Genetic and mechanistic basis for APOBEC3H alternative splicing, retrovirus restriction, and counteraction by HIV-1 protease

Diako Ebrahimi1, Christopher M. Richards1, Michael A. Carpenter1,2, Jiayi Wang1, Terumasa Ikeda1,2, Jordan T. Becker1, Adam Z. Cheng1, Jennifer L. McCann1, Nadine M. Shaban1, Daniel J. Salamango1, Gabriel J. Starrett1,4, Jairam R. Lingappa3, Jeongsik Yong1, William L. Brown1 & Reuben S. Harris1,2

Human APOBEC3H (A3H) is a single-stranded DNA cytosine deaminase that inhibits HIV-1. Seven haplotypes (I–VII) and four splice variants (SV154/182/183/200) with differing antiviral activities and geographic distributions have been described, but the genetic and mechanistic basis for variant expression and function remains unclear. Using a combined bioinformatic/experimental analysis, we find that SV200 expression is specific to haplotype II, which is primarily found in sub-Saharan Africa. The underlying genetic mechanism for differential mRNA splicing is an ancient intronic deletion [del(ctc)] within A3H haplotype II sequence. We show that SV200 is at least fourfold more HIV-1 restrictive than other A3H splice variants. To counteract this elevated antiviral activity, HIV-1 protease cleaves SV200 into a shorter, less restrictive isoform. Our analyses indicate that, in addition to Vif-mediated degradation, HIV-1 may use protease as a counter-defense mechanism against A3H in >80% of sub-Saharan African populations.

1 Department of Biochemistry, Molecular Biology and Biophysics, Masonic Cancer Center, Institute for Molecular Virology, Center for Genome Engineering, University of Minnesota, Minneapolis, MN 55455, USA. 2 Howard Hughes Medical Institute, University of Minnesota, Minneapolis, MN 55455, USA. 3 Departments of Global Health, Medicine and Pediatrics, University of Washington, Seattle, WA 98104, USA. 4Present address: Center for Cancer Research, National Cancer Institute, National Institutes of Health, Bethesda, MD 20892, USA. These authors contributed equally: Diako Ebrahimi, Christopher M. Richards. Correspondence and requests for materials should be addressed to R.S.H. (email: rsh@umn.edu)
ost humans have seven APOBEC3 (apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like 3, A3) genes organized in tandem on chromosome 22 and located between conserved chromobox genes CBX6 and CBX7. These genes, A3A-D and A3F-H, encode members of the APOBEC family of polynucleotide cytosine deaminases (reviewed in refs. 2,3). The hallmark biochemical activity of this family is single-stranded DNA cytosine-to-uracil deamination, although the family namesake (APOBEC1) and at least one A3 enzyme (A3A) also have the capability to edit RNA cytosines. These enzymes have diverse functions in biology ranging from transcriptome editing (APOBEC1 and A3A), antibody gene diversification (AID), and restriction of a broad number of DNA-based parasites (APOBEC1, AID, and A3A-H) (reviewed by refs. 2,3,6–8). The principle function of A3 enzymes in innate immunity is evidenced by high degrees of genetic variation including large and small changes both between and within species. For instance, gene copy numbers vary from 1 gene in rodents (encoding 2 deaminase domains),7 7 genes in most humans (11 deaminase domains),9,10 and up to 14 genes in bats (18 deaminase domains).11 Many humans lack the entire A3B coding sequence due to a 3 bp deletion removing an essential amino acid.13 These deletions due to a 26.5 kb deletion12 and/or (A3A) also have the capability to edit RNA cytosines.4,5 These enzymes are important for the global frequency of the A3H haplotype II reported splice variants (SV154, SV182, SV183, SV200).39 SV154 is predicted to be corrupted structurally and non-functional, SV182 and SV183 are formed by utilization of tandem alternative splice acceptors and have been shown to be similar functionally in prior experiments, and SV200 has a C-terminal extension that could potentially influence multiple activities relevant to HIV-1 restriction.23,39 Here we endeavor to connect genotype to phenotype by asking whether a SNP (or multiple SNPs) within the A3H gene is responsible for alternative splicing and production of SV200. Genetic association and reporter gene studies combined to demonstrate that a single non-coding variation, a 3 nucleotide deletion in intron 4 (Δctc), is embedded within the haplotype II block and responsible for SV200 mRNA expression. Immuno-blots of lymphoblastoid cell lines (LCLs) and peripheral blood mononuclear cells (PBMCs) that showed the SV200 enzyme is expressed in cells with haplotype II genotypes. HIV-1 infectivity experiments showed that the A3H haplotype II SV200 enzyme is at least fourfold more restrictive than other splice variants and, interestingly, that this enhanced activity is counteracted by a single cleavage event catalyzed by HIV-1 protease inside virus particles. Thus, our studies have uncovered a novel protease-dependent mechanism used by HIV-1, in addition to Vif, to counteract A3H-mediated restriction. This cleavage mechanism may be required by transmitting viral isolates in pandemic regions such as sub-Saharan Africa to counteract the increased restriction potency of cells with haplotype II genotypes.

