

Sequence editing by Apolipoprotein B RNA-editing catalytic component-B and epidemiological surveillance of transmitted HIV-1 drug resistance

Robert J. Gifford^a, Soo-Yon Rhee^a, Nicolas Eriksson^b, Tommy F. Liu^a, Mark Kiuchi^a, Amar K. Das^c and Robert W. Shafer^a

Design: Promiscuous guanine (G) to adenine (A) substitutions catalysed by apolipoprotein B RNA-editing catalytic component (APOBEC) enzymes are observed in a proportion of HIV-1 sequences *in vivo* and can introduce artifacts into some genetic analyses. The potential impact of undetected lethal editing on genotypic estimation of transmitted drug resistance was assessed.

Methods: Classifiers of lethal, APOBEC-mediated editing were developed by analysis of lentiviral *pol* gene sequence variation and evaluated using control sets of HIV-1 sequences. The potential impact of sequence editing on genotypic estimation of drug resistance was assessed in sets of sequences obtained from 77 studies of 25 or more therapy-naive individuals, using mixture modelling approaches to determine the maximum likelihood classification of sequences as lethally edited as opposed to viable.

Results: Analysis of 6437 protease and reverse transcriptase sequences from therapy-naive individuals using a novel classifier of lethal, APOBEC3G-mediated sequence editing, the polypeptide-like 3G (APOBEC3G)-mediated defectives (A3GD) index, detected lethal editing in association with spurious 'transmitted drug resistance' in nearly 3% of proviral sequences obtained from whole blood and 0.2% of samples obtained from plasma.

Conclusion: Screening for lethally edited sequences in datasets containing a proportion of proviral DNA, such as those likely to be obtained for epidemiological surveillance of transmitted drug resistance in the developing world, can eliminate rare but potentially significant errors in genotypic estimation of transmitted drug resistance.

© 2008 Wolters Kluwer Health | Lippincott Williams & Wilkins

AIDS 2008, **22**:717–725

Keywords: antiretroviral therapy, cytidine deaminase, drug resistance, epidemiology, highly active, HIV-1

Introduction

Since 1991, it has been observed that HIV-1 sequences occasionally contain an excess of guanine (G) to adenine (A) substitutions [1–4]. More recently, it has become clear that this 'G-to-A hypermutation' reflects the activity of host enzymes belonging to the apolipoprotein B RNA-editing catalytic component (APOBEC) family of cytidine deaminases, most notably apolipoprotein B mRNA-editing catalytic polypeptide-like 3G (APOBEC3G) and

apolipoprotein B mRNA-editing catalytic polypeptide-like 3F (APOBEC3F). These enzymes can incorporate into budding HIV-1 particles, and on subsequent infection of target cells, catalyse the deamination of cytidine (C) to uridine (U) in nascent viral reverse transcripts, which can be manifested as G-to-A substitutions in plus-stranded proviral DNA (reviewed in [5,6]). In general, HIV-1 avoids APOBEC-mediated sequence editing through the activity of the virally encoded Vif protein [7]. The occasional detection of hypermutated HIV-1 sequences *in vivo* [2–4]

From the ^aDivision of Infectious Diseases, Department of Medicine, the ^bDepartment of Biostatistics, and the ^cDivision of Biomedical Informatics, Department of Medicine, Stanford University, Stanford, California, USA.

Correspondence to Dr Robert Gifford, Division of Infectious Diseases, Stanford University, Stanford, CA 94305, USA.

Tel: +1 650 725 2946; fax: +1 650 725 2088; e-mail: rjmg@stanford.edu.

likely reflects natural variation among viruses in their capacity to suppress cytidine deamination, with some viruses carrying defective *vif* alleles [8]. Although it has been suggested that some degree of sublethal editing by APOBEC enzymes may contribute to HIV-1 evolution [8–10], extensive G-to-A editing generally leads to mutational impairment of viruses [11]. Genetic data support a long-standing role for primate APOBEC enzymes in innate immune defence against retroviruses [12], and mutational impairment of retroviral nucleic acids by sequence editing forms one component of this defence [6].

Sequence variation in lethally edited viruses may reflect qualitatively different biological processes to variation in viable viral genomes (i.e. sequence editing as opposed to purifying selection). It is, therefore, important that lethally edited sequences are identified in analyses that assume data to represent viable genetic material under selection, such as genotypic estimation of drug resistance. In this report, we develop a robust classifier, the ‘APOBEC-mediated defectives index’ (AD index), suitable for efficiently screening lethally edited sequences from large and diverse HIV-1 *pol* gene datasets. We show that APOBEC-mediated substitutions can introduce spurious ‘drug-resistance mutations’ into sequences that have been inactivated through lethal editing, and that spurious drug resistance in association APOBEC-mediated sequence editing is observed in a proportion of sequences obtained from both blood and plasma. The implications of this finding for surveillance of transmitted drug resistance (TDR) are discussed.

Materials and methods

Sequence data

To identify conserved APOBEC target sites in HIV-1 protease and reverse transcriptase genes, sequences of representative lentiviruses were obtained from Genbank and aligned. These included equine infectious anaemia virus (EIAV), AF327877; caprine arthritis encephalitis virus (CAEV), AF322109; ovine maedi-visna virus (OMVV), M34193; feline immunodeficiency virus (FIV), M36968; and bovine immunodeficiency virus (BIV), M32690. Alignments of protease and reverse transcriptase genes from representative primate lentiviruses were obtained from the Los Alamos Sequence Database (www.hiv.lanl.gov).

Three control datasets were used to assess methods for discriminating APOBEC-mediated sequence editing from naturally occurring and antiretroviral drug-selected HIV-1 sequence variation: a ‘wildtype control set’ of 20 nearly full-length reference sequences; a ‘hypermutated control set’ of 28 nearly full-length sequences that were annotated in Genbank as displaying G-to-A hypermuta-

tion (these previously collated datasets [11] included representatives of each of the established group M subtypes); and a ‘drug-resistant control set’ of 697 group M sequences obtained from the Stanford HIV Drug Resistance Database. Sequences in this dataset were generated by population sequencing of plasma virus from patients who were reported as having had two or more different reverse transcriptase inhibitors and two or more protease inhibitors. All of these sequences had at least four major drug-resistance mutations [13] and represented the following subtypes and circulating recombinant forms (CRFs): B (508), C (51), G (39), D (25), A (23) F (23), CRF01 (16), CRF02 (12).

