

# Two Genetic Determinants Acquired Late in *Mus* Evolution Regulate the Inclusion of Exon 5, which Alters Mouse APOBEC3 Translation Efficiency

Jun Li<sup>1</sup>, Yoshiyuki Hakata<sup>1\*</sup>, Eri Takeda<sup>1‡</sup>, Qingping Liu<sup>2</sup>, Yasumasa Iwatani<sup>3</sup>, Christine A. Kozak<sup>2</sup>, Masaaki Miyazawa<sup>1\*</sup>

**1** Department of Immunology, Kinki University School of Medicine, Osaka, Japan, **2** Laboratory of Molecular Microbiology, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, Maryland, United States of America, **3** Department of Infection and Immunology, Clinical Research Center, Nagoya Medical Center, Nagoya, Japan

## Abstract

Mouse apolipoprotein B mRNA-editing enzyme catalytic polypeptide-like editing complex 3 (mA3), an intracellular antiviral factor, has 2 allelic variations that are linked with different susceptibilities to beta- and gammaretrovirus infections among various mouse strains. In virus-resistant C57BL/6 (B6) mice, mA3 transcripts are more abundant than those in susceptible BALB/c mice both in the spleen and bone marrow. These strains of mice also express mA3 transcripts with different splicing patterns: B6 mice preferentially express exon 5-deficient ( $\Delta 5$ ) mA3 mRNA, while BALB/c mice produce exon 5-containing full-length mA3 mRNA as the major transcript. Although the protein product of the  $\Delta 5$  mRNA exerts stronger antiretroviral activities than the full-length protein, how exon 5 affects mA3 antiviral activity, as well as the genetic mechanisms regulating exon 5 inclusion into the mA3 transcripts, remains largely uncharacterized. Here we show that mA3 exon 5 is indeed a functional element that influences protein synthesis at a post-transcriptional level. We further employed *in vitro* splicing assays using genomic DNA clones to identify two critical polymorphisms affecting the inclusion of exon 5 into mA3 transcripts: the number of TCCT repeats upstream of exon 5 and the single nucleotide polymorphism within exon 5 located 12 bases upstream of the exon 5/intron 5 boundary. Distribution of the above polymorphisms among different *Mus* species indicates that the inclusion of exon 5 into mA3 mRNA is a relatively recent event in the evolution of mice. The widespread geographic distribution of this exon 5-including genetic variant suggests that in some *Mus* populations the cost of maintaining an effective but mutagenic enzyme may outweigh its antiviral function.

**Citation:** Li J, Hakata Y, Takeda E, Liu Q, Iwatani Y, et al. (2012) Two Genetic Determinants Acquired Late in *Mus* Evolution Regulate the Inclusion of Exon 5, which Alters Mouse APOBEC3 Translation Efficiency. *PLoS Pathog* 8(1): e1002478. doi:10.1371/journal.ppat.1002478

**Editor:** Sara Sawyer, University of Texas at Austin, United States of America

**Received:** July 5, 2011; **Accepted:** November 26, 2011; **Published:** January 19, 2012

This is an open-access article, free of all copyright, and may be freely reproduced, distributed, transmitted, modified, built upon, or otherwise used by anyone for any lawful purpose. The work is made available under the Creative Commons CC0 public domain dedication.

**Funding:** This work was supported by grants-in-aid for Scientific Research from the Ministry of Education, Culture, Sports, Science and Technology of Japan, including the High-Tech Research Center and Anti-Aging Center projects, those from the Japan Society for Promotion of Sciences, KAKENHI (B), and those from the Ministry of Health, Labor and Welfare of Japan for Research on HIV/AIDS (all given to MM), and by a Grant-in-Aid for Young Scientists (B) from the Japan Society for the Promotion of Sciences, KAKENHI 21790450, given to YH. This work was also supported by the Intramural Research Program of the NIH, NIAID, given to CAK. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

\* E-mail: hakata@med.kindai.ac.jp (YH); masaaki@med.kindai.ac.jp (MM)

‡ Current address: Department of Emerging Infectious Diseases, Nagasaki University Institute of Tropical Medicine, Nagasaki, Japan

## Introduction

The family of apolipoprotein B mRNA-editing enzyme catalytic polypeptide-like editing complex 3 (APOBEC3) proteins consists of cytidine deaminases that function as cellular restriction factors against various exogenous and endogenous viruses [1–17]. Seven APOBEC3 paralogues have been identified on human chromosome 22, while only a single copy of the *Apobec3* gene is found in the mouse genome [10,18,19]. Among the human APOBEC3 enzymes, APOBEC3G (hA3G) is the best characterized member and is known to inhibit HIV-1 replication when the virus lacks the functional accessory protein, viral infectivity factor (Vif) [reviewed in 20]. In the absence of Vif, hA3G is incorporated into newly generated virions budding from virus-producing cells and exhibits its antiviral effect in subsequently infected cells. Thus, during reverse transcription in the target cells, the virion-incorporated hA3G catalyzes C-to-U deamination on the minus strand of

nascent viral DNA, resulting in G-to-A mutations on the plus strand of the double-stranded viral DNA, which can be detrimental to viral replication [7,9,10,21,22]. In addition, a deaminase-independent antiviral mechanism exerted by hA3G has also been reported [23,24].

In contrast to its human counterparts, mouse APOBEC3 (mA3) restricts HIV-1 regardless of the presence of Vif, as well as mouse mammary tumor virus (MMTV), ecotropic murine leukemia viruses (MuLVs), Friend MuLV (F-MuLV) and Moloney MuLV (M-MuLV), along with endogenous mouse retroviruses including the AKR ecotropic virus (AKV) [5,25–30]. This suggests that APOBEC3 enzymes protect host genomes from the retroviruses they commonly encounter, although some retroviruses, like HIV-1, have evolved to counter the intracellular restriction mechanisms of their natural hosts.

Friend virus (FV) is an acutely leukemogenic retroviral complex composed of replication-competent F-MuLV and replication

## Author Summary

Susceptibility to acutely leukemogenic Friend virus (FV) retrovirus infection varies among different mouse strains and is governed by several genetic factors, one of which is allelic variations at the mouse *Apobec3* locus. FV-resistant C57BL/6 (B6) mice express higher amounts of *Apobec3* transcripts than susceptible BALB/c mice. We previously showed that the differences in N-terminal amino acid sequences between B6 and BALB/c APOBEC3 proteins partly account for the distinct antiretroviral activities. In addition, B6 and BALB/c mice express major *Apobec3* transcripts of different sizes: the exon 5-lacking and the full-length transcripts, respectively. Here we asked if exon 5 has any role in the antiviral activity of mouse APOBEC3 and found that the presence of this exon resulted in a profound decrease in the efficiency of protein synthesis without affecting the mRNA expression levels. We also identified two genomic polymorphisms that control the inclusion of exon 5 into the *Apobec3* message: the number of TCCT repeats in intron 4 and a single nucleotide polymorphism within exon 5. The distribution of these functional polymorphisms among *Mus* species and wild mouse populations indicates that the exon 5 inclusion occurred recently in *Mus* evolution, and the full-length variant may have selective advantages in some mouse populations.

-defective spleen focus-forming virus (SFFV). Susceptibilities to FV-induced disease development differ among various inbred strains of mice, and these are controlled by several host factors that either directly affect FV replication or influence host immune responses to the viral antigens [31,32]. We and others have reported that the mouse *Apobec3* locus is polymorphic, and its genotypes are associated with the levels of viremia after F-MuLV or FV inoculation [26,27]. Mice of the prototypic FV-resistant strains C57BL/6 (B6) and C57BL/10 exhibit restricted replication of F-MuLV and earlier production of FV-neutralizing antibodies, while FV-susceptible BALB/c and A strains are less restrictive of F-MuLV replication and show delayed production of neutralizing antibodies, all of which are linked with *Apobec3* genotypes [26,27,33,34], although the production of neutralizing antibodies is also influenced by genotypes at the major histocompatibility complex and the *Tnfrsf13c* loci, the latter of which encode the receptor for B-cell activating factor belonging to the tumor necrosis factor family [31–33,35].

Mouse APOBEC3 and hA3G contain two cytidine deaminase domains (CDDs), each harboring the conserved zinc-coordinating motif; however, deaminase activity is exerted only by the N-terminal CDD of mA3 and the C-terminal CDD of hA3G [36]. We showed that the increased efficiency of B6 mA3 in inhibiting F-MuLV replication is associated with differences in the primary amino acid sequence within the active N-terminal CDD [27]; the functional importance of these residues was further implicated by the demonstration that they have been under positive selection in *Mus* [37]. In addition to the above sequence differences in the protein-coding regions, efficient virus restriction is also associated with higher levels of mA3 transcripts in FV-resistant B6 than in -susceptible BALB/c mice [27,29,38,39]. This enhanced transcription was linked with the presence of the long terminal repeat (LTR) of an endogenized xenotropic MuLV in the B6, but not in the BALB/c, *Apobec3* locus [37]. A third factor associated with virus resistance is the presence or absence of exon 5, which encodes a 33-amino acid segment separating the C-terminal and N-terminal CDDs [27]. The mA3 isoform of the FV-resistant

strains of mice lacks exon 5, while the predominant transcript in FV-susceptible mice is the exon 5-containing isoform [27,29,37]. Genetic sequences controlling mA3 splicing have not been identified, although it has been pointed out that polymorphisms between the B6 and BALB/c *Apobec3* alleles at the end of intron 4 include a putative splice acceptor site and possible mRNA branch selection site structures [38]. It has also not been shown whether the well-confirmed differences in transcript levels result in altered expression levels of mA3 protein in FV-resistant and -susceptible strains of mice.

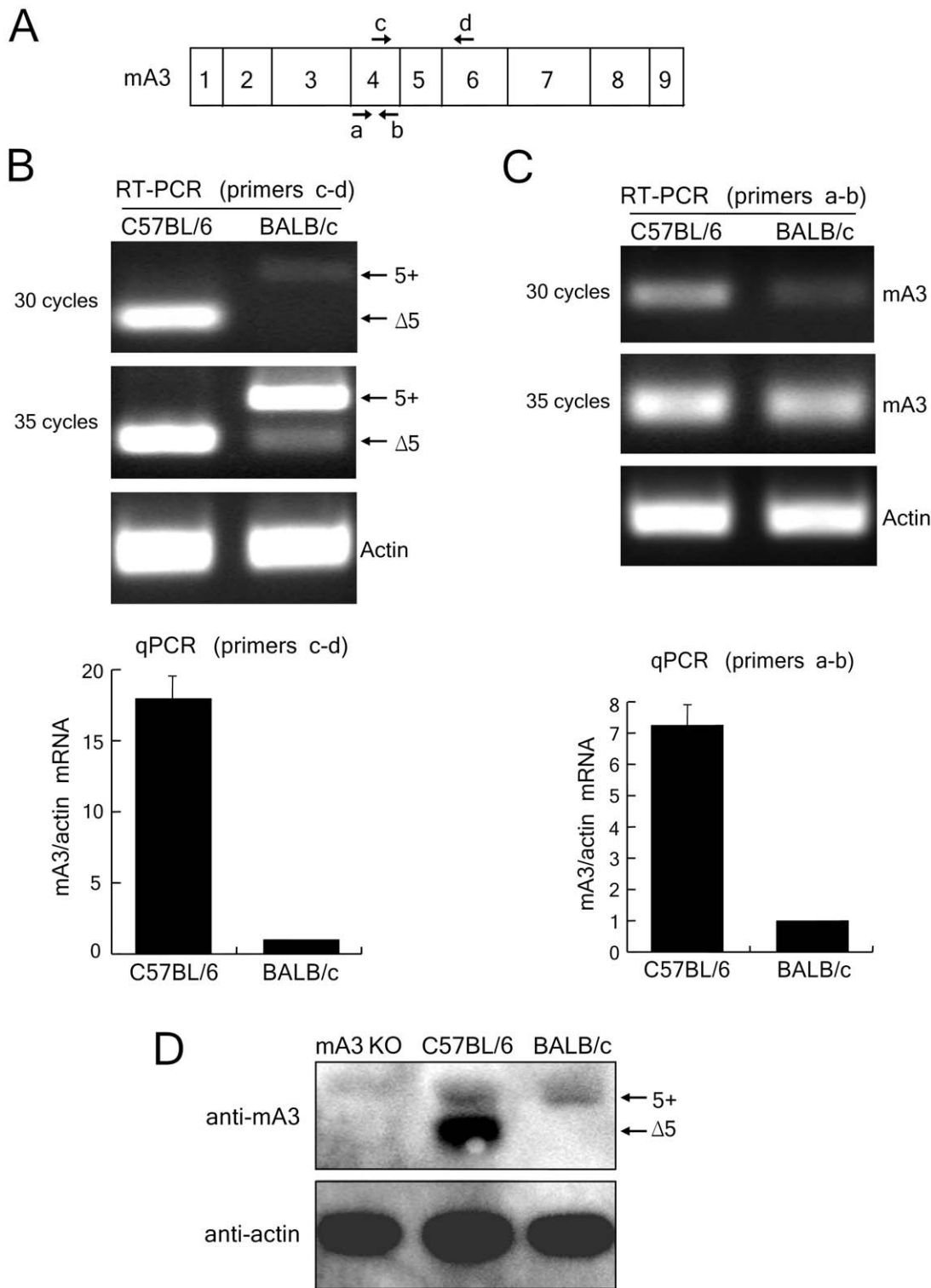
In the present report, we show that mA3 protein is indeed more abundant in B6 than in BALB/c mice. This difference is due in part to more efficient translation of the exon 5-deficient message. We further show extensive functional evidence that two distinct polymorphisms within the *Apobec3* locus regulate exon 5 inclusion during its splicing: the previously predicted [38] TCCT repeat numbers in intron 4 and a newly identified single nucleotide polymorphism (SNP) within exon 5. We also describe the linkage between these splicing regulatory sequences in wild mouse species, their acquisition in *Mus* evolution, and their distribution in wild mouse populations.

## Results

### APOBEC3 expression is higher in B6 than in BALB/c spleens at both transcriptional and protein levels

It has been reported that mA3 mRNA expression is higher in B6 than in BALB/c mice, and the mA3 transcripts detected in B6 mice are predominantly the exon 5-lacking  $\Delta 5$  isoform, while the majority of mA3 transcripts in BALB/c mice contain exon 5 (5+) [27,29,38]. To more accurately describe the above quantitative differences in mA3 transcripts, we performed two types of PCR analyses with two different sets of specific primers (Figure 1A). The primer set c–d, which is the same as the one previously used [27], detected both the 5+ and  $\Delta 5$  transcripts in reverse-transcription PCR (RT-PCR) assays; however, mA3 transcripts were barely detectable after 30 cycles of amplification in BALB/c mice while the  $\Delta 5$  transcript was readily detectable in B6 mice (Figure 1B). After 35 cycles of amplification, both the 5+ and  $\Delta 5$  transcripts became detectable in BALB/c mice, although the 5+ mRNA was more abundant. Quantitative real-time PCR (qPCR) assays revealed mA3 expression levels that were approximately 18-times higher in B6 than in BALB/c spleens, although the difference in transcript lengths might have influenced the efficiencies of amplification in the real-time PCR reactions. To more precisely quantify the mA3 transcripts, we utilized the second primer set, a–b, which generates amplicons of the same size from both alleles (Figure 1A). The qPCR assay performed by using this latter primer set clearly demonstrated that B6 mice expressed >7-times higher amounts of mA3 transcript than BALB/c mice did (Figure 1C), consistent with the previous report [27]. The normalization of mA3 mRNA levels with TATA box-binding protein or GAPDH transcripts instead of  $\beta$ -actin gave similar results (data not shown).

