there is at least 50% transcript left after 4 hours.

This is interesting, given that it has been demonstrated that Apobec1 directly binds transcripts with fast turnover, increasing their stability, but in this case, edited transcripts are long-lived. Perhaps this dynamic is dependent on co-factor utilization. Interestingly, the stability of *Apobec1-/-* transcripts is very similar to their *wildtype* counterparts, indicating that editing does not modify transcript stability.

3.2. miRNA regulation through editing

APOBEC1-dependent 3′UTR edits were previously described to be preferentially located in regions of substantial phylogenetic conservation[82]. Conserved untranslated regions are also regions of abundant Argonaute (Ago) occupancy and resultant miRNA targeting[83], raising the possibility that APOBEC1 editing might influence transcript regulation by miRNAs. Our group previously investigated whether the alteration of 3'UTRs leads to the change of microRNA binding sites using the CLIP-seq[84] technique. This consists on immunoprecipitation of the RNA binding protein Argonaute, followed by protein digestion and library preparation of the resulting fragments. The two types of RNA obtained after this procedure are miRNAs and mRNAs that are interacting with

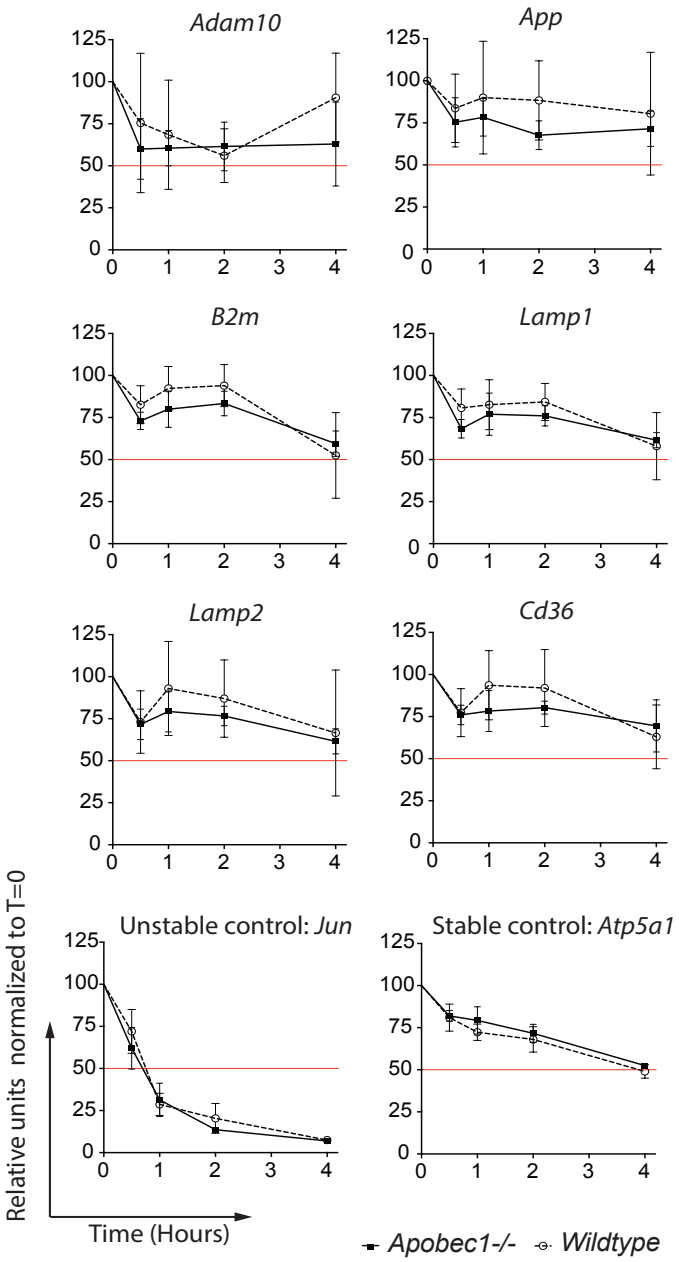

Figure 3.1.3 | Effect of editing in transcript stability. A time course of transcript abundance of the indicated edited transcripts after actinomycin-D treatment of *wildtype* and *Apobec1*$^{-/-}$ BMDMs is shown. *Atp5a1* and *Jun* were used as controls of known stable and labile transcripts, respectively. The red line marks a 50% reduction (n=3).

Argonaute at the moment of cell lysis. The mRNA fragment represents only the fraction protected by Argonaute, therefore it is possible to bioinformatically identify the possible miRNA target sites within this footprint and predict which miRNAs are active.

Indeed, high-throughput sequencing of the mRNA and miRNA fragments revealed that Argonaute significantly bound the subset of transcripts edited by APOBEC1, even though the comparison of *wildtype* and *Apobec1-/-* revealed there was no differential binding of Argonaute. Therefore, we focused on the subset of 3'UTRs where Ago occupancy overlapped with edited sites and assigned to these a set of likely miRNA targets, selected based on miRNA abundance and recently defined rules for canonical as well as non-canonical binding[85]. Upon examination of the mRNA from the footprint, we bioinformatically identified several miRNA binding sites that were potentially created or destroyed by C to U editing, which are candidates for miRNA regulation through APOBEC1 editing (Figure 3.2.1).

In order to determine whether a single nucleotide change is sufficient for the disruption of a miRNA binding site, we performed luciferase assays using a fusion of the luciferase reporter to the edited version of several 3'UTRs. We cloned edited and unedited 3′UTRs into dual-luciferase expression vectors and co-transfected them with their putative miRNA in HEK-293T cells (or an irrelevant control miRNA) in the absence of APOBEC1. We identified 3′UTRs (e.g. Sptssa and Rac1) where editing disrupts predicted miRNA-UTR interactions, resulting in de-repression of luciferase levels in the edited construct, as compared with the unedited construct (Figure 3.2.2). Two of the four miRNA tested were indeed disrupted by editing. This demonstrated that for some miRNAs, a C-T change in the binding sequence is sufficient to eliminate the

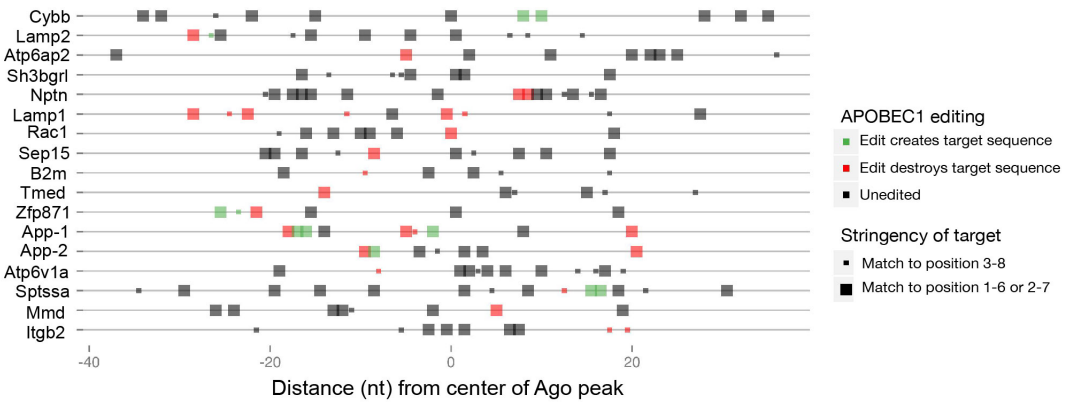

Figure 3.2.1 | Putative miRNA targets in APOBEC1-edited regions that overlap with Ago footprints. "Edited" (with C-to-T mutations reflecting APOBEC1-dependent changes) and "Unedited" (reflecting the genomic reference) footprint sequences were scanned for miRNA targets regions (match to position 2-7, 1-6 or 3-8 of mature miRNA sequence). miRNA targets that would be created (green) or disrupted (red) by an APOBEC1 editing event are depicted. Figure produced by Dewi Harjanto.

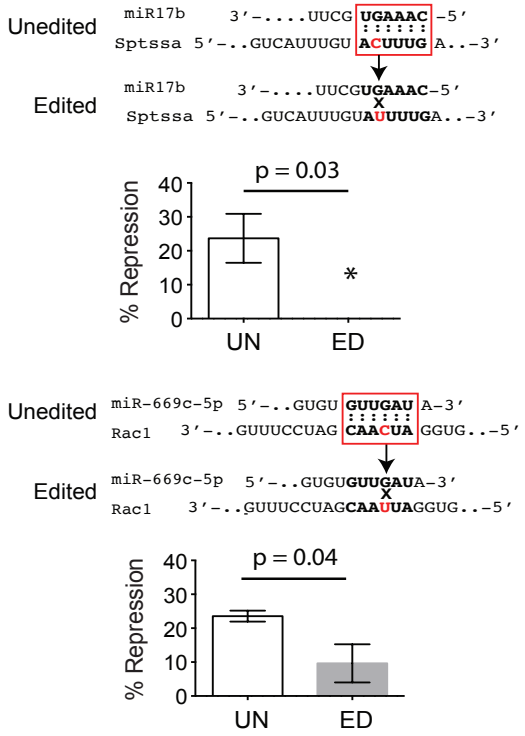

Figure 3.2.2 | APOBEC1 editing disruption of putative miRNA target regions in the *Sptssa* and *Rac1* 3′ UTRs. UN- refers to the "unedited" construct with a sequence consistent with the reference genome. ED- refers to the "edited" construct, mutated to reflect the editing event in question. A schematic depicting miRNA-site deletion or creation by APOBEC1 editing is shown at the top of each graph. The star indicates values below 0 or no relative repression (n=5).

interaction between miRNA and mRNA. This serves as proof of concept that demonstrates that under some circumstances, C to U editing has the ability to disrupt repression through miRNA binding. However, further research is required to figure out the rules that govern transcript de-repression and the extent to which edited transcripts are targets of miRNA binding.

3.3. Editing alters protein production though miRNA independent mechanisms

The overlap between edited sites and Argonaute binding is, however, not complete. The vast majority of edited transcripts have no predicted Argonaute binding. Therefore I tested whether editing events that did not alter a putative miRNA binding sites could also lead to modulation of protein expression. In order to test this, I created luciferase constructs that contain either the unedited or edited version of several 3'UTRs fused to the renilla luciferase gene. Additionally, these constructs contain the firefly luciferase gene, which is co-expressed and used as a transfection control. Each construct was then introduced into Apobec1-/- BMDMs (which contain the physiologically relevant milieu of RNA binding factors), to avoid further editing, and luciferase levels were quantified 24h later. Figure 3.3.1.A shows that from the singly edited constructs, only the edited 3' UTR of *Cd36* showed a decrease in protein product. On the other hand, all hyperedited transcripts showed different levels of regulation (Figure 3.3.2A), albeit small. This experiment also allows us to determine the effect of editing separate from any alternative activity of APOBEC1, and confirms that the effect in protein modulation is the direct result of C to U modifications.

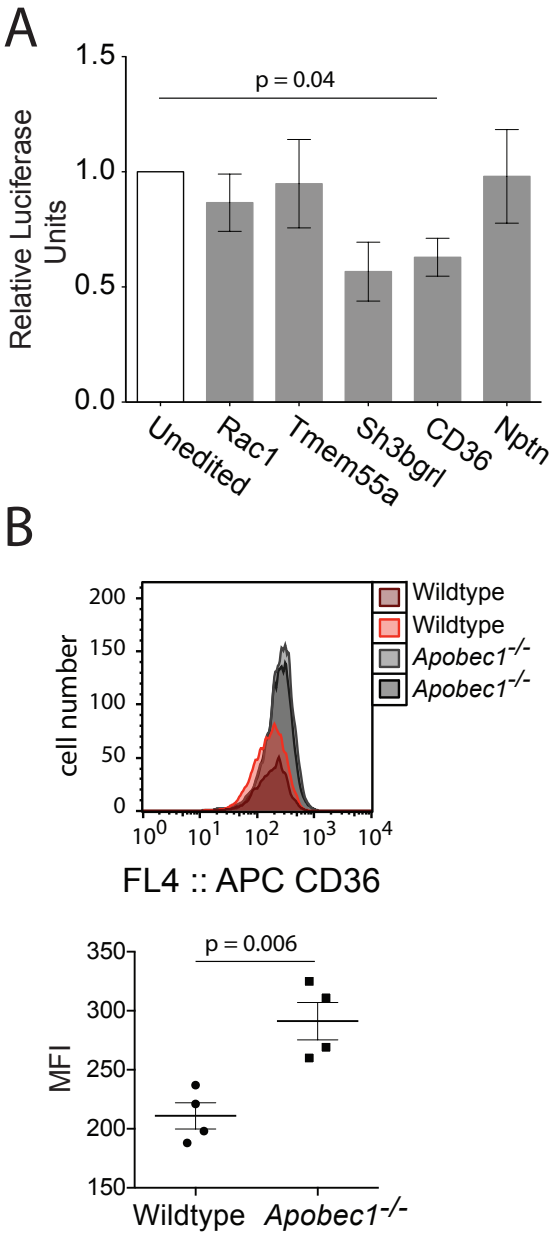

Figure 3.3.1 | Editing can modulate protein expression. (A) Luciferase experiments of single-edited constructs. (B) Expression of CD36 in the surface of BMDMs (FACS).

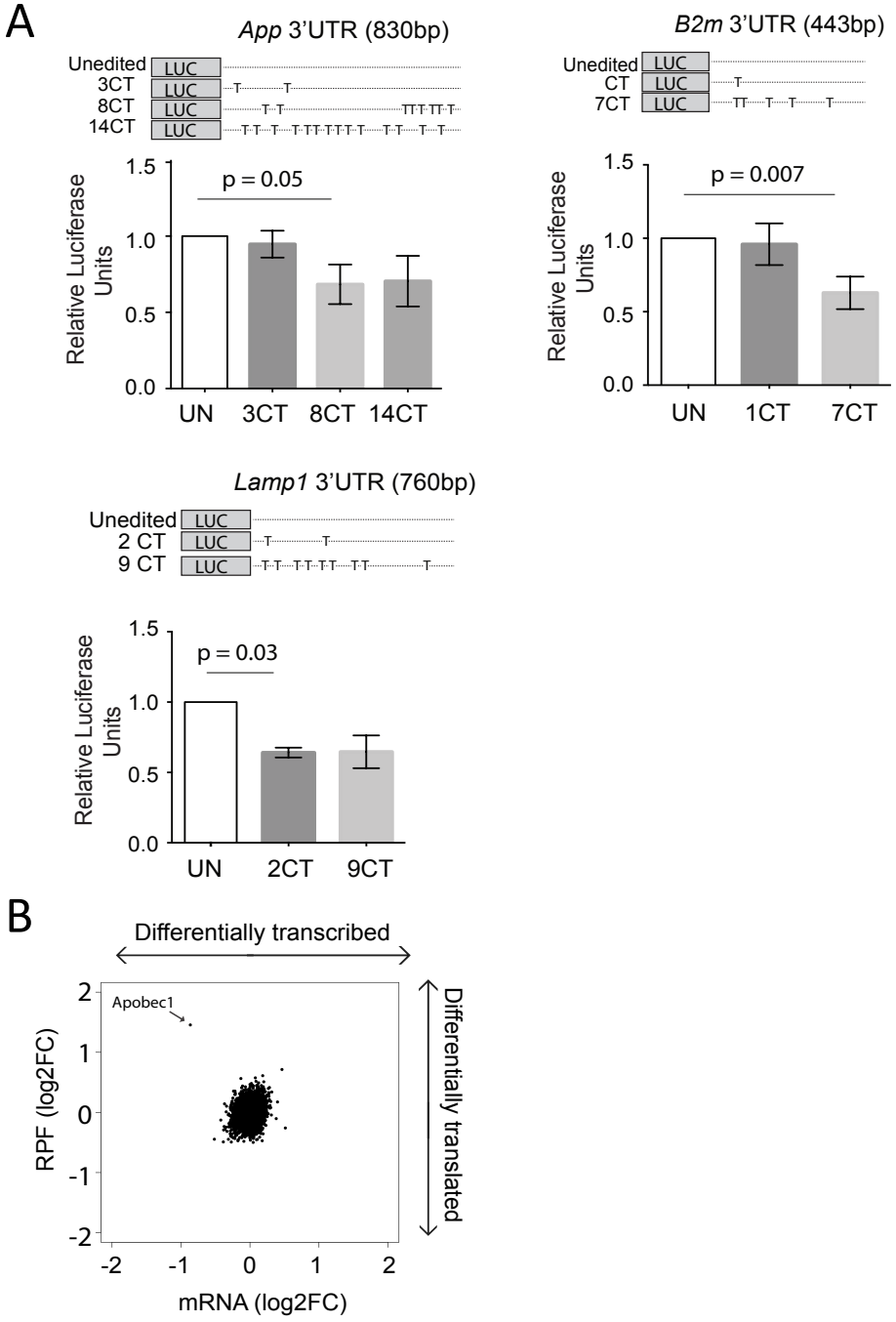

Figure 3.3.2 | Editing can modulate protein expression. (A) Luciferase experiments on different constructs containing hyperedited 3' UTRs. (B) Expression vs. ribosome binding plot of the ribo-seq results, using only the top 25% most highly expressed genes.

We then tested whether we could observe differences in the endogenous protein levels of some of the transcripts tested via luciferase assays. We compared the levels of LAMP1, B2M and CD36 in *wildtype* and *Apobec1*-/- BMDMs, via flow cytometry. We only observed differences in total and cell surface levels of CD36 (Figure 3.3.1.B), which is consistent with the luciferase experiments. This particular transcript has the special characteristic that it is both highly edited (80% frequency of editing) and robustly translated. We therefore speculate that in order to see differences in the expression of protein products of editing trascripts in the cell, a transcript would need to be both highly translated and highly edited, as is the case of CD36.

In order to test globally whether edited transcripts result in differences of protein production, we considered two methods, SILAC and Ribo-seq. SILAC is a method that uses mass spectrometry to determine the differences between two cell populations, one of which is labeled with non-radioactive isotopic labeling[86]. However, the sensitivity of detection is only very good for highly expressed proteins, and thus cannot access the entire proteome. Many of the APOBEC1 modified transcripts are not highly expressed, which is an obstacle for using SILAC.

Ribo-seq is a technique that surveys transcriptome-wide ribosome binding. It uses high throughput RNA sequencing after polyribosome pool down to determine the ratio between expressed total RNA and the RNA being translated. This measure has been shown to be a much better estimation of protein abundance than total RNA[87]. Riboseq has the advantge over SILAC of being able to detect small changes and access lowly expressed transcripts. Given that editing does not happen at an efficiency of 100%, we

expect that the differences we may find when comparing *wildtype* vs. *Apobec1-/-* ribosome binding are small.