To assess the potential effect of APOBEC-mediated sequence editing on epidemiological surveys of TDR, a set of 6437 reverse transcriptase and protease sequences obtained from distinct antiretroviral therapy (ART)-naive individuals was retrieved from Genbank. This dataset represented 77 studies of 25 or more individuals and included the following subtypes and CRFs: B (2499), C (1285), CRF02 (736), A (556), CRF01 (533), G (189), D (172), G (137), other recombinants/unclassifiable sequences (96). All sequences were at least 600 nucleotides in length and were annotated with respect to the tissue source, date, and country of sampling. A complete list of the publications from which sequences were obtained can be found online (<http://cpr.stanford.edu/cpr/>).

Sequence analysis

To quantify the extent of G-to-A substitution in protease and reverse transcriptase sequences, several previously described approaches were applied. The extent of G-to-A substitution can be estimated through comparison to a suitable reference sequence (e.g. a consensus sequence for the viral subtype under analysis). Two general methods that use this approach, ‘G-to-A preference’ (number of G-to-A substitutions/number of all substitutions) and ‘G-to-A burden’ (number of G-to-A substitutions/number of G nucleotides in the reference sequence), are implemented in the programme Hypermut [14]. A similar measure defined by Pace *et al.* [15] and a recently updated version of the Hypermut program (Hypermut 2.0) exploit the predilection of APOBEC enzymes to induce G-to-A mutations in specific dinucleotide contexts (GG-to-AG for APOBEC3G and AG-to-AA for APOBEC3F) to provide enzyme-specific measures of G-to-A substitution. The ‘consolidated 3G’ and ‘consolidated 3F’ scores are defined by Pace as the number of G-to-A substitutions at target dinucleotides per number of target dinucleotides in the reference sequence [15]. In the Hypermut 2.0 program, Fisher’s exact test is used to define an excess of mutations (hypermutation) within specified dinucleotide contexts relative to a control context [16]. A fifth measure, ‘product–substrate ratio’ (PS ratio), can be calculated without comparison to a reference sequence. PS ratio is defined as the number of

substrate dinucleotides in the query sequence per number of product dinucleotides in the query sequence [11].

Viral subtypes were determined for all sequences using the programme STAR [17]. Drug resistance was assessed using a list of ‘surveillance drug-resistance mutations’ (SDRMs) endorsed by the WHO for global surveillance of transmitted HIV-1 drug resistance (<http://www.who.int/hiv/drugresistance/en/>) and designed to standardize estimates of TDR across geographic regions and viral subtypes [18]. For epidemiological surveillance of drug resistance, the presence of a single SDRM in a sequence qualifies it as resistant.

Statistical analysis

Receiver–operator curves were plotted in R using the ROCR package [19]. For mixture modelling, we assumed two underlying distributions: a Poisson distribution, reflecting the large number of sequences with low A3GD indices and a uniform or normal distribution, reflecting the distribution of the remaining ART-naive set. The expectation–maximization algorithm was applied to identify the parameters of distributions and to determine the maximum likelihood classification of sequences as lethally edited as opposed to viable/wildtype. A bootstrap analysis was run to assess the variability in the data resulting from the small number of sequences with high A3GD indices.

Results

Development of apolipoprotein B mRNA-editing catalytic polypeptide-like 3G-mediated defectives index

Sequence context at the dinucleotide level strongly affects the efficiency with which cytidine deamination is carried out by various APOBEC enzymes, such that 3G tends to be associated with GG-to-AG substitutions, whereas 3F is associated with GA-to-AA substitutions [20–22]. The genomic region of HIV-1 typically sequenced for resistance testing contains nine conserved tryptophan (Trp) residues. The single codon for Trp (TGG) contains at least one 3G target site at which a G-to-A substitution leads to a stop codon (potentially, two 3G target sites, or one 3G target site and one 3F target site can be present, depending on sequence variation at the downstream codon position). As stop codons within open reading frames generally indicate mutational inactivation of proteins, stop codons at conserved tryptophan sites are excellent indicators of lethal sequence editing by APOBEC. To identify other mutations that might similarly serve as indicators of lethal, APOBEC-mediated sequence editing, we defined highly conserved 3G and 3F target sites in an alignment of protease and reverse transcriptase genes from representative exogenous lentiviruses. This analysis identified 34 3G and 27 3F target

sites conserved throughout the HIV-1 M group and at which G-to-A substitutions give rise to rare mutations indicative of APOBEC-mediated editing (Table 1 [18]). Conserved APOBEC target sites identified in this analysis were used to define a classifier of lethal APOBEC-mediated editing, the AD index, which is defined as the number of mutations (excluding major drug-resistance mutations) arising through G-to-A substitutions at conserved APOBEC target sites divided by the number of conserved APOBEC target sites in the query sequence. Like other measures of APOBEC-mediated G-to-A substitution, this index can be configured as general, or as 3G or 3F specific, based on target site preferences. To be considered indicator sites suitable for inclusion in calculation of the index, target motifs were minimally required to occur at sites where sequence editing would result in amino acid substitutions and to be conserved across more than 98% of HIV-1 group M isolates in the reference alignment [23]. Of the 61 sites we identified, however, 34 (56%) were also conserved across all primate lentiviruses and 25 (40%) were conserved across all exogenous lentivirus lineages (Table 1). The rarity of G–A substitutions at these sites in HIV-1 was confirmed by their low prevalence (<0.1%) within the Stanford HIV Drug Resistance Database and by reference to other published studies of HIV-1 sequence variation [24,25]. The mutations used to calculate the AMD index are not necessarily lethal to virus replication, but their rarity in a highly mutable pathogen such as HIV-1 indicates that they are efficiently removed by purifying selection in replicating virus populations [26].

Validation of apolipoprotein B mRNA-editing catalytic component-mediated defectives indices using control datasets

The reliability of mutation-based indices for identifying lethally edited sequences was assessed using control datasets. The protease–reverse transcriptase region was extracted from control wildtype and hypermutated genome sets and assessed using general and enzyme-specific indices and four previously described measures of G-to-A substitution (see Methods). Five sequences extracted from hypermutated genomes showed no evidence of excessive G-to-A substitution by any of the measures used and were excluded from subsequent analysis. All 3G specific measures performed better than general or 3F-specific measures (data not shown), which may reflect bias in the control set. As previous studies have also indicated a greater role for 3G in hypermutation of HIV-1 [4,10,15], we however restricted our subsequent analysis to this enzyme.