Protein levels of mA3 in the above prototypic FV-resistant and -susceptible strains of mice were also compared. The spleen lysate from the mA3-knockout mice [40] was used as a negative control. In B6 spleens, an immunoreactive protein corresponding to  $\Delta 5$  mA3 was detected as a prominent band, but the higher molecular weight 5+ mA3 was also faintly detectable (Figure 1D), even though the 5+ message was barely detectable by RT-PCR in the present study (Figure 1B) and in the previous reports [27,29]. On the other hand, only the 5+ mA3 protein was detected, with a much lower intensity, in BALB/c spleens. The immunoblotting assays thus demonstrated that the level of total mA3 protein



**Figure 1. Expression of mA3 transcripts and protein in the spleens of B6 and BALB/c mice.** (A) Schematic representation of the mouse APOBEC3 exon composition. The exons are indicated by numbers and their relative sizes are depicted. The arrows indicate the locations of the PCR primers used in the following experiments. (B and C) RT-PCR and quantitative real-time PCR (qPCR) analyses showing splicing patterns and quantities of mA3 mRNA in B6 and BALB/c spleen cells. The primers utilized were c-d for assays shown in (B) and a-b for those in (C). 5+, exon 5-containing mA3; Δ5, mA3 lacking exon 5. Actin was used as an internal control. Quantitative real-time PCR data show averages of three reaction wells and SD. (D) Immunoblot detection of mA3 protein expressed in mouse spleen cells. The spleen cell extracts were tested for their total protein content by Bradford assays and the same amount of protein was loaded in each lane. Anti-mouse APOBEC3 NT antibody reactive with an N-terminal epitope was first pre-absorbed with the spleen lysate prepared from an mA3-knockout (KO) mouse, and was used as the primary antibody. Mouse actin was used as a loading control. Three independent experiments were done for all assays, and the results were consistent with the representative ones shown here.

doi:10.1371/journal.ppat.1002478.g001

expression in the spleens of B6 mice is much higher than that in BALB/c mice.

### Effects of exon 5 inclusion on mA3 protein expression

The mA3 expression data shown in Figure 1 indicate that levels of mA3 protein expression roughly correlate with the corresponding transcript levels in mice with different *Apobec3* alleles. However, we previously observed that even when FLAG-tagged mA3 was expressed under the control of the cytomegalovirus (CMV) promoter, the levels of expression of the 5+ mA3 protein tended to be much lower than those of the  $\Delta 5$  protein, despite similar levels of mRNA expression in transfected cells [27]. This implies that exon 5 might affect either the efficiency of mA3 translation or protein stability. To evaluate these possibilities, 5+ and  $\Delta 5$  mA3 expression plasmids were constructed with either B6- or BALB/c-derived mA3 cDNA (Figure 2A). The mA3 cDNA were fused with a FLAG tag at their N-terminus and the transcription was driven by the CMV promoter. After transfection of 293T cells with either one of the FLAG-tagged expression constructs, we quantified mA3 transcripts by qPCR and mA3 protein by immunoblotting (Figure 2, B–D). Based on the co-expressed luciferase activities, similar transfection efficiencies for the 5+ and  $\Delta 5$  plasmids were confirmed, and the data from the qPCR showed that the 5+ mA3 transcript was more highly expressed than the  $\Delta 5$  mRNA for both alleles. Nevertheless, the expression levels of the 5+ mA3 protein were lower compared to those of the  $\Delta 5$  counterpart regardless of the allelic differences, although B6  $\Delta 5$  protein was expressed more abundantly than BALB/c  $\Delta 5$  (Figure 2, B and C). These results indicate that the inclusion of exon 5 does not compromise mA3 mRNA expression, but decreases the steady-state levels of mA3 proteins for both allelic variants.

### Exon 5 inclusion does not affect mA3 protein stability but influences its translation

As higher mRNA levels of the 5+ mA3 resulted in lower protein levels in comparison with those for the  $\Delta 5$  isoform in transfected cells, we next examined the steps at which the synthesis or degradation of mA3 protein was affected by the presence of exon 5. To this end, we transfected 293T cells with an expression plasmid harboring the 5+ or  $\Delta 5$  mA3 cDNA, and cycloheximide was added to stop protein synthesis. The inhibitor of nuclear factor- $\kappa$ B, I $\kappa$ B $\alpha$ , was utilized as a control, as this protein is known to be in the process of constant degradation and regeneration under serum-containing culture conditions [41]. As expected, immunodetectable I $\kappa$ B $\alpha$  decreased upon cycloheximide treatment, while the solvent alone did not affect the protein content (Figure 3A). On the other hand, mA3 expressed in the same cells did not exhibit any reduction upon cycloheximide treatment regardless of the presence or absence of exon 5. Further, the  $\Delta 5$  protein was again detected at higher levels than the 5+ protein. These results support the conclusion that the effect of exon 5 inclusion in reducing the amount of expressed mA3 protein is not due to accelerated protein degradation.

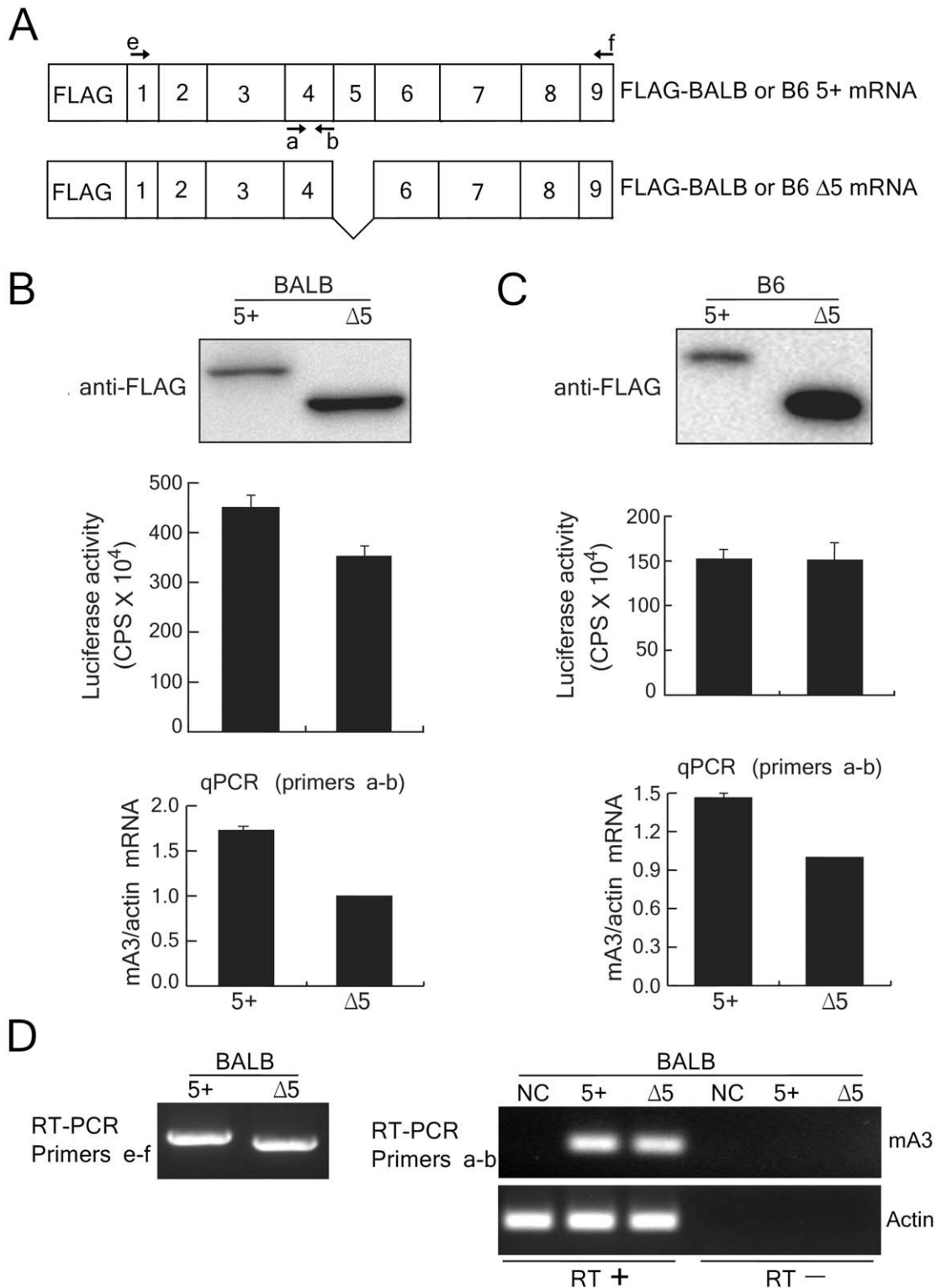
We next utilized an *in vitro* transcription and translation procedure to determine if exon 5 affects the translation of mA3 (Figure 3B). When the 5+ or  $\Delta 5$  mA3 templates were added to the *in vitro* transcription and translation reaction, similar amounts of each transcript were detected by RT-PCR assays both at 30 min and 60 min after the beginning of incubation. However, the tempos of appearance and amounts of the 5+ mA3 protein were different from those of the  $\Delta 5$  counterpart: a large amount of the  $\Delta 5$  protein was detected at as early as 30 min after the beginning of incubation, while the 5+ mA3 was undetectable at the same time-point. The 5+ mA3 protein became detectable after 60 min

of incubation, but its protein level was still markedly lower than that of the  $\Delta 5$  counterpart. These results collectively indicate that the inclusion of exon 5 modulates the translation efficiency of mA3 rather than its protein degradation.

### The TCCT repeat number variation upstream of exon 5 is partially responsible for exon 5 inclusion into the mA3 transcript

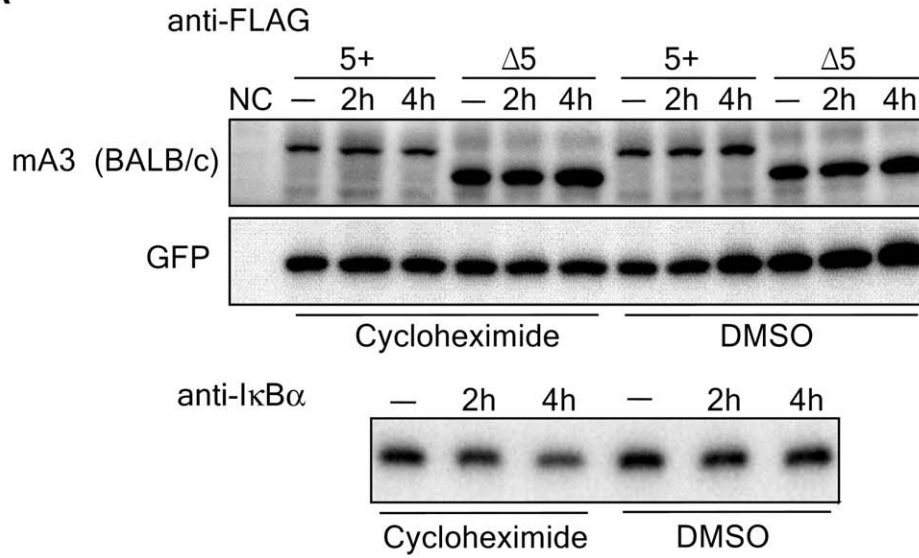
As the inclusion of exon 5 is associated with a reduced level of mA3 protein, we next attempted to identify genetic polymorphisms that affect the splicing patterns of mA3 transcripts in terms of exon 5 inclusion. One possible allelic difference in the mouse *Apobec3* locus putatively associated with its splicing patterns is the possible pre-mRNA branch site polymorphisms found in the intron upstream of exon 5: a T/C SNP that lies within a preferred branch site sequence, TA(T/C)CAAC, and TCCT repeat numbers between this and the acceptor site [38]. As described previously [38], the intron 4 of the BALB/c allele contains a tandem repeat of the TCCT sequence near the intron 4/exon 5 boundary, while the B6 allele contains only a single TCCT copy, changing the length of the putative pyrimidine-rich lariat intermediate. The adjacent T/C SNP at position 741 from the first nucleotide of exon 4 is in linkage with the TCCT repeat number. To directly investigate the possible effect of these polymorphisms, we constructed splicing assay plasmids that harbor the *Apobec3* genomic fragment encompassing exons 4 through 7 from either the B6 or BALB/c allele. The resultant plasmids were designated B6 exon 4–7 and BALB exon 4–7 (Figure 4). Three mutants of each of these constructs were made with different combinations of the position 741 T/C SNP and the TCCT copy number variation as depicted in Figure 4B.

Because of the possible presence of species-specific regulatory factors, the resultant genomic constructs were transfected into BALB/3T3 instead of 293T cells along with the luciferase expression plasmid as a control for transfection efficiency. To avoid the amplification of the endogenous mA3 message expressed in BALB/3T3 cells [27], we utilized the primers g and h (Figure 4B) for RT-PCR assays, which were designed to hybridize to the T7 promoter and V5 tag regions of the expression vector. Transfection with the B6 exon 4–7 plasmid resulted in the expression of only the  $\Delta 5$  transcript, while the BALB exon 4–7 generated both the 5+ and  $\Delta 5$  transcripts with much higher intensity of the 5+ one (Figure 4C), reproducing the splicing patterns observed in B6 and BALB/c spleens, respectively (Figure 1B). The ratio between the 5+ and  $\Delta 5$  transcripts was reduced in the samples transfected with BALB  $\Delta$ TCCT or BALB C741T  $\Delta$ TCCT plasmid harboring only a single copy of TCCT, which was further confirmed by utilizing the primer i hybridizing to the sequence within exon 5 along with primer h (Figure 4C). Quantitative real-time PCR analyses were done by utilizing another primer (primer j), designed to hybridize with the sequence within exon 5, and further confirmed reduced levels of the exon 5-containing message expressed from the modified BALB exon 4–7 constructs lacking the TCCT repeat. These results imply that the TCCT repeat observed in the BALB/c allele, but not the T/C substitution at position 741, at least partly facilitates the exon 5 inclusion. The lack of an effect of the T/C SNP within the putative branch site sequence was further confirmed by the abundant expression of the 5+ message from the BALB C741T construct. However, both the 5+ and  $\Delta 5$  transcripts were still produced from the BALB  $\Delta$ TCCT plasmid, suggesting that the TCCT repeat number is not the only determinant controlling the exon 5 inclusion. In fact, despite the presence of the repeated TCCT, B6 +TCCT and B6 T741C +TCCT did not generate the 5+ message,

