**SUPPLEMENTARY MATERIALS ACCOMPANYING:**

# APOBEC3 proteins mediate the clearance of foreign DNA from human cells

Mark D. Stenglein, Michael B. Burns, Ming Li, Joy Lengyel & Reuben S. Harris

Department of Biochemistry, Molecular Biology and Biophysics, Institute for Molecular Virology, Center for Genome Engineering

University of Minnesota, Minneapolis, MN 55455, USA

Correspondence should be addressed to R.S.H. (rsh@umn.edu).

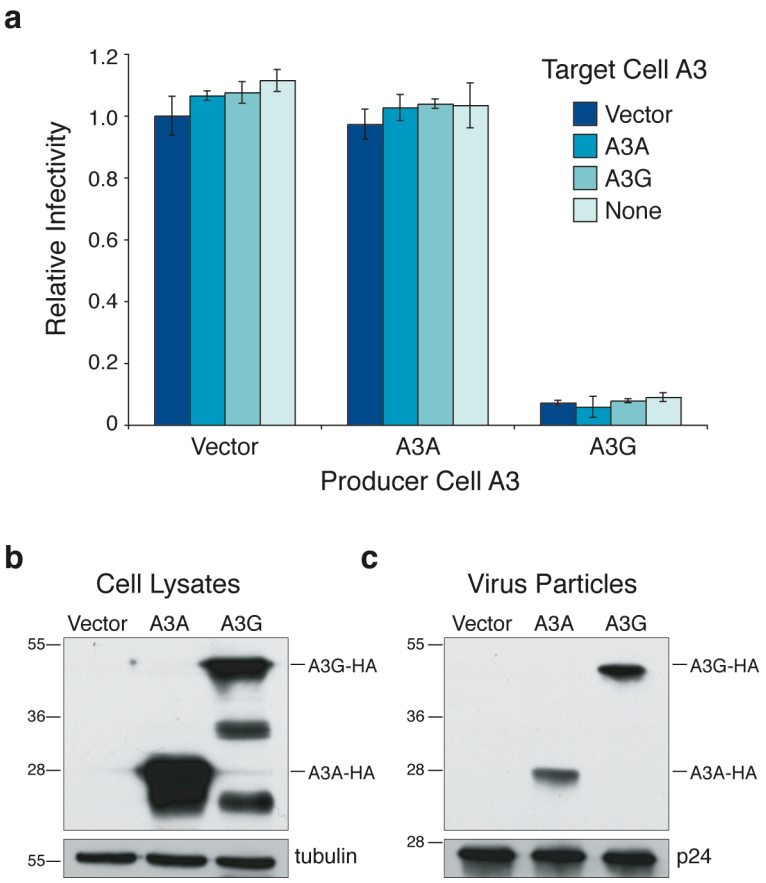
**This document contains:**

**Supplementary Figures 1-7**

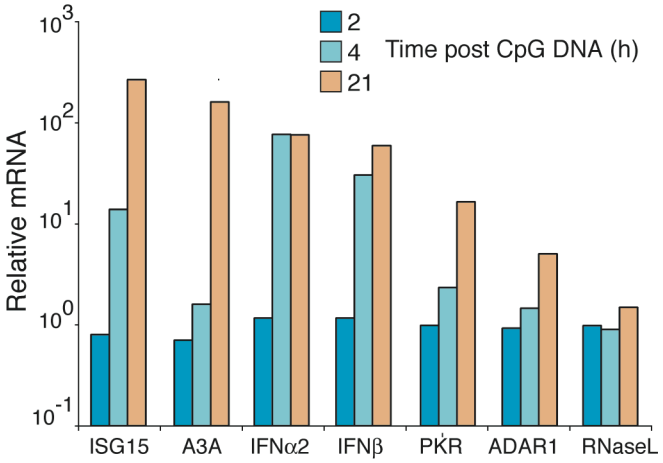
**Supplementary Tables 1-4**

**Full Methods**

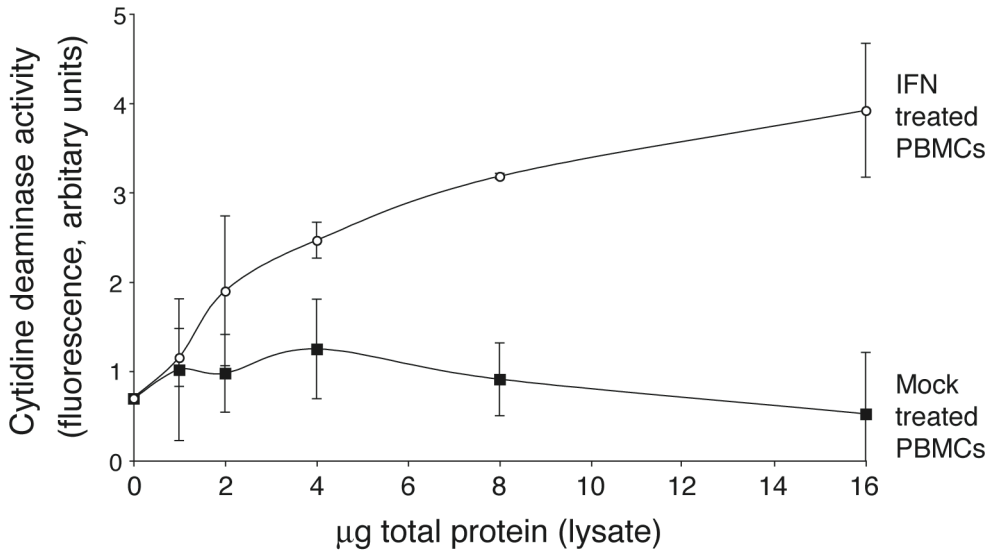
**Additional References**



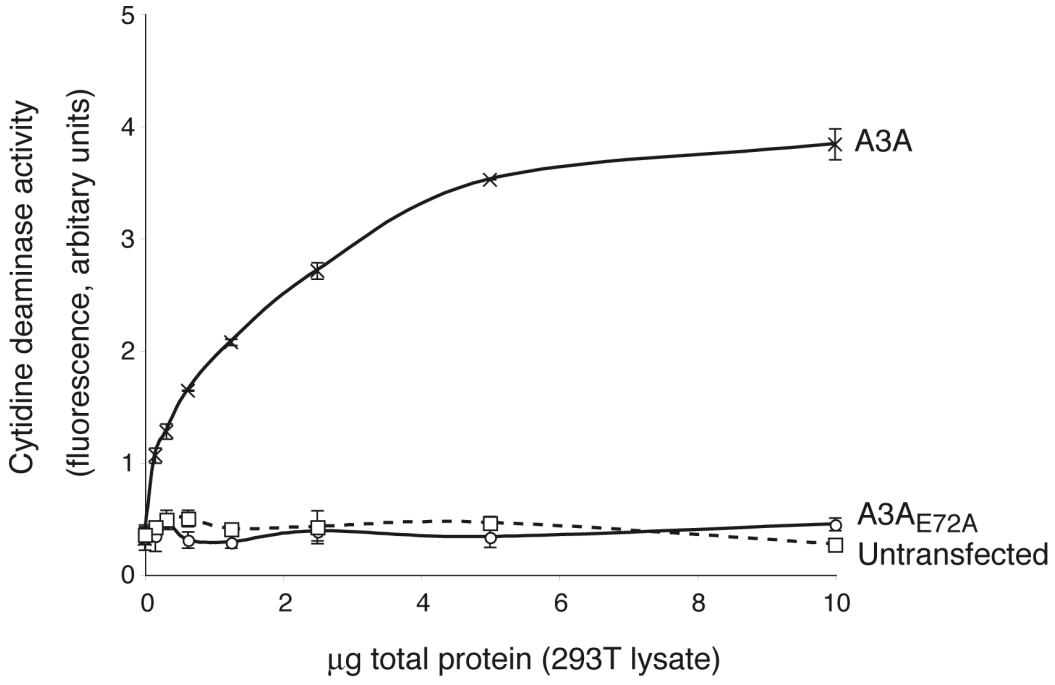

**Supplementary Figure 1** Producer or target cell A3A has no impact on HIV-1 infectivity. (**a**) The relative infectivity of Vif-deficient HIV-GFP. The data are the average and s.d. of three experiments, with the vector control infection normalized to 1. Viruses were produced by transient transfection of HEK-293T cells