J Antimicrob Chemother
doi:10.1093/jac/dkz417

Journal of Antimicrobial Chemotherapy

Integrase strand transfer inhibitor (INSTI)-resistance mutations for the surveillance of transmitted HIV-1 drug resistance

Philip L. Tzou [1]*, Soo-Yon Rhee[1], Diane Descamps[2], Dana S. Clutter[3], Bradley Hare[4], Orna Mor[5], Maxime Grude[6], Neil Parkin[7], Michael R. Jordan[8], Silvia Bertagnolio[9], Jonathan M. Schapiro[10], P. Richard Harrigan[11], Anna Maria Geretti[12], Anne-Geneviève Marcelin[6] and Robert W. Shafer[1] on behalf of the WHO HIVResNet Working Groups

[1]Division of Infectious Diseases, Department of Medicine, Stanford University, Stanford, CA, USA; [2]Université de Paris, IAME, INSERM, F-75018, Paris, France; AP-HP, Hôpital Bichat, Laboratoire de Virologie, F-75018, Paris, France; [3]Kaiser-Permanente Medical Care Program – Northern California, South San Francisco, CA, USA; [4]Kaiser-Permanente Medical Care Program – Northern California, San Francisco, CA, USA; [5]Central Virology Laboratory, Sheba Medical Center, Ministry of Health, Ramat-Gan, Israel and Sackler School of Medicine, Tel-Aviv University, Ramat-Aviv, Israel; [6]Sorbonne Université, INSERM, Institut Pierre Louis d'Epidémiologie et de Santé Publique (IPLESP), AP-HP, Hôpital Pitié-Salpêtrière, Department of Virology, F-75013, Paris, France; [7]Data First Consulting Inc., Sebastopol, CA, USA; [8]Tufts University School of Medicine, Boston, MA, USA; [9]Department of HIV and Global Hepatitis Programme, WHO, Geneva, Switzerland; [10]National Hemophilia Center, Shaba Medical Center, Ramat Gan, Israel; [11]Department of Medicine, University of British Columbia, Vancouver, British Columbia, Canada; [12]Department of Clinical Infection, Microbiology and Immunology, Institute of Infection and Global Health, University of Liverpool, Liverpool, UK

*Corresponding author. E-mail: philiptz@stanford.edu

Received 19 June 2019; returned 1 August 2019; revised 3 September 2019; accepted 6 September 2019

Background: Integrase strand transfer inhibitors (INSTIs) are expected to be widely adopted globally, requiring surveillance of resistance emergence and transmission.

Objectives: We therefore sought to develop a standardized list of INSTI-resistance mutations suitable for the surveillance of transmitted INSTI resistance.

Methods: To characterize the suitability of the INSTI-resistance mutations for transmitted HIV-1 drug resistance (TDR) surveillance, we classified them according to their presence on published expert lists, conservation in INSTI-naive persons, frequency in INSTI-treated persons and contribution to reduced *in vitro* susceptibility. Mutation prevalences were determined using integrase sequences from 17302 INSTI-naive and 2450 INSTI-treated persons; 53.3% of the INSTI-naive sequences and 20.0% of INSTI-treated sequences were from non-B subtypes. Approximately 10% of sequences were from persons who received dolutegravir alone or a first-generation INSTI followed by dolutegravir.

Results: Fifty-nine previously recognized (or established) INSTI-resistance mutations were present on one or more of four published expert lists. They were classified into three main non-overlapping groups: 29 relatively common non-polymorphic mutations, occurring in five or more individuals and significantly selected by INSTI treatment; 8 polymorphic mutations; and 22 rare mutations. Among the 29 relatively common INSTI-selected mutations, 24 emerged as candidates for inclusion on a list of INSTI surveillance drug-resistance mutations: T66A/I/K, E92G/Q, G118R, F121Y, E138A/K/T, G140A/C/S, Y143C/H/R/S, S147G, Q148H/R/K, N155H, S230R and R263K.

Conclusions: A set of 24 non-polymorphic INSTI-selected mutations is likely to be useful for quantifying INSTI-associated TDR. This list may require updating as more sequences become available from INSTI-experienced persons infected with HIV-1 non-subtype B viruses and/or receiving dolutegravir.

Introduction

Monitoring transmitted HIV-1 drug resistance (TDR) is performed genotypically by testing for drug-resistance mutations (DRMs) highly specific for antiretroviral (ARV) drug selective pressure. A consensus list of surveillance DRMs (SDRMs) makes it possible to compare the prevalence of TDR over time and across different

regions. However, developing such a consensus list of DRMs is challenging because many mutations contribute to reduced drug susceptibility, including some that occur as natural variants or polymorphisms in ARV drug-naive individuals.

We previously proposed a set of SDRMs for HIV-1 protease and reverse transcriptase (RT) based on the following criteria: (i) recognized as DRMs by HIV drug-resistance experts; (ii) being non-polymorphic regardless of subtype; and (iii) not being exceedingly rare (i.e. mutations resulting exceedingly rarely from drug pressure were excluded).[1,2] This set of SDRMs has been widely used to facilitate meta-analyses of TDR surveillance data generated by different public health and research groups, particularly in low- and middle-income countries.[3,4]

Dolutegravir, one of the drugs within the integrase strand transfer inhibitor (INSTI) class, has been recommended by the 2018 WHO ARV guidelines as the preferred first-line ART.[5,6] Several low- and middle-income countries have already transitioned to dolutegravir and many more are in the planning phase; it is therefore important to report the emergence or transmission of resistance to this class of drugs using standardized approaches that generate comparable findings.[7] Although INSTI resistance has been rare in population-based studies,[8–14] there have now been multiple case reports of INSTI-associated TDR.[15–19] However, because of the many mutations reported to be associated with reduced INSTI susceptibility, there is no standardized SDRM list to use when quantifying INSTI-associated TDR. We analysed published data on INSTI-associated DRMs to develop a consensus list of candidate INSTI-associated surveillance DRMs using an approach similar to that used for an earlier list of HIV-1 protease and RT SDRMs.

Methods

Established INSTI-resistance mutations

Previously recognized (or established) INSTI-resistance mutations were defined as mutations present on one or more of the following four mutation lists: Stanford HIV Drug Resistance Database (HIVDB), version 8.8; National Agency for AIDS Research (ANRS), version 29; REGA Institute, version 10; and International Antiviral Society (IAS)-USA, version 2017. INSTI-resistance positions were also ranked according to the number of mutations at each position on these mutation lists, the extent of conservation of the position and the literature on the mechanistic importance of the mutation to reduced INSTI susceptibility.

Non-polymorphic INSTI-resistance mutations

The prevalence of each mutation in INSTI-naive and INSTI-treated persons was determined using HIV-1 group M integrase sequences in the HIVDB generated by direct PCR dideoxynucleotide sequencing. Sequences containing two or more signature APOBEC mutations were excluded. Mutations occurring in more than one sequence from the same persons were counted just once.

Mutation prevalences were calculated using two approaches. In the primary approach, amino acids were counted regardless of whether they occurred alone or as part of an electrophoretic mixture (i.e. the amino acid was present in combination with another amino acid—usually the WT consensus). In a secondary analysis, mutation prevalence was determined counting only those amino acids present in pure unmixed form. Mutations that had large differences in prevalence among INSTI-naive or INSTI-treated individuals between the two approaches were noted.

HIV-1 subtype was determined using the HIVDB subtyping program. The prevalence of each amino acid at each integrase position was determined among all INSTI-naive and INSTI-treated persons and for each HIV-1 subtype and circulating recombinant form (CRF). Non-polymorphic mutations were defined as mutations occurring at a prevalence <0.2% in all INSTI-naive persons, <0.5% in all subtypes for which >1000 persons were available and <1.0% in all subtypes for which >200 persons were available.

INSTI-selected mutations

Each integrase amino acid variant was examined for its association with INSTI selection pressure by comparing its proportion in INSTI-treated individuals with its proportion in INSTI-naive individuals using Fisher's exact test. Holm's method was used to control the family-wise error rate for multiple hypothesis testing at an adjusted P value ≤ 0.05.[20] This analysis was performed separately for established INSTI-resistance mutations and for all remaining integrase amino acids that occurred in five or more INSTI-treated persons and were at least three times as common among INSTI-treated compared with INSTI-naive persons.

Phenotypic susceptibility

To characterize the effects of established INSTI-resistance mutations on raltegravir, elvitegravir and dolutegravir susceptibility, we analysed HIV-1 isolates in the HIVDB for which *in vitro* phenotypic susceptibility data were available for each of these INSTIs. Phenotypic results were expressed as the fold change in susceptibility, defined as the EC_{50} of the tested isolate divided by the EC_{50} of the standard WT control isolate used for the assay. Viruses from the same individual that contained the same pattern of INSTI-resistance mutations were excluded.

