



**HAL**  
open science

## **APOBEC3G Generates Nonsense Mutations in HTLV-1 Proviral Genomes In Vivo.**

Jun Fan, Guangyong Ma, Kisato Nosaka, Junko Tanabe, Yorifumi Satou, Atsushi Koito, Simon Wain-Hobson, Jean Pierre Vartanian, Masao Matsuoka

► **To cite this version:**

Jun Fan, Guangyong Ma, Kisato Nosaka, Junko Tanabe, Yorifumi Satou, et al.. APOBEC3G Generates Nonsense Mutations in HTLV-1 Proviral Genomes In Vivo.. *Journal of Virology*, 2010, epub ahead of print. 10.1128/JVI.02239-09 . pasteur-00488728

**HAL Id: pasteur-00488728**

**<https://pasteur.hal.science/pasteur-00488728>**

Submitted on 3 Jun 2011

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

1 **APOBEC3G Generates Nonsense Mutations in HTLV-1 Proviral Genomes *In Vivo***

2 **Running title: Nonsense mutations by hA3G in HTLV-1**

3 Jun Fan<sup>1</sup>, Guangyong Ma<sup>1</sup>, Kisato Nosaka<sup>2</sup>, Junko Tanabe<sup>1</sup>, Yorifumi Satou<sup>1</sup>, Atsushi

4 Koito<sup>3</sup>, Simon Wain-Hobson<sup>4</sup>, Jean-Pierre Vartanian<sup>4</sup> and Masao Matsuoka<sup>1\*</sup>

5 <sup>1</sup>*Laboratory of Virus Control, Institute for Virus Research, Kyoto University, 53 Shogoin*

6 *Kawahara-cho, Sakyo-ku, Kyoto 606-8507, Japan*

7 <sup>2</sup>*Department of Hematology and Infectious Diseases, Kumamoto University School of*

8 *Medicine, 1-1-1 Honjo, Kumamoto 860-0811, Japan*

9 <sup>3</sup>*Department of Retrovirology and Self-Defense, Faculty of Medical and Pharmaceutical*

10 *Sciences, Kumamoto University School of Medicine, 1-1-1 Honjo, Kumamoto 860-0811,*

11 *Japan*

12 <sup>4</sup>*Molecular Retrovirology Unit, Institut Pasteur, 28 rue du Dr Roux 75724 Paris cedex*

13 *15, France*

14 Word count for abstract: 232

15 Word count for text: 5012

16 \*Corresponding author. Mailing address for M. Matsuoka: Institute for Virus Research,

17 Kyoto University, 53 Shogoin Kawahara-cho, Sakyo-ku, Kyoto 606-8507, Japan. Phone:

18 +81-75-751-4048, FAX: +81-75-751-4049. E-mail: [mmatsuok@virus.kyoto-u.ac.jp](mailto:mmatsuok@virus.kyoto-u.ac.jp)

19

1 **ABSTRACT**

2 Human T-cell leukemia virus type 1 (HTLV-1) induces cell proliferation after  
3 infection, leading to efficient transmission by cell-to-cell contact. After a long latent  
4 period, a fraction of carriers develop adult T-cell leukemia (ATL). In ATL cells, genetic  
5 changes in the *tax* gene were reported in about 10% of ATL cases. To determine genetic  
6 changes that may occur throughout the provirus, we determined the whole sequences of  
7 HTLV-1 provirus in 60 ATL cases. Abortive genetic changes including deletion,  
8 insertion and nonsense mutations were frequent in all viral genes except in the *HBZ*  
9 gene, which is transcribed from the minus strand of the virus. G-to-A base substitutions  
10 were the most frequent mutations in ATL cells. Sequence context of G-to-A mutations  
11 was in accordance with the preferred target sequence of human APOBEC3G (hA3G).  
12 Target sequences of hA3G were less frequent in the plus strand of the *HBZ* coding  
13 region than in other coding regions of HTLV-1 provirus. Nonsense mutations in viral  
14 genes including *tax* were also observed in proviruses from asymptomatic carriers,  
15 indicating that these mutations were generated during reverse transcription and prior to  
16 oncogenesis. That hA3G targets the minus strand during reverse transcription explains  
17 why the *HBZ* gene, which is encoded by the plus strand of provirus, is not susceptible to  
18 such nonsense mutations. HTLV-1 infected cells likely take advantage of hA3G to  
19 escape from the host immune system by losing expression of viral proteins.

## 1 INTRODUCTION

2 Human T-cell leukemia virus type 1 (HTLV-1) is the causative agent of both adult  
3 T-cell leukemia (ATL) and inflammatory HTLV-1 associated myelopathy/tropical  
4 spastic paraparesis (HAM/TSP) (9, 40). HTLV-1 is a complex retrovirus that encodes  
5 regulatory genes, *tax* and *rex*, and accessory genes, *p12*, *p13*, and *p30* (6, 29). An  
6 accessory gene, *HTLV-1 bZIP factor* (*HBZ*), is transcribed from the 3' long terminal  
7 repeat (LTR) as an anti-sense transcript (10, 47). Previous studies have shown that Tax  
8 expression can immortalize T-lymphocytes *in vitro* (1, 11), and *in vivo* the expression of  
9 Tax in transgenic mice causes various tumors depending on the tissue-specific promoter  
10 that expresses Tax (12, 14, 20). In more than half of human ATLS, the *tax* gene is not  
11 transcribed (41). Three mechanisms have been identified to inactivate Tax expression  
12 (29): 1) abortive genetic changes of the *tax* gene (7, 41), 2) silencing by DNA  
13 methylation in the 5'LTR (18, 43), and 3) deletion of the 5'LTR (31, 42). Since Tax is a  
14 major target of cytotoxic T-lymphocytes (CTLs) *in vivo* (17), the tumor cells might  
15 escape from the host immune system by suppressing Tax expression. However, the  
16 actual mechanism used to create genetic changes in the *tax* gene remains to be  
17 elucidated. Nonsense mutations in the *tax* gene have also been observed in HTLV-1  
18 carriers and the mechanism to generate these mutations is similarly unknown (8).

19 As a host defense against retroviruses, mammalian cells employ the APOBEC3  
20 family, which causes deamination during reverse transcription, resulting in nucleotide  
21 mutations (3, 26). Human APOBEC3G (hA3G) deaminates cytosine residues of  
22 single-stranded DNA during reverse transcription, resulting in high levels of plus-strand

1 G-to-A mutations. The human immunodeficiency virus (HIV) nucleocapsid is critical  
2 for incorporation of hA3G into virions (2, 24), while its accessory protein, Vif,  
3 counteracts hA3G. Vif inhibits hA3G packaging into HIV-1 virion through  
4 ubiquitination and proteasomal degradation of hA3G (39, 49). HTLV-1 does not encode  
5 a protein analogous to Vif that inactivates hA3G. Instead, it has been demonstrated that  
6 a domain of the HTLV-1 nucleocapsid suppresses the incorporation of hA3G into the  
7 virion (4). Consistent with this finding, it has been reported that G-to-A mutations are  
8 rare in HTLV-1 carriers (25).

9 In this study, we analyzed the sequences of whole proviruses in ATL and carrier cells,  
10 and found that most nonsense mutations in the proviruses were caused by deamination.  
11 The sites of nonsense mutation coincided with the preferred target sequences of hA3G.  
12 In ATL cases, non-sense mutations, deletions and insertions were detected in most of  
13 the viral genes except the *HBZ* gene, supporting the critical role of the *HBZ* gene in  
14 ATL.  
15

## MATERIALS AND METHODS

1  
2 **Cell lines and clinical samples.** MT-2 and MT-4 are HTLV-1-transformed cell lines.  
3 MT-1 and TL-Om1 were derived from leukemia cells. Jurkat cells were negative for  
4 HTLV-1. These cells were cultured in RPMI-1640 medium supplemented with 10%  
5 fetal bovine serum, 100 U/ml penicillin, 100 µg/ml streptomycin and 2 mM  
6 L-glutamine at 37°C under a 5% CO<sub>2</sub> atmosphere. Clinical samples were collected from  
7 60 ATL patients and 10 HTLV-1 carriers. Genomic DNA was extracted from peripheral  
8 blood mononuclear cells using standard phenol-chloroform methods. This study was  
9 conducted according to the principles expressed in the Declaration of Helsinki. The  
10 study was approved by the Institutional Review Board of Kyoto University (G204). All  
11 patients provided written informed consent for the collection of samples and subsequent  
12 analysis.