expressing a vector control, A3A-HA, or A3G-HA, and 48h later equivalent amounts of virus-containing supernatants were used to infect target cells expressing a vector control, A3A-HA, A3G-HA, or no pDNA. These target cells had also been co-transfected with a cherry fluorescent protein expression construct to ensure by flow cytometry that a minimum of 50% of the cells were transfected (57, 58, and 59%, respectively). (**b** and **c**) Anti-HA immunoblots of producer cell and viral particle lysates to verify A3A-HA and A3G-HA expression. Anti-tubulin and anti-p24 blots serve as loading controls for producer cells and virus particles, respectively.



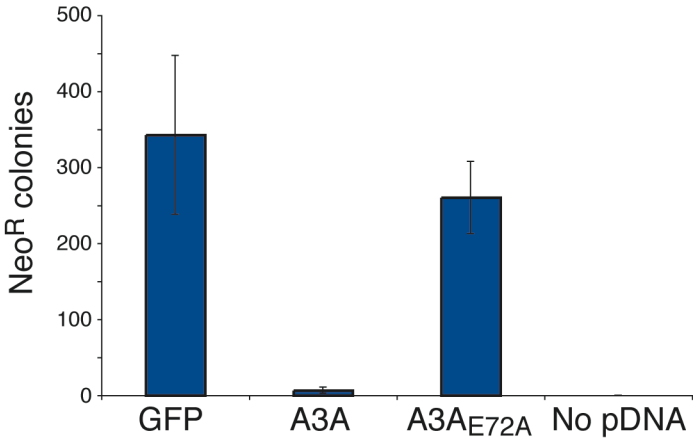

**Supplementary Figure 2** APOBEC3A mRNA induction by CpG DNA relative to the indicated IFN-responsive genes in PBMCs. Cells were treated with CpG DNA for the indicated times, after which mRNA was harvested, used to synthesize cDNA, and quantified as described in the **Online Methods**. mRNA levels are relative to those measured in untreated cells at each time point.