Figure 1a shows the generalized G-to-A measures (G-to-A preference and G-to-A burden), the 3G-specific measures (consolidated 3G score and 3G P/S ratio), and the ‘A3G-mediated defectives index’ (A3GD index) applied to the wildtype and hypermutated control sequence sets. Figure 1b shows the same five measures

Table 1. Mutations at conserved dinucleotide target sites used to define the apolipoprotein B mRNA-editing catalytic component (APOBEC)-mediated defectives indices.

Protease						Reverse transcriptase					
Mutation	Enzyme	Conservation of residue	Mutation	Enzyme	Conservation of residue	Mutation	Enzyme	Conservation of residue	Mutation	Enzyme	Conservation of residue
Trp6 → *	3G	HIV-1/SIV cpz	Glu6 → Lys	3F	HIV-1 M group	Asp110 → Asn	3F	All lentiviruses	Asp110 → Asn	3F	All lentiviruses
Arg8 → Gln	3F	Primate lentiviruses	Gly15 → Arg	3G	Primate lentiviruses	Gly112 → Arg/Ser	3G	HIV-1/SIV cpz	Gly112 → Arg/Ser	3G	HIV-1/SIV cpz
Asp25 → Asn	3F	All lentiviruses	Met16 → Ile	3G	HIV-1/SIV cpz	Asp113 → Asn	3F	All lentiviruses	Asp113 → Asn	3F	All lentiviruses
Gly27 → Arg	3G	All lentiviruses	Asp17 → Asn	3F	HIV-1/SIV cpz	Gly141 → R	3G	Primate lentiviruses	Gly141 → R	3G	Primate lentiviruses
Asp29 → Asn	3F	All lentiviruses	Gly18 → Arg/Ser	3G	All lentiviruses	Gly152 → R	3G	All lentiviruses	Gly152 → R	3G	All lentiviruses
Asp30 → Asn	3F	All lentiviruses	Trp24 → *	3G	All lentiviruses	Trp153 → *	3G	All lentiviruses	Trp153 → *	3G	All lentiviruses
Gly40 → Arg	3G	HIV-1/SIV cpz	Glu28 → Lys	3F	HIV-1/SIV cpz	Gly155 → Arg	3G	Primate lentiviruses	Gly155 → Arg	3G	Primate lentiviruses
Trp42 → *	3G	HIV-1/SIV cpz	Glu29 → Lys	3F	All lentiviruses	Met184 → Ile	3G	All lentiviruses	Met184 → Ile	3G	All lentiviruses
Gly48 → Arg	3G	Primate lentiviruses	Met41 → Ile	3G	HIV-1 M group	Asp185 → Asn	3F	All lentiviruses	Asp185 → Asn	3F	All lentiviruses
Gly49 → Arg	3G	All lentiviruses	Glu42 → Lys	3F	HIV-1/SIV cpz	Asp184 → Asn	3F	All lentiviruses	Asp184 → Asn	3F	All lentiviruses
Gly51 → Arg	3G	All lentiviruses ^a	Glu44 → Lys	3F	HIV-1/SIV cpz	Gly190 → Arg	3G	HIV-1/SIV cpz	Gly190 → Arg	3G	HIV-1/SIV cpz
Gly52 → Ser	3G	All lentiviruses	Glu45 → Arg	3G	All lentiviruses	Asp192 → Asn	3F	HIV-1/SIV cpz	Asp192 → Asn	3F	HIV-1/SIV cpz
Glu65 → Lys	3F	HIV-1 M group	Gly51 → Arg	3G	HIV-1/SIV cpz	Glu194 → Lys	3F	HIV-1 M group	Glu194 → Lys	3F	HIV-1 M group
Gly68 → Arg	3G	HIV-1	Glu53 → Lys	3F	HIV-1/SIV cpz	Trp212 → *	3G	HIV-1/SIV cpz	Trp212 → *	3G	HIV-1/SIV cpz
Gly73 → Ser	3G	HIV-1	Asp67 → Asn	3F	Primate lentiviruses	Gly213 → Arg	3G	HIV-1/SIV cpz	Gly213 → Arg	3G	HIV-1/SIV cpz
Gly78 → Arg	3G	Primate lentiviruses	Trp71 → *	3G	All lentiviruses	Asp218 → Asn	3F	HIV-1/SIV cpz	Asp218 → Asn	3F	HIV-1/SIV cpz
Gly86 → Arg	3G	All lentiviruses	Asp76 → Asn	3F	All lentiviruses	Glu224 → Lys	3F	HIV-1/SIV cpz	Glu224 → Lys	3F	HIV-1/SIV cpz
Gly94 → Arg	3G	HIV-1/SIV cpz	Glu79 → Lys	3F	HIV-1/SIV cpz	Trp229 → *	3G	All lentiviruses	Trp229 → *	3G	All lentiviruses
			Asp86 → Asn	3F	Primate lentiviruses	Met230 → Ile	3G	Primate lentiviruses	Met230 → Ile	3G	Primate lentiviruses
			Trp88 → *	3G	HIV-1/SIV cpz	Gly231 → Arg/Ser	3G	All lentiviruses	Gly231 → Arg/Ser	3G	All lentiviruses
			Glu89 → Lys	3F	All lentiviruses	Glu233 → Lys	3F	HIV-1/SIV cpz	Glu233 → Lys	3F	HIV-1/SIV cpz
			Gly93 → Arg	3G	All lentiviruses	Trp239 → *	3G	All lentiviruses	Trp239 → *	3G	All lentiviruses
			Gly99 → Arg/Ser	3G	All lentiviruses						