13  
14 **Sequencing of complete provirus genomes in ATL cells.** The copy number of  
15 HTLV-1 provirus in ATL cases was determined by inverse PCR (5) and ATL samples  
16 with one complete provirus were selected for direct sequencing. The complete provirus  
17 genome was amplified as two halves (fragment I and II) from genomic DNA samples  
18 derived from ATL patients. Fragment I was amplified using primers  
19 5'-TGACAATGACCATGAGCCCCAAATATCC-3' and  
20 5'-CGGCTATTAAGACCAGGAAG-3'. PCR conditions were as follows: 1 cycle of  
21 94°C for 5 min, 30 cycles of 94°C for 30 s, 64°C for 30 s and 72°C for 5 min, 1 cycle of  
22 72°C for 10 min. Fragment II was amplified using primers 5'-

1 AGAAACAAGCTCAGAAGCTA -3' and 5'-  
 2 TGTACTAAATTTCTCTCCTGAGAGTGC-3'. PCR conditions were as follows: 1  
 3 cycle of 94°C for 5 min, 30 cycles of 94°C for 30 s, 60°C for 30 s and 72°C for 5 min, 1  
 4 cycle of 72°C for 10 min. PCR products were then subjected to nested PCR to amplify  
 5 I-1 and I-1 from fragment I, and II-1 and II-2 from fragment II. Primers for I-1 were  
 6 5'-TGACAATGACCATGAGCCCCAAATATCC-3' and  
 7 5'-GAGCTTAAAGTGATCTTGG-3'. Primers for I-2 were  
 8 5'-TTCCGATAGCCTTGTCTCA-3' and 5'-CGGCTATTAAGACCAGGAAG-3'.  
 9 Primers for II-1 were 5'-TGGTATTATTCAAGCTTCC-3' and  
 10 5'-AAATGCAGGAGTTGGGGATT-3'. Primers for II-2 were  
 11 5'-AACTGTCTAGTATAGCCATC-3' and  
 12 5'-TGTACTAAATTTCTCTCCTGAGAGTGC-3'. PCR conditions for the nested PCR  
 13 were as follows: 1 cycle of 94°C for 5 min, 28 cycles of 94°C for 30 s, X °C for 30 s and  
 14 72°C for 2.5 min, 1 cycle of 72°C for 10 min. X was 60°C for I-1 and II-2, 58°C for  
 15 I-2 and II-1. The final PCR products were gel purified and sequence was determined by  
 16 direct sequencing.

17  
 18 **Sequencing of the *pol* and *tax* regions in HTLV-1 carriers.** Proviral load was  
 19 determined as previously reported (32). Herculase II Fusion DNA Polymerase  
 20 (Stratagene, La Jolla, CA), a Pfu-based DNA polymerase, was used to amplify *pol* and  
 21 *tax* regions from genomic DNA derived from HTLV-1 carriers. Primers to amplify the  
 22 *pol* region were 5'- TACACCTTGCAATCCTATGG-3' and 5'-

1 GCTAGGCTTGCCTAGATGGG-3'. PCR conditions were as follow: 1 cycle of 95°C  
 2 for 5 min, 30 cycles of 95°C for 20 s, 60 °C for 20 s and 72 °C for 50 sec, 1 cycle of 72  
 3 °C for 5 min. Primers to amplify the *tax* region were 5'-  
 4 GGTCTCCGGGCATGACACA-3' and 5'- TCTCCACGCTTTTATAGACT-3'. PCR  
 5 condition were as follow: 1 cycle of 95°C for 5 min, 28 cycles of 95°C for 20 s, 68 °C  
 6 for 20 s and 72 °C for 50 sec, 1 cycle of 72 °C for 5 min. PCR products were purified,  
 7 cloned and at least 20 clones sequenced. DNA polymerase fidelity was studied by using  
 8 a clone carrying an amplified *tax* region as template instead of genomic DNA. The *tax*  
 9 region was amplified using the same PCR conditions described above, cloned and 25  
 10 clones were sequenced.

11

12 **Sequencing of HTLV-1 genome in non-tumor cells derived from ATL samples.**

13 ATL patient-derived tumor cells are considered a mixture of a major population of  
 14 monoclonal expanded tumor cells and a minor population of HTLV-1 infected  
 15 non-tumor cells. To amplify the HTLV-1 genome from non-tumor infected cells, ATL  
 16 samples carrying defective HTLV-1 genome based on direct sequencing were selected.  
 17 Target regions were amplified by setting one of the primers at the deleted region.  
 18 Sequences were determined by direct sequencing.

19

20 **Mutation analysis.** Nucleotide sequences of all ATL samples were aligned using  
 21 GENETYX-MAC Ver.13.0 software. Mutations were determined after filtering for i)  
 22 polymorphisms, defined as a nucleotide substitution occurring in >5 cases; ii) linkers,



1 defined as >7 common nucleotide substitutions occurring in >2 cases. With these  
2 criteria, the sequence of an ATL case, ATL-25, showed the fewest mutations (3 single  
3 nucleotide substitutions with no deletions or insertions) and was selected as a standard  
4 sequence. The nucleotide positions of identified mutations were numbered relative to  
5 this standard sequence, which has been submitted to the DDBJ (DNA Data Bank of  
6 Japan) database under the accession number AB513134.

7 For carrier samples, the sequence of a given region from all clones amplified from a  
8 carrier was aligned and compared with the standard sequence described above with  
9 mutated bases corrected. When a nucleotide substitution occurred in all clones, it was  
10 considered a polymorphism and excluded from the mutation repertoire.

11

12 **cDNA synthesis and quantitative real-time RT-PCR.** Total RNA was extracted  
13 from transfectants using Trizol reagent (Invitrogen, Carlsbad, CA). RNA was treated  
14 with DNase I (Invitrogen) to eliminate the genomic DNA. cDNA was synthesized from  
15 1 µg of total RNA with the Superscript preamplification system (Invitrogen) according  
16 to the manufacturer's protocol. cDNA product was quantified by real-time RT-PCR  
17 with Power SYBR Green PCR Master Mix and 7900HT Fast Real Time PCR System  
18 (Applied Biosystems) according to the manufacturer's instructions. Specific primers for  
19 *hA3G* gene were 5'-CGCGTGCCACCATGAAGATC-3' (forward) and  
20 5'-TGTGGGTGGATCCATCGAGT-3' (reverse). Specific primers for the human *AID*  
21 gene were 5'-TTCACCGCGCGCCTCTACTT-3' (forward) and  
22 5'-GCTGTCTGGAGAGACGAACT-3' (reverse). Target cDNA was normalized to the

1 amount of endogenous mRNA of ACTB. Primers used for ACTB were  
2 5'-AGGCCAACCGCGAGAAGATG-3' (forward) and  
3 5'-CCAGAGGCGTACAGGGATAG-3' (reverse). PCRs were carried out in triplicate.  
4 Data was analyzed by the comparative  $C_t$  method according to the manufacturer's  
5 protocols.  
6  
7 ***In vitro* editing of HTLV-1 genome by hA3G.** QT6 quail cells were transfected with  
8 an infectious molecular clone of HTLV-1 (pX1MT-M) (30) in the presence or absence  
9 of hA3G-expression plasmid. Twenty-four hours post-transfection, HTLV-1-producing  
10 QT6 cells were treated with 200  $\mu$ g/ml of mitomycin C at 37°C for 30 min, well-washed  
11 and then co-cultivated 1:1 with fresh QT6 cells (13). 72 hours later, cells were collected  
12 and genomic DNA was extracted. hA3G-mediated DNA hypermutation were detected  
13 using 3DPCR that allows differential amplification of G-to-A hypermutants using a Taq  
14 DNA polymerase as previously described (25). The first round of PCR was performed  
15 with a denaturation temperature of 94°C using the primer pair  
16 5'-CTGCAGATACAAAGTTAACC-3' (forward) and  
17 5'-TGGAGGAAGGAGGGTGAAT-3' (reverse). PCR conditions were 94°C for 5 min,  
18 followed by 35 cycles of 94°C for 30 sec, 58°C for 30 sec and 72°C for 30 sec, with a  
19 final extension for 10 min at 72°C. Then, 0.5  $\mu$ l of the first-round reaction was used as  
20 template for the second round of PCR using the following primers:  
21 5'-CCATGCTTATTATCAGCCCA-3' (forward) and  
22 5'-GTTGGGGGTTGTATGAGTGA-3' (reverse). PCR program was the same as the

- 1 first round of reaction except that the denaturation gradient ranged from 94-82°C. PCR
- 2 products were purified, cloned and up to 20 clones were sequenced.
- 3

## 1 RESULTS

2 **HTLV-1 proviral sequence changes in ATL.** Previous studies showed that Tax  
3 expression was frequently lost by genetic changes in the *tax* gene, by DNA methylation  
4 of the 5'LTR, and by deletion of the 5'LTR (29). Similarly, although not well studied to  
5 date, the expression of other viral proteins might also be lost through genetic changes in  
6 the provirus sequence. To analyze the genetic changes of the HTLV-1 provirus in ATL  
7 cells, we determined the whole sequence of proviruses (9034bp) in 60 ATL cases. For  
8 this study, we chose ATL cases with HTLV-1 proviruses containing both LTRs. As  
9 shown in Table 1, we detected genetic changes (nonsense mutations, deletions and  
10 insertions) in the coding region of viral genes in 28 of 60 ATL cases (46.7%). Deletions  
11 in the proviruses were detected in 27 cases (45%) while insertions were found in 10  
12 cases (16.7%) (See Table S1 in the supplemental material). These deletions and  
13 insertions resulted in the loss or the truncation of protein(s) in 18 cases (Table 1 and see  
14 Table S1 in the supplemental material).