The effects of the established INSTI-resistance mutations on susceptibility were quantified using least squares regression, an approach previously published for each of the other ARV drug classes.[21–23] The presence or absence of each mutation was an explanatory variable and the $\log_{10}$-fold change in susceptibility was the response variable. In each INSTI regression model, mutation coefficients were proportional to the contribution of the mutation to reduced susceptibility. Five-fold cross-validation was performed on randomly chosen subdivisions of the complete dataset. The complete dataset used for the regression analyses is downloadable (HIVDB Genotype-Phenotype Datasets, https://hivdb.stanford.edu/pages/genopheno.dataset.html).

Principal components analysis

We performed a principal components analysis to jointly analyse the INSTI-associated DRMs, according to six often overlapping characteristics:[24] (i) recognizability: DRMs on four lists were assigned a value of 2 and those on two or three lists were assigned a 1; (ii) positional importance: DRMs at positions 66, 92, 118, 121, 138, 140, 143, 148, 155 and 263 were assigned a 2 and those at positions 145, 146, 147, 149, 151 and 153 were assigned a 1; (iii) non-polymorphism: DRMs with a naive prevalence <0.2 were assigned a 2 and those with a naive prevalence between 0.2 and 1.0 were assigned a 1; (iv) frequency: DRMs occurring in $\geq$10 INSTI-treated persons were assigned a 2 and those in five to nine persons were assigned a 1; (v) statistical association with INSTI therapy: DRMs with a corrected P value $\leq$0.000001 (1.0E−6) were assigned a 2 and those with a corrected P value between 1.0E−6 and 0.05 were assigned a 1; and (vi) statistical association with reduced susceptibility: DRMs with a linear regression coefficient $\geq$1.0 for one or more INSTIs were assigned a 2 and those with a coefficient between 0.5 and 1.0 were assigned a 1. If not otherwise applicable a value of 0 was assigned for each characteristic.

Correlation network analysis

We selected all integrase sequences containing 1 or more of the 59 INSTI-associated DRMs. For persons with more than one integrase sequence, we included only those sequences with a non-redundant pattern of INSTI-

Table 1. HIV-1 integrase mutations reported on four lists of INSTI-resistance mutations

Position	Consensus amino acid	Amino acid	HIVDB	IAS-USA	ANRS	REGA Institute	Number of algorithms containing the mutation
66	T	A	Y	Y	Y	Y	4
66	T	I	Y	Y	Y	Y	4
66	T	K	Y	Y	Y	Y	4
74	L	M	Y	Y	Y	Y	4
92	E	Q	Y	Y	Y	Y	4
97	T	A	Y	Y	Y	Y	4
118	G	R	Y	Y	Y	Y	4
121	F	Y	Y	Y	Y	Y	4
138	E	A	Y	Y	Y	Y	4
138	E	K	Y	Y	Y	Y	4
140	G	A	Y	Y	Y	Y	4
140	G	S	Y	Y	Y	Y	4
143	Y	C	Y	Y	Y	Y	4
143	Y	H	Y	Y	Y	Y	4
143	Y	R	Y	Y	Y	Y	4
147	S	G	Y	Y	Y	Y	4
148	Q	H	Y	Y	Y	Y	4
148	Q	K	Y	Y	Y	Y	4
148	Q	R	Y	Y	Y	Y	4
155	N	H	Y	Y	Y	Y	4
263	R	K	Y	Y	Y	Y	4
74	L	I	Y		Y	Y	3
92	E	G	Y	Y		Y	3
140	G	C	Y		Y	Y	3
143	Y	S	Y		Y	Y	3
145	P	S	Y		Y	Y	3
151	V	L	Y		Y	Y	3
153	S	F	Y		Y	Y	3
153	S	Y	Y		Y	Y	3
155	N	S	Y		Y	Y	3
155	N	T	Y		Y	Y	3
230	S	R	Y		Y	Y	3
51	H	Y	Y			Y	2
74	L	F	Y		Y		2
75	V	I	Y		Y		2
92	E	V	Y			Y	2
138	E	T	Y		Y		2
143	Y	A	Y		Y		2
143	Y	G	Y		Y		2
143	Y	K	Y			Y	2
146	Q	P	Y			Y	2
149	G	A	Y				1
151	V	A	Y			Y	2
157	E	Q	Y		Y		2
163	G	K	Y			Y	2
49	A	G				Y	1
49	A	P				Y	1
95	Q	K	Y				1
114	H	Y				Y	1
146	Q	I				Y	1
146	Q	K				Y	1
146	Q	L				Y	1
146	Q	R				Y	1
148	Q	E			Y		1
148	Q	G			Y		1
148	Q	N	Y				1
163	G	R	Y				1
230	S	G				Y	1
232	D	N	Y				1

associated DRMs. Additionally, positions containing mixtures of a DRM and another amino acid were ignored. The first step in the network analysis was to create a list of positively correlated substitution pairs having a non-parametric Spearman correlation coefficient (ρ) of >0.075, at least three occurrences of each mutation and a P value $\leq$0.00001. We used the R package igraph to create an undirected weighted network graph from the adjacency matrix of positively correlated amino acid substitution pairs.[25] In this network, an edge was created between all correlated amino acid substitution pairs meeting the above criteria. Edge widths were positively correlated with ρ.

Results

Established INSTI-resistance mutations

Table 1 shows that 59 mutations at 26 positions were on one or more mutation lists, including 21 on four lists, 11 on three lists, 12 on two lists and 15 on one list. Positions 66, 92, 118, 121, 138, 140, 143, 148, 155 and 263 were classified as important positions because they either had multiple DRMs (positions 66, 92, 138, 140, 143, 148 and 155) or had a single DRM at a conserved position associated with a published mechanism of INSTI resistance (positions 118, 121 and 263).[26,27] Positions 145–147, 149, 151 and 153 were considered to be in important regions of integrase by virtue of being in the flexible conserved 'catalytic loop' extending between positions 140 and 149 or of being adjacent to the catalytic glutamate at position 152.[26]

Mutation prevalence

Available sequences

The prevalences of the 59 established INSTI-resistance mutations were determined using 17425 sequences from 17302 INSTI-naive persons and 2783 sequences from 2450 INSTI-treated persons. The sequences from INSTI-treated persons were from 1741 (71.1%) raltegravir-treated persons, 285 (11.6%) elvitegravir-treated persons, 152 (6.2%) dolutegravir-treated persons, 196 (8.0%) persons who received more than one INSTI and 76 (3.1%) from persons with uncertain INSTI treatment histories.

Table 2. Prevalence of INSTI-resistance mutations in INSTI-naive and -treated persons