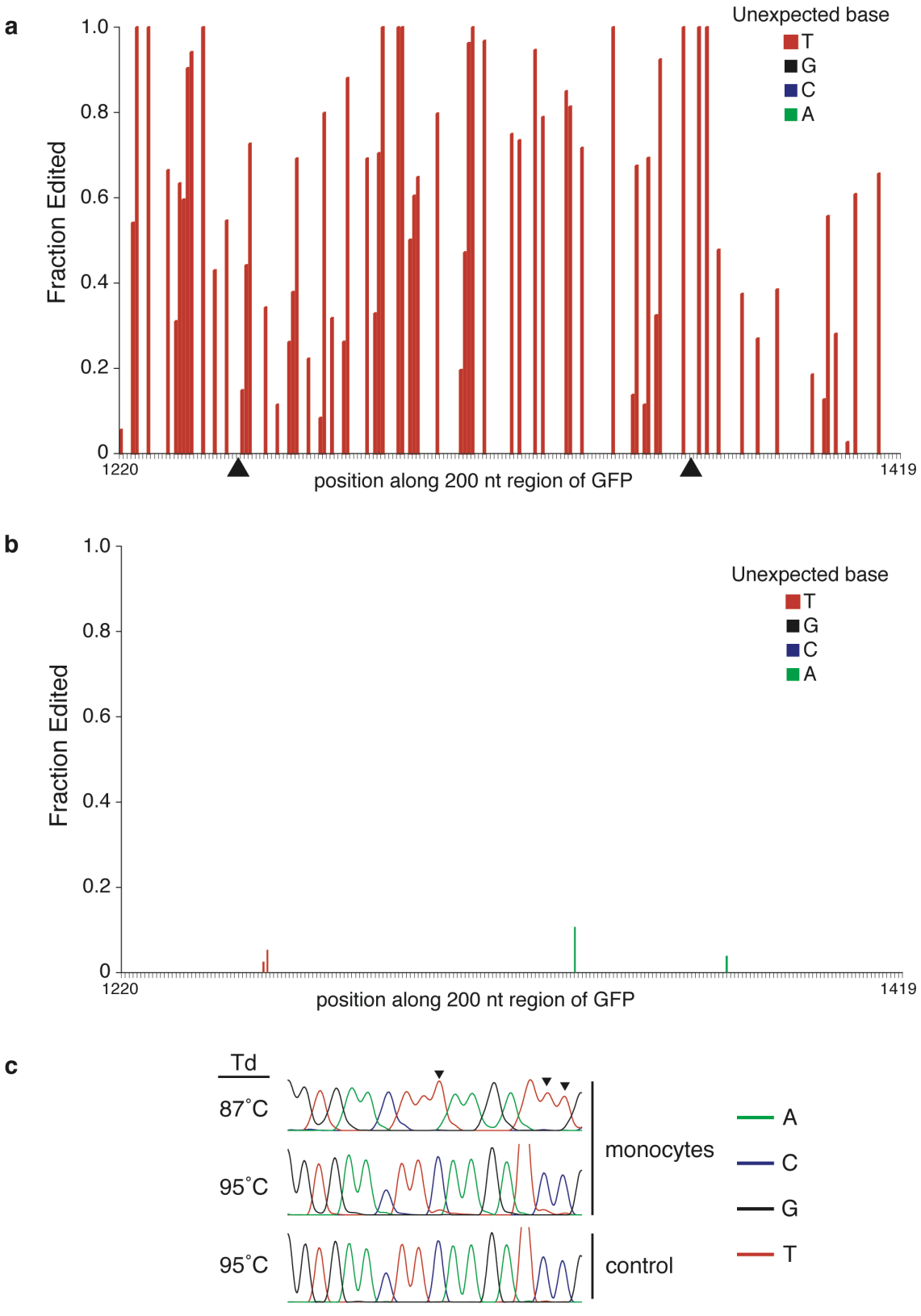

**Supplementary Figure 3** An IFN-inducible DNA cytidine deaminase activity in primary cell extracts. PBMCs were isolated and incubated for 48 hrs in the presence or absence of IFN. Cell extracts were prepared and the indicated amount was incubated with a fluorescently labeled single-stranded DNA deamination substrate. Cytidine deaminase activity was undetectable in untreated cell lyastes, but detectable in IFN treated lysates. The mean and standard deviation of two experiments is shown. See the **Online Methods** for details.



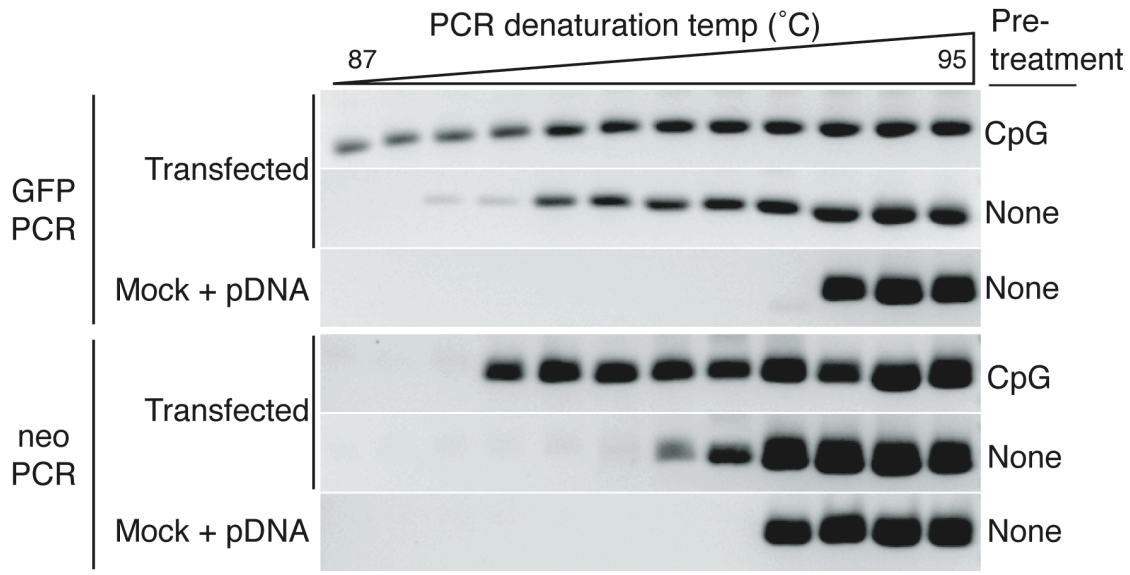

**Supplementary Figure 4** A3A-E72A is inactive in HEK-293 cell extracts. HEK-293 cells were transfected with an A3A or A3A-E72A expression construct. At 24 hrs post-transfection, the cell lysates were prepared and subjected to DNA cytidine deaminase activity assays. The mean and standard deviation of two experiments is shown. See the **Online Methods** for details.




**Supplementary Figure 5** A3A limits gene transfer in another cell type. HeLa cells were transfected as in **Fig. 2a** with a $Neo^R$ reporter plasmid (pcDNA3.1), and gene transfer efficiency was measured by counting the resulting drug resistant colonies.


a
Unexpected base
T
G
C
A
Fraction Edited
1.0
0.8
0.6
0.4
0.2
0
1220
1419
position along 200 nt region of GFP
b
Unexpected base
T
G
C
A
Fraction Edited
1.0
0.8
0.6
0.4
0.2
0
1220
1419
position along 200 nt region of GFP
c
Td
87˚C
95˚C
95˚C
monocytes
control
A
C
G
T

**Supplementary Figure 6** Analysis of populations of hyper-edited molecules. (**a**) Quantification of the editing frequency for each deoxy-cytidine in a representative 200 nucleotide 3D-PCR amplicon derived from plasmid DNA transfected into monocytes. Plasmid DNA was recovered from transfected monocytes, amplified using an 87˚C PCR denaturation temperature, purified by agarose gel electrophoresis, and sequenced. These sequences originated from a population of molecules, so if a sufficiently large fraction differed from the expected nucleotide at each position, then a mixed chromatogram peak would be evident. Multiple instances of cytidines that had been completely or partially converted to uridines were evident, and a 200 nucleotide region of the amplicon, for which there was chromatogram data of sufficiently high quality, was analyzed using phred software (see Online Methods). 70/72 cytidines (97%) in this region displayed some evidence of C-to-T (C-to-U) conversion (*i.e.* had been edited in a fraction of the molecules in the population). The position of the two cytidines for which there was no evidence of editing is marked by black arrowheads (there was no obvious sequence context that explained why those cytidines remained unedited). The position of the 200 nucleotide region in the plasmid pEGFP-N3 (Clontech) is indicated. See also panel **c** below and **Fig. 4c** for representative chromatograms showing full or partially edited cytidines. (**b**) The same quantitative analysis was performed on a control sample, which was generated by mixing GFP-encoding plasmid with total DNA isolated from mock-transfected cells and subjected to the same PCR amplification and sequencing scheme. PCR products from the lowest denaturation temperature at which product amplified (95 ˚C) were sequenced. (**c**) Representative regions of chromatograms are shown. Edited cytidines are indicated with filled arrowheads. The PCR denaturation temperatures used to amplify the populations of molecules and the source of the template DNA are indicated.