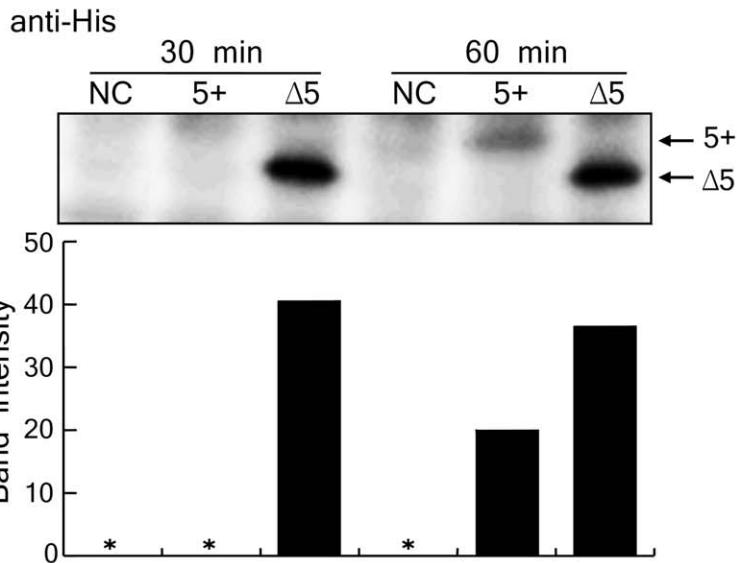
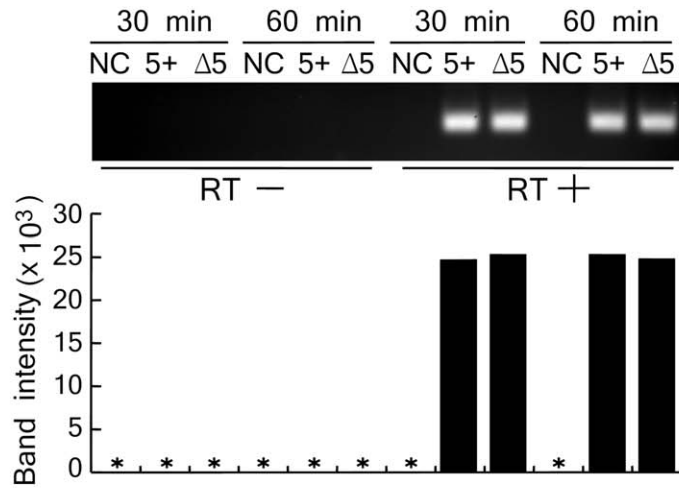


**Figure 2. Protein expression of the full-length and  $\Delta 5$  mA3 in transiently transfected 293T cells.** (A) Exon 5-containing (5+) and  $\Delta 5$  mA3 cDNA were tagged with the FLAG epitope and inserted into the expression vector. The arrows indicate the positions of the PCR primers. The primer set a-b is the same as shown in Figure 1. (B and C) 293T cells were transfected with pFLAG-CMV2-*mA3*<sup>5+</sup> or pFLAG-CMV2-*mA3* <sup>$\Delta 5$</sup>  in B or with pFLAG-CMV2-*mA3*<sup>5+</sup> or pFLAG-CMV2-*mA3* <sup>$\Delta 5$</sup>  in C, which express either the 5+ or  $\Delta 5$  mA3 cDNA cloned from BALB/c or B6 mice, respectively [27]. A luciferase-expressing plasmid, *pluc*, was co-transfected to standardize transfection efficiencies. At 24 hours after transfection, each one-third of the transfected cells was used for immunoblotting, RNA extraction, and luciferase assays. For immunoblotting, the full-length and  $\Delta 5$  mA3 proteins were detected with the anti-FLAG antibody. Quantitative real-time PCR reactions were carried out with primer set a-b and were normalized with the levels of actin transcripts expressed in 293T cells. Data shown are averages of three reaction wells and SD. (D) RT-PCR assays were performed with (RT+) or without (RT-) reverse transcription to demonstrate specific detection of expected mRNA. Both primer sets e-f and a-b were used to detect mA3 mRNA. NC, negative control transfected with the empty vector, pFLAG-CMV2. RT- samples were included to evaluate the possible contamination of transfected DNA. Specific amplification of cDNA generated in the presence of RT was observed in the cells transfected with a plasmid expressing the 5+ or  $\Delta 5$  mA3 cDNA. Similar results were obtained for the corresponding B6-derived cDNA clones. doi:10.1371/journal.ppat.1002478.g002

**A**



**B**



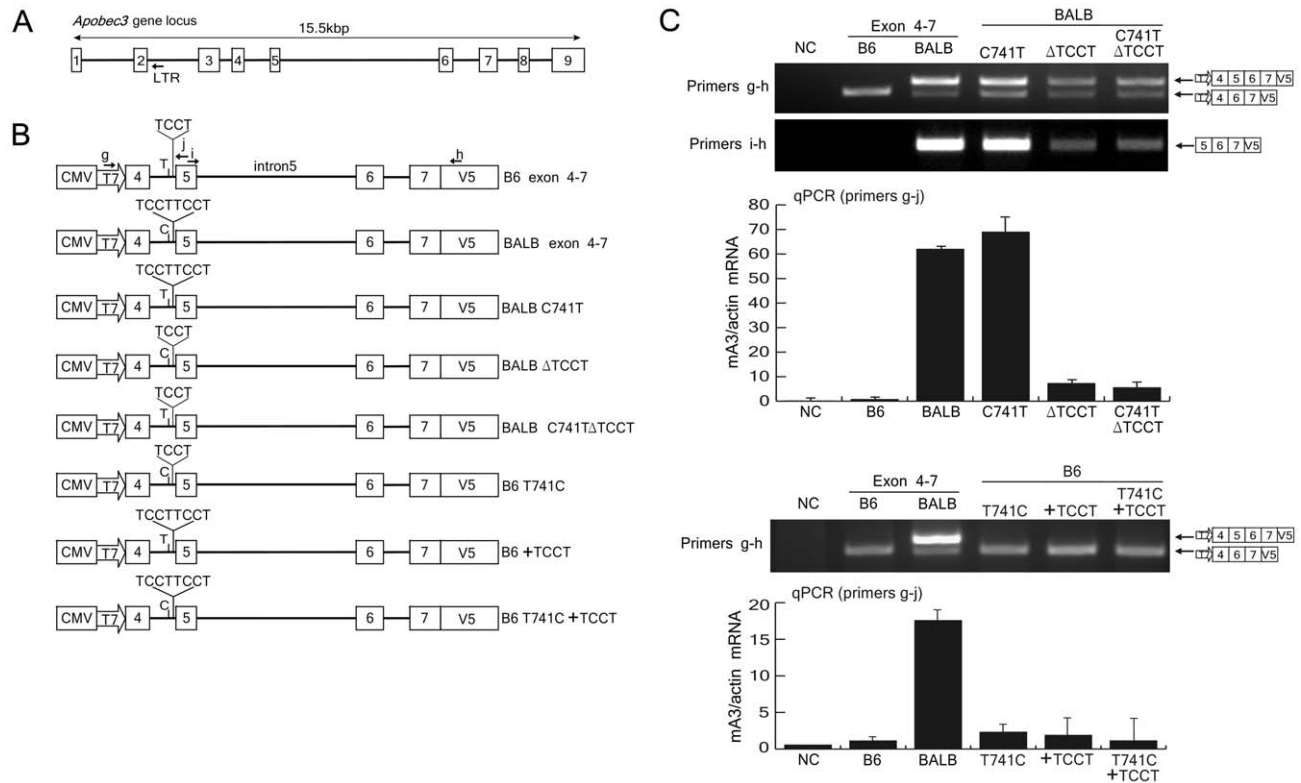
**Figure 3. The presence of exon 5 does not affect protein degradation but exhibits a profound impact on mA3 protein synthesis.** (A) 293T cells were transfected with pFLAG-CMV2-*mA3<sup>Δ5</sup>* or pFLAG-CMV2-*mA3<sup>Δ5</sup>* along with pFLAG-CMV2-GFP, which expresses green fluorescent protein (GFP) as a loading control. The cells were treated with cycloheximide for the indicated duration to stop protein synthesis and then harvested. DMSO was used as a solvent control. The FLAG-tagged mA3 and GFP were detected with the anti-FLAG antibody in immunoblotting. Endogenous IκBα expression was used as a positive control to confirm the effect of cycloheximide. (B) *In vitro* transcription and translation assays. DNA templates containing the entire coding region of the BALB/c full-length or  $\Delta 5$  mA3 cDNA with the T7 promoter and His-tag sequence were subjected to an *in vitro* transcription/translation reaction by incubating for 30 or 60 min. The levels of mA3 protein synthesis and mRNA expression were evaluated by immunoblotting using the anti-His antibody and RT-PCR using the primer set a-b, respectively. Intensities of protein bands on the immunoblot membrane and DNA bands after the RT-PCR reaction and electrophoresis were measured by densitometry and are shown below each corresponding band. \*, signals below detection limits. doi:10.1371/journal.ppat.1002478.g003

while transfection with the corresponding BALB C741T and the wild-type BALB exon 4–7 clearly resulted in the generation of the 5+ mRNA. These results indicate that polymorphisms other than those in intron 4 might play more important roles in determining exon 5 inclusion into mA3 messages.

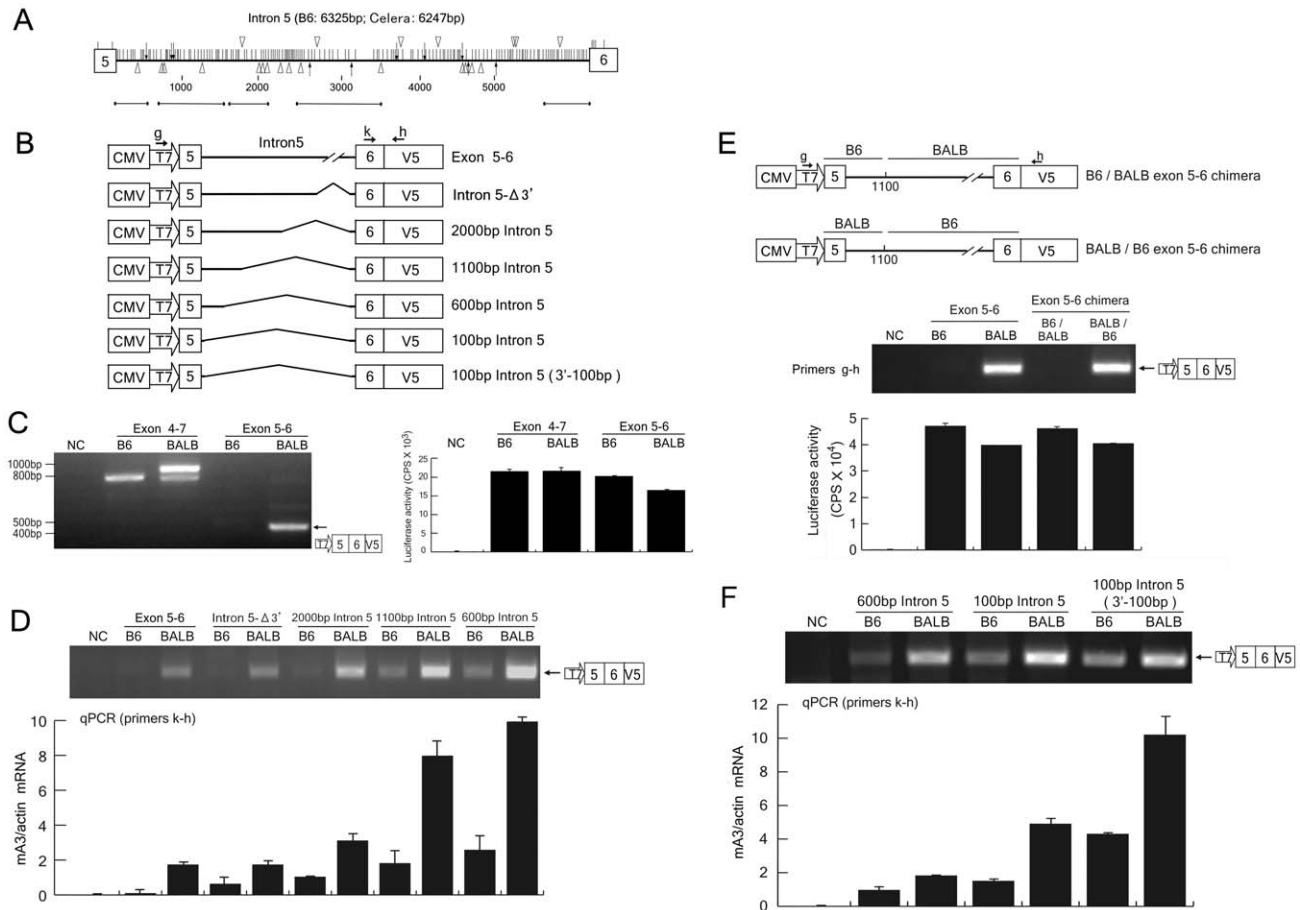
**A region around the exon 5/intron 5 boundary contains a major determinant for exon 5 inclusion**

As abundant expression of the exon 5-containing message was found with the BALB C741T but not with the B6 +TCCT construct, we were prompted to examine other polymorphisms

downstream of intron 4 that might be involved in the splicing of exon 5. To explore other possible sequence variations required for exon 5 inclusion, we focused on intron 5, since this region contains numerous polymorphisms between reported genomic sequences of different mouse strains (Figure 5A). Splicing assay vectors harboring the genomic DNA fragment containing exons 5 and 6 along with the entire intron 5, from either the B6 or the BALB/c allele, were constructed and designated B6 exon 5–6 or BALB exon 5–6 (Figure 5B). BALB/3T3 cells transfected with the BALB exon 5–6 plasmid generated the properly spliced product of the expected size (Figure 5C), indicating that the splice donor site



**Figure 4. The impact of TCCT repeat numbers on exon 5 inclusion in mA3 mRNA splicing.** (A) Schematic representation of the genomic DNA structure of the *Apobec3* gene locus. Boxes indicate exons. LTR, the endogenous retroviral LTR inserted into intron 2 in some strains of mice [37]. (B) Insert structures of the plasmids used for the splicing assays. BALB/c exon 4–7 and BALB/c  $\Delta$ TCCT plasmids harbored the same DNA fragment amplified from BALB/c genomic DNA encompassing exons 4 and 7, except that the BALB/c  $\Delta$ TCCT contains only a single TCCT quadruplet. B6 exon 4–7 plasmid harbored the B6 genomic DNA fragment encompassing exons 4 and 7. C741T and T741C indicate a C to T or reciprocal nucleotide substitution, respectively, within intron 4 at 741-bp downstream from the first nucleotide of exon 4 on the backbone of the BALB/c or B6 exon 4–7 insert. An additional repeat of the TCCT quadruplet was added to the B6 exon 4–6 construct to generate B6 +TCCT. The positions of primers g, h, i, and j used for RT-PCR assays are indicated with the arrows. (C) The plasmid harboring each insert depicted in (B) was transfected into BALB/3T3 cells, total RNA was extracted, and RT-PCR reactions were performed with primer pairs g–h and i–h. A portion of the transfected cells were utilized for luciferase assays to compare transfection efficiencies among the samples. Comparable levels of luciferase activities were observed in all samples in each experiment (data not shown). The predicted splicing products are schematically indicated on the right side of the panel. Quantitative real-time PCR assays were performed with primers g and j, and data are shown here by averages of three reaction wells and SD. doi:10.1371/journal.ppat.1002478.g004