15 Nonsense mutations, deletions, and insertions in the viral genes were found in all  
16 viral genes except the *HBZ* gene (Table 1). Since *p30* and *p13* use an identical coding  
17 frame, the same mutations generate nonsense changes in both *p30* and *p13* (Fig. 1). The  
18 frequency of genetic changes in the *tax* gene was 16.7% (10/60) (Table 1), which was  
19 higher than in previous studies (7, 41). 27 of a total of 29 nonsense mutations  
20 preferentially accumulated in tryptophan codons (27/29: 89.7%) (See Table S2 in the  
21 supplemental material). Since the tryptophan codon is TGG, a G-to-A mutation  
22 generates either a TGA or a TAG stop codon. In one nonsense mutation, the codon for

1 arginine (CGA) was converted to the TGA stop codon by a C-to-T mutation.

2 Interestingly, nonsense mutations tended to accumulate in the same cases (Table 1).

3 It was difficult to discriminate between mutations and polymorphisms in provirus  
4 from ATL patients since the proviral sequences in non-leukemic cells could not be  
5 analyzed in most ATL cases without an internal deletion of provirus. Some mutations  
6 might be characteristic of a subgroup of proviruses (polymorphism). Since nonsense  
7 mutations at the *tax* gene (7469) were detected in 5 ATL cases, we tentatively judged  
8 base substitutions as polymorphisms when more than 5 identical base substitutions were  
9 observed in different cases. Based on this criterion, we observed 591 mutations in all  
10 cases (See Table S3 in the supplemental material). Among them, the G-to-A mutation  
11 was the most frequent, contributing to 28.8% of the total mutations. Other frequently  
12 detected mutations were C-to-T (23.7%), A-to-G (17.1%) and T-to-C (12.9%). Frequent  
13 G-to-A and C-to-T mutations strongly suggested the activity of deamination enzymes in  
14 the generation of these mutations.

15

16 **Mutations in leukemic and non-leukemic cells in ATL cases.** In ATL patients,  
17 non-leukemic HTLV-1 infected cells co-exist with leukemic cells. To analyze the  
18 proviral sequences of non-leukemic cells, we amplified proviral sequences using  
19 primers within regions deleted in leukemic cells to avoid amplification of provirus from  
20 leukemic cells. Nonsense mutations that were observed in leukemic cells were not  
21 detected in non-leukemic cells (Fig. 2), indicating that nonsense mutations generated in  
22 the provirus of leukemic cells were selected for. We compared the whole sequences of

1 HTLV-1 provirus in both leukemic cells and non-leukemic cells from five ATL cases.  
 2 As shown in Table 2, only 37% of base substitutions were considered to be real  
 3 mutations since the nucleotide differed between leukemic cells and non-leukemic cells.  
 4 However, 71% of G-to-A transitions were actually mutations while 73% of other base  
 5 substitutions were polymorphisms (Table 2), indicating that G-to-A mutations are  
 6 predominant in ATL cells.

7 **Association of G-to-A mutations with human APOBEC3G.** A high frequency of  
 8 G-to-A mutations suggests the role of deaminase(s) in generating these mutations.  
 9 Deamination enzymes are known to have individually preferred target sequences for  
 10 deamination (44). Sequences surrounding the nonsense mutations generated by the  
 11 G-to-A mutation in HTLV-1 proviruses showed a predominance of the 5'-GG  
 12 dinucleotide context (target underlined) (Fig. 3A). Among trinucleotides containing GG  
 13 dinucleotides, CGG, TGG and GGG were preferred (Fig. 3A); these are consistent with  
 14 the target sequences of hA3G (48). The AGG sequence was also targeted in HTLV-1,  
 15 unlike in HIV-1. When we checked tetrameric sequences that contain a central GG  
 16 dinucleotide, CGGG and TGGG were the preferred targets in the HTLV-1 provirus,  
 17 similar to those reported for HIV-1 (Fig. 3B). As in HIV-1, a C at the +2 position  
 18 (NGGC) was disfavored in HTLV-1. These findings suggest that the observed G-to-A  
 19 mutations were generated by hA3G. Another deaminase, APOBEC3F (hA3F), which is  
 20 largely coexpressed with hA3G, is also reported to target single-stranded minus DNA  
 21 (46). Unlike hA3G, which favors the 5'-GG dinucleotide, the consensus target sequence  
 22 of hA3F is 5'-GA (21). G-to-A mutations at GA sites contributed 13.4% of all G-to-A

1 mutations (Fig. 3A), suggesting that hA3F might also play a role in these mutations in  
2 the HTLV-1 provirus.

3 Since hA3G targets the minus strand during reverse transcription, nonsense mutations  
4 generated by G-to-A mutations are not found in the *HBZ* gene. A G-to-A mutation in  
5 the proviral sense strand caused a nonsense mutation in the *p12* gene (TGG to TGA),  
6 whereas the same mutation generated a missense mutation in the *HBZ* gene (CCA to  
7 TCA) (Fig. 1). G-to-A mutations occurred most frequently at GG sites with a T or C at  
8 their 5' terminus (Fig. 3A). We therefore counted the TGG or CGG sites within  
9 individual HTLV-1 regions (Table 3). As NGGC (N indicates T or C) was disfavored  
10 by hA3G (Fig. 3B), it was not included. Among HTLV-1 regions, the plus-strand  
11 sequence of the *HBZ* gene had the fewest TGG and CGG sites, even when the number  
12 of sites was normalized by size of the region. This indicates that mutations in the *HBZ*  
13 gene-coding region caused by hA3G are also rare.

14  
15 **Mutations in the HTLV-1 proviral sequences from carriers.** Frequent detection of  
16 G-to-A mutations in HTLV-1 proviral sequences in ATL cases suggests that these  
17 nonsense mutations potentially occur in both HTLV-1 carriers and ATL patients. To  
18 explore this possibility, we next studied the proviral sequences (*pol*, *env*, *p12*, *p13/p30*,  
19 *p27*, *tax*, and *HBZ*) in asymptomatic HTLV-1 carriers. The proviral sequences in  
20 asymptomatic carriers were amplified using a high fidelity DNA polymerase with  
21 proofreading activity, and after subcloning, the sequences were determined for at least  
22 20 subclones. As seen in ATL cases, the mutation frequency differed drastically among

1 individual carriers (representative data are shown in Table S4 and S5 in the  
2 supplemental material), and the distribution of base substitutions among clones varied  
3 within individual carriers. G-to-A mutations accounted for 81% and 72 % of all  
4 mutations in the *pol* and the *tax* genes, respectively (Table 4 and 5). High frequency of  
5 G-to-A mutations was also observed in other viral genes (data not shown). Analyses of  
6 the sequence context of G-to-A mutations showed a predominance of GG sequences,  
7 which is consistent with the finding that hA3G deaminates the viral genome during  
8 reverse transcription (48). In order to compare the frequencies of nonsense mutations in  
9 carriers versus ATL cases, the occurrence of nonsense mutations was further analyzed.  
10 Consistent with findings in ATL cases, nonsense mutations were detected in the *pol*,  
11 *env*, *p30/p13*, *p27* and *tax* genes, but not in the *p12* or *HBZ* genes (Table 6). As in ATL  
12 cases, nonsense mutations in carriers were most frequently observed in the *tax* and *pol*  
13 genes. This result suggests that there is no bias for mutations in specific viral genes in  
14 ATL cases compared with carriers. These results suggest that nonsense mutations in  
15 provirus are not generated during oncogenesis, but are present in the carrier state.

16 In addition to G-to-A mutation, C-to-T mutations were detected in the *pol* and *tax*  
17 genes in carrier cells as in ATL cells (See Table S2 in the supplemental material) as  
18 well as other viral regions (data not shown). Comparative analyses of leukemic cell and  
19 non-leukemic cells showed that 79% of C-to-T base substitutions were polymorphisms  
20 (Table 2). However, after G-to-T mutations, C-to-T mutations were still the most  
21 frequent in ATL patients and HTLV-1 carriers.

22



## 1      **Expression of the *hA3G* gene and activation-induced deaminase gene**

2      In order to analyze the correlation between G-to-A mutation and *hA3G* expression, we  
3      first studied the mRNA level of *hA3G* using real-time PCR assay. The *hA3G* gene was  
4      expressed in normal T cells and the expression level of *hA3G* moderately decreased in  
5      ATL cells. In ATL samples, there was no obvious correlation between the expression  
6      level of *hA3G* and number of G-to-A mutations in the proviral genome.

7      Although only 21% of C-to-T base substitutions were mutations (Table 2), C-to-T was  
8      the second most frequent mutation (See Table S2 in the supplemental material).