**Supplementary Figure 7** Hyperedited plasmid DNA in CD14-enriched monocyte populations. Primary monocytes were isolated from an independent donor as described in **Fig. 1**. The CD14-enriched cell population was transfected with a GFP expression plasmid (pEGFP-N3) or mock-transfected. After 20 hrs, total DNA was recovered from the cells and 3D-PCR was used to detect edited plasmid DNA. Regions of the plasmid containing portions of the GFP and the $Neo^R$ coding sequences were amplified. Primers are listed in **Supplementary Table 4**.

**Supplementary Table 1. Physiological functions of DNA cytidine-to-uridine deamination.**

| | Foreign DNA restriction | Antibody gene diversification | Retroelement restriction | | |
|---|---|---|---|---|---|
| | | | Retroviruses (*e.g.*, HIV-1) | Parvoviruses (*e.g.*, rAAV) | Retro-transposons (*e.g.*, L1) |
| **Enzyme(s)** | A3A & other A3s (not A3G) | AID | A3G & A3F | A3A | A3A, A3B, A3F & A3G |
| **Substrate** | Naked double-stranded DNA (exogenous DNA; *e.g.*, bacterial) | Double-stranded chromosomal DNA (nuclear) | Single-stranded viral cDNA replication intermediates | Undefined DNA replication intermediates | Undefined transposon replication intermediates |
| **Deaminase activity requirement** | Yes | Yes | Yes | No | No |
| **Hypermutation of substrate** | Yes (C/G-to-T/A transitions on both strands) | Yes (all 6 types of base substitutions) | Yes (cDNA strand-specific C/G-to-T/A transitions) | No | No |
| **UNG2 (uracil excision) requirement** | Yes | Yes (and MSH2) | No | No | No |
| **Interferon-inducible** | Strong | None | Weak | Strong | Strong/Weak |
| **Biological function** | Innate immunity to foreign DNA | Adaptive immunity through antibody diversification | Innate immunity to retrovirus infection | Unknown | Innate immunity to endogenous retro-transposition |
| **Relevant Cell Type** | Phagocytes (specialized) & other cells (housekeeping) | B-cells | T-cells & possibly other cell types | Unknown | Germ & somatic cells |
| **Fate of substrate** | Degradation & hypermutation (C/G-to-T/A transitions) | Somatic hypermutation (6 types of base substitutions) & DNA breaks (in class switch recombination) | Degradation / hypermutation | Degradation | Failed integration |
| **References** | This study | 40,61-63 | 46,64-66 | 25,26 | 21-24,56,67,68 |

**Supplemental Table 2: A3A contributes most to DNA restriction in monocytes.**

| | A3A | A3B | A3C | A3DE | A3F | A3G | A3H |
|---|---|---|---|---|---|---|---|
| **Basal monocyte Expression**[a] (relative units) | 230 | 0.2 | 73 | 4.0 | 12 | 26 | 0.1 |
| **IFN inducible**[b] (fold peak induction) | 95 | 2.6 | 1.5 | 1.6 | 2.3 | 1.8 | 1.5 |
| **Dinucleotide preference**[c] | (T/C)C > (A/G)C Refs[25,27,38,39] | (T/C)C > (A/G)C Refs[20,28,39,69,70] | (T/C/G)C > AC(weak) Refs[27,28,39,70,71] | (A/T)C > (C/G)C Refs[72] | TC > GC > (A/C)C Refs[20,28,39,46,71,73,74] | CC > (A/G/T)C Refs[20,28,39,46,64,70,71,73,74] | TC > CC > (A/G)C Refs[27,39,75] |
| **DNA restriction activity in stable gene transfer assay**[d] | 33 | 3.3 | 2.0 | 4.3 | 5.7 | 1.2 | 1.5 |

(a) A3 Basal expression in monocytes in relative units, as in **Fig. 1a**.

(b) Induction by type I IFN in PBMCs, as in **Fig. 1b** (peak fold induction relative to untreated cells).

(c) Preferred deamination dinucleotide context, with the deaminated cytidine underlined. These preferences summarize the consensus of those reported in the references, which vary slightly possibly in part because the templates are different.

(d) Fold reduction in gene transfer efficiency, as in **Fig. 5**.

## Supplemental Table 3. qPCR Primer and Probe Sequences

| Gene Symbol | mRNA NCBI Accession | 5' Primer Name | Seq (5'-3') | 3' Primer Name | Seq (5'-3') | Probe Name | Seq[a] |
|---|---|---|---|---|---|---|---|
| **APOBEC3s** | | | | | | | |
| *APOBEC3A* | NM_145699 | RSH2742 | gagaagggacaagcacatgg | RSH2743 | tggatccatcaagtgtctgg | UPL26 | ctgggctg |
| *APOBEC3B* | NM_004900 | RSH3220 | gacccttttggtccttcgac | RSH3221 | gcacagccccaggagaag | UPL1 | cctggagc |
| *APOBEC3C* | NM_014508 | RSH3085 | agcgcttcagaaaagagtgg | RSH3086 | aagtttcgttccgatcgttg | UPL155 | ttgccttc |
| *APOBEC3D/ APOBEC3DE* | NM_152426 | RSH2749 | acccaaacgtcagtcgaatc | RSH2750 | cacatttctgcgtggttctc | UPL51 | ggcaggag |
| *APOBEC3F* | NM_145298 | RSH2751 | ccgtttggacgcaaagat | RSH2752 | ccaggtgatctggaaacactt | UPL27 | gctgcctg |
| *APOBEC3G* | NM_021822 | RSH2753 | ccgaggacccgaaggttac | RSH2754 | tccaacagtgctgaaattcg | UPL79 | ccaggagg |
| *APOBEC3H* | NM_181773 | RSH2757 | agctgtggccagaagcac | RSH2758 | cggaatgtttcggctgtt | UPL21 | tggctctg |
| **Reference Genes** | | | | | | | |
| *TBP* | NM_003194 | RSH3231 | cccatgactcccatgacc | RSH3232 | tttacaaccaagattcactgtgg | UPL51 | ggcaggag |
| *RPL13A* | NM_012423 | RSH3227 | ctggaccgtctcaaggtgtt | RSH3228 | gccccagataggcaaacttt | UPL74 | ctgctgcc |
| *HPRT1* | NM_000194 | RSH2959 | tgaccttgatttattttgcatacc | RSH2960 | cgagcaagacgttcagtcct | UPL73 | gctgagga |
| *ACTB*[b] | NM_001101 | RSH730 | atcatgtttgagaccttcaa | RSH731 | agatgggcacagtgtgggt | | |
| **Other Genes** | | | | | | | |
| *IFNA2* | NM_000605 | RSH2951 | tcctgcttgaaggacagaca | RSH2952 | tttcagccttttggaactgg | UPL63 | aggaggag |
| *IFNB1* | NM_002176 | RSH2953 | ctttgctattttcagacaagattca | RSH2954 | gccaggaggttctcaacaat | UPL20 | ctggctgg |
| *ISG15* | NM_005101 | RSH3540 | gcgaactcatctttgccagt | RSH3541 | agcatcttcaccgtcaggtc | UPL76 | tggctgtg |
| *EIF2AK2 [PKR]* | NM_002759 | RSH3477 | gacaaagcttccaaccagga | RSH3478 | cggtatgtattaagttcctccatga | UPL39 | ctccacct |
| *ADAR* | NM_001111 | RSH3479 | cgagaatcccaaacaaggaa | RSH3480 | ctggattccacagggattgt | UPL3 | ctgctggg |
| *RNASEL* | NM_021133 | RSH3481 | gggaacatgtgagggactgt | RSH3482 | cacattccgaagcgtcctat | UPL62 | acctgctg |
| **pDNA qPCR assays**[b] | | | | | | | |
| NeoR gene[c] | | RSH951 | gggtagccaacgctatgtcc | RSH2466 | gcttgccgaatatcatggtg | | |
| Sleeping Beauty IR/DR(L)[c] | | RSH2608 | aaaactcgtttttcaactactccac | RSH2609 | ggaaaaatgacttgtgtcatgc | | |
| Sleeping Beauty Transposase[d] | | RSH2606 | cgttatgtttggaggaagaagg | RSH2607 | tgccatctattttgtgaagtgc | | |
| GFP[e] | | RSH1960 | agaacggcatcaaggtgaac | RSH1961 | tgctcaggtagtggttgtcg | | |