**Figure 5. The effect of exon 5 and its downstream sequences on mA3 intron 5 splicing.** (A) Distributions of sequence polymorphisms in the intron 5 of the *ApoBec3* gene locus between the B6 allele [NT\_039621] and the Celera database sequence of mixed mouse DNA [NW\_001030577.1]. Shorter horizontal lines below the thick one representing intron 5 indicate regions of sequenced BALB/c genome [DBJ accession No. AB646261-AB646265], which show nucleotide sequences identical to corresponding Celera database sequences. Spans of the analyzed regions are 1–620, 661–1597, 1643–2058, 2759–3443, and 5558–6247 in base numbers starting from the first nucleotide of intron 5. SNP are shown with vertical lines, single-base indels with arrows, and deletions of  $\geq 2$  bases with triangles. Indels and deletions above the thick horizontal line that represents intron 5 are deletions in the B6 allele relative to the Celera sequence, while those underneath the horizontal line are deletions in the Celera sequence relative to the B6 allele sequence. (B) Plasmid constructions for the splicing assays. The exon 5–6 plasmids harbored either the B6 or BALB/c genomic fragment encompassing exons 5 and 6, including the entire intron 5 of the corresponding allele. Each of the remaining plasmids possessed a sequentially reduced length of the intron 5 as indicated. The precise size of each PCR-generated 5' fragment included was as follows: 3177bp and 3185bp for B6 and BALB/c intron 5- $\Delta 3'$ ; 2106bp and 2086bp for B6 and BALB/c 200bp intron 5; 1120bp and 1110bp for B6 and BALB/c 1100bp intron 5; and 634bp and 620bp for B6 and BALB/c 600bp intron 5, respectively. The primers g–h are the same as shown in Figure 4. (C) RT-PCR detection of spliced messages expressed from the B6 and BALB/c exon 5–6 plasmids along with those from the exon 4–7 plasmids as controls. Primers g and h were used. A lack of expression of the spliced message in cells transfected with the B6 exon 5–6 plasmid was evident. (D and F) Splicing assays using intron 5 deletion plasmids. The intron 5 fragment included in each plasmid is shown in (B). RT-PCR assays were performed with primers g and h. A portion of the transfected cells were utilized for luciferase assays to compare transfection efficiencies among the samples as shown in (C). Comparable levels of luciferase activities were observed in all samples in each experiment (data not shown). Quantitative real-time PCR data show averages of three reaction wells and SD. (E) Reciprocal chimeras were produced between B6 and BALB/c exon 5–6 by exchanging the cloned genomic DNA fragment at position 1100 within intron 5. The exact location of the above position 1100 for B6 and BALB/c intron 5 is described in the legend for (B). RT-PCR detection of spliced messages was performed with primers g and h. A portion of the transfected cells were utilized for luciferase assays to compare transfection efficiencies.  
doi:10.1371/journal.ppat.1002478.g005

downstream of exon 5 and the acceptor site upstream of exon 6 in the BALB/c allele are both functional. In contrast, no spliced product was detectable in cells transfected with the B6 exon 5–6 plasmid despite a higher transfection efficiency. These results indicate that the B6 allele may not carry a functional splice donor site at the exon 5/intron 5 boundary, as the splice acceptor site upstream of exon 6 seems intact and thus can generate a message corresponding to the mA3  $\Delta 5$  when exon 4 and intron 4 are included (Figures 4 and 5C).

To narrow down the region affecting the intron 5 splicing, serial deletion constructs were produced and subjected to the splicing

assay (Figure 5B). The intron 5- $\Delta 3'$  is a deletion construct which lacks the 3' half of intron 5 but retains the 261-bp sequence adjacent to exon 6 to include the putative acceptor site. The cells transfected with the BALB intron 5- $\Delta 3'$  plasmid produced the spliced message as efficiently as those transfected with the parental BALB exon 5–6 construct; however, the B6 intron 5- $\Delta 3'$  plasmid generated a barely detectable band representing spliced message (Figure 5D). Quantitative real-time PCR assays revealed markedly lower levels of properly spliced message generated from the B6 intron 5- $\Delta 3'$  plasmid compared to that generated from the BALB/c counterpart. These data show that the 5' half of intron 5 is



responsible for the observed differences in splice site functions between the B6 and BALB/c alleles. Thus, further deletions progressively closer to exon 5 were introduced into the intron 5 fragment. As expected, the shorter the included intron 5 fragment, the more abundant the spliced product generated was, regardless of the expressed *ApoBec3* alleles, indicating that the intron length of the primary transcript influences the splicing efficiency (Figure 5D). Nevertheless, the spliced product was generated much more efficiently from the BALB/c allele than from the B6 allele with all examined construct pairs, despite comparable levels of transfection efficiency, indicating that each shorter intron fragment still retained the polymorphism responsible for the inefficient splicing of the B6 mRNA.

To exclude the possibility that the 3' region of intron 5 and/or exon 6 from the B6 allele might harbor inhibitory sequences that interfere with splicing, we constructed chimeras between the BALB and B6 exon 5–6 plasmids by exchanging the corresponding genomic DNA fragments at approximately position 1100 within intron 5 (Figure 5E). BALB/B6 chimera harboring the 5' donor site sequence from the BALB/c allele and 3' acceptor site from the B6 allele generated properly spliced mRNA as efficiently as the BALB exon 5–6 plasmid did, while the reciprocal construct harboring the B6 donor and BALB/c acceptor sequences did not (Figure 5E), indicating that there are no inhibitory elements in the fragment harboring the 3' intron 5 and exon 6 from the B6 allele. Thus, we continued to narrow down the sequence affecting the intron 5 splicing toward the 5' end of the genomic constructs.

Even the shortest construct pair, B6 100bp intron 5 and BALB 100bp intron 5, which harbor the 5' intron 5 fragment a mere 100-bp from the exon 5 boundary, still exhibited a readily discernible difference in the amounts of the spliced messages, and this was also true for another pair of deletion constructs that harbor a 100-bp acceptor region fragment of 3' intron 5 (Figure 5F). These results clearly indicate that the region including exon 5 and the 5' 100-bp of intron 5 carries the primary determinants responsible for the different efficiencies in splicing of intron 5 shown by the B6 and BALB/c alleles.

### A single nucleotide polymorphism within exon 5 is primarily responsible for exon 5 inclusion

Within the above narrowed-down region of exon 5 and the 5' 100 bp of intron 5, there are only 4 SNPs between the B6 and BALB/c alleles, as previously described [38]: the T/C SNP at 14-bp downstream from the first nucleotide of exon 5, G/C at 88-bp downstream from the same first nucleotide, G/C at 153-bp downstream within intron 5, and A/C at 163-bp downstream within intron 5 (Figure 6A). We therefore produced a series of point mutants based on the B6 intron 5–100 (3'-100bp) and BALB/c intron 5–100 (3'-100bp) constructs to precisely identify the critical SNP responsible for the functionalities of the splice donor site. In the first set of experiments, shown in Figure 6B, each indicated nucleotide within the BALB/c genomic sequence was substituted with the corresponding nucleotide found in the B6 allele. The substitution of the C at position 14 to the B6-type T slightly reduced the amount of the spliced message; more importantly, however, the similar substitution of C at position 88 to the B6-type G totally abrogated the intron 5 splicing. Simultaneous substitutions of C to G at position 153 and G to A at 163 did not affect the generation of the spliced message. Transfection efficiencies were essentially equivalent for all samples and even higher than those for the C88G construct as confirmed by cotransfecting the luciferase expression plasmid (data not shown). These results indicate that the G/C SNP at position 88 in exon 5 is most critical for the splice donor function.

In a reciprocal set of experiments, we attempted to alter splicing of the transcript of the B6 allele by replacing the polymorphic nucleotides with the BALB/c-types (Figure 6C). When all 4 SNPs in the B6 intron 5–100 fragment were replaced with those corresponding to the BALB/c allele, abundant generation of the spliced message was observed (Figure 6C). A combined substitution of T to C and G to C at positions 14 and 88, respectively, also resulted in efficient splicing of intron 5, and a single substitution of T to C at position 14 alone slightly increased the generation of the spliced message from the plasmid harboring the truncated intron. Importantly, a single substitution of G to C at position 88 resulted in as abundant expression of the spliced message as that observed with the BALB/c allele. Thus, these results collectively indicate that the G at position 88 is critical for the exclusion of exon 5, although the C at position 14 may also play a minor role in the inclusion of this exon.

### Site-directed mutagenesis of the plasmids harboring the entire intron 5

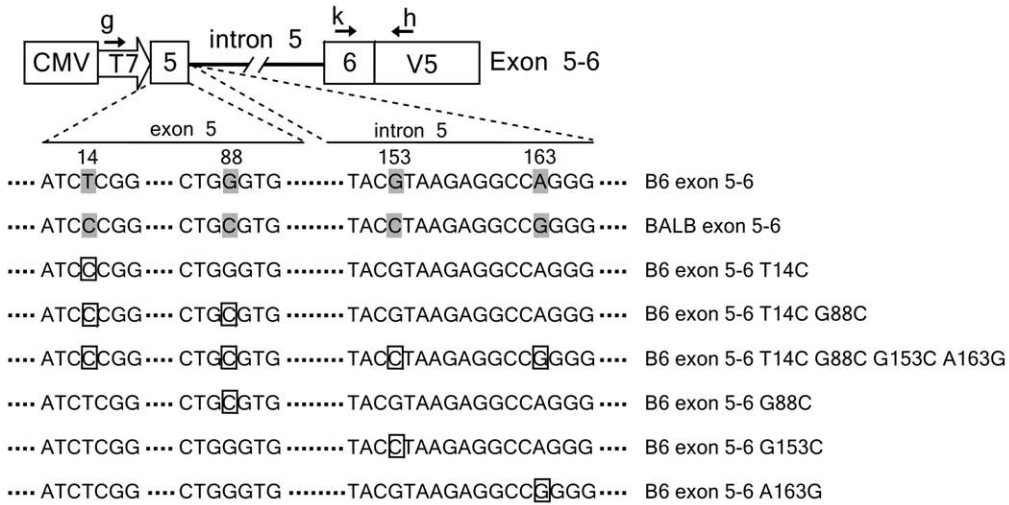
To determine the association of the above 4 SNPs with splicing efficiency under more physiological conditions, we further introduced point mutations into the B6 exon 5–6 plasmid harboring the entire intron 5 of 6kb in length, and the resultant plasmids were subjected to the splicing assays (Figure 7A). In contrast to the result with the B6 100bp intron 5 T14C mutant (Figure 6C), we could not detect the generation of spliced mRNA after transfection of the B6 exon 5–6 T14C mutant (Figure 7B). However, the single substitution at position 88 from the B6-type G to BALB/c-type C led to readily detectable production of the spliced mRNA, despite the general inefficiency in splicing of the full-length intron 5 depicted in Figure 5. Quantitative real-time PCR assays revealed levels of expression of the exon 5-containing spliced message from the B6 exon 5–6 G88C mutant that were about 70% of those expressed from the BALB exon 5–6 plasmid. Single nucleotide substitutions at two other positions did not result in a detectable generation of the spliced mRNA from the B6 allele (Figure 7B). Combined mutations at positions 14 and 88 within exon 5 on the B6 exon 5–6 construct (B6 exon 5–6 T14C G88C) and further addition of the nucleotide replacements at positions 153 and 163 also resulted in detectable generation of the spliced message, but did not noticeably enhance the splicing over what resulted from the G88C substitution alone (Figure 7C).

### Combined effects of the intron 4 TCCT repeat number and exon 5 G/C SNP on exon 5 inclusion into mA3 mRNA

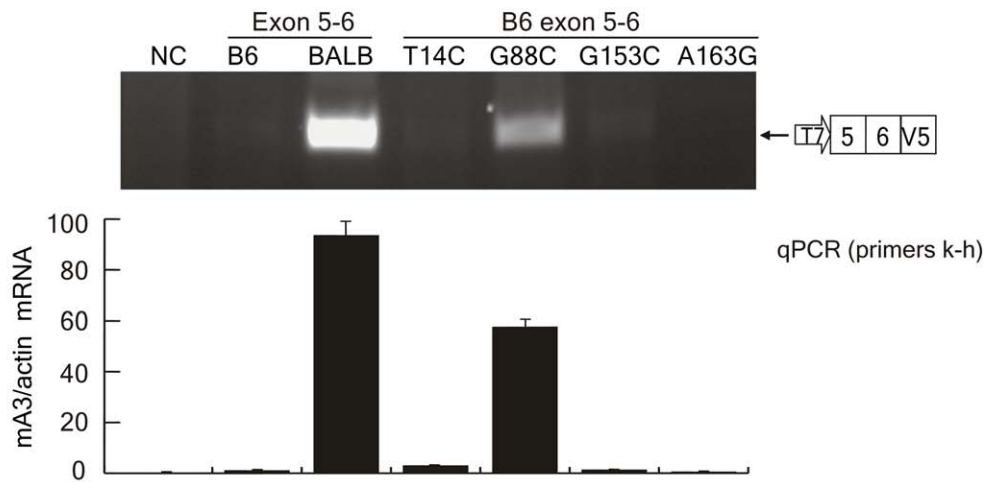
To further compare the effects of the TCCT repeat number and exon 5 G/C SNP at position 88, we constructed B6 and BALB exon 4–7 plasmids that harbored reciprocal substitutions at these two polymorphic sites (Figure 8). Introduction of an additional copy of TCCT into the B6 exon 4–7 construct did not result in the generation of 5+ message, while the combination of the additional TCCT and G88C substitution resulted in the generation of exon 5-containing message at levels comparable to those expressed from the BALB exon 4–7 plasmid. On the other hand, a deletion of a TCCT copy from the BALB exon 4–7 construct resulted in much reduced expression of the 5+ message, and the combination of  $\Delta$ TCCT and C88G substitution totally abrogated the generation of exon 5-containing message. Thus, these results clearly demonstrate that the most critical polymorphism for exon 5 inclusion into the mA3 mRNA is the G/C SNP at position 88 within exon 5, but for the full-level expression of the 5+ mRNA as observed in BALB/c mice the intron 4 TCCT repeat is also required.



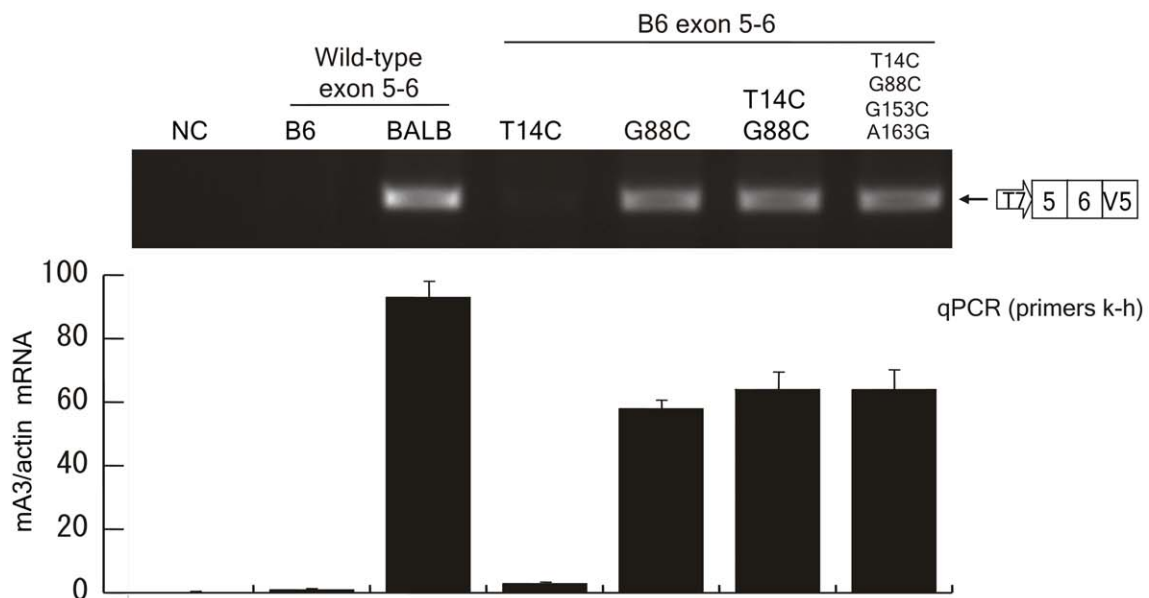
A



B



C



**Figure 7. The G88C nucleotide substitution in exon 5 is most critical for exon 5 inclusion into mA3 mRNA.** (A) Polymorphic nucleotides indicated with the shades were exchanged between the B6 and BALB/c alleles at the sites indicated with the boxes. Plasmids used here carry the entire intron 5 instead of the deleted one used in the experiments shown in Figure 6. (B and C) The results of splicing assays using the plasmids depicted in panel (A). The predicted splicing product is shown on the right side of each panel. RT-PCR assays were performed with primers g and h. A portion of the transfected cells were utilized for luciferase assays to compare transfection efficiencies among the samples. Comparable levels of luciferase activities were observed in all samples in each experiment (data not shown). Quantitative real-time PCR data show averages of three reaction wells and SD.  
doi:10.1371/journal.ppat.1002478.g007