Position	Amino acid	Prevalence in INSTI-naive persons: all subtypes (*n*=17302), no. of persons (%)	A (*n*=1447)	B (*n*=8040)	C (*n*=3194)	D (*n*=484)	F (*n*=268)	G (*n*=244)	01 (*n*=2019)	02 (*n*=978)	highest	Treated prevalence: all subtypes (*n*=2450), no. of persons (%)	Corrected *P* value
Non-polymorphic													
51	Y	4 (0.0%)	0.1%	0.0%	0.0%	0.0%	0.0%	0.0%	0.1%	0.0%	0.1%	12 (0.5%)	<1.0E−6
66	A	7 (0.0%)	0.1%	0.0%	0.1%	0.0%	0.0%	0.0%	0.0%	0.1%	0.1%	17 (0.7%)	<1.0E−6
66	I	8 (0.0%)	0.1%	0.0%	0.1%	0.2%	0.0%	0.0%	0.0%	0.0%	0.2%	36 (1.5%)	<1.0E−6
66	K	0 (0.0%)	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	6 (0.2%)	0.0001
75	I	14 (0.1%)	0.1%	0.1%	0.1%	0.4%	0.0%	0.0%	0.0%	0.0%	0.4%	22 (0.9%)	<1.0E−6
92	G	4 (0.0%)	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.1%	0.1%	10 (0.4%)	0.00001
92	Q	1 (0.0%)	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	137 (5.6%)	<1.0E−6
95	K	24 (0.1%)	0.1%	0.0%	0.5%	0.0%	0.4%	0.0%	0.0%	0.3%	0.5%	27 (1.1%)	<1.0E−6
114	Y	4 (0.0%)	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	5 (0.2%)	0.05
118	R	0 (0.0%)	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	9 (0.4%)	<1.0E−6
121	Y	1 (0.0%)	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	6 (0.2%)	0.0006
138	A	1 (0.0%)	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	47 (1.9%)	<1.0E−6
138	K	32 (0.2%)	0.1%	0.3%	0.1%	0.4%	0.0%	0.0%	0.0%	0.0%	0.4%	127 (5.2%)	<1.0E−6
138	T	2 (0.0%)	0.1%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.1%	15 (0.6%)	<1.0E−6
140	A	0 (0.0%)	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	40 (1.6%)	<1.0E−6
140	C	0 (0.0%)	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	9 (0.4%)	<1.0E−6
140	S	5 (0.0%)	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	379 (15.5%)	<1.0E−6
143	C	0 (0.0%)	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	81 (3.3%)	<1.0E−6
143	H	3 (0.0%)	0.0%	0.0%	0.1%	0.0%	0.0%	0.0%	0.0%	0.0%	0.1%	44 (1.8%)	<1.0E−6
143	R	0 (0.0%)	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	134 (5.5%)	<1.0E−6
143	S	2 (0.0%)	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	9 (0.4%)	0.000009
146	R	0 (0.0%)	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	8 (0.3%)	0.000002
147	G	5 (0.0%)	0.0%	0.0%	0.0%	0.2%	0.0%	0.0%	0.1%	0.0%	0.2%	45 (1.8%)	<1.0E−6
148	H	4 (0.0%)	0.0%	0.0%	0.0%	0.0%	0.0%	0.4%	0.0%	0.0%	0.4%	344 (14.0%)	<1.0E−6
148	K	1 (0.0%)	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.1%	0.1%	15 (0.6%)	<1.0E−6
148	R	7 (0.0%)	0.0%	0.0%	0.0%	0.2%	0.0%	0.0%	0.0%	0.0%	0.2%	159 (6.5%)	<1.0E−6
155	H	0 (0.0%)	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	575 (23.5%)	<1.0E−6
230	R	7 (0.0%)	0.1%	0.0%	0.1%	0.0%	0.0%	0.0%	0.0%	0.0%	0.1%	50 (2.1%)	<1.0E−6
263	K	22 (0.1%)	0.2%	0.2%	0.0%	0.0%	0.8%	0.0%	0.0%	0.2%	0.8%	24 (1.0%)	<1.0E−6
Polymorphic													
49	P	42 (0.2%)	0.1%	0.4%	0.0%	0.8%	0.4%	0.0%	0.1%	0.0%	0.8%	16 (0.7%)	0.01
74	I	1259 (7.3%)	24.1%	4.1%	5.6%	3.9%	4.5%	11.9%	2.2%	17.0%	24.1%	176 (7.2%)	1
74	M	262 (1.5%)	2.0%	0.8%	0.8%	1.4%	1.5%	4.5%	1.3%	8.3%	8.3%	133 (5.4%)	<1.0E−6
97	A	305 (1.8%)	6.5%	0.6%	0.8%	5.2%	6.0%	2.9%	0.6%	6.3%	6.5%	292 (11.9%)	<1.0E−6
157	Q	400 (2.3%)	1.2%	3.2%	0.7%	3.9%	0.4%	1.6%	0.7%	6.4%	6.4%	132 (5.4%)	<1.0E−6
163	K	40 (0.2%)	0.1%	0.2%	0.1%	0.2%	5.6%	0.0%	0.1%	0.3%	5.6%	32 (1.3%)	<1.0E−6
163	R	81 (0.5%)	0.2%	0.3%	0.6%	0.2%	7.8%	0.4%	0.2%	0.0%	7.8%	116 (4.7%)	<1.0E−6
232	N	72 (0.4%)	1.1%	0.5%	0.2%	0.2%	0.0%	0.8%	0.0%	0.2%	1.1%	119 (5.0%)	<1.0E−6
Rare													
49	G	4 (0.0%)	0.2%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.2%	4 (0.2%)	0.1
74	F	0 (0.0%)	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	1 (0.0%)	1
92	V	0 (0.0%)	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	1 (0.0%)	1
143	A	1 (0.0%)	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	2 (0.1%)	0.6
143	G	0 (0.0%)	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	4 (0.2%)	0.006
143	K	0 (0.0%)	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	1 (0.0%)	1
145	S	4 (0.0%)	0.0%	0.0%	0.0%	0.2%	0.0%	0.0%	0.0%	0.0%	0.2%	4 (0.2%)	0.2
146	I	0 (0.0%)	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0 (0.0%)	1

Continued

Table 2. *Continued*

Position	Amino acid	Prevalence in INSTI-naive persons: all subtypes (*n*=17302), no. of persons (%)	A (*n*=1447)	B (*n*=8040)	C (*n*=3194)	D (*n*=484)	F (*n*=268)	G (*n*=244)	01 (*n*=2019)	02 (*n*=978)	highest	Treated prevalence: all subtypes (*n*=2450), no. of persons (%)	Corrected *P* value
146	K	1 (0.0%)	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	4 (0.2%)	0.02
146	L	1 (0.0%)	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	1 (0.0%)	1
146	P	2 (0.0%)	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	1 (0.0%)	1
148	E	0 (0.0%)	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0 (0.0%)	1
148	G	0 (0.0%)	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0 (0.0%)	1
148	N	2 (0.0%)	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	3 (0.1%)	0.3
149	A	8 (0.0%)	0.0%	0.1%	0.0%	0.0%	0.0%	0.0%	0.0%	0.1%	0.1%	2 (0.1%)	1
151	A	2 (0.0%)	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	4 (0.2%)	0.06
151	L	2 (0.0%)	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.1%	0.1%	1 (0.0%)	1
153	F	11 (0.1%)	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.2%	0.0%	0.2%	2 (0.1%)	1
153	Y	1 (0.0%)	0.0%	0.0%	0.0%	0.0%	0.4%	0.0%	0.0%	0.0%	0.4%	3 (0.1%)	0.1
155	S	0 (0.0%)	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	3 (0.1%)	0.04
155	T	1 (0.0%)	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.1%	0.1%	2 (0.1%)	0.6
230	G	28 (0.2%)	0.2%	0.3%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.3%	3 (0.1%)	1

The parenthetical numbers in the headers (i.e. the denominators for mutation prevalence) represent the total number of persons from whom a sequence was available. However, not all sequences encompassed all positions. We include the range in the number of sequences here: INSTI-naive all subtypes, *n*=17113–17302; INSTI-naive A, *n*=1426–1477; INSTI-naive B, *n*=7976–8040; INSTI-naive C, *n*=3156–3194; INSTI-naive D, *n*=478–484; INSTI-naive F, *n*=265–268; INSTI-naive G, *n*=235–244; INSTI-naive CRF01_AE, *n*=2012–2019; INSTI-naive CRF02_AG, *n*=943–978; and INSTI-treated all subtypes, *n*=2226–2450.

The sequences from INSTI-naive individuals included 8136 (46.7%) subtype B, 3194 (18.3%) subtype C, 2019 (11.6%) CRF01_AE, 1456 (8.4%) subtype A, 978 (5.6%) CRF02_AG, 486 (2.8%) subtype D, 268 (1.5%) subtype F, 244 (1.4%) subtype G and 644 (3.7%) sequences belonging to other CRFs and unique recombinant forms. The sequences from INSTI-treated persons included 2226 (80.0%) subtype B sequences; 557 (20.0%) sequences belonged to other subtypes, CRFs and unique recombinant forms.

Non-polymorphic INSTI-selected DRMs

Table 2 shows the prevalence of each of the 59 established INSTI-resistance mutations among INSTI-naive and INSTI-treated persons overall and among INSTI-naive persons according to subtype. The mutations are divided into: (i) 29 non-polymorphic mutations present in five or more INSTI-treated persons; (ii) 8 polymorphic mutations present in five or more INSTI-treated persons; and (iii) 22 rare mutations present in fewer than five INSTI-treated persons.

Of the 29 non-polymorphic mutations present in five or more INSTI-treated persons, each was significantly selected by INSTI treatment with corrected *P* values $\leq$0.05: H51Y, T66A/I/K, V75I, E92G/Q, Q95K, H114Y, G118R, F121Y, E138A/K/T, G140A/C/S, Y143C/H/R/S, Q146R, S147G, Q148H/K/R, N155H, S230R and R263K (Table 2). However, H51Y, Q95K and H114Y were often present as part of an electrophoretic mixture in combination with the consensus amino acid and would not have met statistical significance had mutations been restricted to those occurring in an unmixed form (Table S1, available as Supplementary data at *JAC* Online). All of the non-polymorphic INSTI-selected mutations were at highly conserved positions, with the exception of S230R because S230N was present in 5.0% of INSTI-naive sequences.

Among the 29 non-polymorphic INSTI-selected mutations, Q95K, E138K and R263K were the only mutations with a prevalence in naive persons >0.1% (0.14% for Q95K, 0.18% for E138K and 0.13% for R263K). Nineteen of the 22 R263K mutations and 10 of the 32 E138K mutations occurred as part of a mixture with WT (Table S2). Although E138K and R263K could arise as a result of APOBEC-mediated DNA editing, none of the sequences with either of these mutations contained a stop codon or had other evidence for G-to-A hypermutation.

Polymorphic established INSTI-resistance DRMs

The eight polymorphic established INSTI-resistance mutations were A49P, L74I/M, T97A, E157Q, G163K/R and D232N. With the exception of L74I, which was the most polymorphic, each was significantly associated with INSTI treatment. E157Q occurred in 2.3% of all naive sequences with prevalences of 3.2% in naive subtype B and 6.4% in naive CRF02_AG sequences. L74M and T97A had overall naive prevalences of 1.5% and 1.8%, respectively. However, L74M occurred in 8.3% of naive CRF02_AG sequences and T97A occurred in >5% of naive subtype A and CRF02_AG sequences. G163K/R had prevalences just below 0.5% overall, but had prevalences >5% in subtype F sequences. A49P and D232N had overall prevalences of 0.2% and 0.4%, respectively.