**Results**

**A3H genetic variation in humans.** Human A3H is polymorphic at five codons, and these variants are organized non-randomly into seven different haplotype blocks (reviewed by refs. 6,7,17). Four different A3 enzymes, A3D, A3F, A3G, and A3H, have the capacity to package into assembling viral particles, track within particles until a new host cell becomes infected, and then target virus replication intermediates, e.g., 18–20. Each A3 enzyme can interfere with the overall reverse transcription process to varying degrees by at least two distinct mechanisms. The first is a deamination-independent process by which A3 enzymes bind viral genomic RNA and directly impede reverse transcription, e.g., 21–23. The second is a deamination-dependent process in which viral cDNA cytosines are converted to uracils, which become immortalized by reverse transcription into genomic strand G-to-A mutations, e.g., 24–26. Left unchecked, mono or combinatorial A3 restriction activity can fully suppress virus replication, e.g., 27–29. Therefore, it is not surprising that HIV-1 and related lentiviruses have an A3 counter-defense mechanism governed by the viral protein Vif, which nucleates the formation of an E3 ubiquitin ligase complex to degrade cellular A3 enzymes and prevent access to susceptible viral replication intermediates (reviewed by refs. 6,7,17). Thus, during every round of virus replication, there is a continuous interaction between A3 enzymes attempting to restrict virus replication and Vif working to neutralize the A3 enzymes and preserve viral genomic integrity. The balance between restriction and counteraction is most likely to be sensitive to perturbation as proviral sequences from patients often have G-to-A mutations, e.g., 30–31, and A3 mutagenesis has been linked to virus evolution including immune escape and drug resistance. 32–36.

A3H is the most variable A3 gene in the human population. Four SNPs and one indel have been reported within A3H coding regions: N15/AN15, R18/L18, G105/R105, K121/D121, and E178/D178.13,37 These natural variations have been organized non-randomly into seven different haplotype blocks (I–VII)13,38 and two in particular have a major impact on A3H function. The deletion of N15 common to haplotypes III, IV, and V1 results in a non-functional protein, likely due to loss of structural integrity and rapid degradation (i.e., undetectable by immunoblots and activity assays).13,19,23,38 G105 occurs in haplotypes I and V1 (although inconsequential in the latter due to ΔN15) and compromises antiviral activity in large part due to poor expression and altered subcellular localization.13,19,38–41. Only haplotypes II, V, and VII have demonstrated potent restriction activity against Vif-deficient HIV-1 in model systems13,20,38,39,42 and, due to a high global allele frequency and strong geographic biases, only A3H haplotype II is likely to contribute to HIV-1 restriction and mutagenesis in vivo.19,20,43,44. For instance, although the global frequency of the A3H haplotype II is 20.9%, the allele frequency in HIV-1 pandemic areas of sub-Saharan Africa is 57%, which means 82% of this population has at least one allele of this HIV-1 restrictive haplotype.

A3H is further diversified by alternative splicing, with four reported splice variants (SV154, SV182, SV183, SV200).39 SV154 is predicted to be corrupted structurally and non-functional, SV182 and SV183 are formed by utilization of tandem alternative splice acceptors and have been shown to be similar functionally in prior experiments, and SV200 has a C-terminal extension that could potentially influence multiple activities relevant to HIV-1 restriction.23,39 Here we endeavor to connect genotype to phenotype by asking whether a SNP (or multiple SNPs) within the A3H gene is responsible for alternative splicing and production of SV200. Genetic association and reporter gene studies combined to demonstrate that a single non-coding variation, a 3 nucleotide deletion in intron 4 (Δctc), is embedded within the haplotype II block and responsible for SV200 mRNA expression. Immuno-blots of lymphoblastoid cell lines (LCLs) and peripheral blood mononuclear cells (PBMCs) showed that the SV200 enzyme is expressed in cells with haplotype II genotypes. HIV-1 infectivity experiments showed that the A3H haplotype II SV200 enzyme is at least fourfold more restrictive than other splice variants and, interestingly, that this enhanced activity is counteracted by a single cleavage event catalyzed by HIV-1 protease inside virus particles. Thus, our studies have uncovered a novel protease-dependent mechanism used by HIV-1, in addition to Vif, to counteract A3H-mediated restriction. This cleavage mechanism may be required by transmitting viral isolates in pandemic regions such as sub-Saharan Africa to counteract the increased restriction potency of cells with haplotype II genotypes.

**A3H haplotypes I and II are spliced differentially.** Our analyses of 1000 Genomes Project data confirmed the prevalence of haplotypes I, II, III, and IV in human populations13,38 (i.e., allele frequencies > 1%; Table 1 and Fig. 2a). A3H haplotypes III and IV are not detectable at the protein level, even upon exogenous overexpression13,19,38,39. This is due to a single-amino acid deletion, AN15, which destabilizes the overall protein structure.23 Therefore, because A3H ΔN15 haplotypes III and IV are unlikely to have functional importance, we focused further computational
and biological studies on haplotypes I and II, which have global allele frequencies of 46.4 and 20.9%, respectively. Moreover, due to geographic biases in A3H haplotype distributions, average allele frequencies vary dramatically in different regions of the world with, for instance, haplotype I constituting 65% in Asia and haplotype II 57% in sub-Saharan Africa (Fig. 2a).

Human A3H has 4 known splice variants (SV) that translate into proteins with 182, 183, 200, and 154 amino acids (Fig. 1a, b). Genetic polymorphisms often underlie alternative splicing events. We therefore asked whether A3H haplotypes I and II are spliced differentially. To answer this question, we first surveyed RNAseq read depths of A3H exons in EBV-immortalized lymphoblastoid cell lines (LCLs) from 20 homozygous haplotype I and 20 homozygous haplotype II donors selected randomly from the 1000 Genomes Project. These RNAseq profiles revealed a major difference between mRNA species with respect to exon 4b (2 representative profiles in Fig. 2b and 40 in Supplementary Fig. 1). Exon 4b is mostly excluded from mRNA species of A3H haplotype I LCLs but is frequently included in mRNAs of A3H haplotype II LCLs. On average, exon 4b is expressed fivefold more frequently in A3H haplotype II versus haplotype I LCLs (15.8 ± 1.2 versus 3.0 ± 0.2; n = 20 LCLs representing each haplotype; p = 6 × 10⁻⁹, two-tailed Student’s t-test).

