Enhanced surveillance of human immunodeficiency virus type 1 drug resistance in recently infected MSM in the UK

Emma CUNNINGHAM[1], Yuen Ting CHAN[1], Adamma AGHAIZU[2], David F. BIBBY[1,3], Gary MURPHY[1], Jennifer TOSSWILL[1], Ross J. HARRIS[4], Richard MYERS[1], Nigel FIELD[2], Valerie DELPECH[2,3], Patricia A. CANE[1,3], O. Noel GILL[2,3] and Jean L. MBISA[1,3]*.

[1]Virus Reference Department, National Infection Service, Public Health England, UK.

[2]HIV and STI Department, National Infection Service, Public Health England, UK.

[3]National Institute for Health Research Health Protection Research Unit (NIHR HPRU) in Blood Borne and Sexually Transmitted Infections, University College London, UK.

[4]Statistics, Modelling and Economics Department, National Infection Service, Public Health England, UK.

*Corresponding author

Keywords. HIV-1 transmitted drug resistance, HIV-1 subtypes, low frequency variants, antiretroviral therapy, next generation sequencing

Running title. HIV-1 low frequency TDR variants

Corresponding author. Jean L. Mbisa, Antiviral Unit, Virus Reference Department, Public Health England, 61 Colindale Avenue, London NW9 5EQ, UK (tamyo.mbisa@phe.gov.uk).

Abstract: 250 words

Manuscript: 3,525 words

SYNOPSIS

Objectives: To determine the prevalence of low inferred frequency HIV-1 transmitted drug resistance (TDR) in men who have sex with men (MSM) in the UK and their predicted effect on first-line therapy.

Methods: The HIV-1 *pol* gene was amplified from 442 newly diagnosed MSM identified as likely recently infected by serological avidity testing in 2011 to 2013. The PCR products were sequenced by next generation sequencing with a mutation frequency threshold of >2% and TDR mutations defined according to the 2009 WHO surveillance drug resistance mutations (SDRM) list.

Results: The majority (75.6%) were infected with subtype B and 6.6% with rare complex or unique recombinant forms. At mutation frequency threshold of >20%, 7.2% [5.0 – 10.1%] of the sequences had TDR and this doubled to 15.8% [12.6 – 19.6%] at >2% mutation frequency (p<0.0001). The majority (26/42; 62%) of low frequency variants were against protease inhibitors (PIs). The most common mutations detected at >20 and 2-20% mutation frequency differed for each drug class, these being: L90M (n=7) and M46IL (n=10) for PIs, T215rev (n=9) and D67GN (n=4) for nucleos(t)ide reverse transcriptase inhibitors (NRTIs), K103N (n=5) and K101E/G190E (n=2 each) for non-nucleoside RTIs (NNRTIs), respectively. Combined TDR was more frequent in subtype B than non-B (OR=0.38; 95%CI=0.17-0.88; p=0.024) and had minimal predicted effect on recommended first-line therapies.

Conclusions: The data suggest differences in the types of low frequency compared to majority TDR variants that requires a better understanding of the origins and clinical significance of low frequency variants. This will better inform diagnostic and treatment strategies.

INTRODUCTION

Drug resistance mutations identified in newly diagnosed, treatment-naïve patients are presumed to be a result of the transmission of drug resistant variants. Transmitted drug resistance (TDR) can limit the treatment options available to newly diagnosed HIV-1 positive patients and is associated with an increased risk of virologic failure.[1, 2] The prevalence of TDR in the UK has been estimated by analyzing the earliest available sequences from all treatment-naïve patients or from seroconverters submitted to the UK HIV drug resistance database matched to the national new HIV diagnoses database.[3-7] The sequences are generated using Sanger capillary sequencing which has a limit of detection of approximately 20% of variant frequency present in the viral population. Using these data, the prevalence of HIV-1 TDR in the UK appeared to have peaked at ~13% in 2002, before declining to a nadir of 6.6% in 2013.[8, 9] However, the level of TDR has been consistently higher in men who have sex with men (MSM) and has been shown to be more likely in MSM infected with subtype B compared to other exposure groups.[8, 10] The prevalence of TDR in the MSM group in the UK was estimated at 15.2% at its peak in 2002 before declining to its lowest level in 2013 at 7.5%.[11] The MSM risk group also bears the highest burden of the infection accounting for 54% of new diagnoses and 41% of people living with HIV in the UK in 2013.[11]

Following transmission, and in the absence of drug pressure, drug resistance mutations may revert to wild-type or an intermediate form due to the reduced replicative capacity of viruses with particular mutations.[12] Some drug resistance mutations may also persist in latently infected cells or become compartmentalized, reappearing later in the presence of antiretroviral (ARV) drugs.[13, 14] A number of studies have been conducted to determine the persistence of TDR.[15-18] However, the persistence of TDR mutations has been shown to vary between particular types of mutations.[19] In addition, some mutations may fall below the limit of detection of Sanger capillary sequencing and this may have an impact on the response to ARV therapy as these low frequency variants could reemerge upon initiation of therapy resulting in treatment failure.[20-22]