9      Activation-induced deaminase (AID) was found to be a key factor for the switch  
10     recombination of immunoglobulin (Ig), and is also implicated in hypermutations of Ig  
11     genes (15). AID deaminates cytosine to uracil. Aberrant expression of AID driven by  
12     NF- $\kappa$ B activation is found in gastric cancers and hepatomas, and AID expression has  
13     been invoked to explain induced mutations observed in cancer cells (28). Therefore, we  
14     studied *AID* expression in 7 ATL cases with different proviral mutation frequencies. As  
15     shown in Fig.4, the *AID* gene transcription increased in ATL-15. In this case, we found  
16     six C-to-T mutations in the provirus, and the sequence context showed more AID  
17     preferred sequences in C-to-T mutated sites compared to other cases without AID  
18     activation (data not shown). Thus, AID might also play a role in the deamination of the  
19     HTLV-1 genome in at least some ATL cases. Since AID is thought to target double  
20     strand DNA of provirus, it likely generates C-to-T mutations in the plus strand of  
21     provirus. However, C-to-T base substitutions contained many polymorphisms, as shown  
22     in Table 2. Therefore, correlation between C-to-T mutations and AID expression

1 remains to be studied.

2

### 3 ***In vitro* editing of the HTLV-1 genome by hA3G**

4 Previous reports have suggested that HTLV-1 is relatively resistant to the antiviral  
5 effect of hA3G (4), consistent with our finding that the frequency of G-to-A mutations  
6 throughout the HTLV-1 proviral genome was low. In order to provide direct evidence  
7 that G-to-A mutations observed in the proviruses are the result of hA3G-mediated  
8 genome editing, we studied *in vitro* the editing effect of hA3G on the HTLV-1 genome.  
9 HTLV-1 viruses were generated in the presence of exogenous hA3G to allow packaging  
10 into budding viral particles, a step that is required for exerting editing activity of the  
11 cytidine deaminase. After a round of virus infection, the HTLV-1 DNA was analyzed  
12 using a highly sensitive PCR-based protocol, referred to as 3D PCR, that is capable of  
13 amplifying G-to-A hypermutated genomes as described in the Materials & Methods.  
14 Differential amplification of a tax region from cells infected with viral particles  
15 produced in the presence of hA3G was achieved when the denaturation PCR  
16 temperature was lowered to between 84-86°C (data not shown). All sequences of  
17 differentially amplified PCR clones exhibited extensive G-to-A hypermutation, and the  
18 number (percentage) of G-to-A transitions ranged from 8 (13%) to 29 (48%) per clone  
19 (Fig. 5). In ATL cases in the same region, G-to-A mutations at four nucleotide positions  
20 resulted in premature stop codons in the *p30* and *tax* genes (Fig. 1). The occurrence of  
21 such mutations was observed *in vitro* when viruses were produced together with hA3G.  
22 G-to-A mutations at two sites, in which nonsense mutations were more frequently

1 detected in ATL cases, were also frequent among 3D PCR clones. As expected, a  
2 preference for GG dinucleotide context in hA3G-mediated G-to-A editing was observed  
3 (Fig. 5C). Our data suggested that hA3G is indeed able to actively deaminate HTLV-1  
4 viral genome during reverse transcription although with low editing frequency, leading  
5 to G-to-A mutations in the plus-strand of the provirus.  
6

## DISCUSSION

1  
2 In this study, in order to acquire a full mutation spectrum of HTLV-1 virus,  
3 sequences of proviral genomes were determined in 60 ATL cases as well as 10  
4 asymptomatic carriers. In addition to deletion and insertion, we identified base  
5 substitutions, and found that G-to-A mutations were most prevalent. In addition to the  
6 observation that G-to-A mutations in ATL and carriers occurred at the preferred target  
7 sequence of hA3G, we also experimentally confirmed that hA3G actively introduced  
8 G-to-A mutations to the HTLV-1 viral genome. It has been reported that hypermutation  
9 in the HTLV-1 genome was a rare event (35), and HTLV-1 was relatively resistant to  
10 the antiviral activity of hA3G (33). Unlike HIV, which counteracts hA3G by producing  
11 Vif, HTLV-1 does not express an accessory protein that interferes with hA3G. A  
12 peptide motif in the C-terminus of the HTLV-1 nucleocapsid has been shown to inhibit  
13 hA3G packaging into nascent virions (4). However, inhibition of hA3G induced G-to-A  
14 mutations by HTLV-1 was partial (37). Previous studies analyzed small regions of  
15 provirus amplified by PCR in carriers or HAM/TSP patients, while proviral sequences  
16 of ATL cells were determined by direct sequencing. Since previous studies used  
17 error-prone Taq polymerase, it was very difficult to identify mutations after cloning (36).  
18 The number of ATL cases was so limited in the previous studies that mutations could  
19 not be discriminated from polymorphisms (34). This study first analyzed the whole  
20 sequence of 60 ATL cases by direct sequencing, and also determined the proviral  
21 sequences of clones derived from HTLV-1 carriers using a precise DNA polymerase,  
22 which enabled us to identify mutations in HTLV-1 genomes. Consistent with previous

1 studies (34, 35), our results showed that mutations were relatively rare. However,  
2 among the rare mutations in HTLV-1 genomes, G-to-A mutations in the target  
3 sequences of hA3G were predominant. In addition to G-to-A mutations, other mutations  
4 were also observed in ATL and carriers, indicating the role of other deaminases  
5 including AID and error-prone reverse transcriptase (27).

6 This study demonstrated important roles for cytidine deaminase(s) in generating  
7 mutations in HTLV-1 proviruses. So far, at least 12 cytidine deaminases have been  
8 identified, many of which share significant sequence homology. They are considered to  
9 have evolved through gene transpositions and duplications (45). Although some  
10 members, including APOBEC1 and AID, target cellular genes, most human APOBEC  
11 proteins are thought to defend the host against retroviruses. Our study suggests that  
12 hA3G is responsible for G-to-A mutations in HTLV-1. hA3G binds to single-stranded  
13 DNA, and preferentially deaminates CCCA or CCCG sequences during reverse  
14 transcription (48). This strategy of hA3G can induce nonsense mutations in the  
15 plus-strand coding sequence since TGG is a target of hA3G, resulting in nonsense  
16 mutations like TAG or TGA. However, the *HBZ* gene is encoded by the minus strand of  
17 the provirus. The *HBZ* gene is therefore much less susceptible to nonsense mutations  
18 generated by hA3G. Furthermore, there are few target sequences for hA3G in the plus  
19 strand of the *HBZ* coding RNA. In addition, the coding sequence of the *HBZ* gene  
20 overlaps that of *p12* gene. Thus, *HBZ* further avoids missense mutation in the minus  
21 strand.

22 The frequencies of G-to-A changes in the HTLV-1 provirus (0.21% for the *pol* gene

1 and 0.11% for the *tax* gene) in the carriers are slightly lower than that for vif positive  
2 HIV-1 (0.57%) (48), indicating that HTLV-1 is resistant to hA3G during reverse  
3 transcription. Why are the few nonsense mutations that do occur retained in the  
4 HTLV-1 provirus? The difference between HTLV-1 and HIV-1 is related to their  
5 distinct strategies of propagation. HIV-1 replicates vigorously *in vivo*, producing  
6 tremendous numbers of viral particles. Viruses with nonsense mutations cannot  
7 replicate, and thereby disappear *in vivo*. On the other hand, HTLV-1 promotes the  
8 proliferation of the infected cells themselves by the action of its regulatory and  
9 accessory genes. Therefore, HTLV-1 infected cells can proliferate despite nonsense  
10 mutations occurring in most of the viral genes, provided the cells retain the minimum  
11 set of viral genes that relate to proliferation. This might be a reason why so many  
12 nonsense mutations in various viral genes remain in the provirus. The only gene without  
13 nonsense mutation is *HBZ*. It is likely that the *HBZ* gene is indispensable for  
14 proliferation of ATL cells, as our previous studies reported (38).

15 Previous studies suggest that Tax is critical for proliferation of HTLV-1 infected cells  
16 and oncogenesis. However, we reported that Tax expression is frequently disrupted by  
17 three mechanisms (29). It has been speculated that Tax expression is not necessary in  
18 the late stages of ATL. Rather, since Tax is a major target of CTLs, ATL cells without  
19 Tax expression are selected during leukemogenesis. However, as shown in this study,  
20 nonsense mutations were likely generated by hA3G during reverse transcription ,  
21 indicating that Tax expression was not necessary even in the carrier state. In one carrier,  
22 9 of 21 clones shared a nonsense mutation in the *tax* gene, demonstrating clonal

1 expansion of HTLV-1 infected cells with this mutation (Table S5). Furthermore, 7 of 60  
2 ATL cases contained nonsense mutations in the *tax* gene. HTLV-1 infected cells with  
3 nonsense mutations could proliferate *in vivo* and be transformed to ATL cells. A  
4 nonsense mutation in the *tax* gene has been detected in asymptomatic carriers (8). This  
5 study suggests that this nonsense mutation was generated by hA3G during reverse  
6 transcription. Another explanation is that loss of the expression of some viral gene(s) by  
7 nonsense mutations benefits the cells. Tax generates DNA damage that activates  
8 checkpoints (16, 22). In addition, it has been reported that HTLV-1-infected cells  
9 become cell cycle-arrested due to the activity of Tax (19, 23). Therefore, HTLV-1  
10 infected cells with nonsense mutations in the *tax* gene might have growth-advantage by  
11 losing Tax expression. We previously reported that HTLV-1 provirus lacking the  
12 5'LTR and the second exon of the *tax* gene was detected in ATL cases. By sequencing  
13 the integration sites in such defective provirus, we found 6 bp short repeats generated by  
14 integrase. This finding showed that defective proviruses that could not produce Tax  
15 were generated before integration (31). Taken together, these findings indicate that Tax  
16 is not necessary for oncogenesis at least in some ATL cases.