These results predicted that homozygous A3H haplotype II individuals will express SV182, SV183, and SV200, whereas homozygous A3H haplotype I individuals will only express the former 2 splice variants. To test this prediction, Geuvadis Project RNAseq data sets were analyzed for A3H haplotypes and mRNA levels. As predicted, LCLs from individuals homozygous for A3H haplotype I predominately express SV182 and SV183 at similar frequencies, each ~45% (Fig. 2c). In contrast, LCLs from individuals homozygous for A3H haplotype II express SV182, SV183, and SV200 at statistically indistinguishable frequencies, each ~32% (Fig. 2c). Moreover, A3H haplotype I and II heterozygous LCLs show intermediate levels of all three splice variants. These data strongly indicate that A3H differential splicing has a genetic basis, most likely due to one or more SNPs/indels within the locus itself.

**Table 1 APOBEC3H haplotypes and frequencies from analyses of 1000 Genomes Project data (n = 2504)**

| APOBEC3H haplotype | Amino acid position and variant | Global allele frequency |
|--------------------|--------------------------------|-------------------------|
| I                  | NRGKE 15 18 105 121 178        | 46.4                    |
| II                 | NRDDE 15 18 105 121 178        | 20.9                    |
| III                | ΔRDDE 15 18 105 121 178        | 11.1                    |
| IV                 | ΔLDD 15 18 105 121 178         | 19.1                    |
| V                  | NRDDE 15 18 105 121 178        | 0.36                    |
| VI                 | ΔLGK 15 18 105 121 178         | 0.34                    |
| VII                | NRGKE 15 18 105 121 178        | 0                       |
| VIII               | ΔLDD 15 18 105 121 178         | 0.56                    |
| IX                 | NRGKE 15 18 105 121 178        | 0.50                    |
| X                  | ΔRDDE 15 18 105 121 178        | 0.40                    |
| XI                 | ΔRGKE 15 18 105 121 178        | 0.12                    |
| XII                | ΔLKE 15 18 105 121 178         | 0.12                    |
| XIII               | NRGKE 15 18 105 121 178        | 0.06                    |
The near equal expression of SV182 and SV183 (regardless of
A3H haplotype) is most likely due to a splice acceptor slippage
mechanism, in which the splicing apparatus has an equal chance
of using AG dinucleotide splice acceptor sites arranged tandemly
at the beginning of exon 5 (5′-agc-cag in Fig. 1a). Analogous
acceptor slippage events have been documented for other genes,
e.g., ref. 50. The fourth splice variant, SV154, excludes exon 4 and
includes exon 4b, and it is only observed at low levels regardless
of A3H haplotype (<2–3%). Moreover, SV154 is predicted to
encode a non-functional enzyme due to a loss of multiple
conserved structural elements23 (Fig. 1b). Therefore, studies
hereon focus exclusively on SV182, SV183, and SV200.

**A3H haplotype II expresses three distinct proteins.** To investi-
gate whether these mRNA splicing differences manifest at the
protein level, whole-cell lysates from LCLs homozygous for A3H
haplotypes I and II were immunoblotted with a rabbit anti-A3H
polyclonal antibody (Fig. 2d) and independently with a mouse
anti-A3H mAb P1D8–1 that binds a shared N-terminal epitope19
(Supplementary Fig. 2). Strong signals were evident for SV182/
183 proteins in all A3H haplotype I and II LCLs, and a less
intense but clearly detectable signal was observed for the SV200
enzyme only in LCLs with at least one A3H haplotype II allele
(Fig. 2d). The same antibodies detected all three splice variants in
Vif-deficient HIV-1 infected PBMCs from A3H haplotype II
individuals, but only the shorter two splice variants in haplotype I
specimens (Fig. 2e and Supplementary Fig. 2). Endogenous A3H
haplotype I is expressed at lower levels than haplotype II in both
LCLs and PBMCs as reported19 (compare SV182/3 band intensi-

ties for homozygous specimens in Fig. 2d, e). However, despite
the technical challenge of lower expression levels, A3H haplotype
I samples yielded clear signals for SV182/3 enzymes and no signal
for SV200, even at highest exposure levels (Fig. 2d, e). Impor-
tantly, quantification of the amount of SV200 in immunoblots
from homozygous haplotype II PBMCs and LCLs showed that
this variant comprises 30–35 and 16–43%, respectively, of total
A3H protein (n = 4 and n = 8 independent experiments,
respectively; Fig. 2d, e and additional blots not shown). This
result is concordant with SV200 constituting one-third of total
A3H haplotype II splice products (Fig. 2c) and approximately
one-third of total cellular A3H protein (Fig. 2d, e). Together,
these data demonstrate that the SV200 enzyme is a major
expressed splice variant with similar opportunities (based on
abundance) to contribute to antiviral immunity as SV182 and
SV183.

**SNPs associated with A3H exon 4b alternative splicing.** To investi-
gate the mechanism responsible for A3H alternative splicing,
we first tested the hypothesis that differential splicing may be
due to one of the known coding SNPs G105/R105, K121/D121,
and/or E178/D178 that distinguish A3H haplotypes I and II
(Fig. 1). RNAseq data were used to quantify expression levels of
all 4 A3H splice variants in donors with homozygous haplotype
III or IV genotypes. Haplotypes III and IV are identical to hap-


totype II with respect to these three coding SNPs (Table 1).
 Accordingly, haplotypes III and IV LCLs would be predicted to

---

**Fig. 2** Differential splicing and global distributions of A3H haplotypeI and II. **a** Pie charts representing the frequency of each major A3H haplotype within human populations in the indicated geographic regions. **b** Representative RNAseq profiles of LCLs derived from individuals homozygous for A3H haplotypes I or II with the exon 4b region boxed. **c** Bar charts reporting the average percentage of each splice variant observed in LCLs from donors with the indicated A3H genotype (n values are shown and error bars represent 95% CI, p value is calculated using two-tailed Student's t-test). **d, e** Immunoblots of whole-cell extracts derived from LCL (d) or PBMC (e) with the indicated A3H haplotypes. PBMC were mock-treated or infected by Vif-deficient HIV-1 to induce A3H gene expression. Low and high exposures are shown for anti-A3H blots to show SV200 specificity to the haplotype II genotype. Anti-A3G and anti-TUBULIN (TUB) blots are shown as controls.
express SV200 at levels comparable to haplotype II LCLs. However, these analyses showed that haplotypes III and IV LCLs express much lower levels of SV200 mRNA indicating that none of these coding SNPs is responsible for alternative splicing and expression of SV200 (compare Fig. 2c and Supplementary Fig. 3).