Table 3. Linear regression coefficients for each INSTI-resistance mutation's contribution to the $\log_{10}$-fold reduction in INSTI susceptibility

Position	Amino acid	No. of isolates tested[a]	Linear regression coefficient[b]				Coefficient ≥0.5
			raltegravir	elvitegravir	dolutegravir	maximum[c]	
Non-polymorphic							
66	A	14	−0.12 $_{14}$	0.57 $_{14}$	−0.29 $_{6}$	0.57	yes
66	I	58	0.06 $_{50}$	0.72 $_{55}$	−0.30 $_{25}$	0.72	yes
66	K	7	1.10 $_{5}$	1.56 $_{5}$	0.37 $_{6}$	1.56	yes
92	G	8	0.14 $_{8}$	0.85 $_{8}$	0.16 $_{2}$	0.85	yes
92	Q	96	0.55 $_{78}$	0.98 $_{75}$	0.16 $_{42}$	0.98	yes
118	R	23	0.79 $_{19}$	0.39 $_{18}$	0.97 $_{21}$	0.97	yes
121	Y	26	1.03 $_{21}$	1.30 $_{23}$	0.25 $_{11}$	1.30	yes
138	T	3	0.01 $_{2}$	−0.34 $_{2}$	0.74 $_{3}$	0.74	yes
140	A	30	0.45 $_{27}$	0.47 $_{23}$	0.62 $_{19}$	0.62	yes
140	C	10	0.41 $_{9}$	0.58 $_{9}$	0.27 $_{9}$	0.58	yes
140	S	159	0.37 $_{143}$	0.54 $_{106}$	0.60 $_{115}$	0.60	yes
143	C	45	0.62 $_{44}$	0.06 $_{36}$	0.11 $_{24}$	0.62	yes
143	R	62	1.20 $_{56}$	0.22 $_{43}$	0.07 $_{37}$	1.20	yes
143	S	14	0.77 $_{14}$	0.14 $_{12}$	0.07 $_{2}$	0.77	yes
147	G	49	0.12 $_{33}$	0.51 $_{41}$	0.02 $_{21}$	0.51	yes
148	H	139	1.37 $_{126}$	1.03 $_{92}$	0.25 $_{102}$	1.37	yes
148	K	36	1.41 $_{34}$	1.45 $_{27}$	0.68 $_{24}$	1.45	yes
148	R	118	1.29 $_{100}$	1.22 $_{85}$	0.26 $_{83}$	1.29	yes
155	H	159	0.97 $_{140}$	1.02 $_{112}$	0.28 $_{98}$	1.02	yes
230	R	39	0.43 $_{36}$	0.51 $_{30}$	−0.12 $_{10}$	0.51	yes
263	K	55	0.05 $_{45}$	0.72 $_{38}$	0.57 $_{48}$	0.72	yes
51	Y	24	0.09 $_{16}$	0.32 $_{18}$	0.21 $_{13}$	0.32	
75	I	16	0.12 $_{16}$	0.29 $_{16}$	0.09 $_{14}$	0.29	
95	K	16	−0.08 $_{11}$	0.24 $_{14}$	0.04 $_{10}$	0.24	
114	Y	2	−0.10 $_{2}$	0.13 $_{2}$	NA	0.13	
138	A	20	0.22 $_{15}$	0.29 $_{9}$	−0.06 $_{19}$	0.29	
138	K	93	0.18 $_{78}$	0.05 $_{70}$	0.20 $_{71}$	0.20	
143	H	23	0.21 $_{23}$	−0.11 $_{21}$	0.20 $_{6}$	0.21	
146	R	2	0.16 $_{2}$	0.42 $_{2}$	0.34 $_{2}$	0.42	
Polymorphic							
49	P	3	0.15 $_{3}$	−0.11 $_{3}$	1.10 $_{2}$	1.10	yes
163	K	6	0.61 $_{4}$	0.61 $_{4}$	0.03 $_{6}$	0.61	yes
74	I	69	0.18 $_{57}$	0.09 $_{55}$	0.06 $_{32}$	0.18	
74	M	101	0.24 $_{83}$	0.19 $_{78}$	0.16 $_{61}$	0.24	
97	A	177	0.36 $_{154}$	0.46 $_{131}$	0.17 $_{105}$	0.46	
157	Q	80	0.11 $_{70}$	0.06 $_{62}$	0.17 $_{48}$	0.17	
163	R	42	0.18 $_{32}$	0.11 $_{29}$	0.12 $_{22}$	0.18	
232	N	140	0.07 $_{135}$	0.09 $_{133}$	0.01 $_{13}$	0.09	
Rare							
74	F	13	0.66 $_{13}$	0.64 $_{13}$	0.11 $_{13}$	0.66	yes
92	V	5	0.69 $_{5}$	1.17 $_{5}$	0.28 $_{1}$	1.17	yes
143	A	12	1.09 $_{12}$	0.38 $_{12}$	NA	1.09	yes
143	G	15	1.02 $_{15}$	0.25 $_{13}$	−0.09 $_{3}$	1.02	yes
143	K	2	1.10 $_{2}$	NA	0.56 $_{2}$	1.10	yes
145	S	2	−0.09 $_{2}$	1.58 $_{2}$	−0.25 $_{2}$	1.58	yes
146	I	1	0.52 $_{1}$	1.86 $_{1}$	0.51 $_{1}$	1.86	yes
146	L	3	0.25 $_{3}$	0.89 $_{3}$	0.28 $_{2}$	0.89	yes
146	P	7	0.05 $_{2}$	0.53 $_{7}$	−0.25 $_{2}$	0.53	yes
151	A	5	0.56 $_{4}$	0.45 $_{4}$	0.01 $_{2}$	0.56	yes
151	L	2	1.06 $_{2}$	1.24 $_{2}$	0.54 $_{2}$	1.24	yes

Continued

Table 3. *Continued*

Position	Amino acid	No. of isolates tested[a]	Linear regression coefficient[b] raltegravir	elvitegravir	dolutegravir	maximum[c]	Coefficient ≥0.5
153	Y	18	0.11 $_{15}$	0.37 $_{16}$	0.51 $_{9}$	0.51	yes
155	S	5	0.87 $_{5}$	1.28 $_{4}$	0.31 $_{1}$	1.28	yes
155	T	1	0.73 $_{1}$	1.45 $_{1}$	0.44 $_{1}$	1.45	yes
49	G	3	0.15 $_{3}$	−0.47 $_{3}$	−0.19 $_{3}$	0.15	
146	K	1	−0.08 $_{1}$	0.20 $_{1}$	0.07 $_{1}$	0.20	
148	E	0	NA	NA	NA	NA	
148	G	0	NA	NA	NA	NA	
148	N	4	−0.33 $_{4}$	0.12 $_{4}$	−0.36 $_{4}$	0.12	
149	A	4	0.10 $_{4}$	0.07 $_{4}$	0.21 $_{4}$	0.21	
153	F	9	−0.02 $_{8}$	0.34 $_{9}$	0.20 $_{4}$	0.34	
230	G	6	0.00 $_{6}$	−0.10 $_{6}$	0.22 $_{3}$	0.22	

NA, no phenotypic results were available for this mutation/drug combination.
[a]The number of HIV-1 isolates undergoing susceptibility testing to raltegravir, elvitegravir or dolutegravir. The specific number of tests per INSTI is the subscript in columns 4, 5 and 6.
[b]The effect of INSTI-resistance mutations on susceptibility was quantified using least squares regression.
[c]The highest of the three coefficients for raltegravir, elvitegravir and dolutegravir.

Rare established INSTI-resistance DRMs

Twenty-two of the established INSTI-resistance mutations were rare, occurring in fewer than five INSTI-treated persons. Three of these 22 mutations were significantly selected by INSTI therapy (Y143G, Q146K and N155S). These rare mutations included 13 of the 44 mutations on two or more lists and 9 of the 15 mutations on just one list.

In vitro susceptibility

Phenotypic susceptibility results were available on 1717 virus isolates, of which 988 (57.5%) were tested using the PhenoSense assay. Raltegravir susceptibility results were available on 1555 isolates, of which 970 were tested using the PhenoSense assay. Elvitegravir susceptibility results were available on 1443 isolates, of which 896 were tested using the PhenoSense assay. Dolutegravir susceptibility results were available on 754 isolates, of which 280 were tested using the PhenoSense assay.

Of the 29 non-polymorphic INSTI-associated mutations that occurred in five or more persons, 21 had a regression coefficient $\geq$0.5 $\log_{10}$ to one or more INSTIs, consistent with a $\geq$3.2-fold contribution to reduced susceptibility: T66A/I/K, E92G/Q, G118R, F121Y, E138T, G140A/C/S, Y143C/R/S, S147G, Q148H/K/R, N155H, S230R and R263K (Table 3). The remaining eight non-polymorphic INSTI-associated mutations, H51Y, V75I, Q95K, H114Y, E138A/K, Y143H and Q146R, had regression coefficients <0.5 $\log_{10}$.