Glycine (Gly; GGN where N is any nucleotide) and tryptophan (Trp; TGG) codons are 3G target sites. Methionine (Met; ATG) codons are targets when the downstream codon begins with a conserved guanine nucleotide. 3G substitutions at Gly can produce either arginine (Arg) or serine (Ser) depending on the nucleotide present at the third position in the codon. Only those mutations observed in the lethally edited portion of the sequence dataset from drug naive individuals are indicated in this table but all G-to-A changes occurring at 3G sites were used to calculate the A3GD index. As Gly often contains three consecutive guanines, this position is vulnerable to the activity of apolipoprotein B mRNA-editing catalytic polypeptide-like 3G (APOBEC3G) and apolipoprotein B mRNA-editing catalytic polypeptide-like 3F (APOBEC3F), which could lead to additional changes other than Arg and Ser at these positions. 3G substitutions at Trp and Met targets result in stop codons (*) and isoleucines (Ile), respectively. Mutations in bold are at sites of drug-resistance mutations. Underlined mutations are current surveillance drug-resistance mutations [18]. Gly190 → Ser is a potential drug-resistance mutation at an A3G target site that was not observed in the hypermutated control set. Met41 → Leu and Met230 → L are the drug-resistance mutations reported at reverse transcriptase (RT) positions 41 and 230. These, however, do not result from G-to-A substitutions. Gly27 is a potential target site but neither Gly27 → Arg nor Gly27 → Ser was observed in the hypermutated control set. The enzyme column indicates which editing enzyme the target site is specific to, APOBEC3G (3G) or APOBEC3F (3F). Conservation of the amino acid residue for which the codon contains a target guanine ranges from 'all lentiviruses' (most conserved), through 'all primate lentiviruses', the HIV-1/SIV cpz lineage only, and to the HIV-1 M group only (least conserved).

^aExcept bovine immunodeficiency virus (BIV).

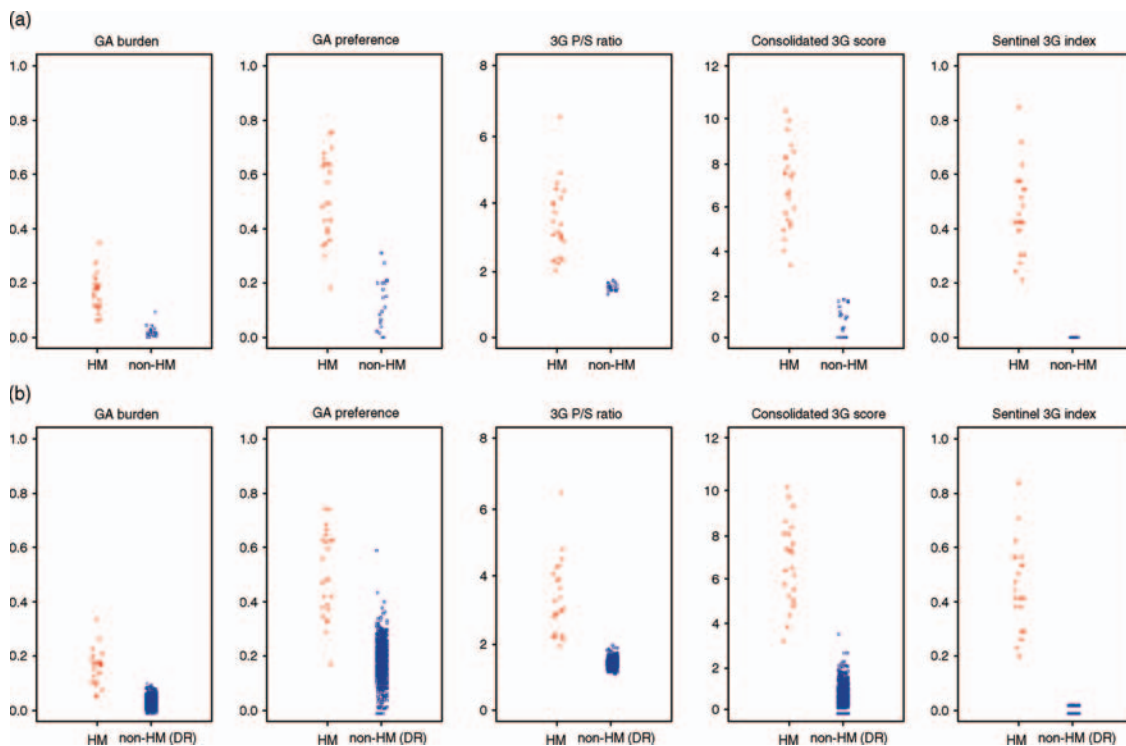


Fig. 1. Comparison of five measures of G-to-A editing using control sequence sets. Two generalized measures ('G-to-A preference' and 'G-to-A burden') and three APOBEC3G specific measures ('consolidated 3G score', '3G P/S ratio', and the 'A3GD index') of G-to-A editing are shown applied to control sequence sets. (a) Plots comparing scores for control hypermutated (HM) sequences with scores for control non-hypermutated (non-HM) sequences. (b) Plots comparing scores for control hypermutated (HM) sequences with scores for control non-hypermutated, drug resistant [non-HM (DR)] sequences.

applied to drug-resistant and hypermutated control sequence sets. The mean and median differences in scores for each of the two groups were highly statistically significant for all five methods ($P < 0.01$); however, overlap between scores for the drug-resistant and hypermutated sets was observed for some measures (Fig. 1). Plotting of receiver–operator curves (data not shown) revealed that the A3GD index performed as well, and in some cases more reliably as a classifier of lethal A3G-mediated editing than other measures. For measures utilizing comparison to a reference sequence, it was observed that natural and drug-selected variation in the query sequence could affect performance, particularly if the reference sequence was not closely related to the sequence under analysis.

Screening of antiretroviral therapy-naive sequence set using the APOBEC3G-mediated defectives index

A3GD indices were calculated for a set of 6203 reverse transcriptase and protease sequences obtained from ART-naive individuals. The heterogeneity in A3GD indices was explored using a mixture model that assumed two underlying distributions: a Poisson distribution, reflecting the large number of sequences with low A3GD indices (i.e. unedited sequences) and a uniform or normal

distribution, reflecting the distribution of the remaining, potentially 3G-edited set (Fig. 2). On the basis of this analysis, an A3GD index of more than 0.08 (equivalent to three or more A3GD index mutations in a full-length protease–reverse transcriptase sequence) indicates a more than 99% probability of mutational inactivation through APOBEC-mediated editing.