We next asked whether a non-coding SNP or indel may be responsible and performed an unbiased association analysis of 5488 SNPs covering the full A3 locus (n = 461; Supplementary Data 1). The top candidate was a 3 bp polymorphism upstream of exon 4b (ctc/Δ, rs149229175; Fig. 3a). The correlation between ctc copy number and SV200 expression is highly significant ($P_{\text{regression}} = 8 \times 10^{-81}$, ANOVA; Fig. 3b). With few exceptions, likely due to rare recombination events, we found that the ctc deletion SNP is highly specific to A3H haplotype II. Cells homozygous for the ctc deletion (mostly also homozygous haplotype II) express the highest levels of SV200. In contrast, cells homozygous for ctc (mostly haplotypes I, III, and IV) have undetectable or low levels of SV200 expression. As expected, haplotype II-speciﬁc ctc deletion are indicated in relevant divergence intervals (not to a speciﬁc feature of the A3H locus). The LINE1 insertion, coding changes, and the haplotype II-speciﬁc ctc deletion are indicated in relevant divergence intervals (not to a specific feature of the A3H locus; Supplementary Fig. 4).

Phylogenetic analyses indicated that the non-coding ctc deletion (Δctc) is a unique and deﬁning feature of the A3H haplotype II block (Fig. 3c). This deletion is absent in other human A3H haplotypes and in the corresponding genomic DNA sequences of non-human primates, chimpanzees and rhesus macaques (Fig. 3c and Supplementary Fig. 4). Moreover, as expected, transcriptomic data from these species failed to show evidence for inclusion of exon 4b or expression of longer A3H transcripts (Supplementary Fig. 4). Interestingly, Δctc is also present in archaic genomes (Neanderthals and Denisovans; Supplementary Fig. 4). Original studies by the Emeran and Malik groups deduced that the ancestral hominid A3H gene encoded a stable protein, and that two different unstable variants emerged independently in humans (R105G in haplotype I and ΔN15 in haplotypes III and IV)\textsuperscript{13}. Our analyses indicate that either Δctc emerged as a result of a similar independent event or was introduced to modern haplotype II humans through recombination with an archaic genome (see Discussion for possible mechanism and origins).

An additional feature of this region of the A3H locus conserved in primates is a 195 bp antisense LINE1 insertion, L1PA1751, which spans the entirety of exon 4b (Fig. 3a). It is therefore tempting to propose that the intronic Δctc may have been the trigger that enabled exonization of this portion of this LINE1 sequence and, simultaneously, fortification of human innate immune function by creating a new A3H enzyme variant with differential antiviral activity.

**Molecular mechanism of A3H exon 4b alternative splicing.** To directly test the hypothesis that the intronic ctc deletion promotes alternative splicing to exon 4b, we constructed an A3H minigene with A3H exons 1–4, intron 4 with and without ctc (both including exon 4b), and exon 5 including the natural 3′ untranslated region (Fig. 4a). Transient minigene expression in HeLa and MCF7 cell lines yielded all 3 splice variants including SV200 (Fig. 4b). SV200 was only expressed from the Δctc construct demonstrating that this 3 bp deletion is essential for exon 4b inclusion and SV200 expression. The deleted ctc is part of an extended polypyrimidine tract, 5′-ttctctctc-3′, which is a predicted binding site for several splicing factors including PTBP1, U2AF2, hnRNP C, and SR52. A reduction from 10 to 7 consecutive pyrimidines is likely to change the repertoire of splicing factors bound to this site, influence interactions with other splice factors, and enable inclusion of exon 4b or, alternatively, weaken an exon 4b exclusion mechanism. It is additionally notable that the ratio of SV200 to SV182/3 protein produced from the minigene is lower than that of the same

![Fig. 3](https://example.com/fig3.png)

---

**Fig. 3** An intronic Δctc associates with A3H haplotype II and SV200 expression. **a** A3H intron 4 alignment with ctc in haplotype I (green) and Δctc in haplotype II (blue) (rs149229175). Additional relevant cis elements include a polypyrimidine tract (purple) and a LINE1 (gray). **b** Box and Whisker plot showing the percentage of SV200 transcripts in cells with the indicated A3H intron 4 genotypes. In this plot, boxes, middle lines and broken lines display interquartile ranges (IQRs), medians, and 1.5 IQR, respectively. Outliers (>1.5 IQR) are shown and the p value is obtained using one-way ANOVA. **c** A schematic showing the phylogenetic relationships between the major human A3H haplotypes (frequencies >1%) and selected non-human primate homologs. The LINE1 insertion, coding changes, and the haplotype II-specific ctc deletion are indicated in relevant divergence intervals (not to a specific time scale).
inhibitor treatments were robust as they also prevented Gag protein (23 kDa), and the virus particle-specific cleavage (R186A; Fig. 5c). Additional single-amino acid substitution mutants and a panel of SV200 truncation mutants (stop codons at positions 185, 186, 187, and 188) provided further support for HIV-1 protease activity at this site (Supplementary Fig. 5). In addition, A3H haplotype II SV200 cleavage by HIV-1 protease is likely to be a conserved function because representative lab strains (subtype B) and clinical HIV-1 isolates from subtypes B and C (transmitted founders and chronic circulating isolates), but not SIVmac239, showed the same cleavage event within particles (Fig. 5d).