We therefore performed Ribo-seq using *wildtype* and *Apobec1-/-* BMDMs. First, cell lysate was split in two samples. The first one was treated with nuclease to degrade RNA unprotected by ribosomes. Then the polyribosomes were selected through column separation and the protected RNA was extracted for ribosomal RNA depletion, reverse transcription and library preparation. The second sample was used for total RNA extraction, followed by ribosomal depletion and library preparation. Both samples were then sequenced using a lane in the Illumina nextseq and reads were aligned to the genome. Ribosomal quantification and gene expression were calculated using DEseq. Translational efficiency was calculated using the coefficient of RPF / mRNA[87], and this was used to asses differential loading between *wildtype* and *Apobec1-/-* samples. We found that the ribo-seq libraries had a lot of variability between samples and were very prone to batch effects, therefore we decided to analyze the top 25% most expressed genes. These included the majority of edited transcripts. Figure 3.3.2.B shows there are no differences in ribosomal binding between *wildtype* and *Apobec1-/-* except for Apobec1; and confirms there are no transcriptional differences between the two genotypes either.

The lack of translational differences observed in the Ribosome capture assay indicates that the effect of editing in protein expression is either very minimal, or that at the level of population, the changes in individual cells with higher degree of editing get lost in the mix. In the next chapter, the possiblility that cells are heterogeneous in terms of editing is addressed. On the other hand, an analysis of either protein abundance or

ribosome binding at the single cell level are experiments which require technology not available at the moment, so differences at this level will remain to be tested.

Chapter 4. APOBEC1 editing contributes to cell heterogeneity

APOBEC1 targets 68 different transcripts in resting BMDMs, mostly through single site editing. These transcripts are targeted with varying frequencies (from ~10%, our arbitrary lower threshold, to close to 100%). Given that APOBEC1 editing has a moderate effect on protein levels, we were curious as to whether the varying editing efficiencies of transcripts at the population level were the result of uniform editing in all cells, or of high levels of editing in a subset of cells (both possibilities would generate similar population-wide RNA-seq profiles; Figure 4.1.A). This has been a long-standing question in the editing field because most of known editing occurs at low frequency for both ADAR and APOBEC1 editing, rising questions about the biological significance of editing in aggregate. To explain this phenomenon, two hypotheses have been created. The first one, proposed by Gommans and Maas, states that such low frequency editing events may be a way of creating "noise" that perhaps may fulfill a biological function as an alternative mechanism to genomic-level mutations for probing potentially advantageous adaptations. Therefore, in their theory, editing at the population level is an accurate representation of editing in individual cells[88]. An alternative hypothesis, presented by Pullirsch and Jantsch, is that RNA editing can diversify cell populations by actually occurring at a very high frequency in specific subsets of cells[89]. To distinguish between them, we used single-cell RNA-Seq to profile editing in individual resting BMDMs and compared these to a bulk population sample.

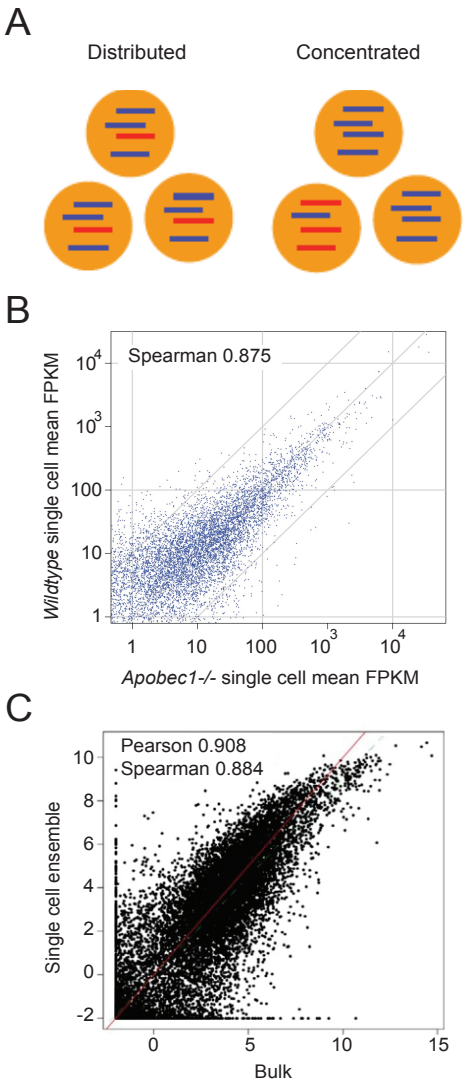

Figure 4.1 | Single cell sequencing. (A) Example plot showing high correlation between gene expressions of single *wildtype* vs. *Apobec1*-/- BMDM cells. (B) Comparison of the gene expression of the ensemble of all sequenced single cells vs. the bulk population.

4.1. Single cell sequencing

To determine whether there is heterogeneity in terms of editing within the macrophage population, we performed single-cell high throughput sequencing, using the C1 system from Fluidigm. *Wildtype* and *Apobec1*-/- BMDMs were sorted into C1 chips where cells were lysed and the RNA was reverse transcribed. Next, single cell libraries were made using the Illumina Nextera system and sequenced in a lane of an Illumina Hiseq.

Gene expression of single wildtype cells was compared to their *Apobec1*-/- counterparts. As observed when comparing bulk populations, there is also a high correlation between the two phenotypes (all possible combinations were analyzed; Figure 4.1.B shows an example of one such pair).

The comparison of the gene expression levels of bulk RNA-Seq (of a population of BMDMs) and the ensemble of all single cell replicates showed a tight correlation ($r>0.99$; Figure 4.1.C), indicating that the gene expression signature obtained from the single cells is a good representation of the bulk population. This also shows an apparent lack of heterogeneity at the transcript expression level in un-stimulated single cells. This has precedent in bone marrow-derived DCs, where resting cells start off as transcriptionally homogeneous, with variation at the level of expression introduced shortly after stimulation[90].

Furthermore, highly expressed transcripts, such as *B2m* showed little variability in expression levels in single cell libraries. However, that is not the case for most edited transcripts (Figure 4.2.1, left); the mid and lowest expressed transcripts showed the most variability between cells, which is consistent with previous reports[91].

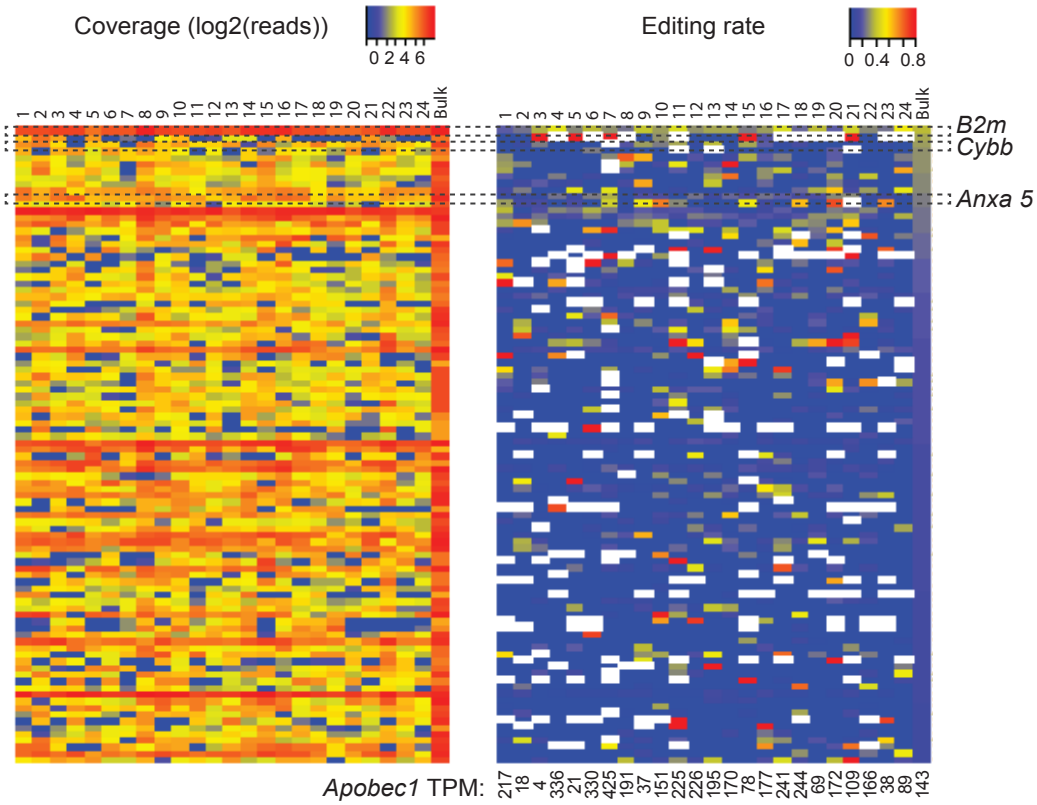

Figure 4.2.1 | Editing distribution in single BMDMs. Coverage (left) and editing rates (right) for high confidence C-to-U edited sites in the 24 single-cell and bulk macrophage RNA-seq data sets. Bottom: expression levels of APOBEC1 in transcript per million. Sites are sorted in order of descending bulk editing rate. Produced by Dewi Harjanto.

4.2. Analysis of the variability of editing rates per site across single cells

Single cell sequencing has been shown to have several characteristics. Due to the small amount of starting material and the efficiency of reverse transcription, it is not possible to access the entire transcriptome, which results in an underrepresentation of lowly expressed transcripts. Also, the reverse transcription is primed with polyA oligos, therefore the reads are enriched at the 3' end of transcripts. These characteristics are not problematic when these reads are used for gene expression analysis[92]. However, editing detection requires sites to be well covered and reads to be properly mapped. The low capture efficiency in single cell sequencing, estimated at around 20%[93], is the major obstacle for using editing detection algorithms, especially for lowly edited or expressed transcripts.

Due to these sampling issues, if a site is not edited, there could be two possibilities: that the site is genuinely not edited in the cell's transcriptome; or that the edited transcripts were present but were not captured during library preparation and sequencing. When examining editing at the high confidence sites identified from the bulk RNA-seq, across the single cells, we observed that, editing is present in at least one of the cells (402 of the 410 sites were covered in at least one cell and 22 of those sites was edited in at least one cell; Figure 4.2.1, left). However, editing levels did not necessarily correlate with APOBEC1 levels (Figure 4.2.1, right-bottom). Within these sites, we can also observe a wide range of editing rates. For example, the site in the *Cybb* transcript (indicated in Figure 4.2.1, right), which is covered uniformly in all cells, is edited in 61.5% of the reads in cell 6 while it isn't edited at all in 11 other cells (compared to just 23% in the bulk experiment).

Whereas it would seem from looking at Figure 4.2.1 that there is indeed heterogeneity in terms of editing among the single cells, the stochasticity inherent in the single cell RNA-seq method has to be accounted for. To ascertain variability within the APOBEC1-mediated editing rates from single cells, in comparison to the bulk editing rates recovered form populations of cells, a more sophisticated statistical method had to be developed. In order to do this, we collaborated with a group of statisticians at the University of Warwick* to model the level of variability of editing rates per site across single cells and in bulk samples. They created a hierarchical Bayesian method[80] to model the behavior of variable and homogeneous cells, and determined the profiles that each scenario would result in (Figure 4.2.2). Essentially, a population with high variability would have a high variability in the posterior distribution and a wide range of p values (Figure 4.2.2, top), whereas a homogeneous population would have a low posterior variance and narrow p values (Figure 4.2.2, bottom). When comparing these to the experimental data, they determined that there is a wide range of variability: they could find instances of transcripts edited with low, mid and high variance across cells (Figure 4.2.3. A, B and C, respectively).

4.3. Experimental validation of editing heterogeneity

To validate the predictions made by the model, we designed an RT-PCR amplification method to specifically look at different regions of transcripts that are editable. In order to carefully do molecular quantification of the edited transcripts, I designed gene specific RT-primers with a 6 nucleotide barcode, which are sufficient to

* Theodore Papamarkou, Chris J. Oates & Anastasia Papavasiliou; Department of Statistics, University of Warwick, UK.

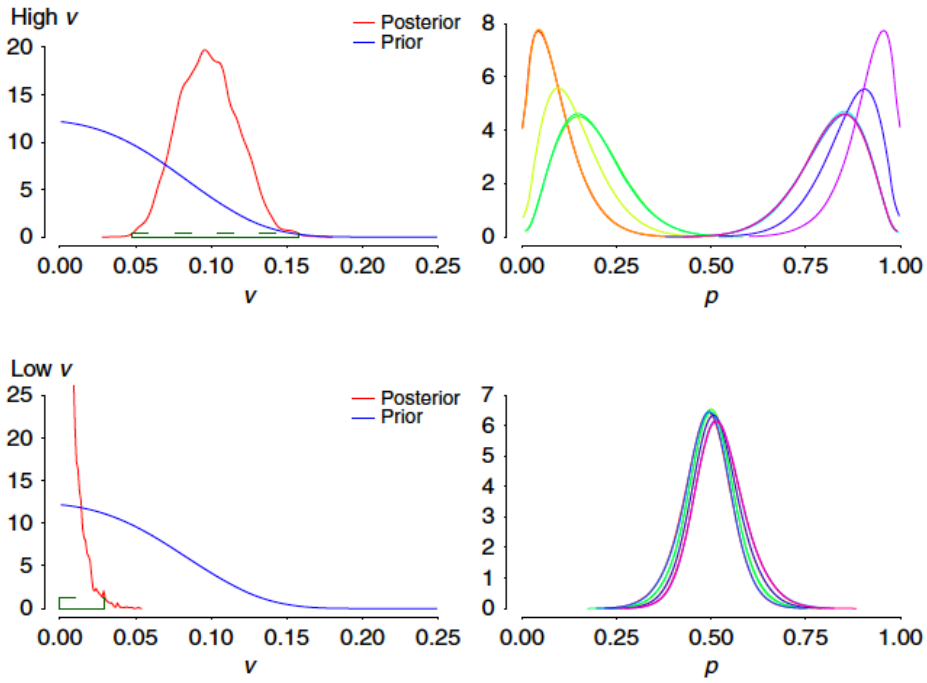

Figure 4.2.2 | Effect of changing the levels of variance on model. Histograms of variance of editing rates (left), editing rates (middle) for two artificial data sets, one with high editing rate variance (top); and the other with low editing rate variance (bottom). Produced by Dewi Harjanto.

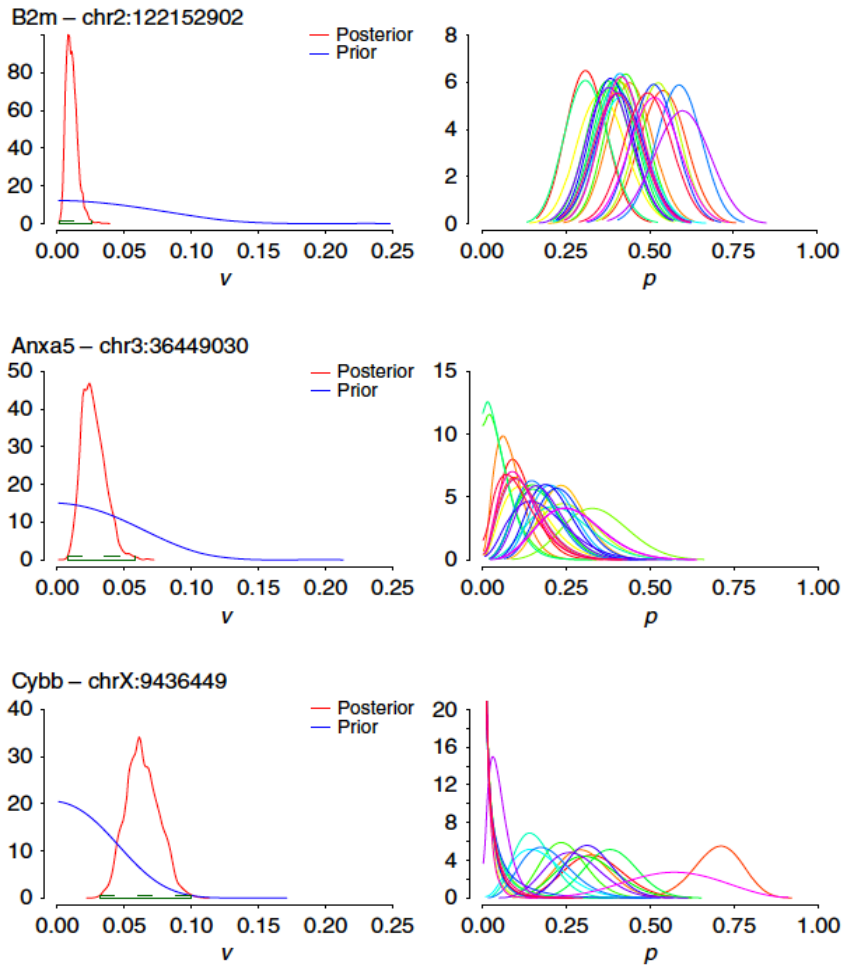

Figure 4.2.3 | Application of the Bayesian model to macrophages. Simulations ran on 24 single-cell data for *B2m*, *Anxa5* and *Cybb*. Left: histograms of variance of editing rates, across all 24 cells; right: histograms of editing rates, denoting the distributions of editing rates for each cell (each cell is labeled with a different color). Produced by Dewi Harjanto.

capture 4096 unique transcripts, a more than adequate amount to characterize the diversity within a single cell[93]. The fragments were then cloned into bacteria and sequenced using Sanger sequencing. Using the barcodes, I could discriminate fragments that are the result of PCR duplication and remove them from the analysis. This allowed us to examine the different sites with substantial depth, which is the main limitation of the high-throughput approach. Figure 4.3.1 shows the location of the edited sites in several single cells. The edited transcripts are labeled in red. The results from this experiment align well with the model predictions. In the case of *Cybb*, for which the posterior distribution in the model suggests high variability, we observed that there exist cells that have very high levels of editing (Figure 4.3.1.C, cells #1, 2 and 6) along with ones very low levels (Figure 4.3.1.C, cells #4, 5 and 7). Similarly, in the case of *B2m* and *Anxa5*, for which the model predicted lower variability, we can see that cells are roughly uniformly edited (Figure 4.3.1.A and B).

When we examined the whole of the amplified region (300nt fragments; Figure 4.3.2), we were able to identify editing that occurred at different sites from the bioinformatically determined ones. In the case of *Anxa5*, we observed that there is substantial editing in these alternative sites, which was not apparent from the bulk sequencing. This is perhaps because of the stringent filters in the bioinformatics algorithms used to identify editing. Transcripts where editing exists in an alternative site are also labeled in red on Figure 4.3.2.