(a) It is not known whether the UPL probes correspond to the coding or template DNA strands of their target sequences (Roche proprietary information).
(b) These assays used SYBR Green I dye (Qiagen) and thus did not require a specific probe.
(c) Amplicon in pT2-SV-neo
(d) Amplicon in pCMV-SB-10
(e) Amplicon in pEAK8-GFP-based constructs

## Supplemental Table 4. 3D-PCR Primer Sets

| Gene | | 5' Primer Name | Seq (5'-3') | 3' Primer Name | Seq (5'-3') |
|---|---|---|---|---|---|
| GFP[a] | First PCR | RSH1358 | gccaccatggtgagcaagggcgagga | RSH2872 | aataggggttccgcgca |
| GFP[a] | Nested PCR | RSH1960 | agaacggcatcaaggtgaac | RSH550 | gcaagtaaaacctctacaaatgtggt |
| GFP[a] | First PCR | RSH3813 | actacaacarccacaacrtctatatc | RSH3814 | crctttcttcccttcctttc |
| GFP[a] | Nested PCR | RSH3815 | cttcaaratccrccacaac | RSH3816 | ctrraacaacactcaaccctatc |
| Neo[a] | First PCR | RSH1337 | cagttccgcccattctccg | RSH266 | ttgtttgcaagcagcagatta |
| Neo[a] | Nested PCR | RSH2388 | attgaacaagatggattgcacg | RSH3566 | atactttctcggcaggagca |
| mCherry[b] | First PCR | RSH1052 | cgctgttttgacctccatag | RSH3467 | tggcatatgttgccaaactc |
| mCherry[b] | Nested PCR | RSH2842 | agggcgagatcaagcagag | RSH3626 | tcatgagggtccatggtg |
| *MDM2*[c] | First PCR | RSH3723 | ccttccatcacattrcaaca | RSH3725 | catrracaatrcaaccattt |
| *MDM2*[c] | Nested PCR | RSH3724 | ccaaactrraaaactcaacacaa | RSH3726 | tcraccttracaaatcacacaa |
| *IFNA2*[d] | First PCR | RSH3727 | tttarrctcacccatttcaa | RSH3729 | tccatrttraaccarttttcat |
| *IFNA2*[d] | Nested PCR | RSH3728 | tccarcaratcttcaatctcttc | RSH3730 | tctrctctracaacctccc |

(a) In plasmid pEGFP-N3 (Clontech)
(b) In plasmid pTRE2-mCherry
(c) Ensembl gene id ENSG00000135679 (http://www.ensembl.org)
(d) Ensembl gene id ENSG00000188379

## FULL METHODS

**DNA constructs.** An A3A expression plasmid was obtained from B. Cullen[22]. The A3A coding region (NM_145699) was amplified by PCR using 5'-GAG CTC GGT ACC ACC ATG GAA GCC AGC CCA GC-3' and 5'-GTC GAC CAT CCT TCC GTT TCC CTG ATT CTG GAG-3', digested with SacI/SalI and ligated into similarly digested pEGFP-N3 (Clontech) and pcDNA3.1-HA vectors[54]. The A3A-GFP fusion was cut out of pEGFP-N3-A3A by digesting with NheI/NotI and ligated into an XbaI/NotI-digested pEF, a derivative of pEAK8 (Edge Biosystems)[55]. The other pEF-A3-GFP expression plasmids were created similarly, first by creating the pEGFP-N3-A3 plasmids, and then subcloning them into pEF. pEGFP-N3-A3B, A3DE, A3F, and A3G are as described[56]. The A3C (NM_014508) and A3H (NM_181773) coding sequences were amplified using primers 5'-GAG CTC GGT ACC ACC ATG AAT CCA CAG ATC AGA AAC CCG-3', 5'-GTC GAC TCC CTG GAG ACT CTC CCG TAG C-3', 5'-GAG CTC GGT ACC ACC ATG GCT CTG TTA ACA GCC GAA AC-3', 5'-GTC GAC TCC GGA CTG CTT TAT CCT CTC AAG CC-3', and cloned as above for A3A. To create pcDNA3.1-A3A-Myc-His, the A3A coding region was amplified from pEGFP-N3-A3A using primers 5'-CAT AGA ATT CAA GGA TGG AAG CCA GCC CAG CA-3', and 5'-GTA TAA GCT TGT TTC CCT GAT TCT GGA G-3'. The PCR product was digested with EcoRI and HindIII and ligated into a similarly digested pcDNA3.1(-)/MycHis (Invitrogen). To create pmCherry-N3, the mCherry reading frame was amplified from pRSETB-mCherry (a gift from Roger Tsien) using primers 5'-GCG GCC GCT TAC TTG TAC AGC TCG TCC ATG CCG-3' and 5'-GGA TCC ATC GCC ACC ATG GTG AGC AAG GGC GAG G-3'. This PCR product was digested with NotI and BamHI and ligated into a similarly digested pEGFP-N3. pmCherry-N3-A3A and -A3G were created by cutting the A3 reading frames out of pEGFP-N3-A3A and -A3G by SacI/SalI restriction digestion, respectively, and ligating them into a similarly digested pmCherry-N3.