Two of the eight polymorphic mutations (A49P and G163K) and 14 of the 23 very rare mutations (L74F, E92V, Y143A/G/K, P145S, Q146I/L/P, V151A/L, S153Y and N155S/T) also had a regression coefficient $\geq$0.5 $\log_{10}$.

Non-established INSTI-selected mutations

Thirty-two additional non-polymorphic mutations that occurred in five or more INSTI-treated persons were significantly selected by INSTI therapy (corrected P value $\leq$0.05), including four mutations at established INSTI-resistance positions (H51D, E92A, Q95R and N155D) and one mutation (N142T) in the highly conserved catalytic loop extending between positions 140 and 149 (Table 4).

Principal components analysis

We jointly analysed each of the 59 established INSTI-resistance mutations according to six often overlapping characteristics—recognizability, positional importance, non-polymorphism, frequency, statistical association with INSTI therapy and statistical association with reduced INSTI susceptibility—as outlined in the Methods section. For each feature, each DRM was scaled to have a high (2), intermediate (1) or low (0) association (Table 5). As the six mutational features were often correlated with one another, we performed a principal components analysis to cluster the mutations in two dimensions. This analysis yielded three main groupings of mutations (Figure 1). One contained 24 of the 29 non-polymorphic INSTI-selected mutations, another contained the 22 rare mutations and a third contained the 8 polymorphic mutations.

Of the 29 non-polymorphic INSTI-selected mutations, 20 were tightly clustered. Three mutations, T66K, F121Y and Y143S, were somewhat less tightly clustered because they occurred in just six to eight treated persons. S230R was also less tightly clustered in part because it was at a position containing the common polymorphism S230N. Five additional non-polymorphic INSTI-selected mutations (H51Y, V75I, Q95K, H114Y and Q146R) did not cluster with the remaining 24 mutations. None of these five mutations was significantly associated with reduced INSTI susceptibility and each was relatively uncommon or owed its association with therapy to usually being present as part of a mixture with WT.

Table 4. Previously unrecognized non-polymorphic mutations with significantly higher prevalence in INSTI-treated compared with INSTI-naive persons

		Prevalence in INSTI-naive persons										Treated prevalence: all subtypes (*n*=2450), no. of persons (%)	Corrected *P* value
Position	Amino acid	all subtypes (*n*=17302), no. of persons (%)	A (*n*=1447)	B (*n*=8040)	C (*n*=3194)	D (*n*=484)	F (*n*=268)	G (*n*=244)	01 (*n*=2019)	02 (*n*=978)	highest		
170	A	2 (0.0%)	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	23 (0.9%)	<1.0E−6
79	I	31 (0.2%)	0.0%	0.3%	0.1%	0.2%	0.0%	0.0%	0.0%	0.1%	0.3%	40 (1.6%)	<1.0E−6
51	D	3 (0.0%)	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	20 (0.9%)	<1.0E−6
70	R	11 (0.1%)	0.1%	0.1%	0.1%	0.0%	0.0%	0.0%	0.0%	0.0%	0.1%	17 (0.7%)	<1.0E−6
112	S	9 (0.1%)	0.0%	0.1%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.1%	14 (0.6%)	0.000002
92	A	1 (0.0%)	0.0%	0.0%	0.0%	0.0%	0.0%	0.4%	0.0%	0.0%	0.4%	9 (0.4%)	0.000002
39	R	30 (0.2%)	0.2%	0.3%	0.1%	0.0%	0.0%	0.0%	0.0%	0.1%	0.3%	20 (1.0%)	0.000003
95	R	10 (0.1%)	0.0%	0.1%	0.1%	0.0%	0.0%	0.0%	0.0%	0.0%	0.1%	14 (0.6%)	0.000004
253	N	8 (0.0%)	0.0%	0.1%	0.0%	0.0%	0.0%	0.0%	0.0%	0.1%	0.1%	12 (0.5%)	0.00001
96	N	1 (0.0%)	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	8 (0.3%)	0.00001
142	T	3 (0.0%)	0.0%	0.0%	0.0%	0.2%	0.0%	0.0%	0.0%	0.0%	0.2%	8 (0.3%)	0.0002
171	R	5 (0.0%)	0.1%	0.0%	0.1%	0.0%	0.0%	0.0%	0.0%	0.0%	0.1%	9 (0.4%)	0.0002
212	L	17 (0.1%)	0.4%	0.1%	0.0%	0.2%	0.8%	0.0%	0.0%	0.1%	0.8%	13 (0.5%)	0.0006
264	R	15 (0.1%)	0.0%	0.1%	0.2%	0.4%	0.0%	0.0%	0.0%	0.1%	0.4%	12 (0.5%)	0.0007
195	H	7 (0.0%)	0.0%	0.1%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.1%	9 (0.4%)	0.0008
196	T	13 (0.1%)	0.0%	0.1%	0.0%	0.2%	0.0%	0.0%	0.0%	0.0%	0.2%	11 (0.5%)	0.001
253	Y	21 (0.1%)	0.3%	0.2%	0.0%	0.2%	0.0%	0.0%	0.0%	0.2%	0.3%	13 (0.6%)	0.002
155	D	3 (0.0%)	0.1%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.1%	6 (0.2%)	0.005
76	V	2 (0.0%)	0.0%	0.0%	0.0%	0.0%	0.4%	0.4%	0.0%	0.0%	0.4%	5 (0.2%)	0.01
160	S	3 (0.0%)	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	5 (0.2%)	0.02
91	V	8 (0.0%)	0.0%	0.0%	0.1%	0.0%	0.0%	0.0%	0.0%	0.0%	0.1%	7 (0.3%)	0.02
177	R	3 (0.0%)	0.0%	0.0%	0.0%	0.0%	0.4%	0.0%	0.0%	0.1%	0.4%	5 (0.2%)	0.02
195	N	11 (0.1%)	0.0%	0.1%	0.0%	0.2%	0.0%	0.0%	0.0%	0.0%	0.2%	8 (0.3%)	0.02
229	N	5 (0.0%)	0.0%	0.0%	0.1%	0.0%	0.0%	0.0%	0.0%	0.0%	0.1%	6 (0.3%)	0.02
195	R	15 (0.1%)	0.1%	0.1%	0.0%	0.0%	0.0%	0.0%	0.1%	0.0%	0.1%	9 (0.4%)	0.02
272	E	3 (0.0%)	0.0%	0.0%	0.0%	0.2%	0.0%	0.0%	0.1%	0.0%	0.2%	5 (0.2%)	0.02
277	R	3 (0.0%)	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.1%	0.0%	0.1%	5 (0.2%)	0.02
241	F	8 (0.0%)	0.0%	0.1%	0.0%	0.2%	0.0%	0.0%	0.0%	0.0%	0.2%	7 (0.3%)	0.02
136	E	6 (0.0%)	0.1%	0.0%	0.0%	0.0%	0.4%	0.0%	0.1%	0.0%	0.4%	6 (0.2%)	0.02
56	Y	19 (0.1%)	0.0%	0.1%	0.2%	0.0%	0.0%	0.0%	0.0%	0.2%	0.2%	10 (0.4%)	0.02
53	K	11 (0.1%)	0.1%	0.1%	0.0%	0.0%	0.0%	0.4%	0.0%	0.2%	0.4%	5 (0.2%)	0.04
68	F	5 (0.0%)	0.0%	0.0%	0.0%	0.0%	0.4%	0.0%	0.0%	0.0%	0.4%	5 (0.2%)	0.05

The parenthetical numbers in the headers (i.e. the denominators for mutation prevalence) represent the total number of persons from whom a sequence was available. However, not all sequences encompassed all positions. We include the range in the number of sequences here: INSTI-naive all subtypes, *n*=17113–17302; INSTI-naive A, *n*=1426–1477; INSTI-naive B, *n*=7976–8040; INSTI-naive C, *n*=3156–3194; INSTI-naive D, *n*=478–484; INSTI-naive F, *n*=265–268; INSTI-naive G, *n*=235–244; INSTI-naive CRF01_AE, *n*=2012–2019; INSTI-naive CRF02_AG, *n*=943–978; and INSTI-treated all subtypes, *n*=2226–2450.

Correlation network analysis

Figure 2 shows that of the 59 INSTI-resistance mutations, 25 frequently co-occurred with one or more other mutations, including 20 non-polymorphic INSTI-selected mutations (H51Y, E92Q, H114Y, E138A/K/T, G140A/C/S, Y143C/H/R/S, S147G, Q148H/K/R, N155H, S230R and R263K), 3 polymorphic INSTI-selected mutations (L74M, T97A and D232N) and 2 rare mutations (A49G and Y143G). Among the known INSTI-resistance mutations at position 148, Q148H was strongly linked with G140S and E138A/T; Q148R with E138K, G140A/C and S147G; and Q148K with E138K and G140A. Y143C/G/H/R were each significantly linked with T97A. Y143C was also strongly linked to S230R.