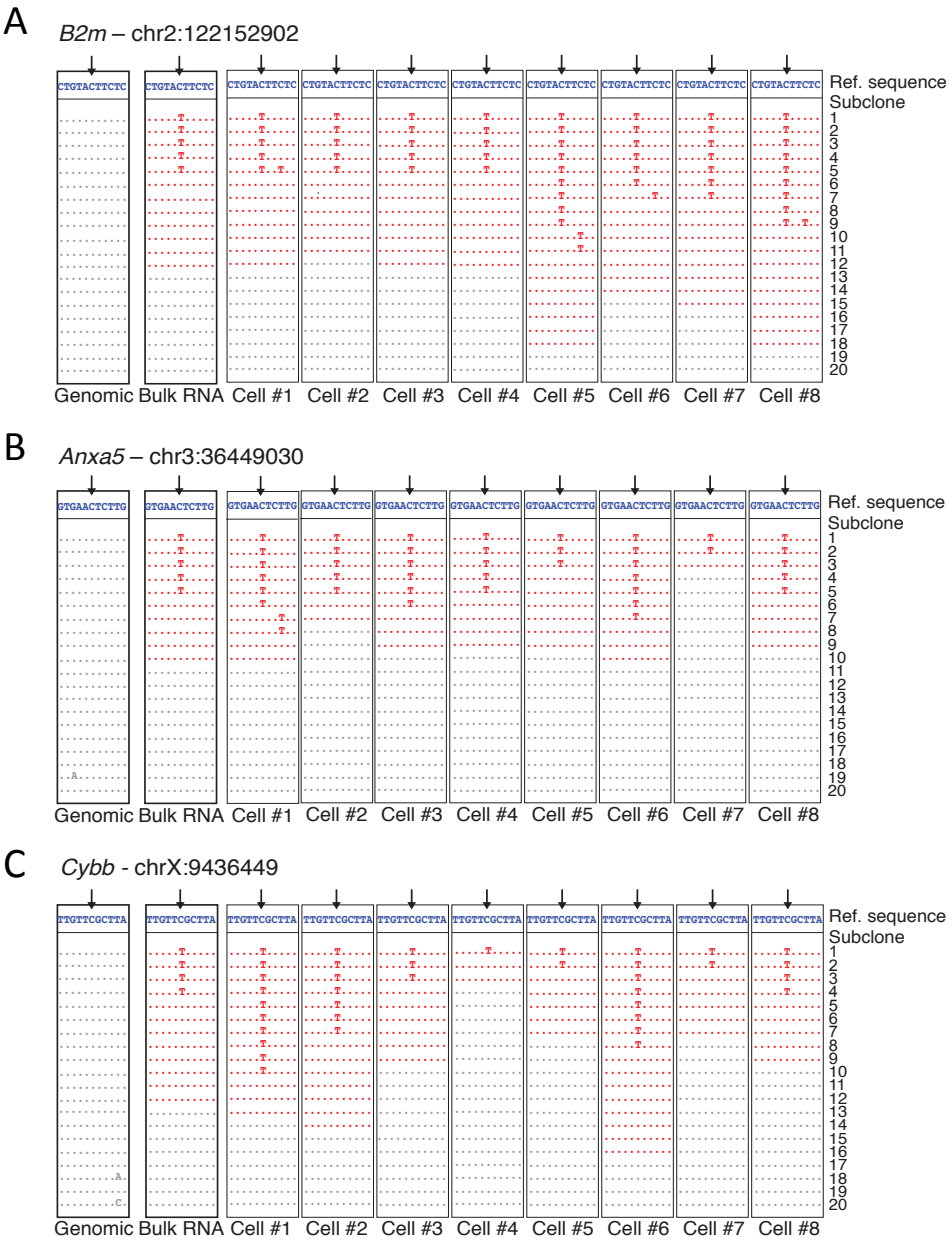

Figure 4.3.1 | Validation of model predictions using targeted amplification of editable sites from single cells. Sequence alignments from targeted RT–PCR amplification and Sanger sequencing of bacterial colonies for *B2m*, *Anxa5* and *Cybb* transcripts from gDNA and cDNA from a bulk sample and cDNA of single cells. Alignments are colour-coded to indicate whether the sequence aligned contained (red) or lacked (grey) editing in the length of the amplicon.

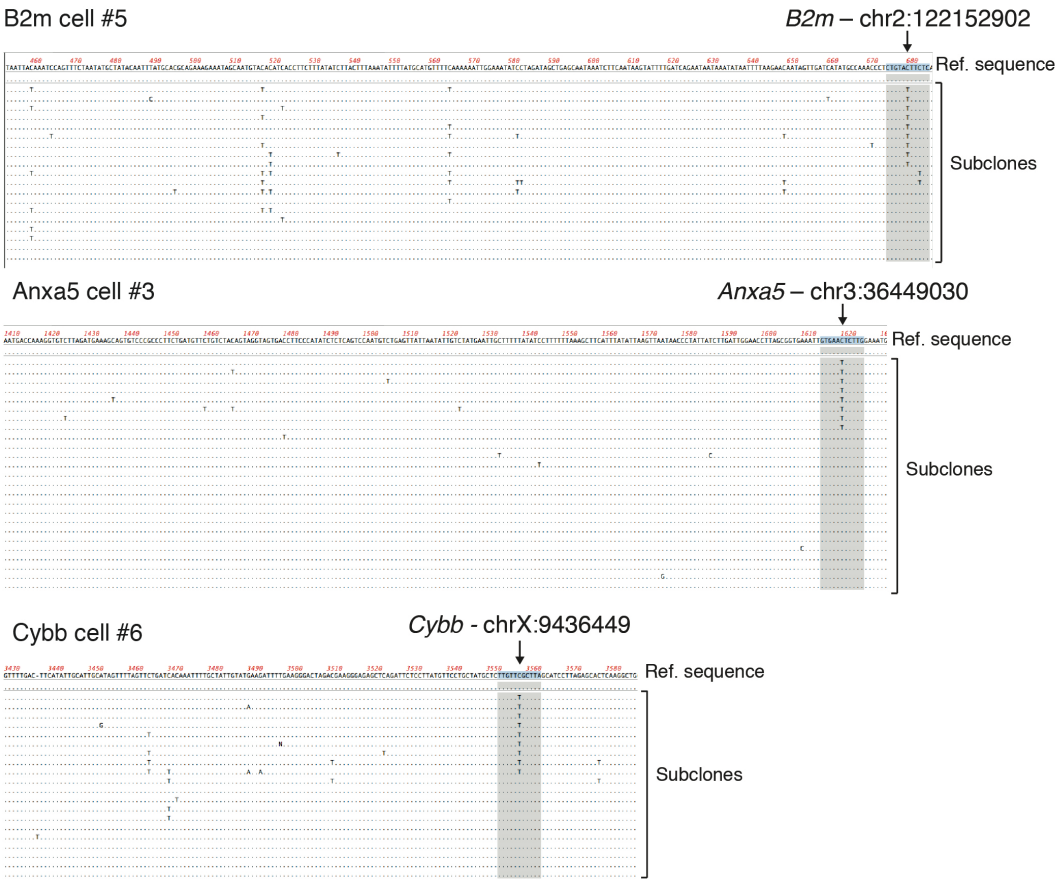

Figure 4.3.2 | Editing at sites alternative to the bioinformatically identified site. The alignment of a single cell is shown for B2m, Anxa5 and Cybb. The site shown in figure # is shown in grey. Alternative C to T changes are shown as T's within the alignments.

CHAPTER 5. FUNCTIONAL CONSEQUENCES OF APOBEC1 EDITING IN MONOCYTES

At a molecular level, editing is involved in the modulation of protein expression, either through the modification of miRNA binding or through alternative mechanisms. This is significant because even though editing occurs in low frequency at the population level, there is substantial variability between cells, raising the possibility that subsets of the population have specific characteristics. This variability between cells is reminiscent of the variability that has been described for monocyte populations and their plasticity, which allows them to respond to fluctuating environmental cues[94,95]. This is crucial to the maintanance of homeostasis as monocytes perform many crucial roles for the organism such as removal of debris and the production of tissue maintanance factors.

However, they also have the ability to react towards foreign antigens and help orchestrate an inflammatory response. Finally, in order to protect the body from exaggerated inflammatory responses, they are also invoved in the repression of the response and tissue repair. It has been long though that these responses are performed by subsets of macrophages that have one of two phenotypes, M1 (inflammatory or classically activated macrophages) or M2 (homeostatic or alternatively activated macrophages). However, it is becoming clear that there is more of a spectrum rather than a binary classification for monocytes and that these cells are very plastic, being ablet to change from one phenotype to another. We therefore evaluated whether the lack of editing had consequences in the activity of macrophages.

Previous investigations of the functional consequences of RNA editing have focused on instances where transcripts are edited with very high frequency and with clear biological implications (e.g. APOBEC1 editing of Apolipoprotein B[96,97]). However, the collective functional consequence of targeting a large number of transcripts has never been examined. In order to find relevant processes that could be modulated by editing, I looked for pathway enrichment of edited transcripts using the online database DAVID (Broad Institute). Because this database is a repository of all genes regardless of origin, it identifies non macophage-specific pathways, although many of the pathways identified are pertinent to immune cells. APOBEC1 edited transcripts in BMDMs encode proteins important for lysosome maturation (e.g. Lamp1, Lamp2, Atp6ap1) and phagocytosis, as well as proteins important for migration (e.g. Rac1, Kras, Pak2, Brb2; Table 5.1). Given that editing can affect protein abundance, we asked whether small alterations in the levels of such proteins could collectively alter cell physiology. Therefore, we assessed the relative performance of *wildtype* and *Apobec1-/-* BMDMs in both phagocytosis and migration.

5.1. Editing deficient BMDMs display altered migration and phagocytic capacities

One of the early consequences of chemokine signaling in BMDMs is directed cell migration. In order to test whether migration is altered in editing-deficient BMDMs, we designed a migration test where BMDMs are placed in the upper chamber of a two chamber culture dish. We determined from the RNA-seq data that the highest expressed chemokine receptor in BMDMs is CXCR4, whose preferential ligand is CXCL12 (also known as stromal-derived-factor-1). Expression of CXCR4 was confirmed via flow

Table 5.1 | Pathway enrichment analysis using the subset of edited transcripts. Sites found in BMDM were fed to the pathway identification website DAVID from the Broad Institute. Significant pathways related to immune cell activity are presented.

Term	PValue	Genes	Fold Enrichment	FDR
Lysosome	3.0E-04	CTSL, LAMP2, SLC17A5, AP1S2, IDS, GNPTAB, TPP1, ATP6AP1, ASAH1, CTSF	4.5	0.338
B cell receptor signaling pathway	6.0E-04	NRAS, KRAS, GRB2, RAC1, PPP3CB, PPP3R1, PPP3CA, NFATC3	5.4	0.681
VEGF signaling pathway	2.5E-03	NRAS, KRAS, RAC1, PPP3CB, PPP3R1, PPP3CA, NFATC3	5.0	2.849
T cell receptor signaling pathway	5.6E-03	NRAS, KRAS, PAK2, GRB2, PPP3CB, PPP3R1, PPP3CA, NFATC3	3.7	6.248
Natural killer cell mediated cytotoxicity	6.7E-03	NRAS, KRAS, GRB2, RAC1, PPP3CB, PPP3R1, PPP3CA, NFATC3	3.5	7.438
Sphingolipid metabolism	6.9E-03	SGPL1, SPTLC1, SGPP1, ASAH1, DEGS1	6.4	7.670
MAPK signaling pathway	2.2E-02	NRAS, KRAS, PAK2, GRB2, RAC1, TGFBR2, PPP3CB, PPP3R1, PPP3CA, PRKACB, GADD45A	2.2	22.920
Chemokine signaling pathway	4.8E-02	NRAS, KRAS, GRB2, GNG10, PREX1, RAC1, PRKACB, CCL5	2.4	43.365
N-Glycan biosynthesis	5.1E-02	STT3B, MAN2A1, MGAT2, FUT8	4.7	45.240
PPAR signaling pathway	5.6E-02	LPL, CD36, GYK, ACSL4, CPT1A	3.4	48.305
Apoptosis	7.4E-02	CASP6, PPP3CB, PPP3R1, PPP3CA, PRKACB	3.1	58.746
Cell cycle	8.5E-02	E2F4, HDAC2, RAD21, RBL2, MCM4, GADD45A	2.5	63.660

cytometry in both wildtype and Apobec1-/- BMDMs, which didn not display differences in cell surface expression of the receptor. Therefore we chose to teste migration towards this chemokine. Briefly, different concentrations of the chemokine CXCL12 were placed in the lower chambers of a two chamber culture dish and the BMDMs were placed in the top chambers. Media without serum was used as a control for basal migration and the amount of migrated cells was determined 4h later using a fluorescent dye that binds to DNA. The median fluorescence intensity is directly correlated to the amount of DNA in each well, which is a good reflection of the number of migrated cells. Migration is reported as the chemokine index, which is a relative measurement that is calculated normalizing to the basal migration. As shown in Figure 5.1.A, APOBEC1-deficient macrophages have a very reduced migration capability. We speculate that because there are no differences in CXCR4 receptor expression, the altered migration phenotype of *Apobec1*-/- BMDMs is the result of differences in downstream signalling molecules (e.g. Ras) or cytoskeleton remodelation (e.g. Rac), both of which are good starting points for future studies.

To test phagocytosis, we added pHrodo-labeled *S. aureus* particles at two different multiplicities of infection (MOI) to BMDMs cultures. These particles are colorless until they reach the acidic environment of the lysosome, where they fluoresce. Phagocytosis can thus be directly quantified by flow cytometry 20 min later. Surprisingly, *Apobec1*$^{-/-}$ BMDMs showed an increased amount of particle uptake, compared to *wildtype* counterparts (Figure 5.1.B), therefore, lack of editing leads to increased phagocytosis. It has been shown before that an over-abundance of LAMP1 and of other proteins involved in phagosome maturation leads to an increase in phagosome

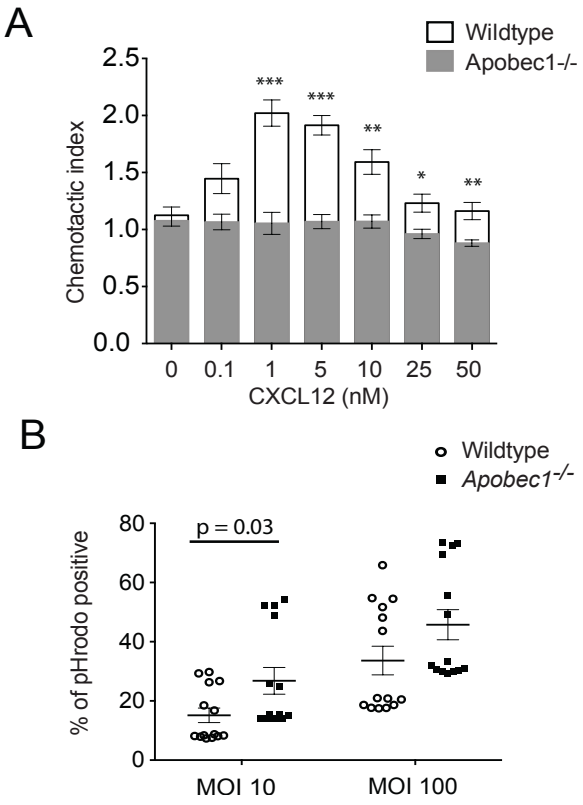

Figure 5.1 | APOBEC1 is required for the proper phagocytosis and migration *in vitro*. (A) Quantification of migration towards CXCL12 (also known as SDF-1). Error bars represent the standard error of the mean, statistical analysis was performed using the multiple measured one-way analysis of variance, followed by a t test with Bonferroni's correction; n=3; *p< 0.05, **p< 0.01, ***p< 0.0001). (B) Phagocytosis of *S. aureus* pHrodo particles (aggregate of n=5). Error bars represent the standard error of the mean; statistical significance was obtained using a t-test. MOI stands for multiplicity of infection. (A) was done in collaboration with Yamina Berchiche.

maturation[98]. We speculate that this may be one of the reasons why Apobec1-/- cells (or a subset of them) that potentially have increased levels of *Lamp1*, have an increased capacity to uptake the beads.

Interestingly, *Apobec1-/-* BMDMs are poor at an essential activity of macrophages but not at an equally fundamental one, specially because the two activities, even though independent from each other, are very related (e.g. cells need to migrate to sites of infection in order to ingest bacterial invaders). We can speculate that this could be the result of the variability of editing between the cells and how this affects specific pathways in different cell subsets. However, this remains to be tested.

5.2. Editing is necessary for proper monocyte differentiation

The differences in macrophage activity observed *in vitro* suggest that *Apobec1-/-* mice ought to have alterations in monocyte activity. In order to find clues for Apobec1-mediated regulation in the physiology if mice, I performed a phenotypic characterization analysis of *Apobec1-/-* mice, that included a complete blood count, blood chemistry and organ dissection and staining* (results not shown). I also characterized the different immune populations in bone marrow, blood, and spleen via flow cytometry (not shown). With all of this information, I concluded that, besides the already published increase in the low density lipoprotein fraction and a reduction in high density lipoprotein cholesterol resulting from impaired *ApoB* editing[60,99], there are no additional physiological differences in the *Apobec1-/-* mice when compared to their *wildtype* counterparts.

---

* This work was done by the Laboratory of Comparative Pathology at Memorial Sloan-Kettering Cancer Center

However, any differences in monocytes *in vivo* could be obscured as a result of compensation mechanisms. In order to study monocytes *in vivo* and have a direct comparison of *wildtype* and *Apobec1-/-* monocytes, which have been differentiated in the same milieu, we turned to a competitive reconstitution experiment. This experiment allows us to look at monocytes developing before compensations mechanisms can kick in and also has the advantage that reconstituting *wildtype* mice allows for proper APOBEC1-editing in intestine and liver. This allows us to isolate cell-autonomous, APOBEC1-dependent effects on the monocyte lineage in the absence of other known APOBEC1 related phenotypes (*e.g.* intestinal editing of *Apolipoprotein B*).

The competitive reconstitution experiment consists of placing *Apobec1-/-* monocytes along with their wildtype counterparts in a *wildtype* environment *in vivo,* thus having a direct comparison of their activity when challenged. One of the ways in which we can obtain both populations within *wildtype* mice is through the creation of chimeras. This can be achieved through bone marrow reconstitution of irradiated mice with bone marrow progenitor cells. After these chimeric mice establish the monocyte populations, we can challenge them and test macrophage differential activity.

Specifically, I injected equal numbers of bone marrow progenitors from *wildtype* (CD45.1) and *Apobec1-/-* (CD45.2) bone marrow into lethally irradiated syngeneic hosts and tested for reconstitution of monocytic populations in spleen via flow cytometry (Figure 5.2.1). The reconstitution of monocytes is one of the earliest events after bone marrow transplantation; we observed monocyte populations in spleen as early as 3 weeks (Figure 5.2.2.A). However, other cells, such as B and T cells have much longer reconstitution times.