17 This study suggests that nonsense mutations in the HTLV-1 provirus are generated by  
18 hA3G in both ATL cases and HTLV-1 carriers. That hA3G targets the minus strand  
19 during reverse transcription explains why the *HBZ* gene is not susceptible to such  
20 nonsense mutations. On the contrary, HTLV-1 infected cells take advantage of hA3G to  
21 escape from the host immune system by losing expression of other viral proteins, while  
22 the *HBZ* gene remains intact.

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27

**ACKNOWLEDGMENTS**

We thank Kuan-Teh Jeang for helpful discussion, Aaron Coutts for his valuable suggestions and kind revision of the manuscript, and David Derse for providing pX1MT-M. This study was supported by a Grant-in-aid for Scientific Research from the Ministry of Education, Science, Sports, and Culture of Japan, a grant from the Naito Foundation, and a grant from the Sumitomo Foundation to M.M., a grant from Japan Leukemia Research Fund to Y.S. Y.S. is supported by JSPS Research Fellowships for Young Scientists.

**REFERENCES**

1. **Akagi, T., H. Ono, and K. Shimotohno.** 1995. Characterization of T cells immortalized by Tax1 of human T-cell leukemia virus type 1. *Blood* **86**:4243-9.
2. **Cen, S., F. Guo, M. Niu, J. Saadatmand, J. Deflassieux, and L. Kleiman.** 2004. The interaction between HIV-1 Gag and APOBEC3G. *J Biol Chem* **279**:33177-84.
3. **Chiu, Y. L., and W. C. Greene.** 2008. The APOBEC3 cytidine deaminases: an innate defensive network opposing exogenous retroviruses and endogenous retroelements. *Annu Rev Immunol* **26**:317-53.
4. **Derse, D., S. A. Hill, G. Princler, P. Lloyd, and G. Heidecker.** 2007. Resistance of human T cell leukemia virus type 1 to APOBEC3G restriction is mediated by elements in nucleocapsid. *Proc Natl Acad Sci U S A* **104**:2915-20.
5. **Etoh, K., S. Tamiya, K. Yamaguchi, A. Okayama, H. Tsubouchi, T. Ideta, N. Mueller, K. Takatsuki, and M. Matsuoka.** 1997. Persistent clonal proliferation of human T-lymphotropic virus type I-infected cells in vivo. *Cancer Res* **57**:4862-7.
6. **Franchini, G., R. Fukumoto, and J. R. Fullen.** 2003. T-cell control by human



- 1 T-cell leukemia/lymphoma virus type 1. *Int J Hematol* **78**:280-96.
- 2 7. **Furukawa, Y., R. Kubota, M. Tara, S. Izumo, and M. Osame.** 2001.
- 3 Existence of escape mutant in HTLV-I tax during the development of adult
- 4 T-cell leukemia. *Blood* **97**:987-93.
- 5 8. **Furukawa, Y., M. Tara, S. Izumo, K. Arimura, and M. Osame.** 2006.
- 6 HTLV-I viral escape and host genetic changes in the development of adult T cell
- 7 leukemia. *Int J Cancer* **118**:381-7.
- 8 9. **Gallo, R. C.** 2005. The discovery of the first human retrovirus: HTLV-1 and
- 9 HTLV-2. *Retrovirology* **2**:17.
- 10 10. **Gaudray, G., F. Gachon, J. Basbous, M. Biard-Piechaczyk, C. Devaux, and**
- 11 **J. M. Mesnard.** 2002. The complementary strand of the human T-cell leukemia
- 12 virus type 1 RNA genome encodes a bZIP transcription factor that
- 13 down-regulates viral transcription. *J Virol* **76**:12813-22.
- 14 11. **Grassmann, R., C. Dengler, I. Muller-Fleckenstein, B. Fleckenstein, K.**
- 15 **McGuire, M. C. Dokhelar, J. G. Sodroski, and W. A. Haseltine.** 1989.
- 16 Transformation to continuous growth of primary human T lymphocytes by
- 17 human T-cell leukemia virus type I X-region genes transduced by a Herpesvirus
- 18 saimiri vector. *Proc Natl Acad Sci U S A* **86**:3351-5.
- 19 12. **Grossman, W. J., J. T. Kimata, F. H. Wong, M. Zutter, T. J. Ley, and L.**
- 20 **Ratner.** 1995. Development of leukemia in mice transgenic for the tax gene of
- 21 human T-cell leukemia virus type I. *Proc Natl Acad Sci U S A* **92**:1057-61.
- 22 13. **Harris, R. S., and M. T. Liddament.** 2004. Retroviral restriction by APOBEC
- 23 proteins. *Nat Rev Immunol* **4**:868-77.
- 24 14. **Hasegawa, H., H. Sawa, M. J. Lewis, Y. Orba, N. Sheehy, Y. Yamamoto, T.**
- 25 **Ichinohe, Y. Tsunetsugu-Yokota, H. Katano, H. Takahashi, J. Matsuda, T.**
- 26 **Sata, T. Kurata, K. Nagashima, and W. W. Hall.** 2006. Thymus-derived
- 27 leukemia-lymphoma in mice transgenic for the Tax gene of human
- 28 T-lymphotropic virus type I. *Nat Med* **12**:466-72.
- 29 15. **Honjo, T., H. Nagaoka, R. Shinkura, and M. Muramatsu.** 2005. AID to
- 30 overcome the limitations of genomic information. *Nat Immunol* **6**:655-61.
- 31 16. **Jin, D. Y., F. Spencer, and K. T. Jeang.** 1998. Human T cell leukemia virus
- 32 type 1 oncoprotein Tax targets the human mitotic checkpoint protein MAD1.
- 33 *Cell* **93**:81-91.

- 1 17. **Kannagi, M., S. Harada, I. Maruyama, H. Inoko, H. Igarashi, G.**  
2 **Kuwashima, S. Sato, M. Morita, M. Kidokoro, M. Sugimoto, and et al.** 1991.  
3 Predominant recognition of human T cell leukemia virus type I (HTLV-I) pX  
4 gene products by human CD8<sup>+</sup> cytotoxic T cells directed against HTLV-  
5 I-infected cells. *Int Immunol* **3**:761-7.
- 6 18. **Koiwa, T., A. Hamano-Usami, T. Ishida, A. Okayama, K. Yamaguchi, S.**  
7 **Kamihira, and T. Watanabe.** 2002. 5'-long terminal repeat-selective CpG  
8 methylation of latent human T-cell leukemia virus type 1 provirus in vitro and in  
9 vivo. *J Virol* **76**:9389-97.
- 10 19. **Kuo, Y. L., and C. Z. Giam.** 2006. Activation of the anaphase promoting  
11 complex by HTLV-1 tax leads to senescence. *EMBO J* **25**:1741-52.
- 12 20. **Lairmore, M. D., L. Silverman, and L. Ratner.** 2005. Animal models for  
13 human T-lymphotropic virus type 1 (HTLV-1) infection and transformation.  
14 *Oncogene* **24**:6005-15.
- 15 21. **Liddament, M. T., W. L. Brown, A. J. Schumacher, and R. S. Harris.** 2004.  
16 APOBEC3F properties and hypermutation preferences indicate activity against  
17 HIV-1 in vivo. *Curr Biol* **14**:1385-91.
- 18 22. **Liu, B., S. Hong, Z. Tang, H. Yu, and C. Z. Giam.** 2005. HTLV-I Tax directly  
19 binds the Cdc20-associated anaphase-promoting complex and activates it ahead  
20 of schedule. *Proc Natl Acad Sci U S A* **102**:63-8.
- 21 23. **Liu, M., L. Yang, L. Zhang, B. Liu, R. Merling, Z. Xia, and C. Z. Giam.**  
22 2008. Human T-cell leukemia virus type 1 infection leads to arrest in the G1  
23 phase of the cell cycle. *J Virol* **82**:8442-55.
- 24 24. **Luo, K., B. Liu, Z. Xiao, Y. Yu, X. Yu, R. Gorelick, and X. F. Yu.** 2004.  
25 Amino-terminal region of the human immunodeficiency virus type 1  
26 nucleocapsid is required for human APOBEC3G packaging. *J Virol*  
27 **78**:11841-52.
- 28 25. **Mahieux, R., R. Suspene, F. Delebecque, M. Henry, O. Schwartz, S.**  
29 **Wain-Hobson, and J. P. Vartanian.** 2005. Extensive editing of a small fraction  
30 of human T-cell leukemia virus type 1 genomes by four APOBEC3 cytidine  
31 deaminases. *J Gen Virol* **86**:2489-94.
- 32 26. **Malim, M. H., and M. Emerman.** 2008. HIV-1 accessory proteins--ensuring  
33 viral survival in a hostile environment. *Cell Host Microbe* **3**:388-98.