### Distribution of the TCCT repeat and G/C88 polymorphism among *Mus* species

Because the inclusion of exon 5 into mA3 message depends on the intron 4 TCCT repeat and the exon 5 G/C polymorphism at position 88 (G/C88), we screened additional inbred strains and wild mouse species for these sequence variations and for the presence or absence of exon 5 in the mA3 message. We sequenced exon 5 and segments of the flanking introns from 39 mice that represent different taxa or members of the same species trapped in different geographic locations (Table S1), as well as those from the inbred laboratory strains B10.A and A/WySn, prototypic strains with FV-restrictive and -permissive phenotypes [26,27,31,32]. The complete dataset is shown in Figure S1. Genomic sequencing showed that A/WySn mice share the TCCT duplication and exon 5 C88 SNP with BALB/c, while B10.A mice are identical to B6 mice at both sites (Figure 9A); this is consistent with previous observations on mA3 expression levels and inclusion of exon 5 in these strains of mice [27]. Among the 39 wild mice, the TCCT duplication was found in 10 mice that possessed the exon 5 C88 SNP, and all of the 29 other mice with G88 had a single TCCT copy as representatively shown in Figure 9A. The observed linkage between the repeated TCCT and C88 is reasonable, as both are required for efficient expression of 5+ message from the BALB/c allele (Figures 4, 6, and 8). RNA samples from 23 of the above mice were typed for the splicing phenotype. The presence or absence of exon 5 in mA3 mRNA correlated with the above identified sequence polymorphisms: exon 5 was present only in 7 mice, all of which had the linked TCCT duplication and C88, consistent with our functional assays. On the other hand, there were at least 7 discrepancies with respect to exon 5 inclusion and the C14 SNP confirming that this SNP has no major role in exon 5 splicing.

The coding sequence for exon 5 was present in all 39 *Mus* DNA specimens, as well as in rat DNA specimens (GenBank EDM15775) as described previously [29]. There were frameshift mutations within exon 5 in *M. setulosus* and both *M. cervicolor* subspecies, all of which lack exon 5 in mA3 mRNA, and the exon 5-containing splice variant was found only in mA3 mRNA of *M. spretus* and 3 of the house mouse species (*M. domesticus*, *M. musculus*, *M. castaneus*) (Figure 9B). The wild mice were also tested for the presence or absence of another genetic variation associated with variable expression levels of mA3, the exon 2-associated xenotropic MuLV LTR insertion [37]. Results indicated that the presence of this LTR (LTR<sup>+</sup>), like the inclusion of exon 5 due to the above allelic variation (exon 5<sup>+</sup>), was introduced into the *ApoBec3* locus at about the time of the house mouse radiation (Figure 9B). These two features are not found together in any mouse, and the LTR<sup>+</sup> and exon 5<sup>+</sup> variants are both found in mice classed as *M. musculus* or *M. domesticus*, while *M. castaneus* mice have both variants as well as the ancestral  $\Delta 5$  LTR<sup>-</sup> mA3 type.

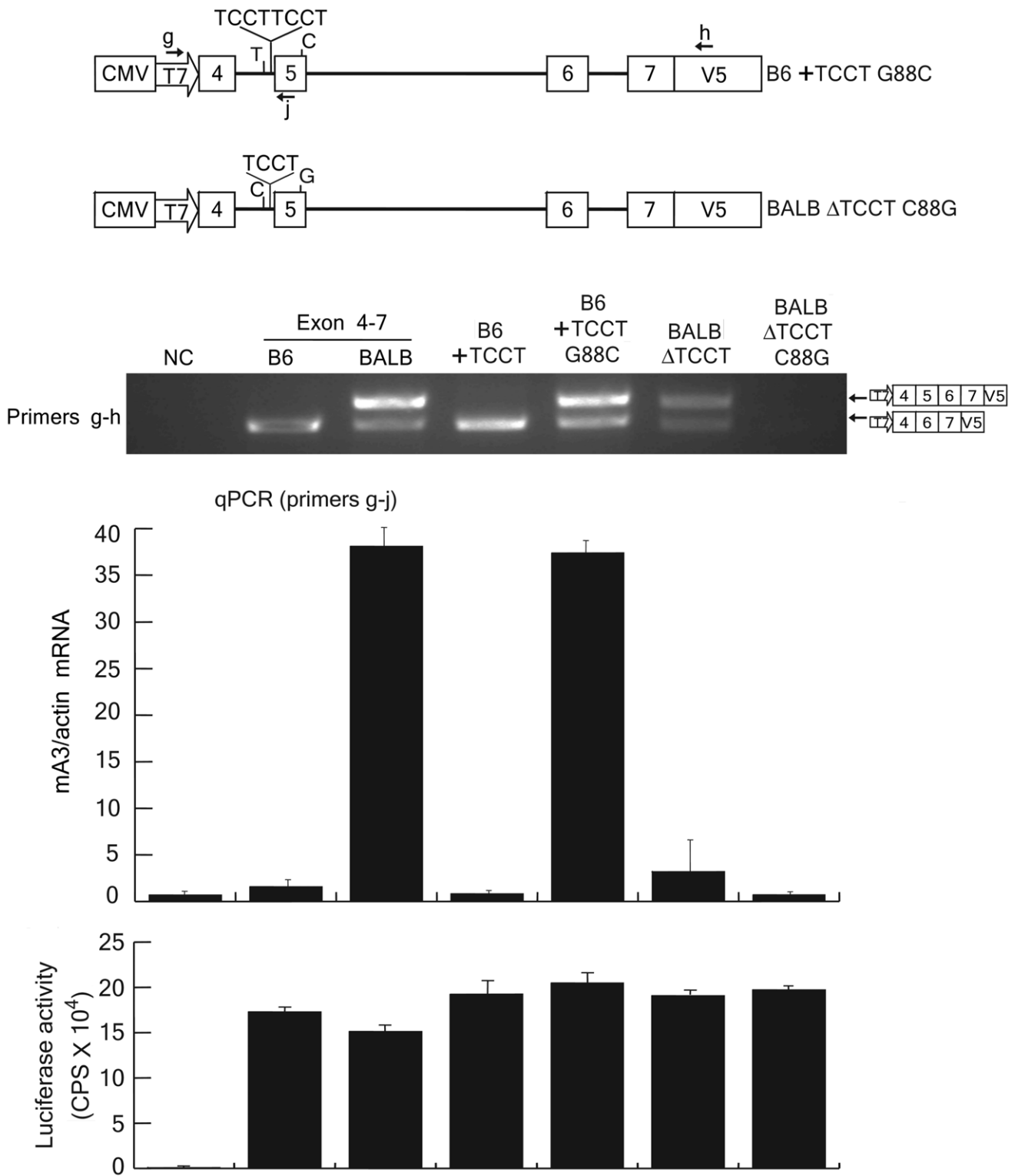
To better understand this species distribution, we typed additional samples of European *M. domesticus* and *M. musculus* for the presence of the LTR and exon 5 (Figure 9C). The geographic distribution of the trapping sites for these 19 house mice and for the exon 5<sup>+</sup> *M. spretus* indicated that the exon 5<sup>+</sup> mA3 was

characteristic of *M. domesticus* of north Africa and western Europe, whereas LTR<sup>+</sup> mA3 was found in eastern Europe ( $P=0.049$  by Fisher's exact test). Four of the 9 eastern European *M. musculus* mice carried the *M. domesticus* variant, which is consistent with previous reports that gene flow across the hybrid zone is biased in the direction from *M. domesticus* to *M. musculus* ([42], for example). Additional samples identified as *M. domesticus* were trapped in the Americas (HAF, CL, WSA, PGN2, SC1, BQC, SAF, JJD, PERA; see Table S1), and these carried either the exon 5<sup>+</sup> mA3 (like *M. domesticus*) or the LTR<sup>+</sup> mA3 (like *M. musculus*); this is likely due to the fact that *Mus* are non-native species that were introduced into the Americas by passive transport, and although generally classed as *M. domesticus*, these mice show evidence of hybridization with other introduced house mouse species [43,44].

### Discussion

Polymorphisms in mA3 are responsible for the fact that the B6 and C57BL/10 mouse strains are more restrictive to the replication of both beta- and gammaretroviruses than the BALB/c and A strains [25–29]. These differences in mA3 antiviral activities have been associated with sequence differences in the N-terminal region of mA3 [27], different levels of its mRNA expression [27,29,38,39], and the presence or absence of exon 5 in mA3 mRNA and protein [27], although the relative importance of these three factors for antiviral activity has not been established. In the present study, we focused on the role of exon 5 and have shown that the  $\Delta 5$  mRNA directs more efficient mA3 protein synthesis than the exon 5-containing message. We have also identified 2 genetic determinants responsible for the inclusion of this exon into mA3, a TCCT repeat in intron 4 and a C88 substitution in exon 5 relative to the B6 allele. We further showed that these two determinants were coordinately acquired within the last 0.5 million years by house mouse species of *Mus*.

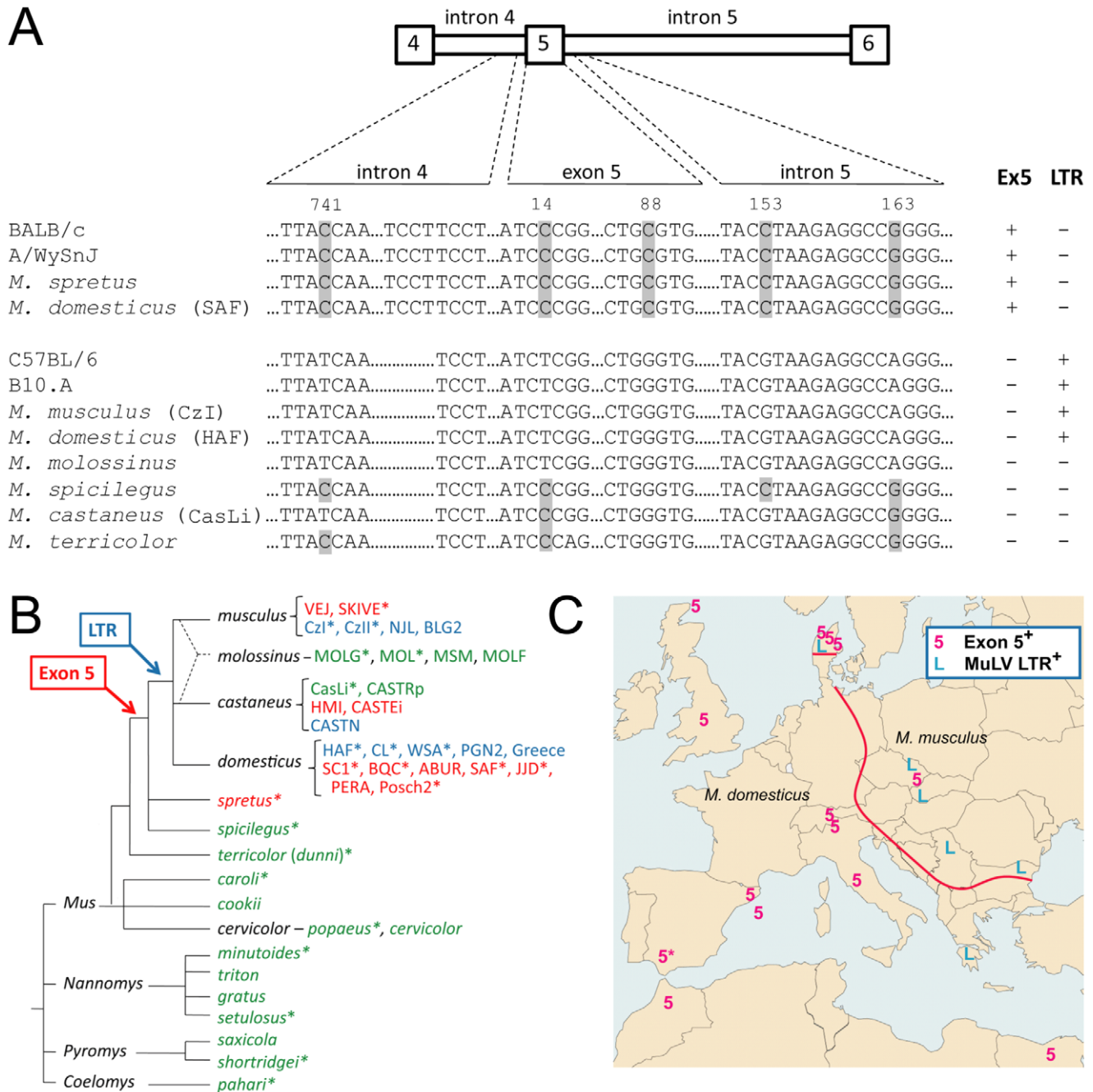
Previous studies have recognized differences in mRNA expression levels of mA3 between B6 and BALB/c mice [27,29,38,39]. Higher constitutive expression of mA3 mRNA is thought to contribute to better antiviral activity as suggested by the facts that higher levels of protein production result in more efficient incorporation of mA3 into viral particles [45] and that lipopolysaccharide (LPS)-induced enhancement of mA3 expression results in better restriction of MMTV *in vivo* [46]. It has also been shown that increased APOBEC3G and APOBEC3F expression is associated with lower viral load in rhesus macaques infected with simian immunodeficiency virus [47]. Higher protein levels of hA3G in CD14<sup>+</sup> monocytes are also associated with HIV-1-exposed but uninfected status in humans [48]. The increased level of *ApoBec3* gene transcription has been associated with the presence in the B6 allele of an intact MuLV LTR, a sequence capable of driving enhanced transcription [37]. While increased levels of the mA3 transcript could be solely responsible for the presently observed increase in mA3 protein levels in B6 mice (Figure 1D), it was also reported that, when driven by the same strong promoter, the B6 mA3 cDNA lacking exon 5 produced more protein product than the exon 5-containing BALB/c cDNA and resulted in more efficient incorporation into viral particles



**Figure 8. Effect of the combination of intron 4 TCCT repeat and exon 5 G/C SNP on exon 5 inclusion into the mA3 transcripts.** B6 exon 4–7 + TCCT and BALB/c exon 4–7 ΔTCCT plasmids, as described in Figure 4, were further modified so that the exon 5 G/C SNP nucleotide at position 88 was reciprocally exchanged. The resultant plasmids were used for *in vitro* transcription assays as described for Figures 4–7. RT-PCR assays were performed with primers g and h. A portion of the transfected cells were utilized for luciferase assays to compare transfection efficiencies among the samples. Quantitative real-time PCR data show averages of three reaction wells and SD. doi:10.1371/journal.ppat.1002478.g008

[27,45]. Thus, it was unclear whether the experimentally observed increased translation was due to the lack of exon 5 in the B6 cDNA or to the other sequence differences that distinguish the two

*Apobec3* alleles [27,37,38]. Here we have clearly shown that the Δ5 mRNA directs much more efficient protein synthesis than the 5+ mA3 mRNA regardless of other allelic differences (Figures 2 and



**Figure 9. Distribution of sequence variations in and around the *Apobec3* exon 5 in *Mus* species and strains.** (A) The *Apobec3* genomic sequences of the indicated strains and representative species are shown. The five SNP tested for their effects on exon 5 splicing are indicated with shaded bars. Also shown are the number of TCCT repeats and typing results indicating the presence or absence of exon 5 (Ex5) in spliced messages and the presence or absence of an MuLV LTR in intron 2. (B) The phylogenetic tree is a synthetic tree adapted from other sources [66–68]. Species are grouped into the 4 *Mus* subgenera and are listed to the right of each branch. Subspecies and stock or inbred line designations are given where multiple samples of a given species were tested. Red, exon 5<sup>+</sup>/MuLV LTR<sup>-</sup>; blue, Δ5/MuLV LTR<sup>+</sup>; green, Δ5/MuLV LTR<sup>-</sup>. \*, mice tested for exon 5 splicing pattern. All were sequenced across exon 5 (Figure S1); PERA and MOLFG sequence information was taken from Okeoma et al. [38]. (C) Distribution of expression variants of m3 in mice trapped in Europe. The red line represents the 20 km-wide hybrid zone separating the ranges of *M. domesticus* and *M. musculus*. Symbols representing m3 allelic variants are placed at trapping sites. doi:10.1371/journal.ppat.1002478.g009

3), and the combination of the higher mRNA expression levels with preferential generation of the Δ5 isoform in mRNA splicing results in much higher m3 protein levels in B6 in comparison with BALB/c mice (Figure 1).