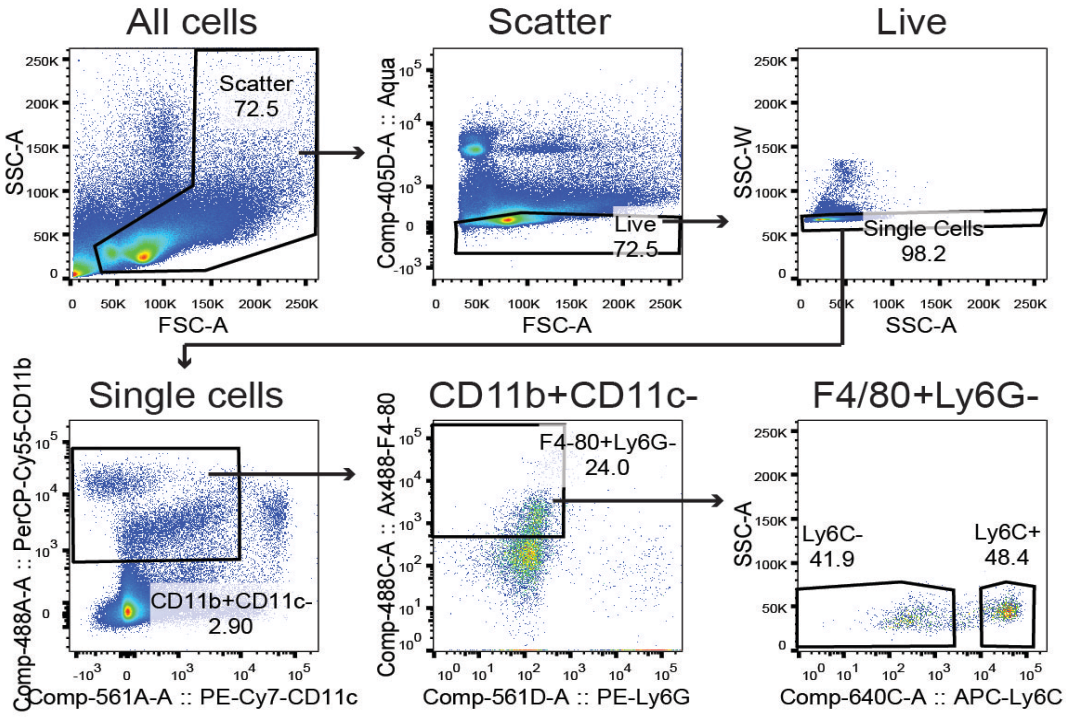

Figure 5.2.1 | Analysis of spleen chimeras. Splenocytes were stained with a cocktail of antibodies and analyzed via flow cytometry. The gating used to separate the populations are indicated.

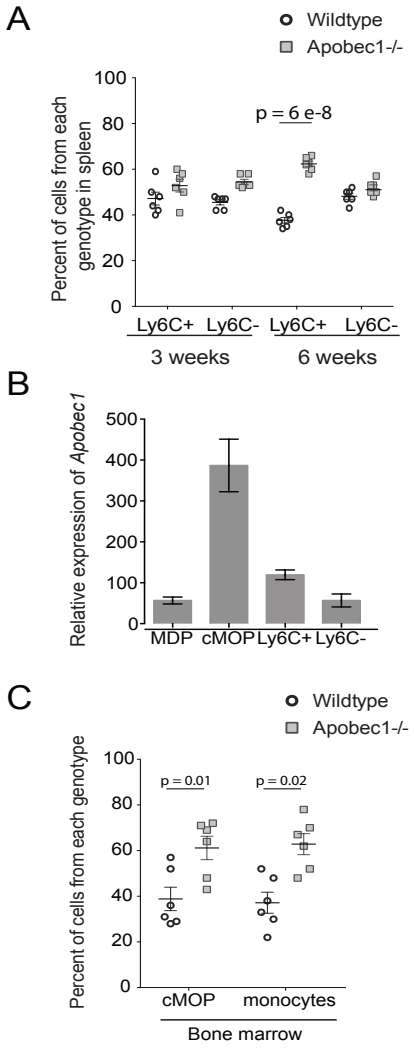

Figure 5.2.2 | APOBEC1 deficient mice have altered monocyte populations. (A) Analysis of spleen in radiation chimeras reconstituted with an equal number of *wildtype* and *Apobec1*$^{-/-}$ lineage-negative bone marrow progenitors (n=6). (B) APOBEC1 expression in sorted monocyte progenitors, relative to *Gapdh* expression. MDP: monocyte dendritic cell progenitor; cMOP: common monocyte progenitor. (E) Analysis of bone marrow monocyte progenitor populations in radiation chimeras reconstituted with an equal numbers of *wildtype* and *Apobec1*$^{-/-}$ progenitor cells, 6 weeks after transplant (n=6).

Surprisingly, when we compared the ability of each progenitor to reconstitute monocyte populations in spleen, we found that, even though the progenitors were mixed in a 1:1 ratio, the *Apobec1*-/- progenitors outcompeted their *wildtype* counterparts and gave rise to significantly increased numbers of M1-like (Ly6C+) monocytes in the spleen at 6 weeks, a trend that initiated at 3 weeks (Figure 4.2.2.A). Meanwhile, the frequency of the other major type of monocyte, M2-like (Ly6C-), was not altered (Figure 5.2.2.A). In a separate experiment, we also irradiated *Apobec1*-/- and *Apobec1*+/- mice and reconstituted them with a 1:1 mix of progenitors. Under these conditions, the Apobec1-/- progenitors also generated increased monocytic populations in spleen (not shown). This demonstrates there are no significant numbers of radiation-resistant cells at 3 weeks post-transplant, which could bias the results of the reconstitution. Also, this indicates that the effect is cell intrinsic.

To test whether the increase in Ly6C+ monocytes is due to an increase in their proliferative capacity, I tested for expression of Ki-67, a nuclear marker specifically found in proliferating cell[100], via flow cytometry. I did not observe Ki-67 expression in Ly6C+ monocytes (not shown), which suggests these cells are not proliferating. Alternatively, the increased number of *Apobec1*-/- Ly6c+ cells in spleen could be the result of increased output from the progenitor cells, increased cell division of mature monocytes or increased migration of monocytes into the spleen. We had previously observed that BMDMs have decreased migratory ability towards CCL12; however, this seems not to be the case *in vivo*. This could be the result of differences between monocytes and BMDMs, which are an artificially derived mature form of macrophages, as compared to monocytes, which are a more immature subset. However, it could also be

that the migration phenotype is restricted to signaling through CXCR4, which has been shown to mediate retention of monocytes in the bone marrow[101]. If this mechanism of retention is indeed impaired in *Apobec1*-/- monocytes, it is possible they would exhibit increased exit from the bone marrow.

Alternatively, APOBEC1 could be regulating the developmental progression of specific subsets of cells in the bone marrow. In order to test this I quantified the expression of APOBEC1 in monocyte subsets and bone marrow monocyte progenitors via qPCR, in sorted populations. APOBEC1 is expressed early in monocyte differentiation, and its transcript levels are highest in the common monocyte progenitor (cMOP) cells (Figure 5.2.2.B). This could mean that in vivo, APOBEC1's function is involved in monocyte differentiation. To test whether the increase of Ly6C+ monocytes is the result of increased progenitor cells, I quantified the number of these progenitors in bone marrow after reconstitution. *Apobec1*$^{-/-}$ cMOP cells outcompeted their *wildtype* counterparts at 6 weeks after transplant (Figure 5.2.2.C), confirming that APOBEC1 plays a role in properly generating monocyte populations at various points during differentiation. Further experiments to catalog edited targets in cMOP cells are required to identify the mechanism through which editing modifies these progenitors.

5.3. Editing is dispensable for initiating a strong inflammatory response

Given that we observed preferential phagocytosis *in vitro* and creation of inflammatory-like monocytes *in vivo*, we speculated APOBEC1 editing could be involved in the establishment of the inflammatory response in general. To test this, I assayed the ability of *wildtype* and *Apobec1*-/- macrophages to respond to

lipopolysaccharide (LPS) and *in vitro* and *in vivo*. LPS is an endotoxin consisting of a lipid and a polysaccharide with an O-antigen, found in the outer layer of Gram-negative bacteria. LPS binds to the CD14/TLR4/MD2 receptor complex in several immune cells including macrophages. This results in the production of pro-inflammatory cytokines, nitric oxide and eicosanoids.

To test the ability of *wildtype* and *Apobec1*-/- BMDMs to mount an inflammatory response, I primed cells with IFNγ for 2 hours, followed by addition of LPS, and quantified the expression of APOBEC1 and several pro-inflammatory cytokines at several time points via qPCR. I also tested for the production of nitrogen reactive species, via the quantification of nitric oxide in the media after IFNγ + LPS stimulation. I observed that the expression of *Apobec1* increased 2 fold after 18h of LPS stimulation and continued to increase over time (Figure 5.3.A). However, this was not accompanied by differential expression of *Nos2*, nor the pro-inflammatory cytokines *Ccl5*, *Il12b* or *Il1a* (not shown). The expression of the pro-inflammatory cytokine *Tnfa* seemed to be transiently induced at an early time point in Apobec1-/- BMDMs (Figure 5.3.B).

To test whether there is differential production of cytokines, I performed a quantification assay using the Luminex technology. This consists of capturing free cytokines in the supernatant of the cell culture via beads coupled with specific antibodies. Next, these beads are incubated with fluorophores that aid in the determination of cytokine concentration. Luminex assays can be easily multiplexed, giving the ability to test for a large panel of cytokines at a time. I used the Cytokine 20-Plex Mouse Panel (Thermo Fisher Scientific), which includes a wide variety of cytokines as well as chemokines and growth factors (GM-CSF, IFN-γ, Il-1α, IL-1β, IL-2, IL-4, IL-5, IL-6,

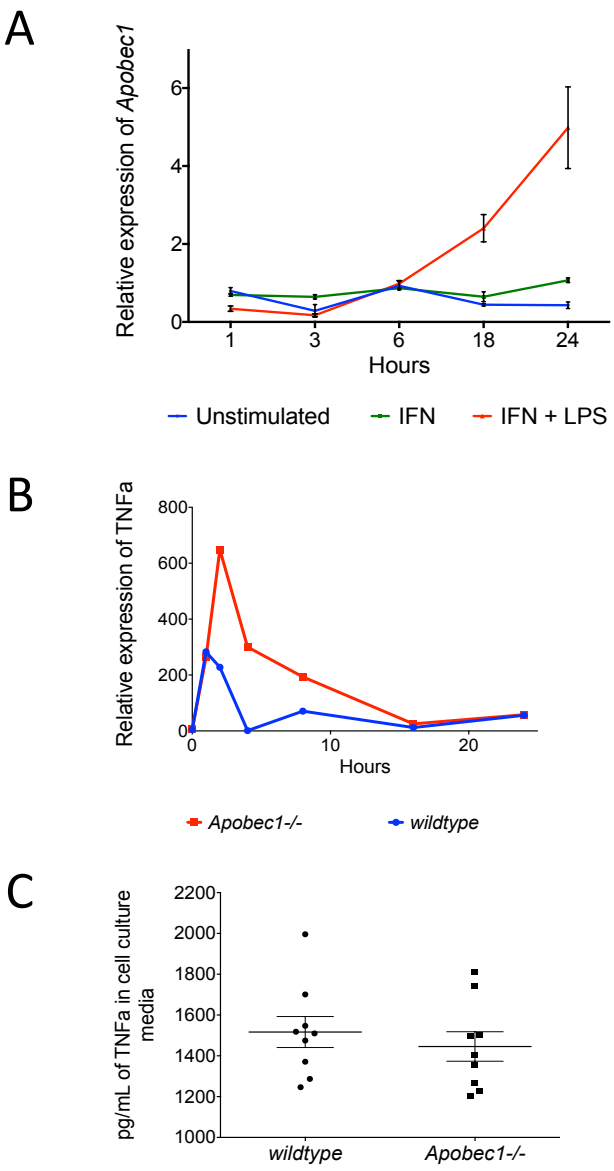

Figure 5.3 | APOBEC1 deficient BMDMs do not have enhanced ability to establish an inflammatory response. (A) Analysis of *Apobec1* expression in *wildtype* BMDMs after IFN + LPS stimulation, via qPCR (n=3). (B) Relative expression of *Tnfa* at different timepoints after stimulation with IFN + LPS. Both *Apobec1* and *Tnfa* were normalized using *Rpl32*. (C) Quantification of TNFa protein in the supernatatnt of cells after 24 h of stimulation with IFNg + LPS (via Luminex; n=9).

IL-10, IL-12 (p40/p70), IL-13, IL-17, TNFα, IP-10, KC, MCP-1, MIG, MIP-1α, FGF-basic and VEGF). Using this technology, I observed that there were no differences in the amount of cytokines secreted by *Apobec1*-/- BMDMs in comparison to their *wildtype* counterparts (e.g. TNFα is presented in Figure 5.3.C). Furthermore, I found that resting *Apobec1*-/- BMDMs did not express pro-inflammatory cytokines.

Additionally, I tested the fitness of monocytes to incite an inflammatory response *in vivo* using two models of acute inflammation: intraperitoneal LPS and zymosan intraperitoneal injection. Both of these are particles found on bacterial outer membranes and yeast, respectively, and are identified by receptors in macrophages. Upon recognition of these particles, macrophages secrete inflammatory cytokines that are able to recruit more cells into the peritoneum. Depending on the dose utilized, the response can be cleared in 48 to 72 hours. I observed that *wildtype* and *Apobec1*-/- mice have a similar ability to induce an inflammatory response in the peritoneum: both mice have equal ability to recruit cells into the peritoneum, produce sufficient levels of inflammatory cytokines, and resolve the inflammation to allow for survival (data not shown). This is consistent with the *in vitro* results presented above. However, they seem inconsistent with the results that show a higher establishment of Ly6C+ pro-inflammatory monocytes in *Apobec1*-/-.

This indicates that either there is no difference between the two genotypes, or that the time scale that I looked at is off. Another alternative is that the overwhelming amount of inflammation produced by the LPS and zymosan very highly activate the system and mask subtle differences.

5.4. Editing in microglia

In an effort to continue with the characterization of inflammation in other organs, I collaborated with Dr. Karen Bulloch's group at the Rockefeller University to study brain alterations*. Dr. Bulloch's group is interested in studying brain immunology and brain pathologies in aging. They examined the brains of 3 month and 1-year-old wildtype and Apobec1-/- mice. They only identified alterations in the brains of aged *Apobec1*-/-. These alterations consisted of clustering of cells that stained positive for Iba-1 (a marker that in the brain is specific of microglia), indicative of activated microglia (not shown). This was accompanied by an increase in pro-inflammatory cytokine expression (from whole cortex and hippocampus extracts; not shown).

In order to relate these findings to APOBEC1 editing, we first looked in the literature for gene expression datasets in brain cell subsets to identify cells expressing APOBEC1 within the brain. We examined the RNA-seq dataset from Zhang and collaborators[102], who purified neurons, astrocytes, microglia, endothelial cells, pericytes, and various maturation states of oligodendrocytes from mouse cortex. We found the highest expression of APOBEC1 is within microglia and astrocytes. However, when we searched for the known APOBEC1 cofactors ACF and RBM47, we found that they were exclusively expressed in microglia, making this cell type the most likely candidate where editing might be found. These microglial cells were isolated from mouse brains using a two step approach, first brains were perfused with PBS in order to eliminate contaminating monocytes from the periphery, then microglia cells were isolated using

* Dan Cole, Youngcheul Chung, Khatuba Gagnidze and Karen Bullock

CD45 antibodies. However, CD45 is a cell marker present in many subtypes of immune cells. Previous work from Dr. Bulloch's lab, using a *CD11c-egfp* transgenic mouse, has demonstrated that mouse olfactory bulb CD45+ subset also includes brain dendritic cells, which can be characterized their expression of CD45 and CD11b, namely P1 ($CD45^{int}$ $CD11b^{hi}$), P2 ($CD45^{hi}$ $CD11b^{hi}$) and P3 ($CD45^{hi}$ $CD11b^{int}$)[103]. Therefore, we sought to characterize the expression of APOBEC1 in two main subsets of CD45 expressing cells ($CD45^{+}CD11b^{+}CD11c^{+}$ and $CD45^{+}CD11b^{+}CD11c^{-}$) within cortex and hippocampus, where the brain alterations were originally described.

In order to identify editing in these subsets, we first sorted $CD45^{+}CD11b^{+}CD11c^{+}$ (CD11c+ MG) and $CD45^{+}CD11b^{+}CD11c^{-}$ (CD11c- MG) cells from wildtype and Apobec1-/- mice, and performed RNA extraction and library preparation to perform high throughput RNA-sequencing. Interestingly, using gene expression analysis, we did not find differential gene expression between *wildtype* and *Apobec1-/-* in either the CD11c+ MG or CD11c- MG subsets (not shown). This is very similar to my observations in BMDMs derived from the same genotypes (as mentioned in Chapter 2).

In order to detect editing in the CD11c+ MG or CD11c- MG datasets, we used our editing detection algorithm. Surprisingly, the algorithms were only able to identify a few sites that were modestly edited. In order to validate these sites, I tested them through PCR amplification and Sanger sequencing, in a similar way as shown in figure 2.3; however, I could not validate any of the sites.

I then decided to query the sites that were identified in BMDMs, to see if they could also be edited in the CD11c+ MG and CD11c- MG populations. To accomplish this, I looked at specific edited sites in several transcripts via PCR amplification and high

throughput sequencing (high throughput amplicon sequencing), which allowed me to access each site with substantially more depth than regular high throughput RNA-sequencing. I chose several sites that are edited in BMDMs and expressed in the CD11c+ and CD11c- MGs (Table 5.2). Indeed I observed that the editing rate at these sites is very small in both populations of microglia, which is consistent with the failure of the algorithm to identify them.

As I have posed before (in chapter 4), low levels of editing are not necessarily the representation of editing in single cells within a population. Given that editing is variable within BMDMs it is possible that it would also be within individual microglial cells. This would suggest that editing is not present in all microglia and that instead, subsets of the microglial population could exist which harbor high levels of editing. In order to test for this possibility, I examined a known subpopulation of CD11c+ cells that are characterized by the expression of the CD103 marker. These cells are highly enriched in olfactory bulb after intranasal infection with VSV[103], therefore the cells were collected after VSV infection and RNA was extracted (this procedure was done previously for a project unrelated to this work). I used the RNA to do a spot check of editing in this population, amplifying two of the most highly edited targets in BMDMs, *B2m* and *Lamp2,* in a similar way as described for figure 2.3. I observed that in contrast to CD11c+ MG, which displayed a very small amount of editing, the subpopulation of CD11c+CD103+ cells had a high amount of editing (Figure 5.4). This shows that in fact different subpopulations of microglia demonstrate different levels of editing, therefore there is a possibility that the lack of APOBEC1 editing of a subset of cells may be involved in the clustering of microglia and inflammation phenotype observed in

Apobec1-/- aged brains. Further work is needed to identify the specific population(s) of microglia, which harbors APOBEC1 editing in aged brains, and to link those specific populations to the phenotype observed.

Table 5.2 | Amplicon sequencing-based validation of putative edited genes. Sites found in BMDM were amplified in both CD11c+ and CD11c- MG and sequenced and the frequency of C to T changes was quantified.