Amino acid substitution mutants were generated using site-directed mutagenesis (QuikChange, Stratagene; primer sequences available on request). The integrity of all constructs was confirmed by restriction digestion and DNA sequencing. This was particularly important for all A3A expression constructs, which we have found can be toxic to *E. coli* and periodically self-inactivate.

The uracil DNA glycosylase inhibitor expression plasmid, pcDNA3.1-UGI, was created by digesting pEF-UGI[40] with EcoRI/NotI and ligating the UGI-encoding fragment into a similarly digested pcDNA3.1 (Invitrogen). The Sleeping Beauty DNA transposon plasmids, pT2/SV-neo and pCMV-SB10, are as described[47,57]. pTre2-mCherry was created by amplifying the mCherry coding region from pRSETB-mCherry using primers 5'-GCG GCC GCT TAC TTG TAC AGC TCG TCC ATG CCG-3' and 5'-GGA TCC ATC GCC ACC ATG GTG AGC AAG GGC GAG G-3'. The PCR product was digested BamHI/NotI and ligated into a similarly digested pTre2pur (Clontech). This plasmid was digested with XhoI, and the larger fragment was purified on an agarose gel and self-ligated. Plasmid pT2/TK-neo was created by digesting pT2/SV-neo with HindIII and pRH180[58] with SpeI, Klenow blunting and ligating these fragments.

**Cell culture.** HEK-293, HEK-293T, and HeLa cells were maintained in Dulbecco's modified Eagle's medium supplemented with 10% fetal bovine serum and 25 units/ml penicillin and 25 μg/ml streptomycin at 37 ˚C and 5% $CO_2$. These lines were transfected with TransIT-LT1 (Mirus Bio Corporation) according to the manufacturer's protocol.

HEK-293T cell stably expressing uracil DNA glycosylase were created by transfecting cells with pEF-UGI, selecting clones with puromycin, and screening for UNG2 inhibition as described[40].

Clonal HEK-293 cell lines harboring a stably integrated TK-neo cassette were generated by transfecting cells with pCMV-SB-10 and pT2-TK-neo. Cells were plated at limiting dilutions in 96 well plates and selected in 1 mg/ml G418 (neomycin). Clonal $Neo^R$ cell lines were expanded and verified to be $TK^+$ by treatment with 5μM gancyclovir (Calbiochem). Clones that failed to yield $Gan^R$ presumably had multiple integrations of the TK-Neo cassette and were discarded.

**Primary cell experiments.** Blood samples were obtained from healthy donors (Memorial Blood Centers, Minneapolis, MN). PBMCs were isolated by density gradient centrifugation using Ficoll Paque Plus according to the manufacturer's directions (GE Healthcare Life Sciences). In some experiments, monocytes were enriched by negative selection of non-monocytes using magnetic

separation (MACS separation, Miltenyi Biotec) or centrifugation (RosetteSep, Stem Cell Technologies), according to the manufacturer's protocols.

PBMCs or monocytes were treated with 2 ng/ml universal type 1 interferon (R&D Systems) or with 3 μM CpG DNA oligonucleotide ("ODN2216" 5'- ggG GGA CGA TCG TCg ggg gg -3', lower case letters designate nucleotides linked by phosphorothioate bonds) for the times indicated in the text and figures.

Cells were analyzed using a Cell Lab Quanta SC flow cytometer (Beckman Coulter). CD14-positive cells were stained with CD14-FITC (Miltenyi Biotec). Cells were incubated with propidium iodide to assess cell viability.

**Immunoblotting and antibodies.** Cells were harvested and total protein extracted in a buffer containing 25 mM HEPES (pH 7.4), 10% glycerol, 150 mM NaCl, 0.5% Triton X-100, 1 mM EDTA, 1 mM $MgCl_2$, 1 mM $ZnCl_2$, and protease inhibitors. The extracts were clarified by centrifugation for 10 minutes at 20,800g at 4˚C. The extracted proteins (15 μg) were fractionated by SDS-PAGE, transferred to a polyvinylidene difluoride membrane (Millipore), and probed with an anti-A3A polyclonal antiserum, an anti-GFP monoclonal antibody (Clontech), or an anti-eEF1alpha monoclonal antibody (Upstate). The anti-A3A polyclonal serum was generated by immunizing a rabbit with a peptide corresponding to A3A residues 171-199 (CPFQPWDGLEEHSQALSGRLRAILQNQGN) mixed with TiterMax Gold adjuvant (Sigma). Primary antibodies were detected by incubation with fluorescently labeled secondary antibodies and imaging on an Odyssey imaging device (LI-COR Biosciences).

**DNA cytidine deaminase activity assays.** PBMC or transfected HEK-293T cell lysates were prepared as above for immunoblotting. The deaminase activity in the lysates was determined using a FRET-based assay essentially as described[59]. Briefly, serial dilutions of lysates were incubated for 2h at 37˚C with a DNA oligonucleotide 5'-(6-FAM)-AAA-TTC-TAA-TAG-ATA-ATG-TGA-(TAMRA). FRET occurs between the fluorophores, decreasing FAM fluorescence. If cytidine deaminase activity is present in the lysates, the single cytidine is converted to uridine, which is then excised by uracil DNA glycosylase (NEB). Resulting abasic sites are cleaved by incubating reactions for 2min at 95˚C. Once cleaved, the FAM and TAMRA labels are

physically separated, FRET diminishes, and FAM fluorescence increases. Fluorescence is measured on the Lightcycler 480 instrument (Roche).

**Quantitative reverse-transcription PCR assays.** Total RNA was isolated from cells (RNeasy, Qiagen). cDNA was reverse transcribed from 1 μg of total RNA using random hexamer primers and AMV reverse transcriptase (Roche). cDNA was used as template in qPCR reactions performed in a Lightcycler 480 instrument (Roche) according to the manufacturer's protocol. Reactions were performed in duplicate and the mean of the duplicates reported. All gene expression data were normalized to the geometric mean of at least two of the following three reference genes, whose levels varied little in our experiments: TATA-box binding protein (TBP), Ribosomal protein 13A (RPL13A), or hypoxanthine-guanine phosphoribosyltransferase (HPRT). A detailed description of these quantitative assays, as applied to multiple cell types and tissues, will be reported elsewhere (Refsland, Stenglein & Harris, in preparation). **Supplementary Table 3** lists all primer and probe sequences.