**A3H-II SV200 hyperactivity is attenuated by HIV-1 protease.**

Prior studies comparing the HIV-1 restriction activities of the different human A3H splice variants were constrained by a need to use epitope tags and additionally limited because A3H amounts and particularly SV200 levels were not quantified per unit incorporated into viral particles. Moreover, experiments above in Fig. 5 indicated similarly strong Vif-deficient HIV-1 restriction activities for SV182/3 and SV200 but the activity of the latter splice variant was likely underestimated due to lower enzyme amounts in viral particles (and potentially also to proteolytic cleavage).

We therefore sought to quantify the Vif-deficient HIV-1 restriction activities of untagged A3H haplotype II SV200, a hypo-cleavable mutant SV200 V185A, a hyper-cleavable mutant SV200 R186A, and SV183 in a series of single-cycle experiments (mutants from Fig. 5c and Supplementary Fig. 5). In these titration experiments, restriction activities were quantified as a function of the amount of encapsidated A3H normalized to p24 (Fig. 6 and Methods). Interestingly, A3H SV200 and the hypo-cleavable SV200 derivative V185Q showed the strongest restriction activities per unit of encapsidated protein, whereas SV183 showed the weakest and the hyper-cleavable variant R186A showed intermediate levels (quantification in Fig. 6a of immunoblots from three biologically independent experiments shown in Fig. 6b). In quantitative terms, A3H SV200 is fourfold more restrictive than SV183 and at least twofold more restrictive than the hyper-cleavable variant R186A (Fig. 6). Thus, altogether, these data indicate that HIV-1 protease functions to attenuate the potent hyper-restriction activity of A3H SV200.

**Discussion**

A3H is the most variable human A3 gene with five biallelic amino acid positions that make up at least 12 different haplotype blocks (including 6 new ones reported here). Additional A3H diversity occurs through four different splice variants. Here we show that the splice variant encoding the longest protein, SV200, is expressed as a result of a non-coding 3 nucleotide deletion (Δctc)
unique to the haplotype II. The other major splice variants, SV182/3, occur at similar frequencies in all major haplotypes and are therefore not haplotype-specific. Quantification of Vif-deficient HIV-1 restriction activity showed that SV200 is at least fourfold more potent than that of the shorter splice variants. Thus, the A3H haplotype II-specific Δctc is the first non-coding polymorphism shown to change the amino acid composition and affect the function of a human APOBEC enzyme. Furthermore, functional characterization of this splice variant led to the surprising finding that HIV-1 protease specifically cleaves SV200 into a shorter 186 amino acid isoform. Additional Vif-deficient HIV-1 restriction experiments with hypo-cleavable and hyper-cleavable single-amino acid mutant derivatives of A3H haplotype II SV200 indicated that this cleavage event attenuates antiviral activity. Thus, these studies combined to show that HIV-1 protease provides an additional layer of protection against restriction by A3H haplotype II, in addition to the established Vif-mediated proteasomal degradation mechanism.

A3H is the only A3 gene with a repetitive element that contributes to the coding sequence (LINE L1PA17, Fig. 1a and Fig. 3a). Phylogenetic comparisons of human and non-human primate A3H genes indicate that this insertion occurred early in the primate lineage, most likely prior to the evolutionary divergence of known primates (Fig. 3c). This conclusion is supported by a lack of a homologous LINE1 insertion in the A3Z3 gene (A3H homolog) of non-primate mammalian lineages (DNA sequence alignments not shown). The specificity of the intronic Δctc to A3H haplotype II (apart from rare recombinants) and the highly significant correlation between the occurrence of this deletion variant and alternative splicing to include exon 4b and ultimately produce the A3H haplotype II SV200 enzyme (Figs. 2c, 3b) combine to support a model in which the deletion itself may have been the final genetic step in creating exon 4b from LINE1 sequence (i.e., birth of a new exon). Further work will be needed to determine if the Δctc promotes an exon inclusion mechanism or suppresses an exon exclusion mechanism because
there are clear precedents for both in the literature\textsuperscript{52–56}. However, to our knowledge, this is the first example of LINE1 exonization of a primate A3 gene. This is ironic given that A3H has been shown to elicit strong LINE1 restriction activity\textsuperscript{13,61,62}.

DNA polymerases are known sources of indels, particularly in repetitive sequence tracts\textsuperscript{63}. The original Δ\textsuperscript{ctc} may therefore have been caused by a DNA polymerase slippage event that removed one of the three possible ctc trinucleotides from the larger polypyrimidine tract 5′-tt(ctc)(ct(c)tc)-3′ (brackets demark possibilities). The timeframe in which this event occurred is difficult to pinpoint but most likely prior to the global radiation of the four most common A3H haplotypes (I, II, III, and IV; Fig. 2a). The original Δ\textsuperscript{ctc} may have independently occurred one or more times in an ancestral A3H haplotype II population, as it does not exist (apart from rare recombinants) in other modern human A3H haplotype populations. An alternative hypothesis is that Δ\textsuperscript{ctc} emerged in modern hapIII humans as a result of recombinant with an archaic genome. An analogous Δ\textsuperscript{ctc} variant exists in Neanderthal and Denisovan genomic DNA sequences (Supplementary Fig. 4). Moreover, within the 7 kb A3H gene there are an additional 13 SNPs specific to archaic and modern humans with the haplotype II genotype. In contrast only one SNP is shared between archaic and modern humans with the haplotype I genotype. Altogether, these data suggest that Δ\textsuperscript{ctc} may have an archaic origin. However, further studies are needed to pinpoint the exact source of this important variant.