The mechanisms by which the presence of exon 5 in the mRNA interferes with protein synthesis are currently unclear; however,

predictions of the secondary structure for the m3 mRNA suggest that the portion encoding exon 5 may form stable stem-loop structures (Figure S2), which might interfere with efficient translation. The predicted stem-loop structure is thermodynamically more stable for B6 than for BALB/c exon 5, and this portion is actually spliced out from the major transcript in B6 mice. It is

possible, therefore, that the elimination of this exon from mA3 mRNA confers a functional advantage to B6 mice through more efficient translation of the  $\Delta 5$  mRNA and resultant higher levels of mA3 protein expression. It should be noted, however, that introduction of this exon between the two CDDs does not physically disrupt the functional domains, as the 5+ product of the BALB/c allele still exerts deaminase activity and restricts AKV replication *in vitro* [29]. The possible effect of this 33-amino acid exon on interactions between the N-terminal and C-terminal CDDs and on mA3 oligomerization remains to be elucidated.

Our results also show that the ancestral *Mus Apobec3* gene locus lacked both the MuLV LTR and the functional polymorphisms that determine exon 5 inclusion (Figure 9). These two genetic features were acquired independently at about the time of the house mouse radiation 0.5–1.0 million years ago, and are found in distinct lineages of wild mice and different inbred strains of laboratory mice. It should be noted that all 4 house mouse species originated and diverged from an ancestral population on the Indian subcontinent that carries many of the alleles found in peripheral Eurasian populations [49,50]. The observed distribution of the LTR<sup>+</sup> and exon 5<sup>+</sup> mA3 variants among the house mouse lineages is consistent with the appearance of these variants in this ancestral population. It is thus likely that *M. domesticus* progenitors carrying the exon 5<sup>+</sup> *Apobec3* allele moved west to north Africa and western Europe, and *M. musculus* (LTR<sup>+</sup>) moved north to Russia, eastern Europe and north China. *M. castaneus*, which carries these two variants as well as the ancestral  $\Delta 5$  LTR<sup>-</sup> mA3, migrated east to Thailand and China. Only the ancestral mA3 type is found in *M. molossinus*, which is a natural hybrid of *M. musculus* and *M. castaneus* [44], and these Japanese mice presumably inherited their *Apobec3* gene from their *M. castaneus* progenitors.

Numerous studies indicate that host antiviral factors co-evolve with the pathogens they restrict, and this “arms race” is responsible for mutational changes in APOBEC3 in primates and mice as evidenced by detectable positive selection [37,51]. The acquisition of the MuLV LTR that is associated with enhanced mA3 transcription in virus-infected mice also makes sense as an evolutionary adaptation to pathogen infection [37]. Thus, the B6 *Apobec3* allele has acquired two advantageous features that contribute to enhanced retrovirus restriction: high mRNA expression directed by the intron 2 LTR that is in linkage with the positively selected, more functional amino acid sequence in the N-terminal deaminase domain [27,37]. The BALB/c allele, in contrast, acquired the polymorphisms that direct exon 5 inclusion as shown in the present study, along with coding sequence polymorphisms associated with lower antiviral activity [27]. The far-flung distribution of exon 5<sup>+</sup> mA3 throughout western Europe (Figure 9C) is surprising, because this allele, having a reduced level of mA3 protein expression with inefficient restriction of MuLV replication at least *in vitro* [27], would seem to be evolutionarily deleterious. The observed distribution of exon 5<sup>+</sup> mA3 suggests either that the exon 5<sup>+</sup> variant provides sufficiently protective levels of antiviral activity in virus-infected mice and is therefore not subject to purifying selection, that these particular mice have not been exposed to significant challenge by mA3-sensitive pathogens, that this mA3 variant provides some other, unrecognized selective advantage, or that this allele is tightly linked to another advantageous polymorphism.

All 4 house mouse species possessing the exon 5<sup>+</sup> genetic variation carry beta- and gammaretroviruses along with other retroelements that are subject to APOBEC3 restriction [25–30,52]. In this regard, although the BALB/c mA3 is less antiviral than the product of the B6 allele, it does have measurable antiviral activity [29]. Further, analysis of 54 germline MuLV proviruses of

three envelope types in the sequenced B6 genome demonstrated evidence of APOBEC3-mediated editing of the polytropic and modified polytropic, but not the xenotropic, proviruses [30]. The mA3-edited polytropic and modified polytropic viruses originated in *M. domesticus* [52,53], mice that carry the exon 5<sup>+</sup> mA3, suggesting that this mA3 variant may be effectively antiviral *in vivo*. Finally, in mA3-deficient B6 mice, the absence of this protein differentially affects FV replication depending on target cell types [54]. As cell type-dependent and pathogen-induced changes in mA3 expression may differ between mice of the ancestral  $\Delta 5$  LTR<sup>-</sup>, exon 5<sup>+</sup>, and LTR<sup>+</sup> variants, mice with the exon 5<sup>+</sup> *Apobec3* allele may still restrict mA3-sensitive pathogens in critical target cells.

It is also possible that some mA3 polymorphisms have been selected for reasons other than their antiviral functions, and these functions may take precedence in mice not threatened with retrovirus assault. It is unlikely that differences in mA3 expression levels and/or its amino acid sequence provide a significant selective advantage in terms of normal development, survival, or fertility, as *Apobec3*-knockout mice show no increased propensity for tumor development or disease, and both male and female mA3-deficient mice were fertile [40]. On the other hand, the presence of active cytidine deaminases can have costs that may become more significant when endogenous and exogenous retroviral activity is low. Transgenic mice overexpressing mouse activation-induced deaminase (AID) or rabbit apolipoprotein B mRNA-editing enzyme catalytic polypeptide 1 (APOBEC1), two other members of the AID/APOBEC family of cytidine deaminases, developed neoplastic diseases and showed evidence of significant editing of various expressed genes [55,56]. This suggests that when the retroviral threat is low, the consequences of possessing highly active mutagenic enzymes, like the  $\Delta 5$  mA3 expressed in B6 mice, can outweigh their advantages as antiviral factors. In this regard, it should be noted that infectious MuLV has not been isolated from European *M. domesticus*, which carry the less highly expressed 5+ mA3, whereas the more active *Apobec3* alleles are found in all 3 of the other species of *Mus*, all of which have been found to carry infectious MuLVs [57].

Finally, it has also been observed that there is a similarly widespread distribution in humans of a deletion of a genomic segment harboring the *Apobec3b* gene locus [58]. At least one copy of this deletion is present in >40% of humans. The selective advantage of this genetic variant has not been determined, but its retention may suggest a possible advantage in reducing the genotoxic activity of this cytidine deaminase, or alternatively, it may be linked to positively selected variants in one or more of the 6 linked human APOBEC3 paralogues [59,60]. Further comparative analyses of the geographic and species distributions of natural mouse pathogens and the various mA3 variants may help define previously unrecognized targets of APOBEC3-mediated restriction and provide further insight into the coevolution of pathogens and this host restriction factor.

## Materials and Methods

### Ethics statement

The studies utilizing laboratory animals were carried out in strict accordance with the Act on Welfare and Management of Animals of the government of Japan and the Regulations for the Care and Use of Laboratory Animals of Kinki University. The protocol was approved by the Institutional Animal Experimentation Committee of Kinki University School of Medicine (Permit Number: KAME-19-029). All surgery was performed under sodium pentobarbital anesthesia, and all efforts were made to

minimize suffering. All studies involving wild mice were performed in compliance with the US Government Principles for the Utilization and Care of Vertebrate Animals used in Testing, Research, and Training; the Public Health Service Policy on Humane Care and Use of Laboratory Animals; The Animal Welfare Act and amendment laws; the Animal Care Policies of the USDA; The Guide for the Care and Use of Laboratory Animals (7th Edition; National Research Council); and the guidelines of the Committee on the Care and Use of Laboratory Animals under an NIAID-approved animal study protocol, and all studies and procedures were reviewed and approved by the Institutional Animal Care and Use Committee of the NIH (Permit Number: ASP LMM 1).

### Mice and mouse genomic DNA

C57BL/6Ncrslc, BALB/cCrslc, and B10A/SgSns1c mice were purchased from Japan SLC, Inc., Hamamatsu, Japan. Breeding pairs of A/WySnJ mice were purchased from The Jackson Laboratory, Bar Harbor, ME. The mA3-deficient strain on the B6 background has been described [27,40]. All laboratory mice were housed and bred in the Experimental Animal Facilities at Kinki University School of Medicine under specific pathogen-free conditions. The isolation of genomic DNA from spleens was carried out with DNeasy blood and tissue kit (Qiagen, Inc., Hilden, Germany) according to the manufacturer's instructions.

DNA and RNA were separately isolated from animals and cell lines developed from wild mice and wild mouse-derived breeding colonies or inbred strains (Table S1). Many wild-derived mice were obtained from M. Potter (National Cancer Institute, Bethesda, MD). CAST/Rp mice were obtained from R. Elliott (Roswell Park Cancer Institute, Buffalo, NY). Cells from some wild mouse species were obtained from J. Rodgers (Baylor College of Medicine, Houston, TX) and from J. Hartley, M. Lander or S. Chattopadhyay (National Institute of Allergy and Infectious Diseases, Bethesda, MD) [61,62]. *M. cervicolor popaeus* mice and tissue samples were obtained from R. Callahan (National Cancer Institute). Mice or DNA samples of inbred lines of *M. castaneus* (CAST/Eij) and *M. molossinus* were obtained from The Jackson Laboratory. *M. musculus* DNA samples were obtained from S. Chattopadhyay and H. Morse (National Institute of Allergy and Infectious Diseases). DNA samples from wild-trapped European *M. domesticus* were provided by M. Nachman (University of Arizona, Tucson). DNA samples from 5 wild-derived strains (BLG2, NJL, MSM, HMI, PGN2) were obtained from the National Institute of Genetics, Mishima, Japan. A set of *Nannomys* DNAs was obtained from Y. Cole and P. D'Eustachio (Departments of Biochemistry and Medicine, New York University, NY); these mice were classed into 4 species on the basis of skeletal features by J. T. Marshall (Smithsonian Natural History Museum, Washington, DC). DNA was isolated from cultured tail biopsies, spleen, or liver by standard protocols, and RNA was isolated from the spleen or liver using TRI-Reagent (Molecular Research Center, Cincinnati, OH) or by a guanidine chloride extraction method [63].

### Mouse *Apobec3* sequences and the prediction of mRNA secondary structures

DNA containing the mouse *Apobec3* exon 5 and associated intron sequences was amplified using either one of the following forward primers: 5'-GGACAATGGTGGCAGGCGATTTC-3', 5'-GCATCTTTGTGGATGGGG-3', and the reverse primer 5'-TCATTCCTCAATGCTCCTCC-3'. PCR products were cloned into pCR2.1-TOPO (Invitrogen, Carlsbad, CA) before sequencing. The *Apobec3* region was amplified from genomic DNA of B6, BALB/c, B10A, and A/WySnJ mice using the forward primer

5'-TTACAAATTTTAGATACCAGGATTCTAAGCTTCAGG-AG-3' and the reverse primer 5'-GTCCTTTATGTGGGTTTCAAGGACC-3'. PCR products were treated with ExoSAP-IT (USB, Cleveland, OH) and directly sequenced with the above reverse primer. To determine the BALB/c genomic sequence of the *Apobec3* intron 5, the BALB exon 5–6 and BALB intron 5- $\Delta 3'$  plasmids (Figure 5) were sequenced by using BigDye Terminator V3.1 Cycle Sequencing Kit (Applied Biosystems, Foster City, CA) and an ABI PRISM 3100 Genetic Analyzer (Applied Biosystems) using the following primers: T7 promoter forward, 5'-TAATAC-GACTCACTATAGGG-3'; and V5 reverse, 5'-CGTAGAATC-GAGACCGAGGAGAGGGTTAGGGATAGGC-3'.

The secondary structure of a portion of the mA3 mRNA exon 5 was predicted with *mfold* [64,65].

### Cell culture and DNA transfection

BALB/3T3 and human 293T cells were cultured in Dulbecco's modified Eagle medium supplemented with 10% heat-inactivated fetal bovine serum (Invitrogen). These cells were seeded at  $1.0 \times 10^5$ /well in a well of 6-well plates one day prior to transfection. DNA transfection was performed by using Lipofectamine 2000 (Invitrogen) according to the manufacturer's protocols. Mouse spleen cells were harvested as described previously for protein and RNA extractions [27]. At least three independent transfection experiments were performed in this study and representative results are shown in the figures.

### Plasmid constructions

The expression plasmids, pFLAG-CMV2-*mA3<sup>b</sup>* $\Delta 5$  harboring the mA3 cDNA derived from the  $\Delta 5$  transcript of the B6 allele, pFLAG-CMV2-*mA3<sup>d</sup>* harboring the cDNA derived from the 5+ transcript of the BALB/c allele, pFLAG-CMV2-*mA3<sup>d</sup>* $\Delta 5$  harboring the cDNA derived from the  $\Delta 5$  transcript of the BALB/c allele, and control pFLAG-CMV2-GFP have been described [27]. The genomic DNA encoding the mA3 exon 5 was amplified from B6 genome with the following primers: 5'-ACCTTGCTACATCTCGGTCCCTTC-CAGC-3' and 5'-CTGCCCTCCACCCAGAACCCTCGTCTC-TGG-3'. The above pFLAG-CMV2-*mA3<sup>b</sup>* $\Delta 5$  was used as a PCR template with primers 5'-GCCAATGGACCCGCTAAGTGAA-GAGG-3' and 5'-CTCAGAATCTCCTGAAGCTTAGAATCC-TGG-3' to amplify the linearized plasmid with a gap between exons 4 and 6. The two PCR products above were treated with T4 polynucleotide kinase (TAKARA Bio, Otsu, Japan) and fused by using the DNA Ligation Kit ver.2.1 (TAKARA Bio) to construct the plasmid pFLAG-CMV2-*mA3<sup>b</sup>*, which expresses the 5+ mA3 cDNA derived from the B6 allele.