Discussion

We analysed 59 INSTI-resistance mutations on four expert panel mutation lists according to six often overlapping characteristics. Despite their large number, the mutations could be classified into three main groups: 29 non-polymorphic mutations significantly

Table 5. Established INSTI-resistance mutations scaled according to six often overlapping mutation characteristics

DRM	Recognizability	Positional importance	Non-polymorphism	Treatment frequency	Association with INSTI therapy	Association with reduced susceptibility	Sum
143R	2	2	2	2	2	2	12
148H	2	2	2	2	2	2	12
148K	2	2	2	2	2	2	12
148R	2	2	2	2	2	2	12
155H	2	2	2	2	2	2	12
66A	2	2	2	2	2	1	11
66I	2	2	2	2	2	1	11
92Q	2	2	2	2	2	1	11
140A	2	2	2	2	2	1	11
140S	2	2	2	2	2	1	11
143C	2	2	2	2	2	1	11
147G	2	2	2	2	2	1	11
263K	2	2	2	2	2	1	11
66K	2	2	2	1	1	2	10
92G	2	2	2	2	1	1	10
118R	2	2	2	1	2	1	10
121Y	2	2	2	1	1	2	10
138A	2	2	2	2	2	0	10
138K	2	2	2	2	2	0	10
138T	1	2	2	2	2	1	10
140C	2	2	2	1	2	1	10
143H	2	2	2	2	2	0	10
143S	2	2	2	1	1	1	9
155S	2	2	2	0	1	2	9
230R	2	0	2	2	2	1	9
143G	1	2	2	0	1	2	8
155T	2	2	2	0	0	2	8
51Y	1	0	2	2	2	0	7
75I	1	0	2	2	2	0	7
92V	1	2	2	0	0	2	7
143A	1	2	2	0	0	2	7
143K	1	2	2	0	0	2	7
145S	2	1	2	0	0	2	7
151L	2	1	2	0	0	2	7
163K	1	0	1	2	2	1	7
49P	0	0	1	2	1	2	6
74M	2	0	0	2	2	0	6
95K	0	0	2	2	2	0	6
97A	2	0	0	2	2	0	6
146R	0	1	2	1	2	0	6
153Y	2	1	2	0	0	1	6
146I	0	1	2	0	0	2	5
146P	1	1	2	0	0	1	5
151A	1	1	2	0	0	1	5
153F	2	1	2	0	0	0	5
157Q	1	0	0	2	2	0	5
163R	0	0	1	2	2	0	5
232N	0	0	1	2	2	0	5
74F	1	0	2	0	0	1	4
74I	2	0	0	2	0	0	4
146L	0	1	2	0	0	1	4
148E	0	2	2	0	0	0	4

Continued

Table 5. *Continued*

DRM	Recognizability	Positional importance	Non-polymorphism	Treatment frequency	Association with INSTI therapy	Association with reduced susceptibility	Sum
148G	0	2	2	0	0	0	4
148N	0	2	2	0	0	0	4
114Y	0	0	2	1	1	0	4
146K	0	1	2	0	0	0	3
149A	0	1	2	0	0	0	3
49G	0	0	2	0	0	0	2
230G	0	0	2	0	0	0	2

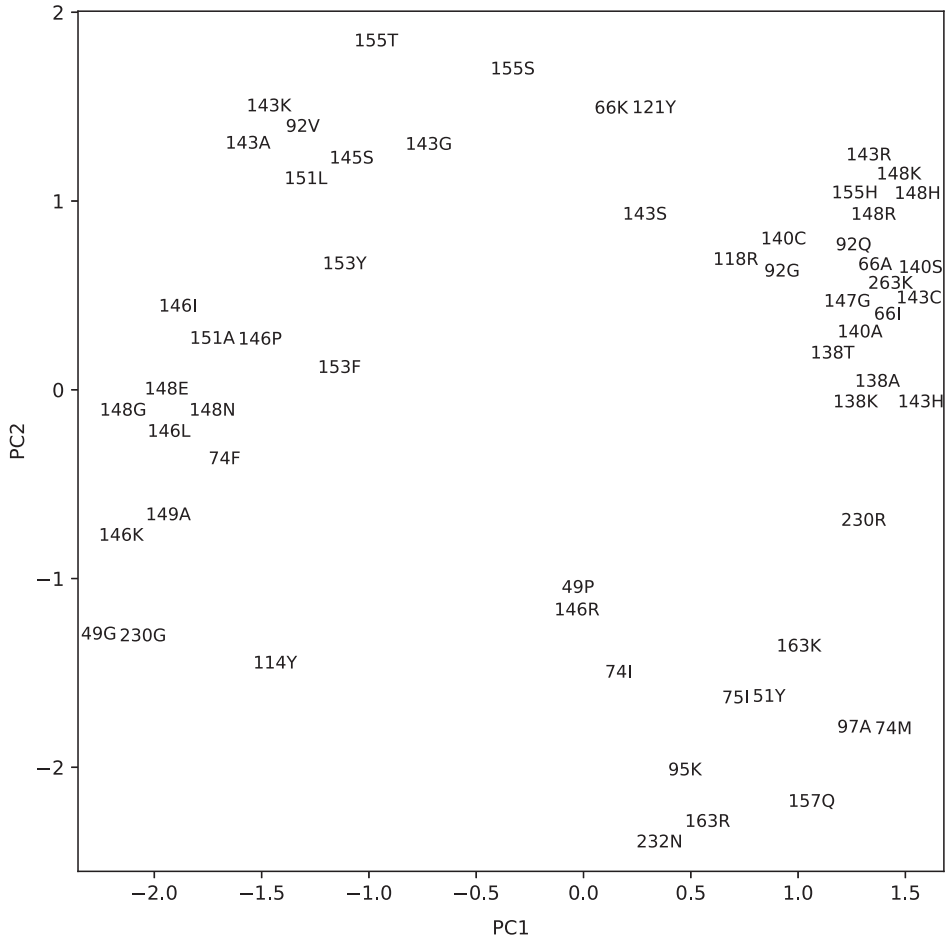

Figure 1. Principal components analysis of the 59 INSTI-resistance mutations on one or more expert lists. The original characteristics or variables were related to the number of expert lists containing the mutation, the positional importance of the mutation, the frequency of the mutation in INSTI-naive and INSTI-experienced persons and the statistical association of the mutation with reduced susceptibility to one or more INSTIs. The mutations at the upper right are non-polymorphic mutations significantly associated with INSTI treatment. Those at the lower right are polymorphic mutations significantly associated with therapy. Those on the left are rare non-polymorphic mutations, which nearly always occurred too infrequently to be associated with therapy.

selected by INSTI treatment; 8 polymorphic mutations (of which 7 were significantly selected by INSTI treatment) and 22 rare mutations of which nearly all were too uncommon to be significantly selected by INSTI treatment. The 29 non-polymorphic INSTI-selected mutations were subjected to more scrutiny and 24 emerged as strong candidates for inclusion on a list of INSTI surveillance DRMs: T66A/I/K, E92G/Q, G118R, F121Y, E138A/K/T, G140A/C/S, Y143C/H/R/S, S147G, Q148H/R/K, N155H, S230R and R263K.