Applying this cut-off to the ART-naive sequence dataset revealed that nine of 77 studies contained at least one hypermutated sequence (Table 2 [4,5,27–34]). Although 12 sequences from four studies were annotated as hypermutated, 16 sequences from five studies were not. Of these 16, 13 were classified as hypermutated by the Hypermut 2.0 program. A total of eight sequences classified as lethally edited by the A3GD index did not contain any stop codons. 75% of lethally edited sequences from drug naive patients had major resistance mutations (including SDRMs) arising through G-to-A changes at conserved APOBEC target sites and no other drug-resistance mutations, suggesting that the apparent TDR in these patients was an artifact introduced by A3G-mediated sequence editing. In these studies, the estimated prevalence of drug resistance within the study population (as defined by the presence of one or more SDRMs) was lower by a median of 2.3% (range: 0.4–10.3%) if the

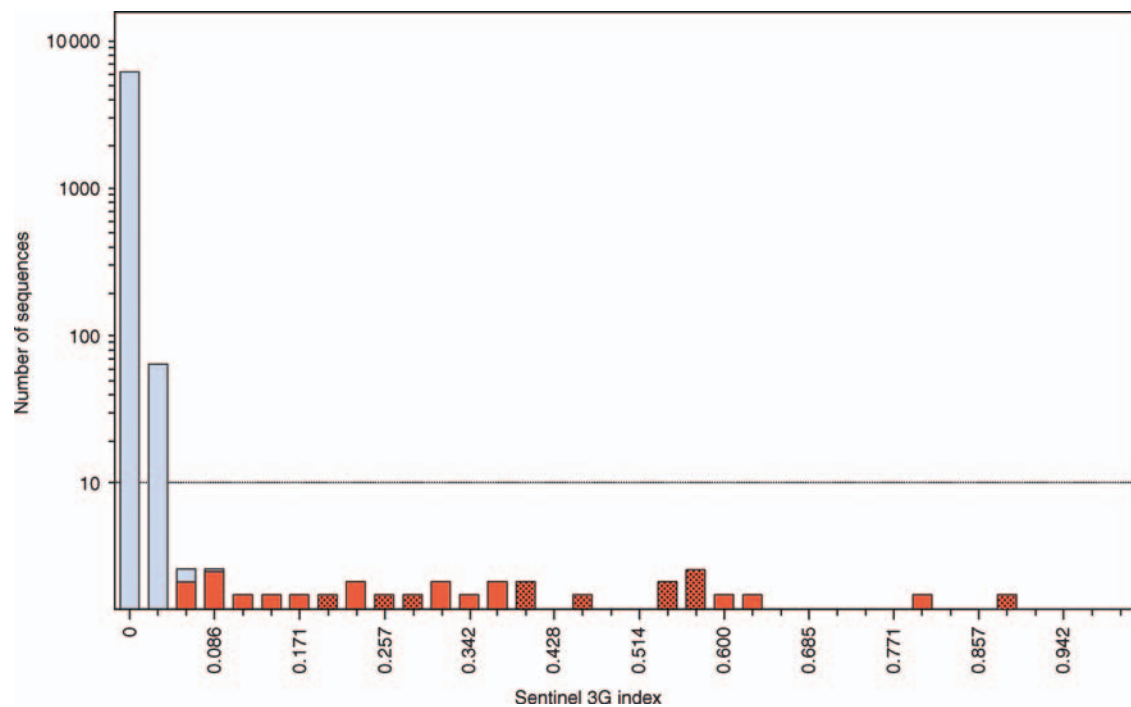


Fig. 2. Distribution of A3G-mediated defectives (A3GD) scores among 6203 sequences obtained from patients reported as treatment naive. The distributions reflect a mixture of Poisson ($\mu=0.0003$) and either uniform or normal distributions (mean=0.4) that were fitted using an expectation-maximization algorithm. The maximum likelihood fit to these distributions suggests that the presence of three A3GD mutations (A3GD index >0.08 or 3/35) in a full-length resistance test sequence (i.e. codons 1–99 of protease and 1–240 of reverse transcriptase) is associated with a more than 99% likelihood of resulting lethal editing by apolipoprotein B mRNA-editing catalytic polypeptide-like 3G (APOBEC3G). The Y-axis has been scaled linearly between 0 and 10 and logarithmically between 10 and 10 000 to optimally illustrate the extremely high numbers of sequences with A3GD indices close or equal to zero, as well as the actual numbers of sequences with A3GD indices of more than 0.08. At each index, the proportion of the histogram coloured red indicates the maximum likelihood proportion of lethally edited sequences. The stippled areas in a histogram indicate those sequences that had been annotated in Genbank as having G-to-A hypermutation.

lethally edited sequences were excluded. From the current International AIDS Society (IAS)-USA major [35] and/or SDRM lists [18], G73S and D30N in protease and D67N and M184I in reverse transcriptase, all of which can arise through G-to-A substitutions at highly (100%) conserved APOBEC dinucleotide target sites, were the mutations most commonly associated with lethal editing. Other drug-resistance mutations could potentially arise through APOBEC-mediated G-to-A substitutions [9]; but in these cases, the introduction of that mutation through a single G-to-A substitution was dependent on rare polymorphisms at synonymous codon positions or within neighbouring codons, or on combined 3G-mediated and 3F-mediated substitutions (Table 1).

In the ART-naive dataset, lethally edited sequences were significantly more likely to be obtained from proviral DNA of peripheral blood mononuclear cells (19/640, 3%) than sequences obtained from plasma (9/5563, 0.2%; $P<0.001$), as might be expected [4]. Viral genetic material that had apparently been extensively edited by APOBEC3G, however, was also identified in a small

proportion of plasma samples (Table 2), and although this observation was unexpected, it was supported by the identification of similarly edited sequences ($\sim 0.1\%$) in a separate dataset comprising 9996 sequences obtained from plasma and submitted to the Stanford Medical Centre for resistance testing. In the Stanford data set, the presence of potentially lethally edited sequences in plasma was always associated with low plasma viral load. It is not clear whether plasma sequences with high A3GD index scores, but no stop codons, represent viruses capable of initiating a new round of infection, and in these cases, editing should be considered potentially rather than necessarily lethal. In theory, genomic RNA expressed from otherwise defective proviruses may be packaged into viral particles provided a second, viable provirus is present to provide the necessary viral proteins in *trans*.

Discussion

Several inexpensive, generic, and fixed-dose triple antiretroviral drug combinations have shown remarkable

Table 2. Reverse transcriptase (RT) and/or protease (PR) sequences from 28 HIV-1-infected persons with an A3GD index of more than 0.8 consistent with lethal, A3G-mediated sequence editing.