**Gene transfer experiments.** 250,000 HEK-293 or HeLa cells were plated into 6-well plates. After 24 h of incubation, the cells were transfected with neo-encoding plasmid pcDNA3.1 (Invitrogen) and pEF-A3A-GFP or pEF-A3A-E72A-GFP. When the A3A expression plasmid was titrated, additional empty pEF vector was added to transfections to keep the pDNA mass constant. Two days post transfection, 100,000 cells were plated into 10cm dishes containing media supplemented with 1 mg/ml G418 (Cellgro). At the same time, to confirm that equivalent numbers of colony-forming cells were plated, equivalent serial dilutions of the cell populations were plated into drug-free media in 24 well plates, at nominal densities of 200, 100, 50, 25, 12, and 6 cells per well. After 12-14 days of additional incubation, colonies were fixed, stained with crystal violet, and counted. Experiments were performed in triplicate and the mean and standard deviation is reported. Essentially identical results were obtained in HEK-293 and HeLa cells. Sleeping Beauty DNA transposon-mediated gene transfer assays were performed as described[47].

**Cell viability, proliferation, and apoptosis assays**. HEK-293 or HeLa cells were transfected as above, and at the indicated time points post transfection, cells were harvested, labeled with propidium iodide, a fluorescent dye to which live cells are impermeable, and analyzed by flow

cytometry. Hydrogen peroxide ($H_2O_2$; 0.5 mM) served as a cytotoxic control in apoptosis experiments. Proliferation was monitored by incubating cells with the Cell Titer 96 aqueous reagent (Promega), according to the manufacturer's instructions.

**Transient GFP expression assays.** HEK-293 or HeLa cells were transfected as above, and at the indicated times post transfection, GFP-fluorescence was analyzed by flow cytometry.

**Plasmid DNA Q-PCR assays.** HEK-293 or HeLa cells were transfected as above, and at the indicated times post transfection, cells were harvested and total DNA extracted (DNeasy, Qiagen). 50 ng of total DNA was used as template for each qPCR reaction. Reactions were performed on an iCycler instrument using SYBR Green I (BioRad) according to the manufacturer's protocol. Reactions were performed in duplicate and the mean of duplicates reported. Plasmid DNA levels were normalized to genomic DNA levels as measured by a qPCR assay specific for the beta actin locus. Primers are listed in **Supplementary Table 3**.

**3D-PCR and DNA sequencing assays to detect hyper-editing in cell lines.** HEK-293 cells, HEK-293T cells, or HEK-293T cells stably expressing UGI were transfected with reporter plasmids pTRE2-Δpuro-mCherry or pEGFP-N3, pcDNA3.1-UGI or pcDNA3.1, and pEF-A3A-GFP or pEF-A3A-E72A-GFP. In instances where UGI expression was titrated into transfection reactions, empty pcDNA3.1 vector was added to keep pDNA mass constant. 24 h post-transfection, cells were treated with DNaseI (Roche) to remove extracellular input plasmid. 24 h later, cells were harvested and total DNA extracted (DNeasy, Qiagen). 50 ng of total DNA was used as input to PCR using primers listed in **Supplementary Table 4** and Taq DNA polymerase (Roche), which amplifies uracil-containing DNA, according to the manufacturer's protocol. Reaction conditions were: 94˚C for 30 seconds, then 25 cycles of 94˚C for 15 seconds, 50˚C for 30 seconds, and 68˚C for 2 minutes, and a final extension at 68˚C for 7 minutes. 0.25 ul of this PCR was used as template in a second PCR using nested primers (**Supplementary Table 4**). Nested PCR reactions used Phusion DNA polymerase (Finnzymes), with other reagents according to the manufacturer's protocol. Reaction conditions were: a gradient of denaturation temperatures ($T_d$) for 30 seconds, then 25 cycles of the $T_d$ gradient for 15 seconds, 52˚C for 30 seconds, and 72˚C for 15 seconds, and a final extension at 72˚C for 7 minutes. $T_d$ gradients are as

indicated in figures. PCR products were separated on agarose gels and detected by ethidium-bromide staining. Experiments to detect edited genomic DNA were performed essentially identically using primers specific to various genomic loci (listed in **Supplementary Table 4**). To generate positive control templates of A3A-deaminated genomic DNA, total DNA was first heat denatured for 15 minutes at 100˚C, rapidly cooled on ice, then incubated with purified A3A (see below) for 1 hr at 37˚C. Reactions contained 12 ng of purified A3A per 500 ng DNA, an amount of enzyme determined to be saturating using the oligo-based deaminase assay described above. Reactions were heated at 100˚C for 3 minutes to denature the A3A and used as PCR templates as described above.

During the course of these studies we observed that the level of A3A-induced deamination is so high that primer design becomes difficult. For instance, since almost every DNA cytidine is a potential A3A substrate, an ideal PCR primer set must be devoid of G's. Since this is not feasible, we elected to minimize the number of G's and, at unavoidable positions, have the primers synthesized with R's (G or A). This partly degenerate approach enabled us to detect C/G-to-T/A transitions in both strands of plasmid substrates recovered from A3A-overexpressing HEK-293 cells or from A3A-induced (IFN or CpG-treated) CD14-positive monocytes (data not shown). The observation that A3A deaminates both DNA strands reinforces our main conclusion that it mediates foreign DNA clearance by a DNA deamination- and uracil excision-dependent mechanism.

**A3A-MycHis protein expression and purification.** Purification of A3A-Myc-His was done as described[60]. Briefly, HEK-293T cells were transfected as above with pcDNA3.1-A3A-Myc-His. 48 h later, $1x10^8$ cells were harvested, washed with PBS and resuspended in 10ml of cell lysis buffer (25mM HEPES, pH7.4, 150mM NaCl, 0.5% Triton X-100, 1mM EDTA, 1mM MgCl2, 1mM ZnCl2, 10% Glycerol, Roche EDTA-free complete protease inhibitor cocktail) supplemented with 50 μg/ml RNase A (Qiagen). The cell suspension was transferred to an ice-cold Dounce homogenizer and homogenized for 10 strokes, then rotated for 1 hr at 4˚C, followed by incubation at 25˚C for 20 min. The lysates were then clarified by centrifugation (12,000 g, 4°C, 10 min). NaCl was then added to the lysates to bring the final concentration to 0.8 M. The lysates were then mixed with 50μl Ni-NTA agarose (Qiagen) by rotating over night at 4˚C. The suspension was then loaded onto a Poly-Prep chromatography column (Bio-Rad). Following

extensive washing with wash buffer (50 mM Tris, pH 8.0, 0.3 M NaCl, 10% glycerol, 0.5% Triton X-100, 50 mM imidazole), A3A-Myc-His was eluted in 6x 200μl of elution buffer (50 mM Tris, pH 8.0, 0.3 M NaCl, 10% glycerol, 0.5% Triton X-100, 150 mM imidazole). Protein purity (over 80%) was assessed by SDS-PAGE and Coomassie blue R250 staining. Protein concentration was determined by Bradford assays (BioRad).