- 1 27. **Mansky, L. M.** 2000. In vivo analysis of human T-cell leukemia virus type 1  
2 reverse transcription accuracy. *J Virol* **74**:9525-31.
- 3 28. **Matsumoto, Y., H. Marusawa, K. Kinoshita, Y. Endo, T. Kou, T. Morisawa,**  
4 **T. Azuma, I. M. Okazaki, T. Honjo, and T. Chiba.** 2007. Helicobacter pylori  
5 infection triggers aberrant expression of activation-induced cytidine deaminase  
6 in gastric epithelium. *Nat Med* **13**:470-6.
- 7 29. **Matsuoka, M., and K. T. Jeang.** 2007. Human T-cell leukaemia virus type 1  
8 (HTLV-1) infectivity and cellular transformation. *Nat Rev Cancer* **7**:270-80.
- 9 30. **Mitchell, M. S., E. T. Bodine, S. Hill, G. Princler, P. Lloyd, H. Mitsuya, M.**  
10 **Matsuoka, and D. Derse.** 2007. Phenotypic and genotypic comparisons of  
11 human T-cell leukemia virus type 1 reverse transcriptases from infected T-cell  
12 lines and patient samples. *J Virol* **81**:4422-8.
- 13 31. **Miyazaki, M., J. Yasunaga, Y. Taniguchi, S. Tamiya, T. Nakahata, and M.**  
14 **Matsuoka.** 2007. Preferential Selection of Human T-Cell Leukemia Virus Type  
15 1 Provirus Lacking the 5' Long Terminal Repeat during Oncogenesis. *J Virol*  
16 **81**:5714-23.
- 17 32. **Miyazato, P., J. Yasunaga, Y. Taniguchi, Y. Koyanagi, H. Mitsuya, and M.**  
18 **Matsuoka.** 2006. De novo human T-cell leukemia virus type 1 infection of  
19 human lymphocytes in NOD-SCID, common gamma-chain knockout mice. *J*  
20 *Virol* **80**:10683-91.
- 21 33. **Navarro, F., B. Bollman, H. Chen, R. Konig, Q. Yu, K. Chiles, and N. R.**  
22 **Landau.** 2005. Complementary function of the two catalytic domains of  
23 APOBEC3G. *Virology* **333**:374-86.
- 24 34. **Okazaki, S., R. Moriuchi, N. Yosizuka, K. Sugahara, T. Maeda, I. Jinnai, M.**  
25 **Tomonaga, S. Kamihira, and S. Katamine.** 2001. HTLV-1 proviruses  
26 encoding non-functional TAX in adult T-cell leukemia. *Virus Genes* **23**:123-35.
- 27 35. **Renjifo, B., I. Borrero, and M. Essex.** 1995. Tax mutation associated with  
28 tropical spastic paraparesis/human T-cell leukemia virus type I-associated  
29 myelopathy. *J Virol* **69**:2611-6.
- 30 36. **Saito, M., Y. Furukawa, R. Kubota, K. Usuku, S. Sonoda, S. Izumo, M.**  
31 **Osame, and M. Yoshida.** 1995. Frequent mutation in pX region of HTLV-1 is  
32 observed in HAM/TSP patients, but is not specifically associated with the  
33 central nervous system lesions. *J Neurovirol* **1**:286-94.

- 1 37. **Sasada, A., A. Takaori-Kondo, K. Shirakawa, M. Kobayashi, A. Abudu, M.**  
2 **Hishizawa, K. Imada, Y. Tanaka, and T. Uchiyama.** 2005. APOBEC3G  
3 targets human T-cell leukemia virus type 1. *Retrovirology* **2**:32.
- 4 38. **Satou, Y., J. Yasunaga, M. Yoshida, and M. Matsuoka.** 2006. HTLV-I basic  
5 leucine zipper factor gene mRNA supports proliferation of adult T cell leukemia  
6 cells. *Proc Natl Acad Sci U S A* **103**:720-5.
- 7 39. **Takatsuki, K.** 2005. Discovery of adult T-cell leukemia. *Retrovirology* **2**:16.
- 8 40. **Takeda, S., M. Maeda, S. Morikawa, Y. Taniguchi, J. Yasunaga, K. Nosaka,**  
9 **Y. Tanaka, and M. Matsuoka.** 2004. Genetic and epigenetic inactivation of tax  
10 gene in adult T-cell leukemia cells. *Int J Cancer* **109**:559-67.
- 11 41. **Tamiya, S., M. Matsuoka, K. Etoh, T. Watanabe, S. Kamihira, K.**  
12 **Yamaguchi, and K. Takatsuki.** 1996. Two types of defective human  
13 T-lymphotropic virus type I provirus in adult T-cell leukemia. *Blood*  
14 **88**:3065-73.
- 15 42. **Taniguchi, Y., K. Nosaka, J. Yasunaga, M. Maeda, N. Mueller, A. Okayama,**  
16 **and M. Matsuoka.** 2005. Silencing of human T-cell leukemia virus type I gene  
17 transcription by epigenetic mechanisms. *Retrovirology* **2**:64.
- 18 43. **Trono, D.** 2004. Retroviruses under editing crossfire: a second member of the  
19 human APOBEC3 family is a Vif-blockable innate antiretroviral factor. *EMBO*  
20 *Rep* **5**:679-80.
- 21 44. **Turelli, P., and D. Trono.** 2005. Editing at the crossroad of innate and adaptive  
22 immunity. *Science* **307**:1061-5.
- 23 45. **Wiegand, H. L., B. P. Doehle, H. P. Bogerd, and B. R. Cullen.** 2004. A  
24 second human antiretroviral factor, APOBEC3F, is suppressed by the HIV-1 and  
25 HIV-2 Vif proteins. *EMBO J* **23**:2451-8.
- 26 46. **Yoshida, M., Y. Satou, J. Yasunaga, J. Fujisawa, and M. Matsuoka.** 2008.  
27 Transcriptional control of spliced and unspliced human T-cell leukemia virus  
28 type 1 bZIP factor (HBZ) gene. *J Virol* **82**:9359-68.
- 29 47. **Yu, Q., R. Konig, S. Pillai, K. Chiles, M. Kearney, S. Palmer, D. Richman, J.**  
30 **M. Coffin, and N. R. Landau.** 2004. Single-strand specificity of APOBEC3G  
31 accounts for minus-strand deamination of the HIV genome. *Nat Struct Mol Biol*  
32 **11**:435-42.
- 33

1 **FIGURE LEGENDS**

2 **Fig. 1. Schema of the distribution of nonsense mutations in HTLV-1 provirus.** The  
 3 *p13* gene has the same reading frame as the *p30* gene. A star indicates a nonsense  
 4 mutation observed in an ATL case.

6 **Fig. 2. Mutations in non-tumor cell-derived proviral sequence in ATL patients.**

7 HTLV-1 genomes from two ATL samples with internal deleted regions were  
 8 schematically demonstrated. As indicated by horizontal arrows, one of the PCR primers  
 9 was set in the deleted region in order to specifically amplify a proviral region originated  
 10 from non-tumor cells (See Materials & Methods). Nested PCR products, represented by  
 11 short lines, were subject to direct sequencing. Nucleotide bases mutated in tumor cells  
 12 according to the criteria mentioned in Materials & Methods were displayed. A base  
 13 substitution is considered as a polymorphism if the same base was found also in the  
 14 non-tumor cells derived from the same patient sample. Positions underlined are those at  
 15 which a G-to-A mutation will result in a nonsense mutation.

17 **Fig. 3. Target sequence preference of G-to-A mutations. (A)** Sequence context of  
 18 G-to-A mutations. In all G-to-A mutations, the percentage of ones at the indicated  
 19 dinucleotide sequence context is shown (left panel). In GG context, the influence of 5'  
 20 (middle panel) and 3' (right panel) nucleotides neighboring the GG dinucleotide on  
 21 G-to-A mutations is displayed. **(B)** Frequency of G-to-A mutations on sites of  
 22 tetranucleotides containing a central GG. HTLV-1 genome sequence from ATL-25 was

1 used as a standard sequence with mutated bases corrected to calculate the total number  
 2 of GG sites that exist in the provirus genome (See Materials & Methods). The  
 3 percentage of GG sites that underwent G-to-A mutation for at least once in 60 ATL  
 4 cases was plotted against each sequence context. Mutated bases are underlined. CGGG  
 5 and TGGG were the preferred targets for G-to-A mutation in the HTLV-1 provirus. A C  
 6 at the +2 position (NGGC) was disfavored.