				Amplicon sequencing on cDNA			RNA -sequencing		Amplicon sequencing on DNA		
Region	Gene	Chromosome	Coordinate	Reads Coverage	Frequency of C to T	Frequency of Non-C to T	Reads Coverage	Frequency of C to T	Reads Coverage	Frequency of C to T	Frequency of Non-C to T
UTR3	*Anxa5*	chr3	36449030	354443	**1.40%**	0.00%	72	2.80%	14530	0.1%	0.0%
UTR3	*B2m*	chr2	122152613	91	**1.10%**	0.00%	31	0.00%	13468	1.6%	0.0%
UTR3	*B2m*	chr2	122152682	452558	**1.20%**	0.00%	17854	0.60%	13680	0.0%	0.0%
UTR3	*B2m*	chr2	122152740	444796	**7.60%**	0.00%	17223	5.70%	13627	0.2%	0.0%
UTR3	*B2m*	chr2	122152742	446438	**1.50%**	0.10%	16235	1.00%	13628	0.1%	0.0%
UTR3	*B2m*	chr2	122152787	669269	**2.90%**	0.20%	8190	1.40%	20889	0.3%	0.0%
UTR3	*B2m*	chr2	122152804	675429	**3.20%**	0.00%	9610	1.90%	20914	0.3%	0.1%
UTR3	*B2m*	chr2	122152902	416611	**6.40%**	0.00%	10425	3.90%	10358	1.0%	0.0%
UTR3	*Lamp2*	chrX	38421317	125796	**1.00%**	0.00%	666	0.50%	30059	0.0%	0.0%
UTR3	*Lamp2*	chrX	38421565	129390	**1.30%**	0.00%	893	2.00%	31356	0.0%	0.0%
UTR3	*Sh3bgrl*	chrX	109160420	22804	**4.10%**	0.00%	16	12.50%	110353	0.0%	0.0%
UTR3	*Sh3bgrl*	chrX	109160354	25647	**3.90%**	0.00%	23	0.00%	112949	0.0%	0.0%
UTR3	*Sh3bgrl*	chrX	109160178	13610	**2.10%**	0.00%	18	0.00%	59802	2.0%	0.0%
UTR3	*Dennd1b*	chr1	139175650	132688	**1.60%**	0.00%	5	0.00%	52069	0.0%	0.0%
exonic	*Med13l*	chr5	118754418	163564	**1.10%**	0.00%	39	0.00%	70500	0.0%	0.0%
exonic	*Pki3r4*	chr9	105650713	2585	**1.30%**	0.00%	12	0.00%	1	0.0%	0.0%

Amplicon sequencing was also performed in 9 other genes with no editing detected (Sec24a, Rapgef5, Tfrc, Semad6, Psmc3ip, Zfp318, Dnajc1, Dcun1d4 and Gab2).

				Amplicon sequencing on cDNA			RNA -sequencing		Amplicon sequencing on DNA		
Region	Gene	Chromosome	Coordinate	Reads Coverage	Frequency of C to T	Frequency of Non-C to T	Reads Coverage	Frequency of C to T	Reads Coverage	Frequency of C to T	Frequency of Non-C to T
UTR3	*Anxa5*	chr3	36449225	84085	**1.80%**	0.00%	256	0.00%	15824	0.0%	0.0%
UTR3	*Anxa5*	chr3	36449030	78293	**1.50%**	0.00%	120	0.00%	14530	0.1%	0.0%
UTR3	*B2m*	chr2	122152902	180151	**7.30%**	0.00%	9715	4.80%	10358	1.0%	0.0%
UTR3	*B2m*	chr2	122152740	192034	**5.90%**	0.00%	15585	6.80%	13627	0.2%	0.0%
UTR3	*B2m*	chr2	122152804	282971	**3.40%**	0.20%	7405	2.70%	20914	0.3%	0.1%
UTR3	*B2m*	chr2	122152787	283924	**2.30%**	0.10%	6187	2.10%	20889	0.3%	0.0%
UTR3	*B2m*	chr2	122152742	193011	**1.40%**	0.00%	14523	1.50%	13628	0.1%	0.0%
UTR3	*B2m*	chr2	122152682	195676	**1.30%**	0.00%	16883	0.90%	13680	0.0%	0.0%
UTR3	*B2m*	chr2	122152871	177490	**1.10%**	0.00%	5569	0.90%	19000	0.2%	0.2%
UTR3	*Lamp2*	chrX	38421565	23541	**1.70%**	0.00%	1041	0.90%	31356	0.0%	0.0%
UTR3	*Lamp2*	chrX	38421618	23633	**1.10%**	0.00%	1649	0.10%	31379	0.0%	0.0%
UTR3	*Lamp2*	chrX	38421317	20690	**1.10%**	0.00%	751	0.80%	30059	0.0%	0.0%
UTR3	*Rac1*	chr5	143505935	97069	**1.60%**	0.00%	365	0.80%	1208	0.6%	0.0%
UTR3	*Jag1*	chr2	137081857	122649	**2.00%**	0.00%	*0*	*nan*	*24*	*4.2%*	*0.0%*

Amplicon sequencing was also performed in 14 other genes with no editing detected (Cluh, Polrmt, Fut8, Alg12, Enpp1, Lpp, Sepn1, Vamp4, Bms1, D7ertd443e, Usp48, Socs4, Ncapd3 and Mast2).

Figure 5.4 | APOBEC1 editing is present in a subset of microglial cells characterized by CD11c+ CD103+ expression. C to T RNA editing events can be seen in the B2M (A) and Lamp2 (B) transcripts of subclones obtained from CD45^{int}Cd11b^{+}EYFP^{+} cells from *Cd11c-EYFP* mice.

CHAPTER 6. DISCUSSION

At the beginning of my work, the role of Apobec1, except for a couple of exceptions, seemed to be constrained to the intestine and exquisitely confined to lipid transport and metabolism through editing of the ApoB transcript. With the advent of high throughput sequencing technologies, our group was able to identify additional targets of Apobec1 within the intestine and in several cells of the immune system, especially in macrophages. However, the significance of these findings was absent, as were the functional consequences of these modifications. This is a common theme for ADAR targets as well, since for most such transcripts, especially the ones that do not result in changes in coding regions, the functional consequences remain murky.

Upon close examination of APOBEC1 targeted transcripts I was able to demonstrate that Apobec1 editing has a significant role outside of the intestine as it confers macrophages the ability to maintain proper homeostasis. These results were presented throughout my thesis and will be discussed in depth here.

6.1. Comparison of *wildtype* vs. *Apobec1-/-* macrophages

My experiments demonstrated that although there are no differences in transcript processing (transport, stability and differential expression) between the *wildtype* and *Apobec1-/-* genotypes, APOBEC1 regulates protein expression in miRNA dependent and independent ways, which are the direct consequence of editing. As noted in figures 3.1.1 and 3.1.2, Apobec1 edited transcripts did not display selective nuclear retention, in comparison to their unedited counterparts. This is a process that for ADAR, has been

proposed to be a molecular consequence of editing, which is mediated through binding of an RNA binding protein (p54), which presumably recognizes the tertiary structure of he dsRNA. In the case of Apobec1 editing, we do not expect edited regions to have specific tertiary structures due to their AU rich nature. However, differential RNA protein binding between edited and unedited transcripts is still to be investigated.

Stabilization of transcripts is the second process that has been proposed for ADAR editing, therefore I decided to test whether this is a general phenomenon that results from editing. Also, Apobec1, in its role as an RNA binding protein, has been shown to increase the stability of several transcripts with AU rich 3'UTRs, such as COX2 and TNF-α. Figure 3.1.3 shows that Apobec1 edited mRNAs have a high stability, even when Apobec1 is absent from the milieu (in the Apobec1-/- sample). This indicates that perhaps Apobec1 transcript stabilization is separate form its role in editing.

I also showed a significant overlap between Argonaute occupancy and APOBEC1 target regions (not shown), and that in principle, the change in a single nucleotide within the miRNA-binding site has the ability to modify miRNA binding (Figure 3.2.2). This agrees with what has been found for ADAR edited miRNAs[49,104]. This is significant because editing of miRNAs results in stringent and tissue-specific gene regulation, which is crucial for development. By editing miRNA target sites, a shift or redirection of target genes could be created, given that the pool of miRNA increases. This is yet to be investigated for both Apobec1 and ADAR editing. Further discussion on miRNA detection and regulation is presented on chapter 7.3.

During the course of my work, Davidson's group at the University of Washington also examined the transcripts of liver and intestine of *wildtype* and *Apobec1-/-* mice, and

observed that in those systems, editing also results in altered protein expression[105]. His group conducted differential gene expression analysis, and in contrast to my data, he found that a few hundred genes were differentially expressed between wildtype and Apobec1-/- enterocytes. Among those, 17 edited genes were differentially expressed. Furthermore, using SILAC (method discussed in section 3.3) they were able to find 26 edited genes differentially translated with a magnitude of two fold changes. This agrees with what I have reported in figures 3.3.1 and 3.3.2A, and points to a conserved function of Apobec1. They also showed differences in expression of CD36, which agrees with my findings, presented on figure 3.3.1B. However, when I analyzed translation globally via ribo-seq (method discussed in section 3.3) I didn't observe significant differences between the *wildtype* and *Apobec1-/-* genotypes. This could be an inherent difference between the two systems: enterocytes vs. macrophages. Alternatively, it could be that the changes in protein expression I observe in Cd36, for example, are result of an alternative mechanism, such as ribosome stalling. Caveats on ribo-seq method are further discussed on section 6.3, below. Blanc and collaborators, however, did not explore further the implication of the observed protein changes in enterocyte biology.

Altogether, changes in protein expression have implications for the fine-tuning of macrophage responses to changing environments. Previous investigations of the functional consequences of RNA editing have focused on instances where transcripts are edited with very high frequency and/or alter decoding by targeting genic sequences, which makes them have clear biological implications (*e.g.* APOBEC1 editing of Apolipoprotein B[58] and ADAR editing of GluR[96,97]). However, the collective functional consequence of targeting a large number of transcripts has never been examined, and is a

long standing question in the field. One of the challenges of such an analysis is the low frequency of editing in the vast majority of sites. Altogether the frequency of APOBEC1 editing and the resulting changes in protein expression we have found in BMDMs are modest. We hypothesized that these small changes would not be sufficient to have biological consequences by themselves. However, it is possible that the sum of small changes in a certain pathway or function would make an important contribution to the resulting activity.

Indeed, when I looked closely at two of the most important activities of macrophages, migration and phagocytosis, I discovered that cells from the two phenotypes are indeed not equivalent (figure 5.1). In the case of phagocytosis, I found that editing targets several structural proteins that are essential for the formation of the mature lysosome, as well as proteins involved in the formation of proton pumps, also necessary for the acidification of the phagosome. Furthermore, some other proteins, which were not specifically identified to be in the same pathway through literature mining, but are nevertheless necessary for uptake of particles, such as cytoskeleton remodeling proteins and a scavenger receptor, are also edited.

Neither of the above mentioned proteins would represent a good candidate for further studies on their own due to the low fraction of edited transcripts at the population level. Therefore, new strategies would need to be devised in order to dissect the exact contribution of each one of them to the overall phenotype. This may involve examination at the single cell level, where I demonstrate variability in terms of editing exists among cells of the same genotype (in figures 4.2.3 and 4.3.1). In such a scenario, a target that is edited at a low frequency in the population could be highly edited in just a subset of cells,

in which the consequences of editing would be magnified. For example, let's take the case of migration, where *Apobec1*-/- macrophages seem to be unfit to migrate whereas their wildtype counterparts do so quite efficiently (figure 5.1): if we were to look at a subset of macrophages that migrate more efficiently and therefore cross the membrane first, we might find that they posses a particular editing signature that is different from the macrophages that migrate at a slower pace. This kind of study would be relevant because low frequency editing is actually quite pervasive, especially in the case of ADAR editing, and insights from this work may be applicable to other scenarios, such as the brain. Additionally, given that other epitranscriptomic marks have such a dynamic nature, I expect that the study of the sum of editing events, C to U as well as A to I, will uncover a wonderfully complex regulatory network.

Furthermore, macrophages have the characteristic that they are very plastic, that is, they are able to change their phenotype in response to varying environmental cues. Here, I provide two key findings that have implications in macrophage plasticity and may contribute to heterogeneity in cells that appear identical (as defined by gene expression measurements): (1) Editing is variable between cells from a seemingly homogeneous population (as shown in figures 4.2.3 and 4.3.1), and (2) *Apobec1*-/- monocyte differentiation is altered in the absence of editing (as shown in figure 5.2.2). These results indicate that perhaps editing is involved in the formation of subsets of cells with different capabilities, some of which are perhaps better equipped for specific environments and may act as "first responders" that eventually signal to the rest of the population in order to create a coordinated response. This possibility needs further investigation but it is a very enticing hypothesis.

Altogether these results predict alterations in the monocyte population in Apobec1-deficient mice that might result in physiological alterations in the mice. Indeed this is the case in brains of aged mice, where our collaborators found alterations in the distribution of microglia, accompanied with inflammation (which under normal conditions is absent in the brain). Even though I did not provide direct evidence that editing is involved in creating this phenotype, I do believe that there is a connection. If editing is indeed creating subpopulations of cells that have "special" characteristics, such a subset in the context of the brain may be playing a modulating role, in charge to maintain homeostasis, preventing unnecessary inflammation. A finer characterization of the heterogeneity of microglia within the cell would be necessary to test for this hypothesis and conditional deletion of APOBEC1 within this subpopulation would be necessary in order to directly link editing to the phenotypes observed.

However, the consequences of APOBEC1 deficiency in the brain do affect the mice significantly. Further research from our collaborators at Dr. Bulloch's lab found that these physiological changes are accompanied by alterations in mouse behavior (personal communication), which highlights the importance of APOBEC1 in modulating homeostasis. The work of our collaborators is likely to be key in recognizing physiological alterations in the *Apobec1*-/- mice, and in establishing a link between APOBEC1 editing and brain disorders.

6.2. Caveats on the molecular characterization at the RNA level

At a first glance we may conclude, upon examination at the molecular level, that the function of APOBEC1 can be attributed to the modulation at the protein level, and

that the transcripts affected undergo no alternative fates than their unedited counterparts. That is, edited and unedited transcripts are equivalent to the cell. However, Davidson and collaborators have reported differential expression between the two genotypes in intestine, which also leads to differential translation of edited transcripts[105]. This discrepancy may be a reflection of the different cell systems used: enterocytes vs. macrophages. But the possibility still exists that true differences are being obscured due to the fact that Apobec1 does not edit with 100% efficiency at every site. Therefore, the currently available genome-wide comparison is of a true editing deficient (*Apobec1*-/-) to an almost deficient (*wildtype*) counterpart. This is indeed one of the major obstacles with the gene expression, stability and localization measurements and in general, to the complete molecular characterization of edited transcripts.

Such characterization is possible, using for example the constructs like the ones employed here for luciferase assays. However, this kind of assay would be cumbersome and low throughput. A better approach would entail the analysis of both edited and unedited transcripts within the *wildtype* sample. Such an experiment would have the added advantage that both transcripts are subjected to the same milieu and have as similar availability to the same RNA binding proteins as possible. Current technology does not allow for such an experiment. However, one of the ways in which I envision one could tackle this problem is through selective tagging of edited transcripts either via complementation with fluorescent probes, or alternatively, by engineering an APOBEC1 enzyme that is fused to a protein that adds an additional tag (e.g. a methyl mark) to APOBEC1 bound (and thus presumably edited) transcripts. Such marked transcripts could then be separated and quantified (at different time points for stability or in the

different fractions) or sequenced independently from the untagged (and presumably unedited) ones (Figure 6.2, below).

A second option is to engineer an APOBEC1 enzyme that is capable of editing each transcript 100%, which would be informative for fine molecular characterization of transcript expression differences and protein production. However, it would not allow us to see what biological consequences are truly the results of fine-tuning. Moreover, it would eliminate the cell diversity created by hyperediting, which in its complexity may

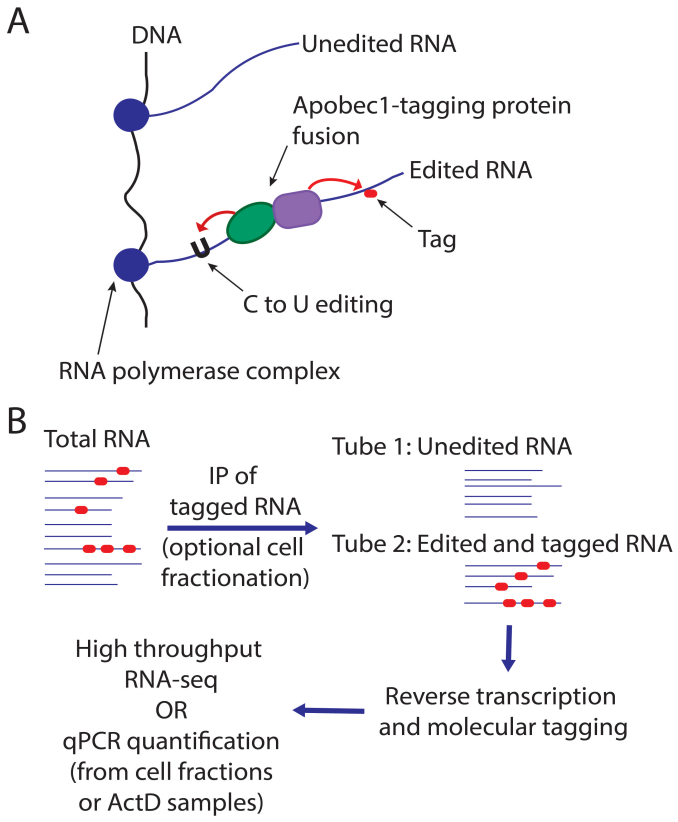

Figure 6.2 | Proposed methodology to follow edited and unedited transcripts within the *wildtype* sample. (A) A modified Apobec1-tagging protein fusion will both edit and tag transcripts during transcription. Cells that express this fusion protein can also be treated with Actinomycin D for stability measurements or any other substance of interest, for example, LPS. (B) RNA from those cells can be retrieved and edited transcripts separated from the pool via any method that is best for selecting the specific tag on the RNA (e.g. immunoprecipitation (IP), column separation, etc.). Once the fractions have been separated, gene expression analysis can be performed.

be the core of APOBEC1 editing, whatever that may be. The main caveat of engineering a protein that has 100% efficiency is that it could have off target effects, such as DNA editing. It has been reported that Apobec1 (especially of rat origin) can induce DNA mutations and promote cancer[106,107]. In a study by Severi and colleagues, rat Apobec1 was shown to target a GFP reporter inserted into DT40 chicken cells[107], which inactivates expression of GFP, a system used to extensively analyze AID-mediated DNA editing[108]. They found that the mutational signature of APOBEC1 resembles that of AID. Furthermore, they demonstrated that a similar process happens in K562 human cells (a model of chronic myeloid leukemia), whose mutational sequence context resembles that of esophageal adenocarcinomas. This shows that off target effects can be related to the level of expression of Apobec1. It has been proposed that the limiting factor preventing high amounts of APOBEC1-mediated ApoB editing is the amount of ACF, which acts as a repressor of Apobec1. This could also be the case of overexpression of Apobec1 leading to DNA methylation, albeit not necessarily through ACF.

One possibility that could be examined is whether APOBEC1 contributes itself to the stability of transcripts. It has been previously reported that APOBEC1 has affinity for AU rich sites that have been shown to be present in several untranslated regions of transcripts which are known to undergo fast degradation[70]. Upon APOBEC1 binding, the half-life of such transcripts apparently increases. Given that the transcripts affected by APOBEC1 in macrophages are also AU rich, the possibility that APOBEC1 itself and not only the editing is responsible for the long half-life still remains to be examined. In order to test for this, Apobec1 conditional mutants would need to be constructed. One of the main obstacles for this is the lack of knowledge of the crucial residues necessary for

editing. A report by MacGinnie et. al. identified several residues responsible for Apobec1 editing of ApoB through mutational analysis. They found that a histidine to arginine (H61R) yielded a catalytically dead enzyme and glutamic acid to glycine (E63G) resulted in binding but not editing of ApoB mRNA *in vitro*[109]. However, when we tested whether these mutations would eliminate editing of Apobec1 targets in macrophage, we found that both mutations resulted in editing.