Study	GenBank accession	Tissue source	Subtype	Hypermut 2.0 <i>P</i> value ^a	A3GD index	3G mutations at drug-resistance positions ^b	
						PR	RT
1 [27]	AF457057	PBMC	A	<0.01	0.82	G73S	M41I, M184I , G190R, M230I
	AF457071	PBMC	A	<0.01	0.59	G48E ^a	M41I, G190R
	AF457074	PBMC	A	<0.01	0.53	G48K ^a , G73S	M184I , G190R, M230I
	AF457060	PBMC	CRF16	0.02	0.46		M41I, M184I , G190R, M230I
	AF457076	PBMC	A	<0.01	0.41	G48R	M184I , M230I
2 [28]	AF457091	PBMC	A	0.07	0.26		M230I
	AF484484	PBMC	A	<0.01	0.56	G48R, G73S	M41I, M184I , M230I
3 [29]	DQ461868 (PR)	PBMC	CRF02	<0.01	0.39		M41I, M184I , G190R
	DQ461920 (RT)						
	DQ461851 (PR)	PBMC	CRF02	0.04	0.23	G48R, G73S	
	DQ461904 (RT)						
	DQ461856 (PR)	PBMC	CRF02	<0.01	0.20	G48R, G73S	
	DQ461909 (RT)						
4 [13]	DQ461871 (PR)	PBMC	CRF02	0.43	0.10	G48R	
	DQ461923 (RT)						
	AF480808 (RT)	PBMC	B	0.02	0.25	NA	M41I, M184I , G190R, M230I
	AF480809 (RT)	PBMC	B	<0.01	0.17	NA	M41I, G190E ^a
	AF480745 (RT)	PBMC	B	0.17	0.09	NA	M230I
	AF273175 (RT)	PBMC	B	0.01	0.53	NA	M41I, M184I
5 [30]	AY358058	PBMC	CRF01	<0.01	0.53	G48R, G73S	M41I, M184I , M230I
	AY358055	PBMC	CRF01	<0.01	0.56	G73S	M41I, M184I , G190R, M230I
6 [31]	AY358053	PBMC	CRF01	<0.01	0.24		M184I , M230I
	AY358054	PBMC	CRF01	<0.01	0.21		M184I , G190R
	AY238223 (PR)	Plasma	CRF01	<0.01	0.44		M41I, M184I , G190R, M230I
	AY238122 (RT)						
	AY238286 (PR)	Plasma	CRF01	<0.01	0.09		M184I , M230I
7 [32]	AY238120 (RT)						
	AY238289 (PR)	Plasma	CRF01	<0.01	0.32		M41I, M184I , G190R, M230I
	AY238124 (RT)						
	DQ164128	Plasma	C	<0.01	0.79	G48R, G73S	M41I, M184I , G190R, M230I
	DQ164125	Plasma	C	<0.01	0.65	G48R, G73S	M184I , G190E ^a , M230I
8 [33]	DQ275665	Plasma	C	<0.01	0.59	G48R	M184I , G190R, M230I
	DQ164123	Plasma	C	0.03	0.35		G190R
	DQ164124	Plasma	C	0.14	0.26		M41I, M230I
	AY845730	Plasma	CRF02	<0.01	0.36	G48R	G190R, M230I

A3GD, APOBEC3G-mediated defectives.

^aG48E, G48K and G190E would not result from APOBEC3G mediated editing alone. They, however, may result from an APOBEC3G-mediated editing followed by editing consistent with APOBEC3F. Protease sequences were not available [4,5]; therefore, the relevant column contains 'NA' for not applicable. PBMC, primary blood mononuclear cells. Details of the Hypermut 2.0 program can be found at <http://www.hiv.lanl.gov/>. *P* values of less than 0.05 are considered to represent a hypermutant and are indicated in bold.

^bMutations in bold indicate surveillance drug-resistance mutations (SDRMs).

success in pilot HIV-1 treatment programmes in low-income and middle-income countries [36–39]. Scaling-up access to ART in low-income and middle-income countries should, therefore, reduce morbidity and mortality in millions of HIV-1-infected persons [40]. Standardized, simplified treatment protocols have been developed to guide clinical decision-making accompanying antiretroviral drug use in areas where resource limitations preclude the intensive clinical management typical of high-income countries [40]. The threat from TDR in this environment is uncertain [41–43]. Accordingly, the World Health Organisation has established a global programme for genotypic surveillance of TDR in representative untreated populations in regions where access to ART is expanding, so that any rise in the prevalence may be detected at any early stage [18,44]. As a relatively low threshold has been set for public health

action in response to TDR (>5% prevalence) [44], it is important that surveillance protocols are implemented scrupulously. Reporting a falsely low prevalence of drug resistance could undermine the success of standard therapies, whereas reporting a falsely high prevalence could result in the abandonment of current effective and relatively inexpensive treatments.

In this report, we found that spurious drug-resistance mutations are occasionally observed in association with patterns of substitution indicative of lethal, APOBEC-mediated sequence editing. Substitutions that can be introduced by sequence editing and by drug selection include mutations associated with each of the three main antiretroviral drug classes: protease inhibitors (G73S, D30N), nucleoside inhibitors (M184I), and nonnucleoside (G190SE) reverse transcriptase inhibitors. Such

artificial TDR within lethally edited sequences, however, is only likely to occur in datasets containing a significant proportion of sequences derived from proviral DNA, such as samples stored using dried blood spot (DBS) filter paper [45].

As DBS filter paper is a highly attractive medium for transporting viral material in resource-limited settings [46], and SDRMs should not be removed from the list until it is shown that this will not significantly diminish the sensitivity of surveillance protocols, we propose that screening for lethal editing should be carried out routinely in surveillance programmes incorporating DBS-derived samples. The resource implications of screening are minimal, as both A3GD index-based and PS ratio-based screening protocols have been integrated into the Calibrated Population Resistance (CPR) tool (an online and publicly accessible programme for estimation of transmitted resistance) on the Stanford HIV drug resistance website (<http://cpr.stanford.edu/cpr/>). An advantage of using A3GD index-based and PS ratio-based approaches (as opposed to other measures of G→A substitution) is that they can be calculated without a reference sequence, thereby obviating the need to accurately characterize natural diversity in the query sequence set, which is likely to be challenging in areas where subtype diversity is high (i.e. mainly West and Central Africa, but increasingly other areas as well [47–49]). CPR also provides PS ratio values for sequences as a second measure of G→A substitution. The A3GD index and PS ratio are also appropriate classifiers for rapidly screening lethally edited sequences from large protease and reverse transcriptase sequence datasets to be used in other analyses, such as molecular epidemiological investigations utilizing phylogeny [47,48]. The observation that a small proportion of sequences obtained from plasma contained patterns of substitution indicative of lethal editing, including stop codons in some sequences, may have implications for the design of studies investigating minority strains in HIV-1 infection, and also suggests that in rare cases, screening for lethal sequence editing may be relevant to procedures based on plasma samples, such as interpretation of drug resistance in sequences obtained during routine genotypic testing.