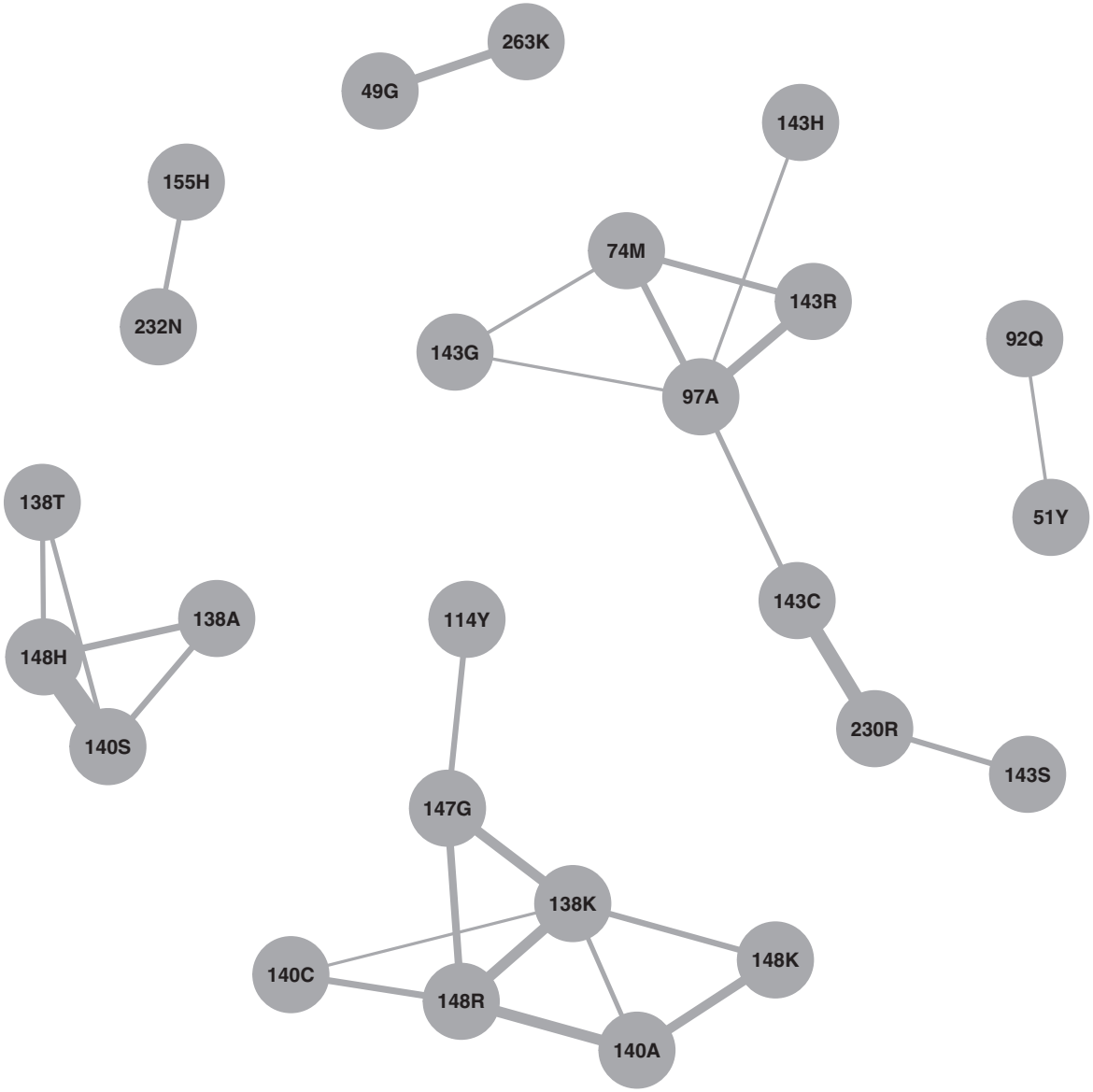

Figure 2. Correlation network analysis of the 25 INSTI-resistance mutations that most frequently co-occurred with one or more other INSTI-resistance mutations. INSTI-resistance mutations having a non-parametric Spearman correlation coefficient (ρ) of >0.075 and a *P* value $\leq$0.00001 are linked with an edge. Edge thickness is proportional to ρ, with the greatest thicknesses for the edge between G140S and Q148H (ρ=0.93), Y143C and S230R (ρ=0.65) and G140A and Q148R (ρ=0.38).

HIV-1 DRMs can be classified according to whether they are non-polymorphic in the absence of ART, whether they are selected *in vitro* during ARV passage experiments or *in vivo* in persons receiving ART, their effects on *in vitro* susceptibility and virological response to ART. Of the 24 suggested INSTI surveillance DRMs, 18 were on all four of the published expert DRM lists, 5 were on three lists and 1 was on two lists. Therefore, the presence of one of these mutations in a person initiating therapy is evidence for pretreatment resistance, regardless of whether it arose from previous therapy or from transmitted resistance.

We used a strict threshold for defining polymorphisms by excluding mutations with an overall prevalence $\geq$0.2%, a prevalence $\geq$0.5% in any subtype or CRF with >1000 persons, or a prevalence $\geq$1.0% in any subtype or CRF with >200 persons. This threshold is twice as high as the threshold we used to define non-polymorphic mutations in a paper published in 2009 to select DRMs for RT inhibitor- and PI-associated TDR surveillance.[2] It was possible to use a lower threshold for INSTI-resistance mutations without sacrificing sensitivity because the rarity of transmitted INSTI resistance may have made it unlikely that non-polymorphic INSTI-resistance DRMs were detected in INSTI-naive persons.[2] Therefore, the presence of an INSTI surveillance DRM in an untreated person strongly suggests that the transmitted virus had previously been exposed to INSTI treatment.

The WHO drug resistance surveillance programme initially classified TDR in recently infected persons using the list of protease

and RT SDRMs. In 2014, surveillance of TDR was deprioritized in favour of estimating the prevalence of drug resistance in all patients initiating first-line ART, referred to as pretreatment drug resistance (PDR). PDR includes those with TDR as well as those in whom ART had been interrupted or who had received ART to prevent mother-to-child transmission.[28,29]

As WHO-recommended surveys of PDR aim to predict susceptibility of virus to ARV drugs used in first- and second-line regimens, in order to inform global and national ART selection, the HIVDB system has been used to interpret the data, because it takes into consideration the contribution of polymorphisms to predicted drug resistance.[29] In contrast, the SDRM list is designed for epidemiological purposes—i.e. documentation of transmission of drug-resistant virus—and therefore is not intended for use when analysing data from surveys of PDR. Nonetheless, for many investigators the list will be useful for distinguishing polymorphic from non-polymorphic INSTI DRMs because polymorphic DRMs that occur in the absence of therapy are often subtype dependent and may influence regional and temporal estimates of TDR.

A set of DRMs designed for public health surveillance as opposed to individual patient management need not contain rare mutations, which are less likely to be recognized as DRMs by experts, public health officials and laboratory personnel. Rare DRMs are also less likely to be significantly associated with ARV drug selection pressure. The 22 rare mutations in this study included nine variants at the well-recognized INSTI-resistance positions 92, 143, 148 and 155 and eight variants at the highly conserved positions 145, 146, 149 and 153. Of these, E92V, V151A/L and N155S/T have been observed during *in vitro* passage with investigational INSTIs and are associated with variably reduced susceptibility to current INSTIs.[30,31] Y143A/K are raltegravir-selected variants associated with reduced raltegravir susceptibility.[32,33] P145S and Q146P are elvitegravir-selected mutations associated with reduced elvitegravir susceptibility.[34,35] S153Y/F have been selected *in vitro* by multiple INSTIs and cause minimal reductions in INSTI susceptibility.[30,36]

Approximately 80% of the sequences from treated persons were subtype B viruses and these were usually from persons treated with raltegravir and elvitegravir. This is not surprising because INSTIs have been primarily used in upper-income countries that have a lower prevalence of non-subtype B viruses,[37–41] with relatively small numbers of reports of INSTI resistance emanating from low- and middle-income countries.[42,43]

WHO's 2018 ARV guidelines recommend the use of dolutegravir as preferred first-line ART.[5,6] It is expected that the new regimen will be adopted widely in many regions of the word, including settings with non-B subtype and where programmatic challenges, such as ARV drug stockouts, might result in the emergence of INSTI resistance and eventually its transmission.[7] As INSTIs become more widely used in areas with a high prevalence of non-B variants this list may need to be updated. In addition, virological failure with emergent resistance has been uncommon in persons receiving the second-generation INSTIs dolutegravir and bictegravir. Several of the rare mutations might eventually be considered SDRM candidates should they prove to be common after more sequences become available, particularly sequences from persons with virological failure on a second-generation INSTI such as dolutegravir, bictegravir or cabotegravir.

Funding

This work was supported by the National Institute of Allergy and Infectious Diseases at the National Institutes of Health (grant no. R24 AI136618 to P.L.T., S.-Y.R. and R.W.S.) and the Agence Nationale de Recherches sur le SIDA et les hépatites virales (grants to D.D., M.G. and A.-G.M.). The funders had no decision-making role regarding the manuscript.

Transparency declarations

None to declare.

Supplementary data

Tables S1 and S2 are available as Supplementary data at *JAC* Online.

References

1 Shafer RW, Rhee SY, Pillay D *et al*. HIV-1 protease and reverse transcriptase mutations for drug resistance surveillance. *AIDS* 2007; **21**: 215–23.

2 Bennett DE, Camacho RJ, Otelea D *et al*. Drug resistance mutations for surveillance of transmitted HIV-1 drug-resistance: 2009 update. *PLoS One* 2009; **4**: e4724.

3 Gupta RK, Gregson J, Parkin N *et al*. HIV-1 drug resistance before initiation or re-initiation of first-line antiretroviral therapy in low-income and middle-income countries: a systematic review and meta-regression analysis. *Lancet Infect Dis* 2018; **18**: 346–55.

4 Rhee SY, Blanco JL, Jordan MR *et al*. Geographic and temporal trends in the molecular epidemiology and genetic mechanisms of transmitted HIV-1 drug resistance: an individual-patient- and sequence-level meta-analysis. *PLoS Med* 2015; **12**: e1001810.

5 WHO. Updated Recommendations on First-Line and Second-Line Antiretroviral Regimens and Post-Exposure and Recommendations on Early Infant Diagnosis of HIV. Interim Guidance. 2018. https://www.who.int/hiv/pub/guidelines/ARV2018update/en/.