Regardless of the exact window in time that the Δ\textsuperscript{ctc} variant entered the human A3H haplotype II population, it was most likely long before HIV-1 became a pandemic virus (i.e., multiple thousands of years ago versus hundreds of years ago for HIV-1 zoonosis\textsuperscript{64}). This suggests that the antiviral activity of A3H haplotype II SV200 is not limited to HIV-1 and may be considerably broader, consistent with the current working model for A3-mediated innate immunity in which multiple enzymes combine to form a broad, overlapping defense (reviewed by refs. \textsuperscript{2,3}). We speculate that the emergence of Δ\textsuperscript{ctc} in ancestral humans may have originally conferred resistance to an ancestral human pathogen. Such resistance may have conferred a selective advantage to ancestral A3H haplotype II humans and promoted their enrichment in the overall population and particularly in sub-Saharan Africa where A3H haplotype II is currently dominant (57% allele frequency equating to 82% of the population with at least 1 copy, versus other parts of the world such as a 7–10% allele frequency in Asians and Europeans, respectively equating to 13–19% of the population with at least one copy; Fig. 2a). Haplotype-specific differential splicing has the potential to have a significant contemporary global impact because haplotype II individuals will have a larger, and potentially stronger, antiviral A3H repertoire, overall A3 protein repertoire, and therefore potential to restrict a broader array of pathogens. It is not hard to imagine a scenario, for instance, in which populations with A3H haplotypes I, III, and IV may be more susceptible to a virus (or DNA-based parasite) than those with A3H haplotype II. A3H heterozygotes may have an even greater advantage given significant amino acid and functional differences between haplotypes.

The increased potency of A3H haplotype II SV200 against Vif-deficient HIV-1 and counteraction by HIV-1 proteolytic cleavage strongly suggests that this variant enzyme poses a significant threat to the virus. It is tempting to speculate that the viral protease-dependent A3 counteraction mechanism may have been part of the overall process of HIV-1 adaptation to the human population (especially given the dominance of this allele in sub-Saharan Africa corresponding with the likely geographical origin HIV-1 in humans\textsuperscript{64}). This idea is supported conservation of this activity amongst multiple different HIV-1 strains/sub-types but not for the rhesus macaque virus, SIVmac239 (Fig. 5d). Moreover, it may be a human-specific fortification of the conserved Vif-dependent A3 degradation mechanism because the Δ\textsuperscript{ctc} variant, intron 4b inclusion, and SV200 production are all unique features of the human A3H gene that do not occur in non-human primate A3H genes. Although this is the first example of a human retroviral protease cleaving a human A3 enzyme, there are at least two other reports of a retroviral protease cleaving a host A3 enzyme (i.e., FIV protease cleaving feline A3\textsuperscript{65} and MLV protease cleaving murine A3\textsuperscript{66}). Thus, we propose that viral proteases may provide an additional layer of protection against restriction by A3 enzymes and, depending on the virus and host species, may constitute either a primary or a secondary counteraction mechanism.

Methods
Bioinformatic analyses. We first determined the A3H genotype of 2054 individuals from the genetic polymorphism data set of the 1000 Genomes Project, Phase 3. 461 of these individuals were represented by RNAseq data reported by the Geuvadis Project\textsuperscript{49}. Using these two data sets we identified and randomly picked 20 homozygous A3H haplotype I and 20 homozygous haplotype II and inspected their RNAseq BAM files in the Integrative Genomics Viewer (IGV) tool (Supplementary Fig. 1). To compare the expression of exon 4b in these 40 samples, we quantified
with a 95% CI of 0.2 in haplotype I cells. Read depth of exon 4b was 15.8 with a 95% CI of 1.2 in haplotype II cells and 3.0 normalized by the read depth of exon 3 for each sample. The average percentage provided to the Cuf such that the sum of FPKM values of all transcripts within an individual sample is equal to 100.

To identify the genetic source of differential A3H haplotype II SV200 expression, we performed a global correlation analysis between the expression of SV200 in 461 donors (Genovis Project LCLs) and the number of copies of each allele in all of the 548 polymorphic positions in the A3 locus (human chromosome 22 region spanning A3-A3H plus 1000 nucleotides upstream of A3A and downstream of A3H).

A phylogenetic analysis was done using a multiple alignment of full-length A3H gene sequences from human haplotypes I (pig reference), II, III, IV, chimpanzee, and gorilla. The A3H genes of each of these species as well as that of rhesus macaque were extracted directly from the UCSC Genome Browser or deduced from known variants using the 1000 Genomes Project Database. The Neandertal and Denisovan A3H sequences were inferred from their corresponding UCSC Genome Browser tracks.

Endogenous A3H immunobots. LCLs were obtained from the NIGMS Human Genetic Cell Repository at the Coriell Institute for Medical Research. One archived PBMC specimen used in Supplementary Fig. 2 came from a HIV-1 uninfected Ugandan participant in the Partners in Prevention HSV/HIV Transmission Study. Regulatory review for this is covered by protocol review and approval through institutional review boards of the Ugandan HIV/AIDS Research Network (NARC), the Ugandan National Council for Science and Technology (UNCST), and the University of Washington (UW), with the University of Minnesota added through modification of the UW approval. Additional PBMC specimens were isolated from mononuclear cells enriched Trimax (Memorial Blood Centers, Minneapolis) by a gradient of Ficoll Paque Premium (GE Healthcare). These additional specimens were classified as exempt from review by the University of Minnesota Institutional Review Board (IRB) because donors are deidentified by Memorial Blood Centers. Cells were stimulated with 0.5 μg/ml anti-CD3 antibody (clone UCHT1, R&D System), 0.1 μg/ml anti-CD28 antibody (clone CD28.2, eBioscience) and 20 U/ml interleukin-2.