References

- Vartanian JP, Meyerhans A, Asjo B, Wain-Hobson S. **Selection, recombination, and G→A hypermutation of human immunodeficiency virus type 1 genomes.** *J Virol* 1991; **65**:1779–1788.
- Janini M, Rogers M, Birx DR, McCutchan FE. **Human immunodeficiency virus type 1 DNA sequences genetically damaged by hypermutation are often abundant in patient peripheral blood mononuclear cells and may be generated during near-simultaneous infection and activation of CD4(+) T cells.** *J Virol* 2001; **75**:7973–7986.
- Koulinska IN, Chaplin B, Mwagagile D, Essex M, Renjifo B. **Hypermutation of HIV type 1 genomes isolated from infants soon after vertical infection.** *AIDS Res Hum Retroviruses* 2003; **19**:1115–1123.
- Kieffer TL, Kwon P, Nettles RE, Han Y, Ray SC, Siliciano RF. **G→A hypermutation in protease and reverse transcriptase regions of human immunodeficiency virus type 1 residing in resting CD4+ T cells in vivo.** *J Virol* 2005; **79**:1975–1980.
- Cullen BR. **Role and mechanism of action of the APOBEC3 family of antiretroviral resistance factors.** *J Virol* 2006; **80**:1067–1076.
- Holmes RK, Malim MH, Bishop KN. **APOBEC-mediated viral restriction: not simply editing?** *Trends Biochem Sci* 2007; **32**:118–128.
- Sheehy AM, Gaddis NC, Malim MH. **The antiretroviral enzyme APOBEC3G is degraded by the proteasome in response to HIV-1 Vif.** *Nat Med* 2003; **9**:1404–1407.
- Simon V, Zennou V, Murray D, Huang Y, Ho DD, Bieniasz PD. **Natural variation in Vif: differential impact on APOBEC3G/3F and a potential role in HIV-1 diversification.** *PLoS Pathog* 2005; **1**:e6.
- Berkhout B, de Ronde A. **APOBEC3G versus reverse transcriptase in the generation of HIV-1 drug-resistance mutations.** *AIDS* 2004; **18**:1861–1863.
- Hache G, Mansky LM, Harris RS. **Human APOBEC3 proteins, retrovirus restriction, and HIV drug resistance.** *AIDS Rev* 2006; **8**:148–157.
- Suspene R, Rusniok C, Vartanian JP, Wain-Hobson S. **Twin gradients in APOBEC3 edited HIV-1 DNA reflect the dynamics of lentiviral replication.** *Nucleic Acids Res* 2006; **34**:4677–4684.
- Sawyer SL, Emerman M, Malik HS. **Ancient adaptive evolution of the primate antiviral DNA-editing enzyme APOBEC3G.** *PLoS Biol* 2004; **2**:E275.
- Rhee SY, Gonzales MJ, Kantor R, Betts BJ, Ravela J, Shafer RW. **Human immunodeficiency virus reverse transcriptase and protease sequence database.** *Nucleic Acids Res* 2003; **31**:298–303.
- Rose PP, Korber BT. **Detecting hypermutations in viral sequences with an emphasis on G→A hypermutation.** *Bioinformatics* 2000; **16**:400–401.
- Pace C, Keller J, Nolan D, James I, Gaudieri S, Moore C, Mallal S. **Population level analysis of human immunodeficiency virus type 1 hypermutation and its relationship with APOBEC3G and vif genetic variation.** *J Virol* 2006; **80**:9259–9269.
- Bruno WJ, Abfalterer WP, Foley BT, Leitner TK, Korber BT. **Detection of hypermutation in HIV sequences using two context positions and avoiding nucleotide content effects (in press).**
- Myers RE, Gale CV, Harrison A, Takeuchi Y, Kellam P. **A statistical model for HIV-1 sequence classification using the subtype analyser (STAR).** *Bioinformatics* 2005; **21**:3535–3540.
- Shafer RW, Rhee SY, Pillay D, Miller V, Sandstrom P, Schapiro JM, et al. **HIV-1 protease and reverse transcriptase mutations for drug resistance surveillance.** *AIDS* 2007; **21**:215–223.
- Sing T, Sander O, Beerenwinkel N, Lengauer T. **ROCR: visualizing classifier performance in R.** *Bioinformatics* 2005; **21**:3940–3941.
- Hache G, Liddament MT, Harris RS. **The retroviral hypermutation specificity of APOBEC3F and APOBEC3G is governed by the C-terminal DNA cytosine deaminase domain.** *J Biol Chem* 2005; **280**:10920–10924.
- Bishop KN, Holmes RK, Sheehy AM, Davidson NO, Cho SJ, Malim MH. **Cytidine deamination of retroviral DNA by diverse APOBEC proteins.** *Curr Biol* 2004; **14**:1392–1396.
- Beale RC, Petersen-Mahrt SK, Watt IN, Harris RS, Rada C, Neuberger MS. **Comparison of the differential context-dependence of DNA deamination by APOBEC enzymes: correlation with mutation spectra in vivo.** *J Mol Biol* 2004; **337**:585–596.
- Leitner T, Korber B, Daniels M, Calef C, Foley B. **HIV-1 subtype and circulating recombinant form (CRF) references sequences, 2005.** In: Leitner T, Foley BT, Hahn B, et al., editors. *HIV Sequence Compendium 2005*. Los Alamos: Theoretical Biology and Biophysics Group; 2005. pp. 41–48.
- Ceccherini-Silberstein F, Erba F, Gago F, Bertoli A, Forbici F, Bellocchi MC, et al. **Identification of the minimal conserved structure of HIV-1 protease in the presence and absence of drug pressure.** *AIDS* 2004; **18**:F11–F19.
- Ceccherini-Silberstein F, Gago F, Santoro M, Gori C, Svicher V, Rodriguez-Barrios F, et al. **High sequence conservation of human immunodeficiency virus type 1 reverse transcriptase under drug pressure despite the continuous appearance of mutations.** *J Virol* 2005; **79**:10718–10729.