B6 or BALB/c genomic fragments harboring exons 4–7 and the intervening introns were amplified by PCR using either the B6 or BALB/c genomic DNA as a template and the common primers 5'-CACCAATTTAAAAAGTGTGGAGAAG-3' and 5'-GTGGGAGGTCCATGACGTCCACCAGGATCCC-3'. Each amplified DNA product was cloned into a pcDNA3.2/V5/GW/D-TOPO cloning vector (Invitrogen) and designated as B6 or BALB exon 4–7, respectively. The above B6 or BALB exon 4–7 was used as a template with the primers 5'-CACCACCTTGC-TACATCTCGGTCC-3' and 5'-GCAGAGATGCTTGACTC-GTTGGTTG-3' or 5'-CACCACCTTGCTACATCCCGGTC-C-3' and 5'-GCAGAGATGCTTGACTCGTTGGTTG-3', respectively, for the amplification of B6 or BALB/c exons 5 and 6 and the intervening intron 5. Each PCR product was cloned into the pcDNA3.2/V5/GW/D-TOPO vector and designated as B6 or BALB exon 5–6. The DNA fragments harboring sequentially deleted *Apobec3* intron 5 were prepared by PCR using either one of the above B6 or BALB exon 5–6 plasmids as a common template



with the primer pairs A–F listed in Table S2. Each amplified DNA product was cloned into the pcDNA3.2/V5/GW/D TOPO vector to generate the expression plasmid shown in Figure 5B.

Reciprocal chimeras between the above B5 and BALB exon 5–6 plasmids were generated by amplifying a linearized plasmid DNA lacking the 3' intron 5 and exon 6 using primer pair G, and by amplifying the insert fragment using primer pair H.

Site-directed mutagenesis was performed by employing the QuikChange Site-Directed Mutagenesis Kit (Stratagene, La Jolla, CA) using the following templates and primers, listed separately in Table S2: BALB 100bp intron 5 (3'-100bp) plasmid was used as a common template for the preparation of BALB C14T, BALB C88G, and BALB C153GG163A with primer pairs I, J, K, respectively: B6 100bp intron 5 (3'-100bp) plasmid was used as a template to make B6 T14C or B6 G88C with primer pair L or M, respectively. The resultant B6 T14C plasmid was used as a template with the same primer pair M employed for the generation of B6 G88C to make B6 T14C G88C. B6 T14C G88C was then used as a template for the generation of B6 T14C G88C G153C A163G by using primer pair O. Similarly, the above-used primer pairs were also utilized for the generation of B6 exon 5–6 T14C, B6 exon 5–6 T14C G88C, and B6 exon 5–6 T14C G88C G153C A163G mutants by using B6 exon 5–6 as a template. The B6 exon 5–6 plasmid was also used as a template with primer pairs M, P, and Q for making B6 exon 5–6 G88C, B6 exon 5–6 G153C, and B6 exon 5–6 A163G, respectively.

To make BALB  $\Delta$ TCCT, 4 consecutive nucleotides within the repeat sequence, TCCT, were deleted from the BALB exon 4–7 plasmid by mutagenesis using primer pair R, listed in Table S2. BALB  $\Delta$ TCCT or BALB exon 4–7 plasmid was used as a template with primer pair S to make BALB C741T  $\Delta$ TCCT or BALB C741T, respectively. To introduce the TCCT sequence by mutagenesis and make B6+TCCT, B6 exon 4–7 plasmid was used as a template with primer pair T. B6+TCCT or B6 exon 4–7 was used as a template with primer pair U to generate B6 T741C+TCCT or B6 T741C, respectively. Similarly, the above produced B6 exon 4–7+TCCT or BALB exon 4–7  $\Delta$ TCCT plasmid was used as a template and G88C or C88G substitution was introduced with primer set M or J, respectively.

All resultant plasmids were entirely sequenced by using BigDye Terminator V3.1 Cycle Sequencing Kit with an ABI PRISM 3100 Genetic Analyzer. In order to normalize the transfection efficiency, a plasmid expressing the luciferase gene, *pluc*, based on the expression vector pGL3 (Promega, Madison, WI), was utilized. All the primers used in this study were purchased from Operon Biotechnologies, Tokyo, Japan.

### Quantitative real-time PCR assays for endogenous mA3 transcripts

For the quantification of mA3 transcripts in mouse spleens, total RNA was extracted from each spleen with RNeasy Mini Kit (Qiagen). The RNA was then subjected to reverse transcription with PrimeScript RT reagent Kit (TAKARA Bio). Real-time PCR reactions were carried out with SYBR Premix Ex Taq II (TAKARA Bio) on an Applied Biosystems 7900HT Fast Real-Time PCR System (Applied Biosystems) with two different sets of *ApoBec3*-specific primers: set 1, 5'-GTGTTGGAAGAAGTTT-GTGG-3' (primer a) and 5'-CCTGAAGCTTAGAATCCTGG-3' (primer b); and set 2, 5'-TTACAAATTTAGATACCAG-GATTCTAAGCTTCAGGAG-3' (primer c) and 5'-TTGGTT-GTAAACTGCGAGTAAAATTCCTCTTAC-3' (primer d). The data were normalized with expression levels of  $\beta$ -actin mRNA to obtain  $\Delta$ Ct values, and  $\Delta\Delta$ Ct values were calculated.

### Detection and quantification of the mA3 transcripts and their splicing products

For the detection of endogenous mA3 transcripts in mouse spleen cells, the RT products and primer sets used for the real-time PCR were also utilized for PCR using KOD Dash DNA polymerase (Toyobo, Osaka, Japan). The PCR products were separated by 1% agarose gel electrophoresis and detected by staining with ethidium bromide. For splicing assays, BALB/3T3 cells were transfected with 1  $\mu$ g of each plasmid harboring a genomic DNA fragment and 0.5  $\mu$ g of *pluc* to normalize the transfection efficiency. Total RNA was extracted from the transfected cells using the RNeasy Mini Kit at 24 hours post-transfection. The total RNA was treated with DNase I, reverse transcribed with SuperScript III First Strand Synthesis System (Invitrogen), and the resultant cDNA was subjected to PCR detection using KOD Dash DNA polymerase with the following primers: 5'-TAATACGACTCACTATAGGG-3' (primer g) and 5'-CGTAGAATCGAGACCGAGGAGGGTTAGGGATAG-GC-3' (primer h), designed to hybridize the T7 promoter and V5 tag regions of the vector, respectively. Primer i, 5'-GGTC-TCCCAGAGACGAGGTTCTG-3', was used to detect only the exon 5-containing transcript. For real-time PCR quantification of the mA3 transcripts containing exon 5, primer j (5'-GTGGAT-GAAGAGCTGGAAGGGACCG-3') was used along with the above primer g. Transcripts containing exon 6 were similarly quantified by using primer k (5'-CAACCAACGAGTCAAG-CATCTCTGC-3') and the above primer h.

For RT-PCR detection of mA3 mRNA in cells transfected with an mA3 expression plasmid, the same RT product as was used for real-time PCR analyses was utilized with the above-described primer set 1 (a–b). The RT products made by PrimeScript RT reagent Kit were supposed to be relatively short in length because a mixture of oligo-dT and random 6-mers were used as primers in the RT reaction. Thus, in order to detect the full-length and  $\Delta$ 5 transcripts, the following primers, 5'-GGGAATTCGATGG-GACCATTCTGTCTGGGATGCAGCCATCGC-3' (primer e) and 5'-GGGTCGACTCAAGACATCGGGGGTCCAAGCTG-TAGGTTTTCC-3' (primer f), were used along with newly synthesized RT products generated by the SuperScript III First Strand Synthesis System (Invitrogen). To quantify the mA3 transcripts in 293T cells transfected with an mA3 expression plasmid, one-third of the transfected cells were used for total RNA isolation using the RNeasy Mini Kit at 24 hours after transfection. The purified RNA was treated with 5 units of DNase I (TAKARA Bio) for 1 hour at 37°C to digest the transfected DNA and then reverse transcribed with the PrimeScript RT reagent Kit. The real-time PCR was performed as described above.

For RT-PCR analyses of mA3 transcripts generated by the *in vitro* transcription/translation system, the reaction was stopped by the addition of lysis buffer included in RNeasy Mini Kit after 30 or 60 min of incubation, and the transcribed products were purified with the above kit. After the treatment of the purified RNA with DNase I, the RT-PCR reaction was carried out with the above primer set 1. After electrophoresis, gel images were recorded with a FluorChem™ IS-8900 transilluminator and band intensities were analyzed with AlphaEase FC Stand Alone software (Alpha Innotech, San Leandro, CA).

*ApoBec3* splicing patterns for wild-derived mice were identified by RT-PCR to amplify a segment of mA3 RNA spanning exon 5 from total RNA using forward primer 5'-GGACCATCTCTG-TCTGGGATGCAGCCATCG-3' and reverse primer 5'-GG-TTGTAAACTGCGAGTAAAATTC-3'.

## Luciferase assays

Luciferase assays for normalization of transfection efficiencies were performed by utilizing the Luciferase Assay System (Promega). The enzymatic activities were measured by Wallac 1420 ARVO™ MX-2 Multilabel Counter (Perkin Elmer).

## Pre-absorption of the antibody

A spleen from an *Apobec3* knock-out mouse [40] was homogenized in 1ml ice-cold phosphate-buffered saline (PBS). Four ml of ice-cold acetone was added to the homogenate, mixed, and incubated on ice for 30 min. The lysate was centrifuged at 10,000×g at 4°C for 10 min. The pellet was washed with ice-cold acetone once and dried completely at room temperature to make spleen extract powder. Two μl of anti-APOBEC3 NT antibody specific for the N-terminal portion of mA3 (Millipore, Billerica, MA) was added to 0.5ml of KBTB buffer (50 mM Tris-HCl, pH 7.5, 150 mM NaCl, 10 mM KCl, 1% Triton X-100) containing 10% (w/v) unimmunized sheep serum as a carrier, into which 3mg of the spleen extract powder was dissolved. The mixture was incubated overnight at 4°C with gentle rotation. After centrifugation at 10,000×g at 4°C for 10 min, the supernatant was used as a pre-absorbed anti-mA3 antibody.

## Immunoblotting

Western blotting analyses were conducted as described previously [27] with some modifications. Briefly, proteins were extracted with a lysis buffer (1% Nonidet P-40, 25 mM Tris-HCl, pH 7.5, 140 mM NaCl, 1 mM EDTA, 10 mM Na<sub>4</sub>P<sub>2</sub>O<sub>7</sub>) containing protease inhibitors from Complete, Mini, EDTA-free Protease Inhibitor Cocktail Tablets (Roche Applied Science, Mannheim, Germany) and a phosphatase inhibitor, PhosSTOP (Roche Applied Science). Total protein concentrations were determined by Bradford assay (Nacalai Tesque, Kyoto, Japan) and the extracts were mixed with sodium dodecyl sulfate (SDS)-polyacrylamide gel electrophoresis (PAGE) sample buffer and heated at 95°C for 5 min. The proteins were separated by SDS-PAGE, transferred to Immobilon-P membrane (Millipore), and the blotted membranes were blocked with 5% (w/v) skim milk (Wako Pure Chemicals, Osaka, Japan) in Tris-buffered saline with 0.05% Tween 20 (TBST). The blocked membranes were incubated with the primary antibody at 4°C overnight. Membranes were then washed with TBST and incubated with horse radish peroxidase (HRP)-conjugated secondary antibody for 2 hours at room temperature, washed again with TBST, and the bound antibodies were detected using ECL plus reagent (GE Healthcare, Tokyo, Japan). The images were captured with a LAS-1000 Plus (Fujifilm, Tokyo, Japan) and the band intensities evaluated with Image Gauge ver. 3.12 (Fujifilm). A 1/5 dilution of the above-mentioned pre-absorbed mA3 antibody was made with IMMUNO SHOT (COSMO BIO, Tokyo, Japan) and was used as a primary antibody. Anti-FLAG M2 monoclonal antibody (mAb) (Sigma-Aldrich), anti-IκBα mAb (L35A5) (Cell Signaling Technology, Beverly, MA), anti-actin antibody (C-11) (Santa Cruz, CA), and the His-probe (H-15) (Santa Cruz) were diluted at 1:1000 with TBST. HRP-conjugated rabbit anti-mouse IgG (Zymed, South San Francisco, CA) and HRP-conjugated Goat anti-rabbit IgG antibodies (Invitrogen) were also diluted at 1:1000 and used as secondary antibodies to detect each appropriate primary antibody.

## Assessment of mA3 protein stability

$1.0 \times 10^5$  of 293T cells were transfected with 0.1 μg of an mA3 expression plasmid, 0.01 μg pFLAG-CMV2-GFP, and 0.1 μg of *pluc*. After 24h, the cells were treated with 10 μg/ml of cycloheximide or its solvent dimethyl sulfoxide (DMSO) as a control for 0, 2, or 4 hours. The cells were washed and resuspended in PBS. One-tenth of the cell suspension was subjected to luciferase assays to normalize transfection efficiencies, and the remaining cells were dissolved in the SDS-PAGE sample buffer. The normalized amount of cell lysates were separated by SDS-PAGE followed by immunoblotting as described above.

## In vitro transcription and translation

The FLAG-mA3 cDNA was amplified with highly proofreading Pfu Turbo DNA Polymerase (Stratagene) from the plasmids pFLAG-CMV-*mA3<sup>d</sup>* and pFLAG-CMV-*mA3<sup>d</sup>Δ5* [27], with a forward primer harboring the T7 promoter sequence, 5'-GGATCCTAATACGACTCACTATAGGGAACAGCTGGGATGGGACCATTCTGTCTGGGATGC-3' and a reverse primer harboring the His-Tag sequence, 5'-TCAATGGTGATGGT-GATGATGAGCAGCAGCAGACATCGGGGGTCCAAGCTGTAGG-3'. The PCR products were subjected to reactions for *in vitro* transcription and translation using TNT T7 Quick for PCR DNA (Promega) according to the manufacturer's protocol. Half of the generated products were mixed with the SDS-PAGE sample buffer and analyzed by immunoblotting. The remaining products were used for RNA purification with RNeasy Mini Kit (Qiagen) followed by RT-PCR to detect mA3 transcripts.

## Supporting Information

**Figure S1 Nucleotide sequence of the genomic region encoding mA3 exon 5 and segments of flanking introns from several laboratory mouse strains and wild mouse species.** Mouse *Apobec3* exon 5 and the flanking introns from 39 mice that represent different taxa or members of the same species trapped in different geographic locations, as well as those from the inbred laboratory strains BALB/c, B10.A, and A/WySn, were sequenced and aligned with the corresponding B6 sequence. The exon 5 and six key polymorphic regions, C/T741 and TCCT repeat in intron4, C/T14 and C/G88 in exon5, and C/G153 and A/G163 in intron 5, are indicated. Accession numbers for all newly obtained sequence data are also provided in this figure. (PDF)

**Figure S2 Possible stem-loop structures predicted from the mA3 intron 5 mRNA sequence.** The mRNA secondary structures of exon 5 encoded by the B6 and BALB/c alleles were predicted by using the mfold [64,65]. Polymorphic nucleotides within this exon, U/C at position 14 and G/C at position 88, are indicated. (PDF)

**Table S1 Designations and sources of wild-derived mice, their cells, and DNA samples.** (PDF)

**Table S2 Primers used to generate intron 5 deletion mutants and chimeras, for exon 5/intron 5 nucleotide substitutions, and for modification of TCCT repeat and T/C 741 SNP in intron 4.** (PDF)

## Acknowledgments

We thank Drs. Sachiyo Tsuji-Kawahara, Eiji Kajiwara, and Shiki Takamura for their helpful discussions and comments and Mr. J. Brian Dowell for critically reading and correcting the manuscript. We thank Alicia Buckler-White for DNA sequencing. We are also grateful to the National Institute of Genetics, Mishima, Japan, for the provision of genomic DNA samples from wild mouse strains.