Minigene alternative splicing experiments. To make the A3H minigene constructs, genomic DNA was extracted from a homzygous A3H haplotype II LCL sample (ID: HG00442; obtained from NIGMS Human Genetic Cell Repository at the Coriell Institute for Medical Research) and A3H intron 4 was amplified by high-fidelity PCR, cloned into pET12a, and verified by DNA sequencing. Next, a gBlock was digested with XmeriXcl and ligated into pET12a and the pET12a-A3H minigene construct was then digested with BamHI/Xhol, and cloned into the base A3H haplotype II expression vector. Subsequently, the 3′-UTR from the HG00442 LCL genomic DNA was amplified, ligated into pET12a, and confirmed by DNA sequencing. The 3′-UTR was then added to both of the expression constructs (±ctc) using compatible Adapter/Aid cut-sites.

Recieved: 20 April 2018 Accepted: 13 September 2018
Published online: 08 October 2018
1. LaRue, R. S. et al. Guidelines for naming nonprimate APOBEC3 genes and proteins. J. Virol. 83, 494–497 (2009).
2. Siriwardena, S. U., Chen, K. & Bhagwat, A. S. Functions and malfunctions of mammalian DNA-cytosine deaminases. Chem. Rev. 116, 12688–12710 (2016).
3. Salter, J. D., Bennett, R. P. & Smith, H. C. The APOBEC protein family: united by structure, divergent in function. Trends Biochem. Sci. 41, 578–594 (2016).
4. Rana, M. T. & Eickbush, T. H. Epitranscriptomic profiling across cell types reveals associations between APOBEC1-mediated RNA editing, gene expression outcomes, and cellular function. Proc. Natl Acad. Sci. USA 114, 13296–13301 (2017).
5. Sharma, S. et al. APOBEC3A cytidine deaminase induces RNA editing in monocytes and macrophages. Nat. Commun. 6, 6881 (2015).
6. Harris, R. S. & Dudley, J. P. APOBECs and virus restriction. Virology 479, 480C, 131–145 (2015).
7. Simon, V., Bloch, N. & Landau, N. R. Intrinsic host restrictions to HIV-1 and mechanisms of viral escape. Nat. Immunol. 16, 546–553 (2015).
8. Casellas, R. et al. Mutations, kategios and translocations in B cells: understanding AID promiscuous activity. Nat. Rev. Immunol. 16, 164–176 (2016).
9. Harris, R. S. & Liddament, M. T. Retroviral restriction by APOBEC proteins. Nat. Rev. Immunol. 4, 868–877 (2004).
10. Conticello, S. G. The AID/APOBEC family of nucleic acid modifiers. Genome Biol. 9, 229 (2008).
11. Haywood, J. A. et al. Differential evolution of antiviral restriction factors in primate b cells as revealed by APOBEC3 gene complexity. Mol. Biol. Evol. 35, 1626–1637 (2018).
12. Kidd, J. M. et al. Population stratification of a common APOBEC gene deletion polymorphism. PLoS Genet. 3, e63 (2007).
13. OhAinle, M. et al. Antiretroviral activity of APOBEC was lost twice in recent human evolution. Cell Host Microbe 4, 249–259 (2008).
14. Meyerson, N. R. & Sawyer, S. L. Two-stepping through time: mammals and viruses. Trends Microbiol. 19, 286–294 (2011).
15. Compton, A. A., Malik, H. S. & Emerman, M. Host gene evolution traces the evolutionary history of ancient primate lentiviruses. Philos. Trans. R. Soc. Lond. B. Biol. Sci. 368, 20120496 (2013).
16. Dregal, N. K. & Emerman, M. Evolutionary conflicts between viruses and restriction factors shape immunity. Nat. Rev. Immunol. 12, 687–695 (2012).
17. Malim, M. H. & Bieniasz, P. D. HIV restriction factors and mechanisms of evasion. Cold Spring Harb. Perspect. Med. 2, a009420 (2012).
18. Hultquist, J. F. et al. Human and rhesus APOBEC3D, APOBEC3F, APOBEC3G, and APOBEC3H demonstrate a conserved capacity to restrict Vif-deficient HIV-1. J. Virol. 85, 11220–11234 (2011).
19. Refsland, E. W. et al. Natural polymorphisms in human APOBEC3H and APOBEC3G, and APOBEC3H demonstrate a conserved capacity to restrict single-stranded DNA scanning ability. J. Biol. Chem. 282, 2587–2595 (2007).
20. Singh, R., Valcarcel, J. & Green, M. R. Distinct binding specificity of the human APOBEC3 family to HIV-1 reverse transcriptase. PLoS One 3, e1004761 (2014).
21. Jiang, P. et al. Human APOBEC3G-mediated editing can promote HIV-1 and virus restriction. Cell 149, 2411–2422 (2012).
22. Konig, J. et al. iCLIP reveals the function of hnRNP particles in splicing regulation. Cell 164, 54–66 (2016).
23. Singh, R. et al. HIV restriction factors and mechanisms of evasion. Cold Spring Harb. Perspect. Med. 4, a024481 (2013).
24. Yokomori, K. & Emerman, M. Host gene evolution traces the evolutionary history of ancient primate lentiviruses. Philos. Trans. R. Soc. Lond. B. Biol. Sci. 368, 20120496 (2013).
25. Kong, J. et al. iCLIP reveals the function of hnRNP particles in splicing regulation. Cell 164, 54–66 (2016).
26. Yokomori, K. & Emerman, M. Host gene evolution traces the evolutionary history of ancient primate lentiviruses. Philos. Trans. R. Soc. Lond. B. Biol. Sci. 368, 20120496 (2013).
27. Lappalainen, T. et al. Transcriptome and genome sequencing uncovers functional variation in humans. Nature 501, 506–511 (2013).
28. Modrek, B. & Lee, C. A genomic view of alternative splicing. Nat. Genet. 30, 13–19 (2002).
29. Monlong, J., Calvo, M., Ferreira, P. G. & Guigo, R. Identification of genetic variants associated with alternative splicing using SqtSIsseekR. Nat. Commun. 5, 4698 (2014).
30. Chan, A. et al. Reduced APOBEC3 variant anti-viral activities are associated with altered RNA binding activities. PLoS ONE 7, e38771 (2012).
31. Starrett, G. J. et al. The DNA cytosine deaminase APOBEC3H haplotype I likely contributes to breast and lung cancer mutagenesis. Nat. Commun. 7, 12918 (2016).
32. Feng, Y. et al. Natural polymorphisms and oligomerization of human APOBEC3 contribute to single-stranded DNA editing and disease. J. Biol. Chem. 290, 27188–27203 (2015).
33. Nakano, Y. et al. HIV-1 competition experiments in humanized mice show that APOBEC3H imposes selective pressure and promotes virus adaptation. PLoS Pathog. 13, e1006348 (2017).
34. Zhang, Z. et al. Stably expressed APOBEC3H forms a barrier for cross-species transmission of simian immunodeficiency virus of chimpanzees to humans. PLoS Pathog. 13, e1006746 (2017).
35. Genomes Project, C. et al. A global reference for human genetic variation. Nature 526, 68–74 (2015).
36. Lalonde, E. et al. RNA sequencing reveals the role of splicing polymorphisms in regulating human gene expression. Genome Res. 21, 545–554 (2011).
37. Singh, R., Valcarcel, J. & Green, M. R. Distinct binding specificity of the human APOBEC3 family to HIV-1 reverse transcriptase. PLoS One 3, e1004761 (2014).
38. Monlong, J., Calvo, M., Ferreira, P. G. & Guigo, R. Identification of genetic variants associated with alternative splicing using SqtSIsseekR. Nat. Commun. 5, 4698 (2014).
39. Konig, J. et al. iCLIP reveals the function of hnRNP particles in splicing regulation. Cell 164, 54–66 (2016).
40. Luo, H., Helfman, D. M., Gagel, R. F. & Berger, S. M. Polypropyrimidine tract-binding protein positively regulates inclusion of an alternative 3′-terminal exon. Mol. Cell. Biol. 39, 78–85 (1999).
41. Singh, R., Valcarcel, J. & Green, M. R. Distinct binding specificities and functions of higher eukaryotic polypropyrimidine tract-binding proteins. Science 268, 1173–1176 (1995).
42. Zarnack, K. et al. Direct comparison between hnrNP C and U2AF65 protects the transcriptome from the exonization of Alu elements. Nucleic Acids Res. 45, 11–20 (2017).
43. Singh, R., Valcarcel, J. & Green, M. R. Distinct binding specificities and functions of higher eukaryotic polypropyrimidine tract-binding proteins. Science 268, 1173–1176 (1995).
44. Zarnack, K. et al. Direct comparison between hnrNP C and U2AF65 protects the transcriptome from the exonization of Alu elements. Nucleic Acids Res. 45, 11–20 (2017).
45. Singh, R., Valcarcel, J. & Green, M. R. Distinct binding specificities and functions of higher eukaryotic polypropyrimidine tract-binding proteins. Science 268, 1173–1176 (1995).
Acknowledgements