6.3. Caveats on the molecular characterization at the protein level

The extent to which protein regulation by APOBEC1 editing is the result of miRNA binding still needs to be further investigated. We identified 19 transcripts in which there were sites that could be created or destroyed, but there were further miRNA binding sites contained within the footprints that were seemingly unaffected. However, we estimate that a larger number of binding sites may be targeted in the cases where hyperediting is pervasive. In these cases, the complexity of the possible combinations makes it difficult to estimate the specific contribution of editing at any site, including sites relevant to microRNA regulation.

The binding of Argonaute and mature mRNA-miRNA complexes leads to their interaction with the RISC complex, which in turn leads to gene inhibition via three main mechanisms: site-specific cleavage, mRNA degradation and translational inhibition. Both site-specific cleavage and mRNA degradation would result in differences, albeit small, in gene expression[110,111]. The latter of the processes occurs commonly in mammals and is established when the decrease in protein production is greater than the reduction in the corresponding mRNA levels[112]. Given that we have not been able to observe differences

in gene expression or stability in transcripts derived from *wildtype* vs. *Apobec1*-/-, we speculate that the process editing is involved is in translational repression. The mechanism for this type of repression is not yet fully understood, but several models have been proposed to induce repression at several different stages of translation, including initiation[113–117], increased premature termination[118] and impaired elongation[112]. Furthermore, it has been reported that Ago2 binds directly to the m7G-cap of the mRNA preventing translation initiation[119]. We speculate that any of these mechanisms is responsible for the miRNA regulation. However, the analysis of whether any one or several of these mechanisms is involved in the APOBEC1-mediated protein regulation is still to be investigated.

Given that the vast majority of edited transcripts are not predicted to contain a miRNA binding site or have one created or destroyed by editing, we decided to consider whether regulation was also possible for these transcripts. We observed through luciferase assays, where the 3'UTRs of edited transcripts were fused to the luciferase reporter, that miRNA independent regulation is also possible. Interestingly, this is also the case for a subset of transcripts, most notably the ones that are hyperedited. A finer dissection of which site or sites are the most important for this regulation is still to be performed. Additionally, it does appear that a higher frequency of hyperediting correlates with higher protein repression. However, in the case of the hyperedited transcripts B2m and Lamp1, we were unable to see differences in endogenous protein levels. We speculate that if only a subset of edited transcripts is responsible for the repression of protein, it could get lost in the bulk analysis. This underscores the importance of analysis of single cells, which would be able to show whether for a subset of cells that are

enriched for a certain type of hyperedited transcripts, we would be able to see differences in the endogenous expression of such proteins. We believe this might be the case because we also found a special example (CD36) where the regulation shown in the luciferase assays is also present in the endogenous protein levels.

Differential ribosomal binding of a subset of Apobec1 edited transcripts was recently reported[105], corroborating my luciferase results (Figure 3.3.1). However, when we looked globally at the translational output through sequencing of the polyribosome-bound fraction, we observed only subtle differences between the translation of edited genes between *wildtype* and *Apobec1-/-*. This is not completely unexpected because thus far, we have not observed big changes except for the case of CD36.

Some of the caveats of Ribo-seq are that it is unable to distinguish stalled ribosomes or for transcripts that are translated with higher or lower efficiency, or those which are stalled[120,121], due to its lack of temporal information. For example, there are higher amounts of native CD36 protein in *Apobec1-/-* BMDMs, but its transcript is not one of the highest differentially ribosome-bound. It is possible that this transcript is one such example of an efficiently translated transcript. Given that the differences we observe are subtle, a temporal analysis would be helpful to determine whether editing has a role in either stalling or translation efficiency. The latter is a process that is highly dependent on the 3'UTR length[122] and binding factors[123–125], whose interaction with the 3'UTR could be inhibited by editing.

Furthermore, Apobec1 editing events cannot be observed in the ribosome footprints, as the footprints only cover the coding region of transcripts and editing events are almost exclusively located in the 3'UTR. In order to test whether edited transcripts

are equally able to bind ribosomes, the proportion of edited transcripts in total vs. ribosome-bound RNA should be quantified. In order to do this, I devised an experiment involving an alternative ribosome capture technology, called translating ribosome affinity purification (TRAP)-seq[126]. TRAP-seq takes advantage of Ribosomes tagged with a GFP tag for purification of the ribosome bound fraction. In the published method, the RNA is collected and ribosomes are immunoprecipitated using GFP-specific antibodies and stringently washed. After ribosome pool down, the entire transcripts are available and library preparation for direct RNA sequencing of both fractions is possible. For the detection of editing, I propose to perform amplicon sequencing of the RNA in both the pool down tagged ribosome and whole RNA fractions, using a similar method as described in section 4.3, which uses molecular tagging of transcripts (see figure 6.3.1 below). I believe this method would give a unique insight into the mechanism by which edited transcripts are translated, that could explain the luciferase results presented in Figure 3.3.1 (see figure 6.3.2 below).

6.4. Steady state versus activated macrophages

This work provides an overview of the possible functions of APOBEC1 in macrophage biology. However, these studies are a mix of steady state and activated macrophages. I characterized the edited sites in steady state BMDMs, and showed that these cells gave rise to altered phagocytosis and migration; and I showed that Apobec1 could get induced after LPS stimulation. Therefore, it is important moving forward to characterize how editing is altered during the activation of cells, which could give insight into Apobec1 function during an inflammatory response and would clue us into which

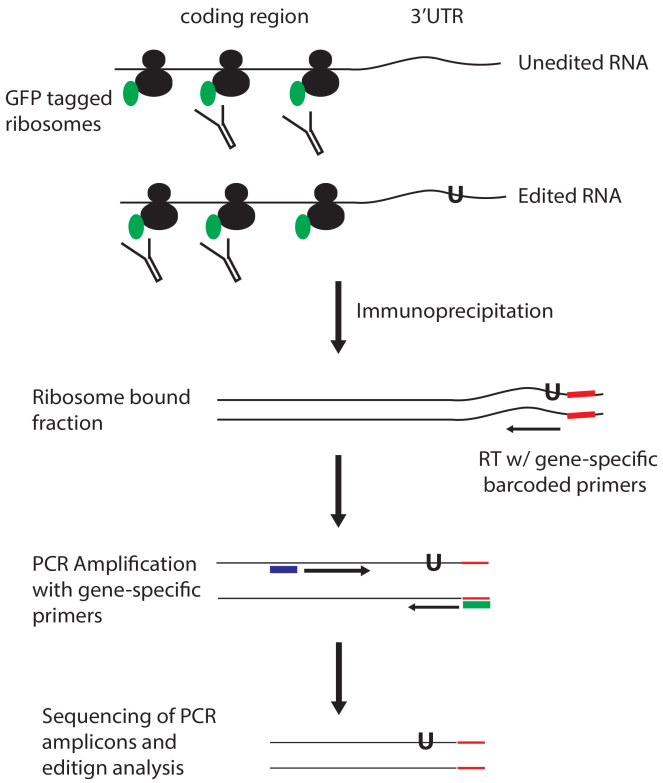

Figure 6.3.1 | Modified TRAP-seq method to identify ribosome-bound edited transcripts. Depicted are the possible method for detecting editing within ribosome-bound transcripts. Using GFP tagged ribosomes, the fraction of RNA being translated can be isolated from the untranslated part. Reverse transcription with barcoded primers can help identify individual RNA molecules after PCR amplification. Finally, the PCR amplicons can be used for library preparation and high throughput sequencing.

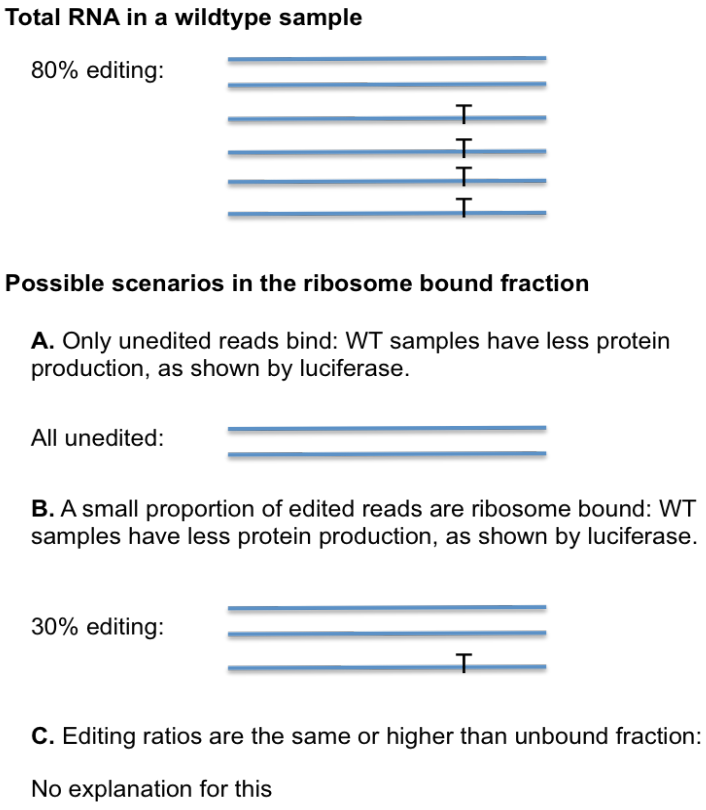

Figure 6.3.2 | Scenarios that would explain differences in protein expression. Depicted are the possible explanations that would result in the differences between *wildtype* and *Apobec1*-/- genotypes shown via luciferase, from looking at the different scenarios of ribosome binding.

aspect of inflammation is being involved, given that the experiments currently available are inconclusive (figure 5.3).

On the other hand, we unexpectedly encountered that APOBEC1 may play a role in monocyte differentiation. Macrophages are a heterogeneous population of cells that have, for research purposes, been classified into different categories given their activation and differentiation states. Even though BMDMs are an excellent model for the study of macrophage responses, they do not represent the functional state of most monocyte populations *in vivo*. The closest relatives of BMDMs *in vivo* are fully differentiated tissue-resident macrophages. Therefore, many of the observations from the *in vitro* studies may or may not apply to these populations. Edited sites within the MDP, cMOP and Ly6+ and – populations remain to be determined. These could shed some light into the role of Apobec1 in the differentiation and also, into which compensatory mechanisms are in place for normal development vs. ones that have developed after grafting.

It is important to consider also that the cells derived from grafting develop in an inflammatory environment that is the result of whole body irradiation. It is possible that the two environments are not equivalent and that Apobec1 only plays a crucial role when cells are exposed to certain inflammatory environments. If this were the case, it would be relevant for the observations in aged brains, where there is an established inflammatory environment. It would also help shed light on the mechanism by which the phenotype of clustered microglia and inflammation in Apobec1-/- aged brains arises: activated microglia provided a pro-inflammatory environment that results in degradation of neurons; or inflammation from some other origin activates otherwise normal microglia and initiated the brain alterations. If inflammation leads to Apobec1 expression (as shown

in figure 5.3) and Apobec1 is involved in the regulation of monocytic cells under inflammatory conditions (as shown in figure 5.2.2), it is possible that the underlying mechanism of Apobec1 is the regulation (through editing or indirectly) of genes that maintain the monocyte program. This could be a potential mechanism underlying the aberrant activation of microglia in the brain of Apobec1-deficient mice: they lack adaptability to the inflammatory environment due to APOBEC1 loss.

6.5 Closing remarks

The substantial phenotypic diversity of monocytes is established epigenetically, and maintained by local environmental cues. Hence, discrete monocyte subsets have unique transcriptomic signatures. However, phenotypic heterogeneity is also present within specific subsets with identical transcriptomic signatures. Here, I demonstrate that APOBEC1-mediated RNA editing generates sequence heterogeneity in macrophages that appear identical by gene expression measures, but which nevertheless can show phenotypic differences.

These epitranscriptomic sequence changes occur mostly in transcript 3' untranslated regions (UTRs) and result in alterations in the quantity of individual proteins through a variety of mechanisms, including miRNA target sequence ablation. The data presented here constitutes the first instance where RNA editing of clusters of transcripts that aggregate in common pathways is shown to specifically affect the output of those pathways, and by extension, cell physiology.

My work suggests that the sequence diversity contributed by RNA editing might provide subsets of cells with distinct informational content. Further, our work implies that

RNA editing might underlie or anticipate the functional heterogeneity apparent in populations of cells of monocytes, the mechanism for which is currently not understood. The data also suggests that editing may contribute to the substantial phenotypic diversity of monocytes, such that phenotypic heterogeneity is also present within specific subsets with identical transcriptomic signatures.

I also demonstrate that changes in protein abundance initiated by RNA editing have functional consequences for the physiology of the monocytes, both *in vitro* and *in vivo*, predicting that slight differences in the maintenance of monocyte subsets might, over time, have critical consequences for the health of the organism.

Given that standard genome-wide association studies (which catalog DNA changes) have not been particularly successful in predicting disease progression or manifestations of neurodegenerative or psychiatric disorders, the development of a link between aberrant RNA editing and brain disorders, might be useful for identifying signatures, shared with peripheral monocytes, that can be used in diagnosis, prevention or treatment.

CHAPTER 7. METHODS

7.1. Mice and isolation of BMDMs

C57BL6 littermate or age-matched mice were used at 6-12 weeks of age. *Apobec1-/-* mice were a gift from N.O. Davidson (Washington University School of Medicine). Mouse femurs were removed and flushed with cold PBS onto a cell strainer. Precursor cells were plated onto untreated 10cm dishes and incubated in a humidified 37°C, 5% CO_2 incubator for 7 days in DMEM, 10% FBS 1% Non-essential amino acids (Invitrogen), 0.1% BME, 20ng/mL M-CSF (Peprotech), replacing half of the media (with complete media plus 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.

7.2. mRNA-Seq

RNA was extracted from macrophage cultures using the Ribopure kit (Ambien). DNAse treatment was performed with Turbo DNAse (Ambien). mRNA libraries for Illumina sequencing were prepared as previously described[82] with the following modifications: 1) polyA selection was performed with Sera Mag oligo-dT magnetic beads (Thermo) 2) mRNA fragmentation in fragmentation buffer was adjusted for higher read length, and was performed for exactly 4 min 45s 3) After double-stranded cDNA libraries were separated by agarose gel electrophoresis and fragments ranging from 275-325nt were excised. Single-end 75nt sequencing was performed on Illumina Genome Analyzer IIx (GAIIx) yielding 28-33 million reads. Trimmed reads were mapped to the C57LBL/6

mouse reference genome (NCBI37/mm9) using Tophat (v1.3.3) and Bowtie (v0.12.8) with the tophat parameters "--solexa1.3-quals -g 1 --coverage-search." Gene expression levels were calculated using Cufflinks (v1.2.1) using the Ensembl gene set.

7.3. RNA editing detection pipeline

A vector consisting of the number of A's, T's, G's and C's that occurred at each coordinate were constructed from the SAMtools pileup33 for both the *wildtype* and *Apobec1-/-* bulk RNA-seq alignments that were de-duplicated using Picard (http://broadinstitute.github.io/picard). For each coordinate that exhibited a C-to-T change only in the *wildtype* sample (that is, not in the knockout alignment, such that only APOBEC1-dependent C-to-T changes are kept) and met a number of stringent quality control thresholds (minimum of five reads covering site, with at least two reads supporting the editing event, excluding sites that showed multiple types of transitions; and discarding reads that contain indels, support an edit in the first or last two base pairs of a read), the angle between the corresponding vectors for the *wildtype* and *Apobec1-/-* were compared. Putative hits were retained if the magnitude of the *wildtype* vector was at least 15 and the angle between the *wildtype* and *Apobec1-/-* vectors was at least 0.11 radians (approximately equivalent to a minimum coverage of 20 reads and an editing rate of 10%). Potential sites were also filtered against genomic DNA-derived SNPs indbSNP138, and removed if they occurred within four base pairs of a splice junction (using the exon junctions compiled by the Zhang lab, using OLego34) or in simple or tandem repeats (softmasked regions by RepeatMasker). Reads supporting edits were run through BLAT35 to ensure that they were not ambiguously mapped. The pipeline was

programmed using Bash and Python with the Pysam (https://github.com/pysam-developers/pysam) pileup engine. RNA editing events were then validated by designing primers proximal to the sites of interest, amplifying those regions from cDNA and genomic DNA from both *wildtype* and *Apobec1*-/- cells, and performing colony sequencing.

7.4. Subcellular fractionation

A total of 10x106 bone marrow derived macrophages were resuspended in 200µL of cold cytoplasmic buffer (0.15% NP-40, 10mM Tris pH 7.5 and150mM NaCl) and incubated on ice 5 minutes, in order to break the cell membrane. The nuclei were separated by centrifugation at 16,000 rcf, 10 minutes at 4ºC. The cytoplasmic supernatant was separated and the nuclei pellet was washed twice with cold PBS. The nuclei were gently resuspended in 100µL of glycerol buffer (20 mM Tris pH 8, 75 mM NaCl, 0.5 mM EDTA and 50% Glycerol). Next, 100µl of nuclei lysis buffer (20 mM HEPES pH7.6, 7.5 mM MgCl2, 0.2 mM EDTA, 0.3M NaCl, 1M Urea, 1% NP40) were added and lysed by vortexing twice for 2 seconds and incubating 2 minutes at 4C. The chromatin fraction was separated with a spin at maximum speed, 2 minutes at 4C. Total RNA was separated from both the nuclear and cytoplasmic fractions using phenol chloroform isoamyl alcohol extraction. The fractionation efficiency was determined by western blot using alpha tubulin (Sigma 4500088) and H3K27Me3 antibodies (Abcam 6002).

RNA half-life estimates after exposure to actinomycin-D.

Cultures of BMDM were exposed to actinomycin-D at a concentration of 10µg/mL for periods of 0, 0.5, 1, 2 and 5 hours. RNA was collected via Trizol and

reverse transcribed using Superscript III and polydT primers. The concentration of edited genes *Cybb, Sptssa, App, Cd36, Lamp1, Lamp2* and *B2m* was determined by RT-PCR, normalizing to GAPDH. Additionally, the genes Atp5a1 and Jun were measured by RT-PCR as controls for long and short-lived transcripts in macrophages, respectively. The half-life of the transcripts was calculated using the formula $t_{1/2}= \ln(2)/k_{decay}$, where k_{decay} is the slope of a semilogarithmic plot of the concentration of mRNA in time.