## References

- Bogerd HP, Wiegand HL, Doehle BP, Lueders KK, Cullen BR (2006) APOBEC3A and APOBEC3B are potent inhibitors of LTR-retrotransposon function in human cells. *Nucleic Acids Res* 34: 89–95.
- Bogerd HP, Wiegand HL, Hulme AE, Garcia-Perez JL, O'Shea KS, et al. (2006) Cellular inhibitors of long interspersed element 1 and Alu retrotransposition. *Proc Natl Acad Sci USA* 103: 8780–8785.
- Chen H, Lilley CE, Yu Q, Lee DV, Chou J, et al. (2006) APOBEC3A is a potent inhibitor of adeno-associated virus and retrotransposons. *Curr Biol* 16: 480–485.
- Esnault C, Heidmann O, Delebecque F, Dewannieux M, Ribet D, et al. (2005) APOBEC3G cytidine deaminase inhibits retrotransposition of endogenous retroviruses. *Nature* 433: 430–433.
- Esnault C, Millet J, Schwartz O, Heidmann T (2006) Dual inhibitory effects of APOBEC family proteins on retrotransposition of mammalian endogenous retroviruses. *Nucleic Acids Res* 34: 1522–1531.
- Esnault C, Priet S, Ribet D, Heidmann O, Heidmann T (2008) Restriction by APOBEC3 proteins of endogenous retroviruses with an extracellular life cycle: *ex vivo* effects and *in vivo* “traces” on the murine IAPe and human HERV-K elements. *Retrovirology* 5: 75.
- Harris RS, Bishop KN, Sheehy AM, Craig HM, Petersen-Mahrt SK, et al. (2003) DNA deamination mediates innate immunity to retroviral infection. *Cell* 113: 803–809.
- Lochelt M, Romen F, Bastone P, Muckenfuss H, Kirchner N, et al. (2005) The antiretroviral activity of APOBEC3 is inhibited by the foamy virus accessory Bet protein. *Proc Natl Acad Sci USA* 102: 7982–7987.
- Mangeat B, Turelli P, Caron G, Friedli M, Perrin L, et al. (2003) Broad antiretroviral defence by human APOBEC3G through lethal editing of nascent reverse transcripts. *Nature* 424: 99–103.
- Mariani R, Chen D, Schrofelbauer B, Navarro F, Konig R, et al. (2003) Species-specific exclusion of APOBEC3G from HIV-1 virions by Vif. *Cell* 114: 21–31.
- Navarro F, Bollman B, Chen H, Konig R, Yu Q, et al. (2005) Complementary function of the two catalytic domains of APOBEC3G. *Virology* 333: 374–386.
- Russell RA, Wiegand HL, Moore MD, Schafer A, McClure MO, et al. (2005) Foamy virus Bet proteins function as novel inhibitors of the APOBEC3 family of innate antiretroviral defense factors. *J Virol* 79: 8724–8731.
- Sasada A, Takaori-Kondo A, Shirakawa K, Kobayashi M, Abudu A, et al. (2005) APOBEC3G targets human T-cell leukemia virus type 1. *Retrovirology* 2: 32.
- Schumacher AJ, Hache G, Macduff DA, Brown WL, Harris RS (2008) The DNA deaminase activity of human APOBEC3G is required for Ty1, MusD, and human immunodeficiency virus type 1 restriction. *J Virol* 82: 2652–2660.
- Sheehy AM, Gaddis NC, Choi JD, Malim MH (2002) Isolation of a human gene that inhibits HIV-1 infection and is suppressed by the viral Vif protein. *Nature* 418: 646–650.
- Turelli P, Mangeat B, Jost S, Vianin S, Trono D (2004) Inhibition of hepatitis B virus replication by APOBEC3G. *Science* 303: 1829.
- Vartanian JP, Guetard D, Henry M, Wain-Hobson S (2008) Evidence for editing of human papillomavirus DNA by APOBEC3 in benign and precancerous lesions. *Science* 320: 230–233.
- Coticello SG, Thomas CJ, Petersen-Mahrt SK, Neuberger MS (2005) Evolution of the AID/APOBEC family of polynucleotide (deoxy)cytidine deaminases. *Mol Biol Evol* 22: 367–377.
- Jarmuz A, Chester A, Bayliss J, Gisbourne J, Dunham I, et al. (2002) An anthropoid-specific locus of orphan C to U RNA-editing enzymes on chromosome 22. *Genomics* 79: 285–296.
- Malim MH (2009) APOBEC proteins and intrinsic resistance to HIV-1 infection. *Philos Trans R Soc Lond B Biol Sci* 364: 675–687.
- Zhang H, Yang B, Pomerantz RJ, Zhang C, Arunachalam SC, et al. (2003) The cytidine deaminase CEM15 induces hypermutation in newly synthesized HIV-1 DNA. *Nature* 424: 94–98.
- Lecossier D, Bouchonnet F, Clavel F, Hance AJ (2003) Hypermutation of HIV-1 DNA in the absence of the Vif protein. *Science* 300: 1112.
- Iwatani Y, Chan DS, Wang F, Maynard KS, Sugiura W, et al. (2007) Deaminase-independent inhibition of HIV-1 reverse transcription by APOBEC3G. *Nucleic Acids Res* 35: 7096–7108.
- Bishop KN, Verma M, Kim EY, Wolinsky SM, Malim MH (2008) APOBEC3G inhibits elongation of HIV-1 reverse transcripts. *PLoS Pathog* 4: e1000231.
- Okeoma CM, Lovsin N, Peterlin BM, Ross SR (2007) APOBEC3 inhibits mouse mammary tumour virus replication *in vivo*. *Nature* 445: 927–930.
- Santiago ML, Montano M, Benitez R, Messer RJ, Yonemoto W, et al. (2008) *ApoBec3* encodes *Rfi3*, a gene influencing neutralizing antibody control of retrovirus infection. *Science* 321: 1343–1346.

## Author Contributions

Conceived and designed the experiments: MM. Performed the experiments: JL YH ET QL CAK. Analyzed the data: YI CAK MM. Contributed reagents/materials/analysis tools: CAK MM. Wrote the paper: YH CAK MM.

- Takeda E, Tsuji-Kawahara S, Sakamoto M, Langlois MA, Neuberger MS, et al. (2008) Mouse APOBEC3 restricts Friend leukemia virus infection and pathogenesis *in vivo*. *J Virol* 82: 10998–11008.
- Low A, Okeoma CM, Lovsin N, de las Heras M, Taylor TH, et al. (2009) Enhanced replication and pathogenesis of Moloney murine leukemia virus in mice defective in the murine APOBEC3 gene. *Virology* 385: 455–463.
- Langlois MA, Kemmerich K, Rada C, Neuberger MS (2009) The AKV murine leukemia virus is restricted and hypermutated by mouse APOBEC3. *J Virol* 83: 11550–11559.
- Jern P, Stoye JP, Coffin JM (2007) Role of APOBEC3 in genetic diversity among endogenous murine leukemia viruses. *PLoS Genet* 3: 2014–2022.
- Chesebro B, Miyazawa M, Britt WJ (1990) Host genetic control of spontaneous and induced immunity to Friend murine retrovirus infection. *Annu Rev Immunol* 8: 477–499.
- Miyazawa M, Tsuji-Kawahara S, Kanari Y (2008) Host genetic factors that control immune responses to retrovirus infections. *Vaccine* 26: 2981–2996.
- Tsuji-Kawahara S, Chikaishi T, Takeda E, Kato M, Kinoshita S, et al. (2010) Persistence of viremia and production of neutralizing antibodies differentially regulated by polymorphic APOBEC3 and BAFF-R loci in Friend virus-infected mice. *J Virol* 84: 6082–6095.
- Santiago ML, Benitez RL, Montano M, Hasenkrug KJ, Greene WC (2010) Innate retroviral restriction by Apobec3 promotes antibody affinity maturation *in vivo*. *J Immunol* 185: 1114–1123.
- Miyazawa M, Nishio J, Wehrly K, Chesebro B (1992) Influence of MHC genes on spontaneous recovery from Friend retrovirus-induced leukemia. *J Immunol* 148: 644–647.
- Hakata Y, Landau NR (2006) Reversed functional organization of mouse and human APOBEC3 cytidine deaminase domains. *J Biol Chem* 281: 36624–36631.
- Sanville B, Dolan MA, Wollenberg K, Yan Y, Martin C, et al. (2010) Adaptive evolution of *Mus Apobec3* includes retroviral insertion and positive selection at two clusters of residues flanking the substrate groove. *PLoS Pathog* 6: e1000974.
- Okeoma CM, Petersen J, Ross SR (2009) Expression of murine *APOBEC3* alleles in different mouse strains and their effect on mouse mammary tumor virus infection. *J Virol* 83: 3029–3038.
- Santiago ML, Smith DS, Barrett BS, Montano M, Benitez RL, et al. (2011) Persistent Friend virus replication and disease in Apobec3-deficient mice expressing functional B-cell-activating factor receptor. *J Virol* 85: 189–199.
- Mikl MC, Watt IN, Lu M, Reik W, Davies SL, et al. (2005) Mice deficient in APOBEC2 and APOBEC3. *Mol Cell Biol* 25: 7270–7277.
- Hatakeyama S, Kitagawa M, Nakayama K, Shirane M, Matsumoto M, et al. (1999) Ubiquitin-dependent degradation of I $\kappa$ B $\alpha$  is mediated by a ubiquitin ligase Skp1/Cull1/F-box protein FWD1. *Proc Natl Acad Sci USA* 96: 3859–3863.
- Teeter KC, Thibodeau LM, Gompert Z, Buerkle CA, Nachman MW, et al. (2010) The variable genomic architecture of isolation between hybridizing species of house mice. *Evolution* 64: 472–485.
- Kozak CA, O'Neill RR (1987) Diverse wild mouse origins of xenotropic, mink cell focus-forming, and two types of ecotropic proviral genes. *J Virol* 61: 3082–3088.
- Orth A, Adama T, Din W, Bonhomme F (1998) Natural hybridization between two subspecies of the house mouse, *Mus musculus domesticus* and *Mus musculus castaneus*, near Lake Casitas, California. *Genome* 41: 104–110.
- Abudu A, Takaori-Kondo A, Izumi T, Shirakawa K, Kobayashi M, et al. (2006) Murine retrovirus escapes from murine APOBEC3 via two distinct novel mechanisms. *Curr Biol* 16: 1565–1570.
- Okeoma CM, Low A, Bailis W, Fan HY, Peterlin BM, et al. (2009) Induction of APOBEC3 *in vivo* causes increased restriction of retrovirus infection. *J Virol* 83: 3486–3495.
- Muñil B, Saueremann U, Motzkus D, Stahl-Henning C, Sopper S (2011) Increased APOBEC3G and APOBEC3F expression is associated with low viral load and prolonged survival in simian immunodeficiency virus infected rhesus monkeys. *Retrovirology* 8: 77.
- Biasin M, Piacentini L, Lo Caputo S, Kanari Y, Magri G, et al. (2007) Apolipoprotein B mRNA-editing enzyme, catalytic polypeptide-like 3G: a possible role in the resistance to HIV of HIV-exposed seronegative individuals. *J Infect Dis* 195: 960–964.
- Boursot P, Din W, Anand R, Darviche D, Dod B, et al. (1996) Origin and radiation of the house mouse: Mitochondrial DNA phylogeny. *J Evolution Biol* 9: 391–415.

50. Din W, Anand R, Boursot P, Darviche D, Dod D, et al. (1996) Origin and radiation of the house mouse: Clues from nuclear genes. *J Evolution Biol* 9: 519–539.
51. Sawyer SL, Emerman M, Malik HS (2004) Ancient adaptive evolution of the primate antiviral DNA-editing enzyme APOBEC3G. *PLoS Biol* 2: E275.
52. Stocking C, Kozak C (2008) Murine endogenous retroviruses. *Cell Mol Life Sci* 65: 3383–3398.
53. Tomonaga K, Coffin J (1998) Structure and distribution of endogenous noncancerous murine leukemia viruses in wild mice. *J Virol* 72: 8289–8300.
54. Ogawa T, Tsuji-Kawahara S, Yuasa T, Kinoshita S, Chikaishi T, et al. (2011) Natural killer cells recognize Friend retrovirus-infected erythroid progenitor cells through NKG2D-RAE-1 interactions in vivo. *J Virol* 85: 5423–5435.
55. Yamanaka S, Balestra ME, Ferrell LD, Fan J, Arnold KS, et al. (1995) Apolipoprotein B mRNA-editing protein induced hepatocellular carcinoma and dysplasia in transgenic animals. *Proc Natl Acad Sci USA* 92: 8483–8497.
56. Okazaki I, Hiai H, Kakazu N, Yamada S, Muramatsu M, et al. (2003) Constitutive expression of AID leads to tumorigenesis. *J Exp Med* 197: 1173–1181.
57. Kozak CA (2010) The mouse “xenotropic” gammaretroviruses and their XPR1 receptor. *Retrovirology* 7: 101.
58. Kidd JM, Newman TL, Tuzun E, Kaul R, Eichler EE (2007) Population stratification of a common APOBEC gene deletion polymorphism. *PLoS Genet* 3: e63.
59. Cascalho M (2004) Advantages and disadvantages of cytidine deamination. *J Immunol* 172: 6513–6518.
60. Pham P, Bransteitter R, Goodman MF (2005) Reward versus risk: DNA cytidine deaminases triggering immunity and disease. *Biochemistry* 44: 2703–2715.
61. Hartley JW, Rowe WP (1975) Clonal cell lines from a feral mouse embryo which lack host-range restrictions for murine leukemia viruses. *Virology* 65: 128–134.
62. Lander MR, Chattopadhyay SK (1984) A *Mus dummi* cell line that lacks sequences closely related to endogenous murine leukemia viruses and can be infected by ecotropic, amphotropic, xenotropic, and mink cell focus-forming viruses. *J Virol* 52: 695–698.
63. Chirgwin JM, Przybyla AE, MacDonald RJ, Rutter WJ (1979) Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease. *Biochemistry* 18: 5294–5299.
64. Zuker M, Mathews DH, Turner DH (1999) Algorithms and thermodynamics for RNA secondary structure prediction: A practical guide. In: Barciszewski J, Clark BFC, eds. *RNA Biochemistry and Biotechnology*. Dordrecht: Kluwer Academic Publishers. pp 11–43.
65. Mathews DH, Sabina J, Zuker M, Turner DH (1999) Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. *J Mol Biol* 288: 911–940.
66. Veyrunes F, Dobigny G, Yang F, O'Brien PCM, Catalan J, et al. (2006) Phylogenomics of the genus *Mus* (Rodentia: Muridae): extensive genome repatterning is not restricted to the house mouse. *Proc R Soc B* 273: 2925–2934.
67. Lundrigan BL, Jansa SA, Tucker PK (2002) Phylogenetic relationships in the genus *Mus*, based on paternally, maternally, and biparentally inherited characters. *Syst Biol* 51: 410–431.
68. Guenet JL, Bonhomme F (2003) Wild mice: an ever-increasing contribution to a popular mammalian model. *Trends Genet* 19: 24–31.