We thank Brett Anderson and Matthew Jarvis for comments, Kate Lauer for technical assistance, and Bruce Torbett for support. The following reagents were obtained through the NIH AIDS Reagent Program, Division of AIDS, NIAID, NIH: anti-human APOBEC3G C-terminal polyclonal antibody (pAb 10201) from Dr. Jaisri Lingappa, pCH058, c/2960 (cot# 11856) from Dr. John Kappes and Dr. Christina Ochsenbauer, pHIGA (cot#12421) from Dr. Beatrice Hahn, HIV-1 Z3618M T/F (cot# 13262) and Z3618F Z3618F (SGA 11) infectious molecular clones (cat# 11856) from Dr. John Kappes and Dr. Christina Ochsenbauer, pRHGA BEC3G C-terminal polyclonal antibody (pAb 10201) from Dr. Jaisri Lingappa, pCH058. the NIH AIDS Reagent Program, Division of AIDS, NIAID, NIH: anti-human APO-

7. Feng, Y., Goubran, M. H., Pollack, T. B. & Chelico, L. Deamination-independent restriction of LINE-1 retrotransposition by APOBEC3H. Sci. Rep. 7, 10881 (2017).
8. Uddin, K., House, N. C. & Freudenberg, C. H. Repeat instability during DNA repair: Insights from model systems. Crit. Rev. Biochem. Mol. Biol. 50, 142–167 (2015).
9. Sharp, P. M., Rayner, J. C. & Hahn, B. H. Evolution. Great apes and zoonoses. Science 340, 284–286 (2013).
10. Yoshikawa, R. et al. Feline immunodeficiency virus evolutionarily acquires two proteins, Vif and Protease, capable of antagonizing feline APOBEC3. J. Virol., 91, e00250-17 (2017).
11. Abuda, A. et al. Murine retrovirus escapes from murine APOBEC3 via two distinct novel mechanisms. Curr. Biol. 16, 1565–1570 (2006).
12. Li, H. et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25, 2078–2079 (2009).
13. Trapnell, C. et al. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat. Protoc. 7, 562–578 (2012).
14. Celum, C. et al. Acyclovir and transmission of HIV-1 from persons infected with HIV-1 and HSV-2. N. Engl. J. Med. 362, 427–439 (2010).