7.5. HITS-CLIP analysis

Argonaute HITS-CLIP was performed in conjunction with the Darnell lab as previously described[84], with a few modifications: BMDMs were isolated from 3 wildtype and *Apobec1*$^{-/-}$ littermate pairs. BMDMs were matured as described above and crosslinked 3x at 200mJ/cm^2 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. Cells were DNAse treated with RQ1 RNAse-free DNAse (Promega). RNAse treatment performed with high (1:100) or low (1:10,000) RNAse A solutions. Protein A beads (first loaded with rabbit anti-mouse IgG bridging antibody (Jackson ImmunoResearch, at 2.3 mg/mL)) were then incubated with 3 uL Ago antibody (2A8 ascites provided by Dr. Z. Mourelatos) per 400 uL of beads and rotated for 3 hours at 4°C. Lysates were incubated with primary antibody-loaded beads for 2 hours at 4°C and washed in a series of high- and low- stringency washes, after which 5'phosphate dephosphorylation was performed, and a radiolabeled linker was ligated to the 3'end of the bound RNA species. Cold L32 RNA linker with 5’ phosphate was added to this

mixture after 1 hr and it was incubated overnight and washed 3x with PNK buffer. 5' ends were re-phosphorylated with T4 PNK. Protein:RNA complexes were eluted off the beads in NuPAGE loading buffer (Invitrogen) and supernatants were run on Novex NuPAGE 8% Bis-Tris gels (Invitrogen) in MOPS running buffer (Invitrogen), 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 and treated with proteinase K (4mg/mL Roche). 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). The ligated reaction was then subjected to DNAse treatment with RQ1 DNAse (Promega) and extracted with phenol-chloroform and ethanol precipitation, 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). Products running between 60-100nt were excised from the lowest cycle number with visual product and gel extracted. Additional PCR amplification followed by gel extraction was performed with fusion primers as previously described to provide the platform for Illumina sequencing. 10-30uL of 10nM DNA was submitted for sequencing on Illumina HiSeq.

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. 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 in 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 (*9*) 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. Clusters made up of ≥8 reads and a 5nt overhang were identified. The peaks of the clusters were identified as previously described (*8*) and the "Ago footprint" around the cluster broadly defined as the region 32nt upstream and 30nt downstream from the peak or narrowly defined as the region 22nt upstream and 24nt downstream of the edited say (region in which Ago is bound 100% of the time). 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). We find this method to be an 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 artificially 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 (with Bedtools) 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. Ago footprints were then overlapped with APOBEC1 editing events to identify regions of potential Ago:APOBEC1 interaction.

miRNA 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. 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 (available upon request) 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 assays (described below).

7.6. Luciferase Assays

For miRNA dependent analysis: site-specific mutagenesis was performed on unedited (matched reference sequence) version target 3′UTRs and these were cloned into luciferase vectors. Renilla luciferase served as an internal control. HEK293T cells were transfected with edited or unedited constructs plus the miRNAs of interest (synthesized by Dharmacon) using the DuoFECT reagent (and according to the manufacturer's recommendations) or an unrelated microRNA control. Cells were returned to the incubator for 48 hours after which luciferase expression was assessed. Cells were lysed with passive lysis buffer by shaking at RT for 15min and subjected to one freeze-thaw cycle at -80°C. Firefly (experimental) and renilla (transfection control) 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 (transfection control) for each construct. We calculated the percentage of repression for both edited and unedited constructs using the reduction in luciferase activity induced by

overexpression of the microRNA of interest relative to the expression of the unrelated microRNA control (following the methods used by Loeb et al.[127]). The statistical significance for the differences was determined using a student's T test with $\alpha = 0.05$ (Prism). The mRNA-miRNA pairs tested were Sptssa with miR17 (miR501 was used as unrelated control), Rac1 with miR669 (miR17 as control).

For miRNA independent analysis: a series of edited and unedited 3′UTRs were cloned using a high fidelity polymerase (Turbo, Invitrogen) from *wildtype* ("edited") and *Apobec1*$^{-/-}$ ("unedited") macrophage cDNA into a subclone vector (Strataclone). Clones were screened and representative clones for each degree of editing (single-site, number of hyper-edited events, etc.) were inserted downstream of a firefly luciferase in a dual-luciferase vector (Promega).

Apobec1$^{-/-}$ BMDMs were transfected with luciferase constructs and pmaxGFP transfection control vector using the Amaxa Mouse Macrophage Nucleofactor kit (Lonza). Cells were returned to the incubator for 24 hours after which luciferase expression was assessed. Cells were lysed and luciferase was determined as described in the previous section. In all cases, significance for the difference between each "edited" and "unedited" pair was calculated by a Mann-Whitney test (Prism). We tested a total of 6 single-site edited transcripts (*Cd36, Tmem55a, Sptssa, Sep15, Anxa5* and *Rac1*) and several combinations of editing in hyper-edited transcripts (*B2m, App, Lamp1* and *Adam10*).

7.7. High-throughput Ribosome profiling

Ribosome profiling libraries were prepared using the Truseq Ribo Profile kit per manufacturer's instructions. Libraries were sequenced with 50 nt single end reads on the Illumina HiSeq 2500. Alignment, and mRNA and RPF read counting were performed as described in 20. Briefly, libraries were trimmed for adapters and quality using Trim Galore. The resulting reads were initially aligned to mouse ribosomal RNA using bowtie (v1.1.2). The remaining unmapped reads were then aligned to mm10 (using the reference sequences and annotations provided by iGenomes (Illumina)) using Tophat2 (v.2.1.0) and the following parameters: –bowtie1, –no- novel-juncs, –read-realign-edit-dist=0 (default settings for the remaining parameters). mRNA reads were counted using HTseq in intersection-strict mode, and RPF reads were counted using a custom script provided by Xiao et al[87]. DESeq2 was then used to assess changes in translational efficiency (TE), defined here as the ratio of RPF to mRNA reads, between the two genotypes. This was achieved with the design ~ assay + genotype + assay:genotype + batch, and analysis run as follows: dds <- DESeq(dds, test='LRT', full=design(dds), reduced = ~ assay + genotype + batch), where assay indicates if the library was mRNA or RPF, genotype indicates if the sample was WT or KO, and batch indicates the batch that the library was prepared with. Differential gene expression analysis was also performed, with DESeq2, using only the total mRNA reads generated from the Ribo-Seq experiment.

7.8. Single cell RNA-Seq

To generate single-cell libraries, wild-type bone marrow-derived macrophages were flowed into a C1 IFC for mRNA seq (10–17 mm) chip using the Fluigidm system.

Lysis, RT and PCR were performed using the SMARTer Kit designed for the C1 (Clontech, Mountain View, CA). The efficiency of chip loading (capture) was confirmed by microscopy and any of the wells that contained either none or more than one cell were noted and discarded from further library preparation and analysis. In all, 24 single cDNA libraries were selected for library preparation on the basis of concentration and size range, as determined via Agilent Bioanalyzer. Sequencing libraries were made from the cDNA using the Nextera XT DNA Sample Preparation Kit (Illumina). To generate conventional bulk RNA-seq libraries, total RNA was extracted using Trizol (Invitrogen, Grand Island, NY) from 500,000 to 1,000,000 macrophages of each genotype (wild type and APOBEC1_/_). A unit of 1 mg of the total RNA collected per condition was then treated with DNase, and then processed using the NEBNext Ultra Directional RNA Library Prep Kit for Illumina (NEB, Ipswich, MA).

Libraries were sequenced on the Illumina HiSeq 2000, generating 100-nucleotide, single-end reads. Reads were trimmed for quality and adapters using Trim Galore!, and then aligned to mm10 (using the reference sequences and annotations provided by iGenomes (Illumina)) using Tophat2, allowing only unique alignments and up to 2 mismatches per 20-nucleotide segment. Gene expression was determined using Cufflinks32. The correlation of gene expression between the ensemble of single cells and the bulk was calculated using the Spearman and Pearson correlation coefficients. Potential PCR duplicates were removed from single cell alignments using Picard (http://broadinstitute.github.io/picard).

7.9. Hierarchical Statistical Model

A lay summary of the method is described here; for full methods, including the mathematical derivation, please see Harjanto el at. 2016[80]. First, an editable site of interest on a given RNA transcript and its associated probability of editing were selected. The transcript can be modeled as a coin with a probability of falling on 'Heads' when tossed equal to the associated editing probability (not necessarily 50%). We can think of each single cell as consisting of a stack of coins (with each coin corresponding to the same site on a different copy of the transcript), each having the same probability of falling on 'Heads'. Finally, we can think of the population of cells as a bag full of stacks of coins, where the probability of 'Heads' for each stack is drawn from an unknown distribution on [0, 1]. The variance of this distribution quantifies the diversity among editing rates of different cells (or stacks of coins). The single-cell experiments correspond to randomly picking any number of stacks from the bag and tossing some of their coins, whereas the bulk experiment corresponds to randomly picking a large number of individual coins from the bag, after emptying all the stacks in and mixing them together. Because we assume that the edited sites can be treated independently, we can restrict attention to just a single genomic location, for which we can estimate a posterior distribution by applying Bayes' rule and the appropriate Markov chain Monte Carlo algorithms. The posterior density of the variance provides a quantification of RNA editing variance.

To test the model, we performed simulations using artificial data sets comprising of 20 cells exhibiting high levels of editing variance for a theoretical site, generated by

randomly sampling the number of edited reads to attain effective editing rates only in the region of 0-5% and 95-100%.

7.10. Targeted single-cell RT–PCR

To analyze editing of specific sites in single cells, single wild-type macrophages were sorted into 96-well PCR plates with 5-ml of lysis buffer, containing 0.45% NP-40, 0.36Uml_1 RNAse Inhibitor and 0.18Uml_1 Superase-In (Ambion). RT–PCR amplification was done with gene specific primers and the OneStep RT–PCR kit (Qiagen), using a modified protocol. Single transcript molecules were tagged with barcoded gene-specific primers that have an additional universal sequence, used in RT. These primers were then digested with 1U of Exonuclease I (30 oC for 30 min; NEB). Afterwards, a mix of universal forward and gene specific reverse primers were added to the PCR mix and 35–40 cycles of PCR were performed. The PCR products were introduced into bacteria using a TOPO TA cloning kit (Invitrogen) and single bacterial colonies were sequenced using Sanger sequencing. The resulting sequences were then aligned to the reference transcriptome (Macvector) and PCR duplicates were eliminated using the barcodes.

Transcript specific primer sequences were as follows:

B2m primers:

RT:

50-

GGCCAGTGAATTGTAATACGACTCACTATAGGNNNNNNAAAGCAGAAGTAG

CCACAGGGTTG-30

PCR-reverse: 50-TTAAGCATGCCAGTATGGCCGA-30

Anxa5 primers: RT:

50-

GGCCAGTGAATTGTAATACGACTCACTATAGGNNNNNNGTCGCCAATGTTTT

GGAT ACTACCATC-30

PCR-reverse: 50-GCGACACATCTGGAGACTATAA GAAGGC-30

Cybb primers: RT:

50-

GGCCAGTGAATTGTAATACGACTCACTATAGGNNNNNNGAGGGTTTGTGCCT

AGTCTTATTGCA-30

PCR-reverse: 50-GCATGCGCTCATCTTGTTTTGACTTC-30

Universal PCR-forward: 50-GGCCAGTGAATTGTAATACGACTCACTATAGG-30

7.11. Migration assay

BMDMs from wildtype or Apobec1-/- mice were starved in serum free DMEM, containing 0.2% BSA and 20 ng/mL mCSF for twelve hours. The lower chambers of the 96 well chemotaxis plate (Neuroprobe 8um, cat. 106-8) were filled with 30uL of DMEM 0.2% BSA in the absence or presence of increasing concentrations of murine CXCL12 (Peprotech), in triplicate. Macrophage suspensions (in phenol red free DMEM (Life Technologies) containing 0.2% BSA) were deposited on the top portion of the migration plate and placed at 37°C, 5% CO2, for 4 hours. Cells that did not migrate were aspirated off the top portion of the filter, which was also wiped with a cotton tip. BMDMs attached to the lower surface of the filter were centrifuged in a 96 well black plate with clear

bottom containing lysis buffer with CyQuant DNA-binding dye (Life Technologies). Fluorescence (excitation 485nm- emission 530nm) was measured using the Synergy Neo plate reader, after a 30-minute incubation at room temperature with gentle shaking. The chemotactic index was calculated by dividing the fluorescence intensity of each sample by the fluorescence intensity in the absence of chemokine. Data was analyzed using Prism 6.0g software (GraphPad Software, San Diego, CA) and statistical significance of the differences between the various conditions was determined using multiple measured one-way analysis of variance, with Bonferroni's correction for the t test.

7.12. pHrodo assays

Wildtype and *Apobec1*-/- BMDM were incubated with 1(MOI 10) or 10mL (MOI 100) of pHrodo Red E. coli BioParticles (Invitrogen) in HBSS at 37C without CO2 FOR 20 min. After the incubation period, cells were washed twice with HBSS and stained with Alexa 488 F4/80. Cells were detached with 5mM EDTA in PBS and analyzed via flow cytometry on a FACS Calibur flow cytometer.

7.13 Bone marrow chimeras

Bone marrow was collected from 8 *Apobec1*$^{-/-}$ CD45.2 and pooled; same with 8 *wildtype* CD45.1 mice. Lineage positive cells were depleted using a cocktail of MAC1, GR-1, B220, CD4, CD8, CD3, Ter119, CD19, CD11c and NK1.1 biotinylated antibodies (BD Pharmingen) together with streptavidin columns (MACS). *Wildtype* CD45.1 mice (Jackson labs) or heterozygote for both CD45.1/CD45.1 and *Apobec1*$^{+/-}$ were lethally irradiated (with two doses of radiation delivered 3 hr apart and totaling 950 rad) and

injected (retro-orbitally) with a 1:1 mix of CD45.1 *wildtype* and CD45.2 *Apobec1*$^{-/-}$ lineage depleted cells. Spleens from chimeric mice were collected 28 days post-transplant in both backgrounds and 48 days in two mice transplanted in the CD45.1 background. For the analysis of macrophages in spleen, cells were stained with Ax488 F4/80, PerCP-Cy5.5 CD11b, APC Ly6C, Aqua death cell exclusion, BV711 CD45.1, BUV395 CD45.2, PE Ly6G, PE-Cy7 CD11c. For the analysis of bone marrow, cells were stained with biotinylated CD3, CD19, NK1.1 and Ly6G; followed by staining with Ax488 CD115, PE-CF594 Streptavidin, PE CD135, BV421 CD117, Aqua dead cell exclusion, BV711 CD45.1 and BUV395 CD45.2. T cells were identified as either CD3+CD4+ or CD3+CD8+, with FITC CD4, PE CD3, APC CD8, Aqua dead cell exclusion, BV711 CD45.1 and BUV395 CD45.2. B cells in spleen were defined as CD19+B220+, stained with Ax488 B220, PE CD19, Aqua dead cell exclusion, BV711 CD45.1 and BUV CD45.2. In bone marrow, B cell progenitors were stained with Alexa 488 B220, PE CD43, PE-Cy7 IgM, Aqua dead cell exclusion, BV711 CD45.1, BUV395 CD45.2. Pro-B cells were defined as IgM-B220+CD45$_{high}$, Large Pre-B cells as IgM-B220+CD43$_{med}$FSC-A$_{high}$, and Pre-B cells as IgM-B220+CD43$_{med}$FSC-A$_{low}$. The quantification of populations was done via flow cytometry using an LSR Fortessa (BD Biosciences) cell analyser.

7.14. High throughput amplicon sequencing of edited fragments

CD11c+ and CD11Cneg cells were isolated from cortex and hippocampus by MACS column separation. Briefly, cells were incubated with biotinylated CD11c antibodies (BD) and the CD11c+ fraction was retained in a MACS (Miltenyi Biotec). The

eluted cells were then incubated with a biotinylated CD11b antibody and the CD11c- (but CD11b+) fraction was retained in a separate MACS column. A 300bp fragment surrounding a predicted edited site was amplified from both CD11c+ and CD11c- cDNA and genomic DNA. Next, all the PCR products were purified and mixed. Library preparation was performed on the mixed sample using the Illumina True-seq kit and was sequenced in an Illumina mi-Seq. Reads were aligned to the genome and editing frequency was determined by counting the number of changes to the reference genome, using a custom script.

REFERENCES

1. Machnicka, M. A. *et al.* MODOMICS: a database of RNA modification pathways—2013 update. *Nucleic Acids Res.* **41,** D262–D267 (2013).
2. Jackman, J. E. & Alfonzo, J. D. Transfer RNA modifications: nature's combinatorial chemistry playground. *Wiley Interdiscip. Rev. RNA* **4,** 35–48 (2013).
3. Durant, P. C., Bajji, A. C., Sundaram, M., Kumar, R. K. & Davis, D. R. Structural effects of hypermodified nucleosides in the Escherichia coli and human tRNALys anticodon loop: the effect of nucleosides s2U, mcm5U, mcm5s2U, mnm5s2U, t6A, and ms2t6A. *Biochemistry (Mosc.)* **44,** 8078–8089 (2005).
4. Dalluge, J. J., Hashizume, T. & McCloskey, J. A. Quantitative measurement of dihydrouridine in RNA using isotope dilution liquid chromatography-mass spectrometry (LC/MS). *Nucleic Acids Res.* **24,** 3242–3245 (1996).
5. Zagryadskaya, E. I., Kotlova, N. & Steinberg, S. V. Key elements in maintenance of the tRNA L-shape. *J. Mol. Biol.* **340,** 435–444 (2004).
6. Moss, B., Gershowitz, A., Stringer, J. R., Holland, L. E. & Wagner, E. K. 5'-Terminal and internal methylated nucleosides in herpes simplex virus type 1 mRNA. *J. Virol.* **23,** 234–239 (1977).
7. Nichols, J. L. & Welder, L. The modified nucleotide constituents of human prostatic cancer cell (MA-160) poly(A)-containing RNA. *Biochim. Biophys. Acta* **608,** 1–18 (1980).
8. Rottman, F. M., Desrosiers, R. C. & Friderici, K. Nucleotide methylation patterns in eukaryotic mRNA. *Prog. Nucleic Acid Res. Mol. Biol.* **19,** 21–38 (1976).

9. Harcourt, E. M., Kietrys, A. M. & Kool, E. T. Chemical and structural effects of base modifications in messenger RNA. *Nature* **541,** 339–346 (2017).

10. Prohaska, K. M., Bennett, R. P., Salter, J. D. & Smith, H. C. The multifaceted roles of RNA binding in APOBEC cytidine deaminase functions. *Wiley Interdiscip. Rev. RNA* **5,** 493–508 (2014).

11. Licht, K. & Jantsch, M. F. Rapid and dynamic transcriptome regulation by RNA editing and RNA modifications. *J Cell Biol* **213,** 15–22 (2016).

12. Meyer, K. D. & Jaffrey, S. R. The dynamic epitranscriptome: N6-methyladenosine and gene expression control. *Nat. Rev. Mol. Cell Biol.* **15,** 313–326 (2014).

13. Li, S. & Mason, C. E. The Pivotal Regulatory Landscape of RNA Modifications. *Annu. Rev. Genomics Hum. Genet.* **15,** 127–150 (2014).

14. Benne, R. *et al.* Major transcript of the frameshifted coxll gene from trypanosome mitochondria contains four nucleotides that are not encoded in the DNA. *Cell* **46,** 819–826 (1986).