7  
 8 **Fig. 4. Expression of the *hA3G* and *AID* gene in ATL cases and in HTLV-1**  
 9 **associated cell lines.** Expression of the *hA3G* (A) and *AID* gene (B) was investigated  
 10 by real-time PCR. Normal resting T cells isolated from 3 healthy blood donors were  
 11 used as a control. *ACTN* was used as an internal control. Relative quantification was  
 12 performed using a comparative Ct method (Applied Biosystems). Expression of *hA3G*  
 13 and *AID* in resting T cells was artificially set as 1. Number of nucleotide mutations in  
 14 the 7 ATL cases was shown below. Among mutated sites, number of AID preferred  
 15 sites was demonstrated in parenthesis. (C) Number of nucleotide mutations in the 7  
 16 ATL cases. Among mutated sites, the number of AID preferred sites was demonstrated  
 17 in parenthesis.

18  
 19  
 20 **Fig. 5. In vitro editing of HTLV-1 genome by hA3G.** (A) Mutation matrix for  
 21 hyperedited *tax* gene sequences derived from cloned 86°C 3DPCR products. n indicates  
 22 the number of bases sequenced. (B) A selection of hypermutated HTLV-1 *tax* sequences

1 from tissue culture. The sequence is given with respect to the viral plus strand. Only  
2 differences are shown. # and % denotes respectively the number of mutations and the  
3 percentage of G-to-A transitions per sequence. The base composition of the sequence is  
4 75 T (24 %), 113 C (36.2 %), 60 G (19.2 %) and 64 A (20.5 %). Position number of the  
5 first base is indicated above the base. Arrows indicate positions where G-to-A-mediated  
6 nonsense mutations were detected in ATL cases. (C) Bulk dinucleotide context of  
7 HTLV tax region DNA strand by hA3G. The ordinate represents the substitution  
8 frequency as a function of the 3' nucleotide. Mutated G residues are underlined. A  $\chi^2$   
9 analysis showed that the observed frequencies by hA3G deviated significantly from the  
10 expected values ( $p < 0.001$ ).

1

Table 1. Abortive genetic changes of HTLV-1 viral genes in 60 ATL cases<sup>a</sup>

	<b>gag</b>	<b>pol</b>	<b>env</b>	<b>p12</b>	<b>p30</b>	<b>p13</b>	<b>p27</b>	<b>tax</b>	<b>HBZ</b>
ATL-2		DEL							
ATL-3								W248*	
ATL-7			DEL						
ATL-8			DEL				W139*		
ATL-10	DEL	DEL	DEL						
ATL-11		W680*	W427*						
ATL-12	DEL		W387*		W165*	W11*	W139*	W56*	
ATL-13	W59*	W680*	DEL		W165*	W11*		W56*	
					W190*	W36*			
ATL-14							DEL	DEL	
ATL-15		W680*							
ATL-27		DEL							
ATL-29			W431*		W165*	W11*		W56*	
ATL-32		W271*							
ATL-36								DEL	
ATL-37		W485*	W50*					W56*	
		W680*							
ATL-39				W87*					
ATL-40					R184*	R30*			
ATL-41		W485*						W56*	
ATL-42			DEL						
ATL-43		DEL							
ATL-44	DEL	DEL		W87*					
ATL-45	DEL							Y286*	
ATL-46	DEL	DEL							
ATL-47	DEL	DEL							
ATL-48	DEL							IN	
ATL-49	DEL	DEL							
ATL-51					W178*	W24*			
ATL-53	DEL								

<sup>a</sup>Only cases with deletion, insertion or non-sense mutations are shown. DEL, deletion.

IN, insertion. W, Tryptophan. \*, stop codon. Number indicates position of non-sense mutation.



Table 2. Concordance rate of mutations between leukemic cells and non-leukemic cells in five ATL cases.

	Total <sup>a</sup>	G-A	C-T	A-G	T-C	A-C	A-T	C-G	C-A	T-G	G-C	G-T	T-A
ATL-8	10 (4)	2 (1)	3 (1)	2 (0)	1 (0)	-	-	1 (1)	-	-	1 (1)	-	-
ATL-12	25 (8)	8 (6)	7 (2)	5 (0)	3 (0)	-	-	2 (0)	-	-	-	-	-
ATL-13	36 (20)	10 (8)	8 (3)	7 (2)	2 (1)	2 (0)	-	-	1 (0)	1 (0)	3 (3)	1 (0)	1 (0)
ATL-14	12 (2)	1 (1)	7 (0)	1 (0)	2 (0)	-	-	-	1 (1)	-	-	-	-
ATL-45	16 (3)	3 (1)	3 (0)	4 (0)	2 (1)	-	3 (0)	-	1 (1)	-	-	-	-
Total	99 (37)	24 (17)	28 (6)	19 (2)	10 (2)	2 (0)	3 (0)	3 (1)	3 (2)	1 (0)	4 (4)	1 (0)	1 (0)
Concordance rate (%)	37%	71%	21%	11%	20%	0%	0%	33%	67%	0%	100%	0%	0%

Each entry  $A(B)$  shows the number  $A$  of each type of mutation in ATL judged by our criteria and the number  $B$  of confirmed mutations. Concordance rate of mutations was calculated by dividing  $B$  by  $A$ .

<sup>a</sup>Number of mutations of all types in each ATL case.

Table 3. Distribution of hA3G target sequence in the plus strand of HTLV-1 genome.

Genome Regions	5'LTR	<i>gag</i>	<i>pol</i>	<i>env</i>	<i>p27</i>	<i>p30</i>	<i>tax</i>	<i>HBZ</i>	3'LTR
Size (bp)	755	1290	2691	1467	570	726	1062	621	755
CGGX <sup>a</sup>	8	12	16	8	3	2	5	1	7
CGGX/1 kb	10.60	9.3	5.95	5.45	5.26	2.75	4.71	1.61	9.27
TGGX	3	10	25	17	10	7	11	1	3
TGGX/1 kb	3.97	7.75	9.29	11.59	17.54	9.64	10.36	1.61	3.97

<sup>a</sup>X indicates A, T or G.

Table 4. Mutations in the *pol* region of 10 HTLV-1 carriers<sup>a</sup>

	Number and type of single nucleotide substitution									Sequence context of G-to-A mutations			
	<u>G</u> -A	C-T	A-G	T-C	A-C	C-G	G-C	G-T	T-A	<u>GA</u>	<u>GC</u>	<u>GT</u>	<u>GG</u>
C72 (25)	5	1					1						5
C79 (20)	8		1		1			1		1	1		6
C63 (28)	1							1				1	
C40 (26)	13		1										13
C32 (22)		2											
C62 (22)	1	2											1
C87 (22)	14									1	1		13
C5 (23)	19	1						1	2				19
C29 (25)	9					1		1			1		8
C82 (27)	7			1							1		6
Total	77	6	2	1	1	1	1	4	2	2	4	1	71

<sup>a</sup>Number of clones studied for each samples was indicated in the brackets. Mutated

bases are underlined.

Table 5. Mutations in the *tax* region of 10 HTLV-1 carriers<sup>a</sup>

	Number and type of single nucleotide substitution								Sequence context of G-to-A mutations			
	G-A	C-T	A-G	A-C	C-A	G-C	G-T	T-A	<u>GA</u>	<u>GC</u>	<u>GT</u>	<u>GG</u>
C72 (24)	3								1			2
C79 (24)	2	4		2	1	1					2	
C63 (21)	19											19
C40 (21)	2											2
C32 (25)	4	1										4
C62 (27)	5	1								1		4
C87 (24)	1	1							1			
C5 (26)	2											2
C29 (25)	3		3					1				3
C82 (22)	5						1					5
Total	46	7	3	2	1	1	1	1	2	1	2	41

<sup>a</sup>Number of clones studied for each samples was indicated in the brackets. Mutated bases are underlined.

Table 6. Nonsense mutations in ten HTLV-1 carriers.

Genes	pol		env						p12	p30/p13	p27 <sup>a</sup>		tax		HBZ
	3955	4540	5465	6158	6342	6362	6482	6494	-	7361	7652	7660	7469	8045	-
C72	1/25	-/25	-/23	-/23	-/23	-/23	-/23	-/23	-/31	-/31	-/31	-/31	1/24	-/24	-/31
C79	2/20	2/20	-/20	-/20	-/20	1/20	-/20	-/20	-/24	-/24	-/24	-/24	1/24	-/24	-/24
C63	-/28	-/28	-/20	-/20	-/20	-/20	-/20	-/20	-/24	2/24	-/24	-/24	9/21	-/21	-/24
C40	-/26	13/26	-/22	-/22	-/22	-/22	-/22	-/22	-/30	-/30	-/30	-/30	2/21	-/21	-/30
C32	-/22	-/22	-/22	-/22	-/22	-/22	-/22	-/22	-/32	-/32	-/32	1/32	-/25	-/25	-/32
C62	1/22	-/22	-/25	-/25	2/25	-/25	-/25	-/25	-/28	1/28	1/28	1/27	-/27	-/28	
C87	1/22	3/22	-/20	-/20	-/20	-/20	-/20	-/20	-/21	-/21	1/21	-/21	-/24	-/24	-/21
C5	-/22	19/22	-/20	-/20	-/20	1/20	-/20	-/20	-/24	-/24	-/24	1/24	1/26	-/26	-/24
C29	-/25	3/25	-/20	1/20	-/20	-/20	-/20	1/20	-/24	2/24	5/24	-/24	1/25	-/25	-/24
C82	3/27	2/27	1/20	-/20	-/20	-/20	1/20	-/20	-/24	1/24	-/24	-/24	1/22	1/22	-/24

Each entry *A/B* in the table shows the number *A* of clones harboring nonsense mutations at a given nucleotide position, from a given carrier, along with the number of clones *B* analyzed in that region from that carrier.