6 Vitoria M, Hill A, Ford N *et al*. The transition to dolutegravir and other new antiretrovirals in low-income and middle-income countries: what are the issues? *AIDS* 2018; **32**: 1551–61.

7 WHO. Global Action Plan on HIV Drug Resistance 2017–2021: 2018 Progress Report. 2018. https://www.who.int/hiv/pub/drugresistance/gap-hivdr-progress2018/en/.

8 Chen I, Zhang Y, Cummings V *et al*. Analysis of HIV integrase resistance in black men who have sex with men in the United States. *AIDS Res Hum Retroviruses* 2017; **33**: 745–8.

9 Ji H, Patterson A, Taylor T *et al*. Prevalence of primary drug resistance against HIV-1 integrase inhibitors in Canada. *J Acquir Immune Defic Syndr* 2018; **78**: e1–3.

10 Tostevin A, White E, Dunn D *et al*. Recent trends and patterns in HIV-1 transmitted drug resistance in the United Kingdom. *HIV Med* 2017; **18**: 204–13.

11 Stekler JD, McKernan J, Milne R *et al*. Lack of resistance to integrase inhibitors among antiretroviral-naive subjects with primary HIV-1 infection, 2007-2013. *Antivir Ther* 2015; **20**: 77–80.

12 Banaez Ocfemia MC, Saduvala N, Oster AM *et al*. Transmitted HIV-1 drug resistance among men who have sex with men - 11 U.S. jurisdictions, 2008-2011. *Twenty-First Conference on Retroviruses and Opportunistic Infections, Boston, MA, USA, 2014*. Abstract 579.

13 Acosta RK, Willkom M, Martin R *et al*. Resistance analysis of bictegravir/emtricitabine/tenofovir alafenamide in HIV-1 treatment-naive patients through 48 weeks. *Antimicrob Agents Chemother* 2019; **63**: e02533-18.

14 Jeong W, Jung IY, Choi H *et al*. Integrase strand transfer inhibitor resistance mutations in antiretroviral therapy-naive and treatment-experienced HIV patients in South Korea. *AIDS Res Hum Retroviruses* 2019; **35**: 213–6.

15 Varghese V, Pinsky BA, Smith DS *et al*. Q148N, a novel integrase inhibitor resistance mutation associated with low-level reduction in elvitegravir susceptibility. *AIDS Res Hum Retroviruses* 2016; **32**: 702–4.

16 Young B, Fransen S, Greenberg KS *et al*. Transmission of integrase strand-transfer inhibitor multidrug-resistant HIV-1: case report and response to raltegravir-containing antiretroviral therapy. *Antivir Ther* 2011; **16**: 253–6.

17 McGee KS, Okeke NL, Hurt CB *et al*. Canary in the coal mine? Transmitted mutations conferring resistance to all integrase strand transfer inhibitors in a treatment-naive patient. *Open Forum Infect Dis* 2018; **5**: ofy294.

18 Volpe JM, Ward DJ, Napolitano L *et al*. Five antiretroviral drug class-resistant HIV-1 in a treatment-naive patient successfully suppressed with optimized antiretroviral drug selection. *J Int Assoc Provid AIDS Care* 2015; **14**: 398–401.

19 Boyd SD, Maldarelli F, Sereti I *et al*. Transmitted raltegravir resistance in an HIV-1 CRF_AG-infected patient. *Antivir Ther* 2011; **16**: 257–61.

20 Aickin M, Gensler H. Adjusting for multiple testing when reporting research results: the Bonferroni vs Holm methods. *Am J Public Health* 1996; **86**: 726–8.

21 Rhee SY, Taylor J, Fessel WJ *et al*. HIV-1 protease mutations and protease inhibitor cross-resistance. *Antimicrob Agents Chemother* 2010; **54**: 4253–61.

22 Melikian GL, Rhee SY, Taylor J *et al*. Standardized comparison of the relative impacts of HIV-1 reverse transcriptase (RT) mutations on nucleoside RT inhibitor susceptibility. *Antimicrob Agents Chemother* 2012; **56**: 2305–13.

23 Melikian GL, Rhee SY, Varghese V *et al*. Non-nucleoside reverse transcriptase inhibitor (NNRTI) cross-resistance: implications for preclinical evaluation of novel NNRTIs and clinical genotypic resistance testing. *J Antimicrob Chemother* 2014; **69**: 12–20.

24 Pedregeosa F, Varoquaux G, Gramfort A *et al*. Scikit-learn: machine learning in python. *J Machine Learning Res* 2011; **12**: 2825–30.

25 igraph - The Network Analysis Package. 2018. https://igraph.org/.

26 Mouscadet JF, Delelis O, Marcelin AG *et al*. Resistance to HIV-1 integrase inhibitors: a structural perspective. *Drug Resist Updat* 2010; **13**: 139–50.

27 Hare S, Smith SJ, Metifiot M *et al*. Structural and functional analyses of the second-generation integrase strand transfer inhibitor dolutegravir (S/GSK1349572). *Mol Pharmacol* 2011; **80**: 565–72.

28 WHO. Guidelines on the Public Health Response to Pretreatment HIV Drug Resistance. 2017. http://www.who.int/hiv/pub/guidelines/hivdr-guidelines-2017/en/.

29 WHO. Surveillance of HIV Drug Resistance in Adults Initiating Antiretroviral Therapy (Pre-Treament HIV Drug Resistance). 2014. https://apps.who.int/iris/bitstream/handle/10665/112802/9789241507196_engpdf?sequence=1.

30 Kobayashi M, Yoshinaga T, Seki T *et al*. In vitro antiretroviral properties of S/GSK1349572, a next-generation HIV integrase inhibitor. *Antimicrob Agents Chemother* 2011; **55**: 813–21.

31 Jones GS, Yu F, Zeynalzadegan A *et al*. Preclinical evaluation of GS-9160, a novel inhibitor of human immunodeficiency virus type 1 integrase. *Antimicrob Agents Chemother* 2009; **53**: 1194–203.

32 Huang W, Frantzell A, Fransen S *et al*. Multiple genetic pathways involving amino acid position 143 of HIV-1 integrase are preferentially associated with specific secondary amino acid substitutions and confer resistance to raltegravir and cross-resistance to elvitegravir. *Antimicrob Agents Chemother* 2013; **57**: 4105–13.

33 Canducci F, Ceresola ER, Boeri E *et al*. Cross-resistance profile of the novel integrase inhibitor dolutegravir (S/GSK1349572) using clonal viral variants selected in patients failing raltegravir. *J Infect Dis* 2011; **204**: 1811–5.

34 Shimura K, Kodama E, Sakagami Y *et al*. Broad antiretroviral activity and resistance profile of the novel human immunodeficiency virus integrase inhibitor elvitegravir (JTK-303/GS-9137). *J Virol* 2008; **82**: 764–74.

35 Smith SJ, Zhao XZ, Burke TR Jr *et al*. Efficacies of cabotegravir and bictegravir against drug-resistant HIV-1 integrase mutants. *Retrovirology* 2018; **15**: 37.

36 Margot NA, Hluhanich RM, Jones GS *et al*. In vitro resistance selections using elvitegravir, raltegravir, and two metabolites of elvitegravir M1 and M4. *Antiviral Res* 2012; **93**: 288–96.

37 Nguyen T, Fofana DB, Le MP *et al*. Prevalence and clinical impact of minority resistant variants in patients failing an integrase inhibitor-based regimen by ultra-deep sequencing. *J Antimicrob Chemother* 2018; **73**: 2485–92.

38 de Souza Cavalcanti J, Ferreira JL, Guimaraes PM *et al*. High frequency of dolutegravir resistance in patients failing a raltegravir-containing salvage regimen. *J Antimicrob Chemother* 2015; **70**: 926–9.

39 Molina JM, Lamarca A, Andrade-Villanueva J *et al*. Efficacy and safety of once daily elvitegravir versus twice daily raltegravir in treatment-experienced patients with HIV-1 receiving a ritonavir-boosted protease inhibitor: randomised, double-blind, phase 3, non-inferiority study. *Lancet Infect Dis* 2012; **12**: 27–35.

40 Hurt CB, Sebastian J, Hicks CB *et al*. Resistance to HIV integrase strand transfer inhibitors among clinical specimens in the United States, 2009-2012. *Clin Infect Dis* 2014; **58**: 423–31.

41 Marcelin AG, Grude M, Charpentier C *et al*. Resistance to integrase inhibitors: a national study in HIV-1-infected treatment-naive and -experienced patients. *J Antimicrob Chemother* 2019; **74**: 1368–75.

42 Pujari SN, Gaikwad S, Joshi K *et al*. Integrase resistance-associated mutations on raltegravir failure in Western India: a preliminary analysis. *J Acquir Immune Defic Syndr* 2018; **77**: e42–5.

43 Ndashimye E, Avino M, Kyeyune F *et al*. Absence of HIV-1 drug resistance mutations supports the use of dolutegravir in Uganda. *AIDS Res Hum Retroviruses* 2018; **34**: 404–14.