26. Smith RA, Anderson DJ, Preston BD. **Purifying selection masks the mutational flexibility of HIV-1 reverse transcriptase.** *J Biol Chem* 2004; **279**:26726–26734.
27. Dowling WE, Kim B, Mason CJ, Wasunna KM, Alam U, Elson L, *et al.* **Forty-one near full-length HIV-1 sequences from Kenya reveal an epidemic of subtype A and A-containing recombinants.** *AIDS* 2002; **16**:1809–1820.
28. Harris ME, Serwadda D, Sewankambo N, Kim B, Kigozi G, Kiwanuka N, *et al.* **Among 46 near full length HIV type 1 genome sequences from Rakai District, Uganda, subtype D and AD recombinants predominate.** *AIDS Res Hum Retroviruses* 2002; **18**:1281–1290.
29. Koizumi Y, Ndembu N, Miyashita M, Lwembe R, Kageyama S, Mbanya D, *et al.* **Emergence of antiretroviral therapy resistance-associated primary mutations among drug-naive HIV-1-infected individuals in rural western Cameroon.** *J Acquir Immune Defic Syndr* 2006; **43**:15–22.
30. Sung H, Foley BT, Bae IG, Chi HS, Cho YK. **Phylogenetic analysis of reverse transcriptase in antiretroviral drug-naive Korean HIV type 1 patients.** *AIDS Res Hum Retroviruses* 2001; **17**:1549–1554.
31. Tovanabutra S, Beyrer C, Sakkhachornphop S, Razak MH, Ramos GL, Vongchak T, *et al.* **The changing molecular epidemiology of HIV type 1 among northern Thai drug users 1999 to 2002.** *AIDS Res Hum Retroviruses* 2004; **20**:465–475.
32. Lan NT, Recordon-Pinson P, Hung PV, Uyen NT, Lien TT, Tien HT, *et al.* **HIV type 1 isolates from 200 untreated individuals in Ho Chi Minh City (Vietnam): ANRS 1257 Study. Large predominance of CRF01_AE and presence of major resistance mutations to antiretroviral drugs.** *AIDS Res Hum Retroviruses* 2003; **19**:925–928.
33. Rousseau CM, Birditt BA, McKay AR, Stoddard JN, Lee TC, McLaughlin S, *et al.* **Large-scale amplification, cloning and sequencing of near full-length HIV-1 subtype C genomes.** *J Virol Methods* 2006; **136**:118–125.
34. Carr JK, Nadai Y, Eyzaguirre L, Saad MD, Khakimov MM, Yakubov SK, *et al.* **Outbreak of a West African recombinant of HIV-1 in Tashkent, Uzbekistan.** *J Acquir Immune Defic Syndr* 2005; **39**:570–575.
35. Johnson VA, Brun-Vezinet F, Clotet B, Gunthard HF, Kuritzkes DR, Pillay D, *et al.* **Update of the drug resistance mutations in HIV-1: 2007.** *Top HIV Med* 2007; **15**:119–125.
36. Ivers LC, Kendrick D, Doucette K. **Efficacy of antiretroviral therapy programs in resource-poor settings: a meta-analysis of the published literature.** *Clin Infect Dis* 2005; **41**:217–224.
37. Severe P, Leger P, Charles M, Noel F, Bonhomme G, Bois G, *et al.* **Antiretroviral therapy in a thousand patients with AIDS in Haiti.** *N Engl J Med* 2005; **353**:2325–2334.
38. Etard JF, Ndiaye I, Thierry-Mieg M, Gueye NF, Gueye PM, Laniece I, *et al.* **Mortality and causes of death in adults receiving highly active antiretroviral therapy in Senegal: a 7-year cohort study.** *AIDS* 2006; **20**:1181–1189.
39. Ferradini L, Jeannin A, Pinoges L, Izopet J, Odhiambo D, Mankambo L, *et al.* **Scaling up of highly active antiretroviral therapy in a rural district of Malawi: an effectiveness assessment.** *Lancet* 2006; **367**:1335–1342.
40. Gilks CF, Crowley S, Ekpini R, Gove S, Perriens J, Souteyrand Y, *et al.* **The WHO public-health approach to antiretroviral treatment against HIV in resource-limited settings.** *Lancet* 2006; **368**:505–510.
41. Blower S, Bodine E, Kahn J, McFarland W. **The antiretroviral rollout and drug-resistant HIV in Africa: insights from empirical data and theoretical models.** *AIDS* 2005; **19**:1–14.
42. Hastings IM, Laloo D, Khoo SH. **Will ART rollout in Africa drive an epidemic of drug resistant HIV?** *AIDS* 2006; **20**:1354–1356 [author reply 1356–1357].
43. Hastings IM, Laloo DG, Khoo SH. **Continued correspondence 'Will ART rollout in Africa drive an epidemic of drug-resistant HIV?'** *AIDS* 2007; **21**:258–259 [author reply 259–260].
44. Bennett DE. **The requirement for surveillance of HIV drug resistance within antiretroviral rollout in the developing world.** *Curr Opin Infect Dis* 2006; **19**:607–614.
45. McNulty A, Jennings C, Bennett D, Fitzgibbon J, Bremer JW, Ussery M, *et al.* **Evaluation of dried blood spots for human immunodeficiency virus type 1 drug resistance testing.** *J Clin Microbiol* 2007; **45**:517–521.
46. Bertagnolio S, Soto-Ramirez L, Pilon R, Rodriguez R, Viveros M, Fuentes L, *et al.* **HIV-1 drug resistance surveillance using dried whole blood spots.** *Antivir Ther* 2007; **12**:107–113.
47. Gifford RJ, de Oliveira T, Rambaut A, Pybus OG, Dunn D, Vandamme AM, *et al.* **Phylogenetic surveillance of viral genetic diversity and the evolving molecular epidemiology of human immunodeficiency virus type 1.** *J Virol* 2007; **81**:13050–13056.
48. Hue S, Pillay D, Clewley JP, Pybus OG. **Genetic analysis reveals the complex structure of HIV-1 transmission within defined risk groups.** *Proc Natl Acad Sci U S A* 2005; **102**:4425–4429.
49. de Oliveira T, Pybus OG, Rambaut A, Salemi M, Cassol S, Ciccozzi M, *et al.* **Molecular epidemiology: HIV-1 and HCV sequences from Libyan outbreak.** *Nature* 2006; **444**:836–837.