<sup>a</sup>Sequences of exon 2 were determined.

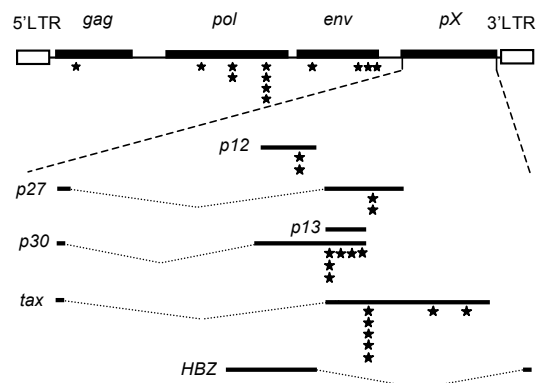


Figure 1

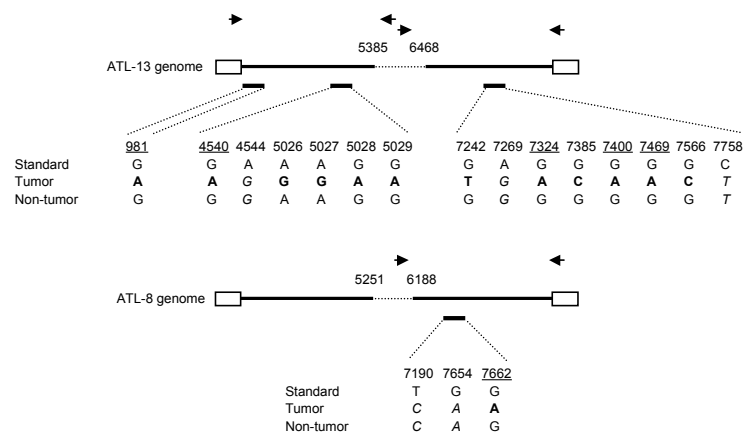


Figure 2

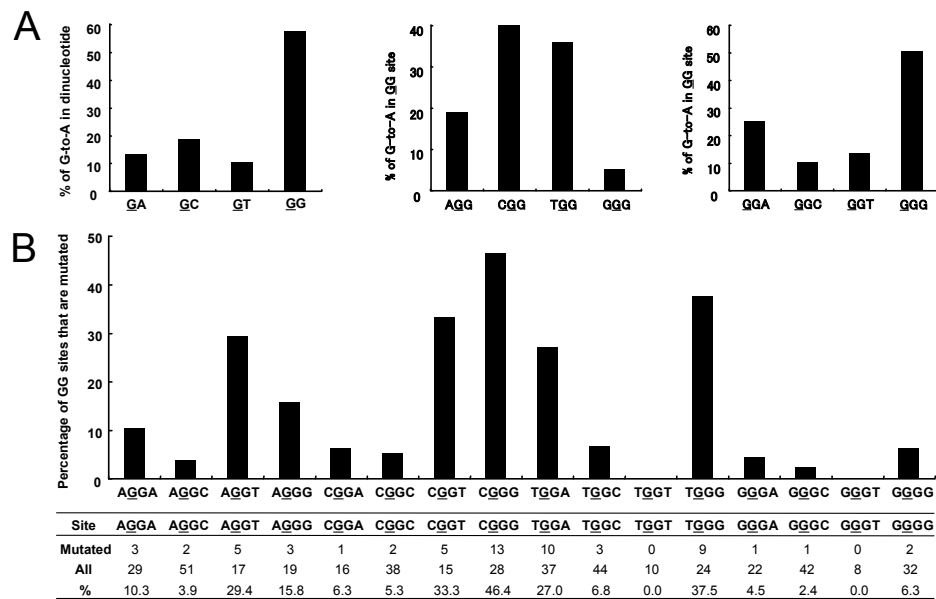


Figure 3



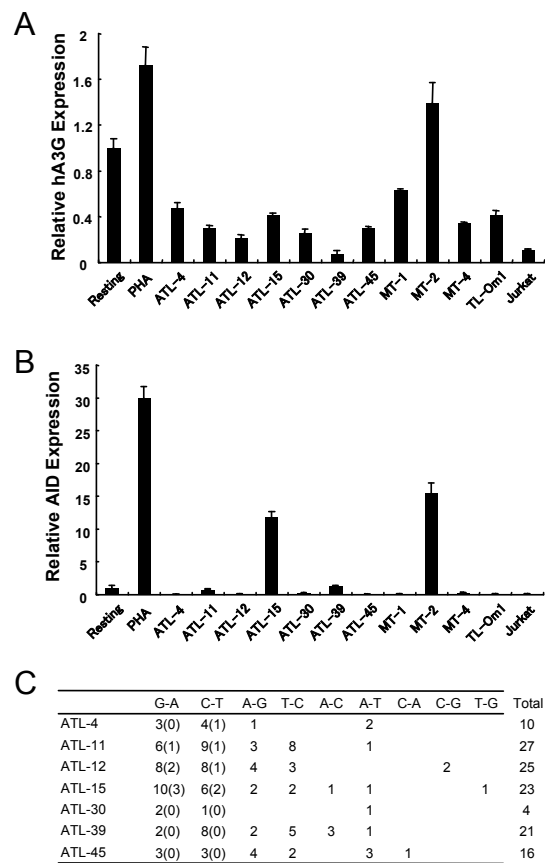


Figure 4

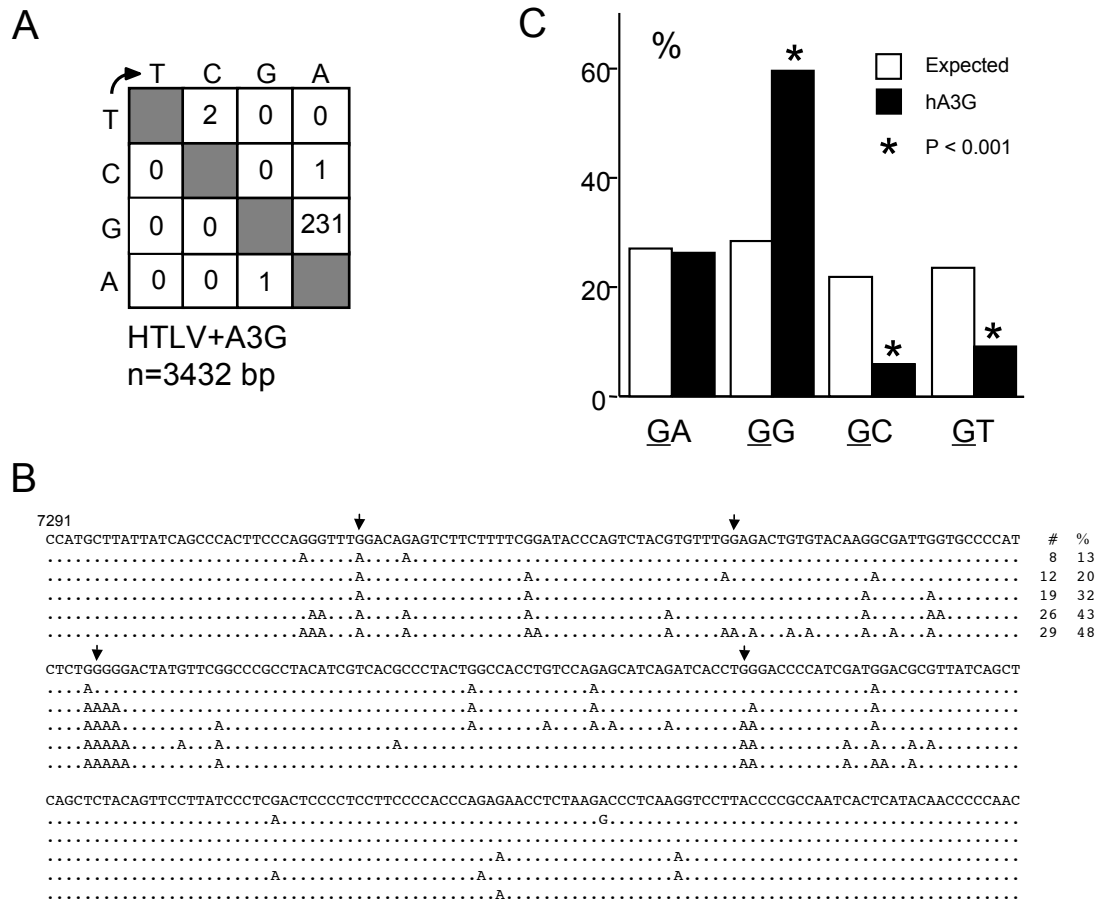


Figure 5