15. Stuart, K. D., Schnaufer, A., Ernst, N. L. & Panigrahi, A. K. Complex management: RNA editing in trypanosomes. *Trends Biochem. Sci.* **30,** 97–105 (2005).

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

17. Gerber, A. P. & Keller, W. RNA editing by base deamination: more enzymes, more targets, new mysteries. *Trends Biochem. Sci.* **26,** 376–384 (2001).

18. Maas, S. & Rich, A. Changing genetic information through RNA editing. *BioEssays* **22,** 790–802 (2000).

19. Valente, L. & Nishikura, K. RNA binding-independent dimerization of adenosine deaminases acting on RNA and dominant negative effects of nonfunctional subunits on dimer functions. *J. Biol. Chem.* **282,** 16054–16061 (2007).

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

21. Slavov, D., Clark, M. & Gardiner, K. Comparative analysis of the RED1 and RED2 A-to-I RNA editing genes from mammals, pufferfish and zebrafish. *Gene* **250,** 41–51 (2000).

22. Hough, R. F., Lingam, A. T. & Bass, B. L. Caenorhabditis elegans mRNAs that encode a protein similar to ADARs derive from an operon containing six genes. *Nucleic Acids Res.* **27,** 3424–3432 (1999).

23. Palladino, M. J., Keegan, L. P., O'Connell, M. A. & Reenan, R. A. dADAR, a Drosophila double-stranded RNA-specific adenosine deaminase is highly developmentally regulated and is itself a target for RNA editing. *RNA N. Y. N* **6,** 1004–1018 (2000).

24. Jin, Y., Zhang, W. & Li, Q. Origins and evolution of ADAR-mediated RNA editing. *IUBMB Life* **61,** 572–578 (2009).

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

26. Melcher, T. *et al.* A mammalian RNA editing enzyme. *Nature* **379,** 460–464 (1996).

27. Chen, C. X. *et al.* A third member of the RNA-specific adenosine deaminase gene family, ADAR3, contains both single- and double-stranded RNA binding domains. *RNA N. Y. N* **6,** 755–767 (2000).

28. Paul, M. S. & Bass, B. L. Inosine exists in mRNA at tissue-specific levels and is most abundant in brain mRNA. *EMBO J.* **17,** 1120–1127 (1998).

29. Nishikura, K. *et al.* Substrate specificity of the dsRNA unwinding/modifying activity. *EMBO J.* **10,** 3523–3532 (1991).

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

31. Higuchi, M. *et al.* RNA editing of AMPA receptor subunit GluR-B: a base-paired intron-exon structure determines position and efficiency. *Cell* **75,** 1361–1370 (1993).

32. Srivastava, D. Making or breaking the heart: from lineage determination to morphogenesis. *Cell* **126,** 1037–1048 (2006).

33. Hartner, J. C. *et al.* Liver Disintegration in the Mouse Embryo Caused by Deficiency in the RNA-editing Enzyme ADAR1. *J. Biol. Chem.* **279,** 4894–4902 (2004).

34. Higuchi, M. *et al.* Point mutation in an AMPA receptor gene rescues lethality in mice deficient in the RNA-editing enzyme ADAR2. *Nature* **406,** 78–81 (2000).

35. Burns, C. M. *et al.* Regulation of serotonin-2C receptor G-protein coupling by RNA editing. *Nature* **387,** 303–308 (1997).

36. Hoopengardner, B., Bhalla, T., Staber, C. & Reenan, R. Nervous system targets of RNA editing identified by comparative genomics. *Science* **301,** 832–836 (2003).

37. Seeburg, P. H. & Hartner, J. Regulation of ion channel/neurotransmitter receptor function by RNA editing. *Curr. Opin. Neurobiol.* **13,** 279–283 (2003).

38. Köhler, M., Burnashev, N., Sakmann, B. & Seeburg, P. H. Determinants of Ca2+ permeability in both TM1 and TM2 of high affinity kainate receptor channels: diversity by RNA editing. *Neuron* **10,** 491–500 (1993).

39. Niswender, C. M., Copeland, S. C., Herrick-Davis, K., Emeson, R. B. & Sanders-Bush, E. RNA editing of the human serotonin 5-hydroxytryptamine 2C receptor silences constitutive activity. *J. Biol. Chem.* **274,** 9472–9478 (1999).

40. Rosenthal, J. J. C. & Bezanilla, F. Extensive editing of mRNAs for the squid delayed rectifier K+ channel regulates subunit tetramerization. *Neuron* **34,** 743–757 (2002).

41. Kawahara, Y. *et al.* Glutamate receptors: RNA editing and death of motor neurons. *Nature* **427,** 801 (2004).

42. Gurevich, I. *et al.* Altered editing of serotonin 2C receptor pre-mRNA in the prefrontal cortex of depressed suicide victims. *Neuron* **34,** 349–356 (2002).

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

44. Kim, D. D. Y. *et al.* Widespread RNA editing of embedded alu elements in the human transcriptome. *Genome Res.* **14,** 1719–1725 (2004).

45. Prasanth, K. V. *et al.* Regulating Gene Expression through RNA Nuclear Retention. *Cell* **123,** 249–263 (2005).

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

47. Iizasa, H. *et al.* Editing of Epstein-Barr virus-encoded BART6 microRNAs controls their dicer targeting and consequently affects viral latency. *J. Biol. Chem.* **285,** 33358–33370 (2010).

48. Kawahara, Y. *et al.* Redirection of silencing targets by adenosine-to-inosine editing of miRNAs. *Science* **315,** 1137–1140 (2007).

49. Ota, H. *et al.* ADAR1 Forms a Complex with Dicer to Promote MicroRNA Processing and RNA-Induced Gene Silencing. *Cell* **153,** 575–589 (2013).

50. Cattaneo, R. *et al.* Biased hypermutation and other genetic changes in defective measles viruses in human brain infections. *Cell* **55,** 255–265 (1988).

51. Ward, S. V. *et al.* RNA editing enzyme adenosine deaminase is a restriction factor for controlling measles virus replication that also is required for embryogenesis. *Proc. Natl. Acad. Sci. U. S. A.* **108,** 331–336 (2011).

52. Taylor, D. R., Puig, M., Darnell, M. E. R., Mihalik, K. & Feinstone, S. M. New Antiviral Pathway That Mediates Hepatitis C Virus Replicon Interferon Sensitivity through ADAR1. *J. Virol.* **79,** 6291–6298 (2005).

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

54. Mannion, N. M. *et al.* The RNA-editing enzyme ADAR1 controls innate immune responses to RNA. *Cell Rep.* **9,** 1482–1494 (2014).

55. Vitali, P. & Scadden, A. D. J. Double-stranded RNAs containing multiple IU pairs are sufficient to suppress interferon induction and apoptosis. *Nat. Struct. Mol. Biol.* **17,** 1043–1050 (2010).

56. Sharma, S. *et al.* APOBEC3A cytidine deaminase induces RNA editing in monocytes and macrophages. *Nat. Commun.* **6,** (2015).

57. Chen, H. *et al.* APOBEC3A Is a Potent Inhibitor of Adeno-Associated Virus and Retrotransposons. *Curr. Biol.* **16,** 480–485 (2006).

58. Hirano, K.-I. *et al.* Targeted Disruption of the Mouse apobec-1 Gene Abolishes Apolipoprotein B mRNA Editing and Eliminates Apolipoprotein B48. *J. Biol. Chem.* **271,** 9887–9890 (1996).

59. Anant, S., MacGinnitie, A. J. & Davidson, N. O. 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 (1995).

60. Nakamuta, M. *et al.* 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 (1996).

61. Smallwood, T. L. *et al.* High-resolution genetic mapping in the diversity outbred mouse population identifies Apobec1 as a candidate gene for atherosclerosis. *G3 Bethesda Md* **4,** 2353–2363 (2014).

62. Lellek, H. *et al.* Purification and Molecular Cloning of a Novel Essential Component of the Apolipoprotein B mRNA Editing Enzyme-Complex. *J. Biol. Chem.* **275,** 19848–19856 (2000).

63. Mehta, A. & Driscoll, D. M. Identification of domains in apobec-1 complementation factor required for RNA binding and apolipoprotein-B mRNA editing. *RNA* **8,** 69–82 (2002).

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

65. Chester, A. *et al.* The apolipoprotein B mRNA editing complex performs a multifunctional cycle and suppresses nonsense-mediated decay. *EMBO J.* **22,** 3971–3982 (2003).

66. Yamanaka, S., Poksay, K. S., Driscoll, D. M. & Innerarity, T. L. Hyperediting 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 (1996).

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

68. Sowden, M., Hamm, J. K. & Smith, H. C. Overexpression of APOBEC-1 Results in Mooring Sequence-dependent Promiscuous RNA Editing. *J. Biol. Chem.* **271,** 3011–3017 (1996).

69. Blanc, V. *et al.* Targeted Deletion of the Murine apobec-1 Complementation Factor (acf) Gene Results in Embryonic Lethality. *Mol. Cell. Biol.* **25,** 7260–7269 (2005).

70. Anant, S. & Davidson, N. O. 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 (2000).

71. Harris, R. S., Petersen-Mahrt, S. K. & Neuberger, M. S. RNA Editing Enzyme APOBEC1 and Some of Its Homologs Can Act as DNA Mutators. *Mol. Cell* **10,** 1247–1253 (2002).

72. Petersen-Mahrt, S. K. & Neuberger, M. S. In Vitro Deamination of Cytosine to Uracil in Single-stranded DNA by Apolipoprotein B Editing Complex Catalytic Subunit 1 (APOBEC1). *J. Biol. Chem.* **278,** 19583–19586 (2003).

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

74. Mukhopadhyay, D. *et al.* 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 (2002).

75. Yamanaka, S., Poksay, K. S., Arnold, K. S. & Innerarity, T. L. 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 (1997).

76. Nelson, V. R., Heaney, J. D., Tesar, P. J., Davidson, N. O. & Nadeau, J. H. Transgenerational epigenetic effects of the Apobec1 cytidine deaminase deficiency on testicular germ cell tumor susceptibility and embryonic viability. *Proc. Natl. Acad. Sci.* **109,** E2766–E2773 (2012).

77. Blanc, V. *et al.* Deletion of the AU-Rich RNA Binding Protein Apobec-1 Reduces Intestinal Tumor Burden in Apcmin Mice. *Cancer Res.* **67,** 8565–8573 (2007).

78. Absence of APOBEC-1 mediated mRNA editing in human carcinomas. *Publ. Online 17 Novemb. 1999 Doi101038sjonc1203039* **18,** (1999).

79. Fossat, N. *et al.* C to U RNA editing mediated by APOBEC1 requires RNA-binding protein RBM47. *EMBO Rep.* (2014). doi:10.15252/embr.201438450

80. Harjanto, D. *et al.* RNA editing generates cellular subsets with diverse sequence within populations. *Nat. Commun.* **7,** 12145 (2016).

81. Chen, L.-L. & Carmichael, G. G. in *Adenosine Deaminases Acting on RNA (ADARs) and A-to-I Editing* (ed. Samuel, C. E.) 111–121 (Springer Berlin Heidelberg, 2011). doi:10.1007/82_2011_149

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

83. Grimson, A. *et al.* MicroRNA Targeting Specificity in Mammals: Determinants beyond Seed Pairing. *Mol. Cell* **27,** 91–105 (2007).

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

85. Helwak, A. & Tollervey, D. Mapping the miRNA interactome by cross-linking ligation and sequencing of hybrids (CLASH). *Nat. Protoc.* **9,** 711–728 (2014).

86. Ong, S.-E. & Mann, M. A practical recipe for stable isotope labeling by amino acids in cell culture (SILAC). *Nat. Protoc.* **1,** 2650–2660 (2007).

87. Xiao, Z., Zou, Q., Liu, Y. & Yang, X. Genome-wide assessment of differential translations with ribosome profiling data. *Nat. Commun.* **7,** (2016).

88. Gommans, W. M., Mullen, S. P. & Maas, S. RNA editing: a driving force for adaptive evolution? *BioEssays* **31,** 1137–1145 (2009).

89. Pullirsch, D. & Jantsch, M. F. Proteome diversification by adenosine to inosine RNA-editing. *RNA Biol.* **7,** 205–212 (2010).

90. Shalek, A. K. *et al.* Single-cell transcriptomics reveals bimodality in expression and splicing in immune cells. *Nature* **498,** 236–240 (2013).

91. Wu, A. R. *et al.* Quantitative assessment of single-cell RNA-sequencing methods. *Nat. Methods* **11,** 41–46 (2014).

92. Tang, F. *et al.* RNA-Seq analysis to capture the transcriptome landscape of a single cell. *Nat. Protoc.* **5,** (2010).

93. Shalek, A. K. *et al.* Single-cell RNA-seq reveals dynamic paracrine control of cellular variation. *Nature* **510,** 363–369 (2014).

94. Mosser, D. M. & Edwards, J. P. Exploring the full spectrum of macrophage activation. *Nat. Rev. Immunol.* **8,** 958–969 (2008).

95. Mosser, D. M. The many faces of macrophage activation. *J. Leukoc. Biol.* **73,** 209–212 (2003).

96. Chen, L. *et al.* Recoding RNA editing of AZIN1 predisposes to hepatocellular carcinoma. *Nat. Med.* **19,** 209–216 (2013).

97. Lomeli, H. *et al.* Control of kinetic properties of AMPA receptor channels by nuclear RNA editing. *Science* **266,** 1709–1713 (1994).

98. Huynh, K. K. *et al.* LAMP proteins are required for fusion of lysosomes with phagosomes. *EMBO J.* **26,** 313–324 (2007).

99. Morrison, J. R. *et al.* Apolipoprotein B RNA editing enzyme-deficient mice are viable despite alterations in lipoprotein metabolism. *Proc. Natl. Acad. Sci. U. S. A.* **93,** 7154–7159 (1996).

100. Scholzen, T. & Gerdes, J. The Ki-67 protein: From the known and the unknown. *J. Cell. Physiol.* **182,** 311–322 (2000).

101. Liu, Q. *et al.* CXCR4 antagonist AMD3100 redistributes leukocytes from primary immune organs to secondary immune organs, lung, and blood in mice. *Eur. J. Immunol.* **45,** 1855–1867 (2015).

102. Zhang, Y. *et al.* An RNA-Sequencing Transcriptome and Splicing Database of Glia, Neurons, and Vascular Cells of the Cerebral Cortex. *J. Neurosci.* **34,** 11929–11947 (2014).

103. D'Agostino, P. M. *et al.* Viral-induced encephalitis initiates distinct and functional CD103+ CD11b+ brain dendritic cell populations within the olfactory bulb. *Proc. Natl. Acad. Sci. U. S. A.* **109,** 6175–6180 (2012).

104. Yang, W. *et al.* Modulation of microRNA processing and expression through RNA editing by ADAR deaminases. *Nat. Struct. Mol. Biol.* **13,** 13–21 (2006).

105. Blanc, V. *et al.* Genome-wide identification and functional analysis of Apobec-1-mediated C-to-U RNA editing in mouse small intestine and liver. *Genome Biol.* **15,** R79 (2014).

106. Saraconi, G., Severi, F., Sala, C., Mattiuz, G. & Conticello, S. G. The RNA editing enzyme APOBEC1 induces somatic mutations and a compatible mutational signature is present in esophageal adenocarcinomas. *Genome Biol.* **15,** (2014).

107. Severi, F. & Conticello, S. G. Flow-cytometric visualization of C>U mRNA editing reveals the dynamics of the process in live cells. *RNA Biol.* **12,** 389–397 (2015).

108. Arakawa, H., Saribasak, H. & Buerstedde, J.-M. Activation-induced cytidine deaminase initiates immunoglobulin gene conversion and hypermutation by a common intermediate. *PLoS Biol.* **2,** E179 (2004).

109. MacGinnitie, A. J., Anant, S. & Davidson, N. O. 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 (1995).

110. Selbach, M. *et al.* Widespread changes in protein synthesis induced by microRNAs. *Nature* **455,** 58–63 (2008).

111. Baek, D. *et al.* The impact of microRNAs on protein output. *Nature* **455,** 64–71 (2008).

112. Gu, S. & Kay, M. A. How do miRNAs mediate translational repression? *Silence* **1,** 11 (2010).

113. Pillai, R. S. *et al.* Inhibition of translational initiation by Let-7 MicroRNA in human cells. *Science* **309,** 1573–1576 (2005).

114. Bhattacharyya, S. N., Habermacher, R., Martine, U., Closs, E. I. & Filipowicz, W. Relief of microRNA-mediated translational repression in human cells subjected to stress. *Cell* **125,** 1111–1124 (2006).

115. Chendrimada, T. P. *et al.* MicroRNA silencing through RISC recruitment of eIF6. *Nature* **447,** 823–828 (2007).

116. Ding, X. C. & Großhans, H. Repression of C. elegans microRNA targets at the initiation level of translation requires GW182 proteins. *EMBO J.* **28,** 213–222 (2009).

117. Mathonnet, G. *et al.* MicroRNA inhibition of translation initiation in vitro by targeting the cap-binding complex eIF4F. *Science* **317,** 1764–1767 (2007).

118. Petersen, C. P., Bordeleau, M.-E., Pelletier, J. & Sharp, P. A. Short RNAs repress translation after initiation in mammalian cells. *Mol. Cell* **21,** 533–542 (2006).

119. Kiriakidou, M. *et al.* An mRNA m7G cap binding-like motif within human Ago2 represses translation. *Cell* **129,** 1141–1151 (2007).

120. Ingolia, N. T., Brar, G. A., Rouskin, S., McGeachy, A. M. & Weissman, J. S. The ribosome profiling strategy for monitoring translation in vivo by deep sequencing of ribosome-protected mRNA fragments. *Nat. Protoc.* **7,** 1534–1550 (2012).

121. Ingolia, N. T., Lareau, L. F. & Weissman, J. S. Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. *Cell* **147,** 789–802 (2011).

122. Tanguay, R. L. & Gallie, D. R. Translational efficiency is regulated by the length of the 3' untranslated region. *Mol. Cell. Biol.* **16,** 146–156 (1996).

123. Szostak, E. & Gebauer, F. Translational control by 3′-UTR-binding proteins. *Brief. Funct. Genomics* **12,** 58–65 (2013).

124. Santhanam, A. N. *et al.* Role of 3′UTRs in the Translation of mRNAs Regulated by Oncogenic eIF4E—A Computational Inference. *PLOS ONE* **4,** e4868 (2009).

125. Wilkie, G. S., Dickson, K. S. & Gray, N. K. Regulation of mRNA translation by 5′- and 3′-UTR-binding factors. *Trends Biochem. Sci.* **28,** 182–188 (2003).

126. Heiman, M., Kulicke, R., Fenster, R. J., Greengard, P. & Heintz, N. Cell-Type-Specific mRNA Purification by Translating Ribosome Affinity Purification (TRAP). *Nat. Protoc.* **9,** 1282–1291 (2014).

127. Loeb, G. B. *et al.* Transcriptome-wide miR-155 binding map reveals widespread noncanonical microRNA targeting. *Mol. Cell* **48,** 760–770 (2012).