**Sequencing to detect edited DNA.** PCR products were separated by agarose gel electrophoresis and purified (QIAquick, Qiagen). 20 ng of purified PCR product was sequenced directly (Genewiz) or PCR products were cloned (Clonejet, Fermentas) and sequenced. The amplicon contained nucleotides 1170-1426 of pEGFP-N3 (Clontech; Genbank accession U57609.1). PCR product sequences were analyzed using phred software version 0.071220.b [www.phrap.org[53]]. Phred calls bases but also calls "secondary bases" if there is a second peak above the background at the same position as the primary peak. At each position in the nucleotide sequence, if the primary or secondary base was not the expected base, the identity of the unexpected base and the fraction editing was calculated as follows: fraction editing = [area of the unexpected peak / (area of the unexpected peak + area of the expected peak)]. All sequences are available upon request.

**Primary cell transfection.** Monocytes were incubated for 20-24 h with 3 μM CpG DNA oligonucleotide or mock treated. $2x10^7$ monocytes were then transfected with 2 μg endotoxin-free pEGFP-N3 DNA preparations (Qiagen) using a Nucleofector 2 electroporation device according to the manufacturers instructions (Amaxa). 3 h post-transfection, cells were treated with DNaseI to remove non-transfected extracellular DNA. At various time points pre- and post-transfection, as indicated in the text, cells were harvested and RNA and protein was isolated. Gene expression was monitored using qRT-PCR, and recovered DNA was analyzed as above.

***TK* selection experiments**. The frequency of cells converting from $TK^+$ to $TK^-$ was used as a measure of genomic DNA mutation. $TK^+$ cells are $Gan^S$ and $TK^-$ cells are $Gan^R$. Therefore, inactivating mutations of the integrated *TK* gene will confer $Gan^R$. Clonal cell lines harboring the TK-neo cassette were transfected, as above, with A3A or control expression plasmids. Two days later, $1x10^6$ cells were plated into gancyclovir-containing (5 μM) medium. $Gan^R$ colonies were allowed to grow for 12-14 days, and then stained with crystal violet and counted.

**HIV infectivity experiments**. HIV-GFP (CS-CG) was produced as described[46]. Briefly, 250,000 293T cells were plated in 2ml DMEM plus 10% FBS into each well of a 6 well plate the day before transfection. After 24 h incubation, the cells were transfected with 0.2 µg pcDNA3.1-A3A-HA, pcDNA3.1-A3G-HA expression construct or the empty pcDNA3.1 vector control and 1 µg of the HIV-GFP cocktail using TransIT LTI as directed by the manufacturer (Mirus Bio). The four plasmid cocktail consists of a HIV-GFP proviral plasmid CS-CG, a Gag-Pol expression plasmid, a Rev expression plasmid, and a VSV-G envelope expression plasmid. 48 h post-transfection, the virus-containing supernatants were collected to infect target cells and to analyze by immunoblotting and the cells were harvested for flow cytometry and immunoblotting.

293T cells were also used as target cells. The same day as the transfection, 375,000 cells were plated in 2 ml of complete DMEM plus 10% FBS in each well of a 6 well plate. After 24 h incubation, the cells were transfected with 0.5 µg A3A-HA, A3G-HA, or a vector control and 0.5 µg of pmCherry-N3 (as a marker for transfection) using TransIT LTI as directed by the manufacturer (Mirus Bio). The next day the transfection was repeated to ensure high efficiency (actual values ranged from 57-59%). 4 h following the second transfection, the cells were harvested, counted and replated at a density of 20,000 cells per well of a 24 well plate. After an additional hour of incubation, the cells were infected with viral supernatants. After an additional 48-72 h, infection was monitored by target cell flow cytometry. Normalized relative infectivity was calculated as the percentage of infected target cells over the percentage of transfected producer cells.

## ADDITIONAL REFERENCES


[53] Ewing, B., Hillier, L., Wendl, M.C. & Green, P. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res* **8**, 175-185 (1998).

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

[55] Harris, R.S., Croom-Carter, D.S., Rickinson, A.B. & Neuberger, M.S. Epstein-Barr virus and the somatic hypermutation of immunoglobulin genes in Burkitt's lymphoma cells. *Journal of Virology* **75**, 10488-10492 (2001).

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

[57] Cui, Z., Geurts, A.M., Liu, G., Kaufman, C.D. & Hackett, P.B. Structure-function analysis of the inverted terminal repeats of the sleeping beauty transposon. *J Mol Biol* **318**, 1221-1235 (2002).

[58] Harris, R.S., Sale, J.E., Petersen-Mahrt, S.K. & Neuberger, M.S. AID is essential for immunoglobulin V gene conversion in a cultured B cell line. *Current Biology* **12**, 435-438 (2002).

[59] Thielen, B.K. *et al.* T cells contain an RNase-insensitive inhibitor of APOBEC3G deaminase activity. *PLoS Pathog* **3**, 1320-1334 (2007).

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

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

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

[63] Rada, C., Di Noia, J.M. & Neuberger, M.S. Mismatch recognition and uracil excision provide complementary paths to both Ig switching and the A/T-focused phase of somatic mutation. *Mol Cell* **16**, 163-171 (2004).

[64] Harris, R.S. *et al.* DNA deamination mediates innate immunity to retroviral infection. *Cell* **113**, 803-809 (2003).

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

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

[67] Kinomoto, M. *et al.* All APOBEC3 family proteins differentially inhibit LINE-1 retrotransposition. *Nucleic Acids Res* **35**, 2955-2964 (2007).

[68] Niewiadomska, A.M., Tian, C., Tan, L., Wang, T., Sarkis, P.T. & Yu, X.F. Differential inhibition of long interspersed element 1 by APOBEC3 does not correlate with high-molecular-mass-complex formation or P-body association. *J Virol* **81**, 9577-9583 (2007).

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

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

[71] Langlois, M.A., Beale, R.C., Conticello, S.G. & Neuberger, M.S. Mutational comparison of the single-domained APOBEC3C and double-domained APOBEC3F/G anti-retroviral cytidine deaminases provides insight into their DNA target site specificities. *Nucleic Acids Res* **33**, 1913-1923 (2005).

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

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

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

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