The development of next generation sequencing (NGS) technologies has allowed the detection of low-level drug-resistant mutations present in a viral population at frequencies as low as 0.3%.[23] Recent studies employing sensitive genotyping assays have shown that the prevalence of TDR can be as high as 30%.[24, 25] However, several studies examining the clinical significance of low frequency variants on treatment outcome have shown that not all low frequency drug-resistant variants detected in treatment-naïve individuals contribute to virologic failure.[26]

This study applied NGS to samples from patients deemed to have been infected within 6 months of sampling by the application of a recent infection testing algorithm (RITA).[27] Not only does this increase the potential to detect TDR mutations before they revert or are archived, as previously done using seroconvertor cohorts,[4-6] but importantly increases the likelihood of detecting low frequency variants. Additionally, since the national reference laboratory in England applies this RITA to ~50% of all newly diagnosed HIV-1 infections, as a population level surveillance of HIV-1 incidence, these samples should also allow a more timely and direct measure of the prevalence of TDR and surveillance of circulating or emerging genotypes compared to samples from all newly HIV-1 diagnosed persons.

The work was conducted as part of the National Institute for Health Research Health Protection Research Unit (NIHR HPRU) at University College London (UCL), a partnership with Public Health England (PHE) in Blood Borne and Sexually Transmitted Infections in collaboration with the London School of Hygiene and Tropical Medicine.

METHODS

Study population. The first plasma specimen from 442 newly HIV-1 diagnosed MSM sampled between July 2011 and December 2013 were analysed. This represents approximately 42% of all samples identified to be likely recent infections during this period and approximately 7% of all new infections among MSM as estimated by parsimonious back-calculation.[8, 28, 29] All patients were treatment-naïve and identified as likely to be recently infected (within 6 months of infection) using RITA, which includes CD4+ count (>200 cells/mm^3), viral load (>1,000 copies/mL) and the AxSYM HIV 1,2gO assay with an avidity index threshold <80% or, for samples taken between September and December 2013, a Limiting-antigen (LAg) avidity assay with an OD index <1.5. The assays differentiate likely recent from long standing infection by the strength of HIV-specific antibody-antigen binding.[30, 31] The assays have a misclassification rate of <5% and samples close to the avidity or OD index cut-off values are more likely to be misclassified.[32] Linked demographic and clinical information was extracted from the HIV & AIDS Reporting System (HARS) held at Public Health England (PHE).

HIV-1 RNA extraction, PCR amplification and sequencing. Two hundred μl of each sample were used to extract viral RNA using QIAamp UltraSens Virus Kit (QIAGEN) as per the kit instructions and the HIV-1 RNA eluted into 60μL of AVE buffer. A 1.3kb region of the HIV *pol* gene (whole of protease and N-terminal half of reverse transcriptase; aa1-320) was amplified as previously described using 10μL of the RNA extract in each PCR reaction.[33] PCR products were purified using QIAQuick kit (QIAGEN) and quantified using both Qubit® dsDNA Broad Range and High Sensitivity Assay Kits and the Qubit® 2.0 Fluorometer (Life Technologies). One ng/μL of the amplified DNA product was used for DNA library preparation with the Nextera XT DNA sample prep kit (Illumina) as per the kit protocol. NGS was performed using the MiSeq reagent kit version 2 (Illumina).

Bioinformatic analysis. A subset of the MiSeq paired-end reads from each FASTQ file was compared to a local database of HIV reference sequences using BLAST to identify an optimum reference sequence for mapping using BWA-MEM version 0.7.5. Utilising SAMTools the resulting files were then converted into BAM format in preparation for in-house developed software, QuasiBAM, which generates consensus sequences of the protease and reverse transcriptase regions and produces detailed information on the frequencies of minority variants present within each sample. These procedures were automated using a computational pipeline developed in-house using Python and C++.

Analysis of HIV-1 subtypes, TDR and predicted drug susceptibility. HIV-1 subtypes were determined using four publically available HIV-1 subtyping tools, these being REGA HIV-1 Subtyping Tool version 3.0, SCUEAL algorithm, jumping profile Hidden Markov Model (jpHMM-HIV) and Context-based Modelling for Expeditious Typing (COMET HIV-1).[34-38] When the subtyping tools were discordant, the subtype or circulating recombinant form (CRF) was called by manual inspection or designated as a URF if it could not be assigned to a particular subtype or CRF. TDR mutations were defined using the WHO 2009 list of surveillance drug resistance mutations.[39] The drug susceptibility of each sample was determined using the Stanford HIV drug resistance database genotypic interpretation algorithm version 7.0.[40]

Statistical analyses. Upper and lower 95% confidence intervals were determined using the exact binomial calculation and estimates of the additional detections provided by the more sensitive 2-20% test via McNemar's chi-squared test for matched data. The association of subtype and other demographic factors (age, geographic region, ethnicity, country of birth and probable country of infection) with TDR rates was determined using univariate analysis involving odds ratio (OR) and Chi-squared tests. Multivariable models were constructed to estimate the independent effects of covariates on TDR rates via logistic regression. Statistical analyses were carried out using Stata 13.1 software (StataCorp. 2013. *Stata Statistical Software: Release 13*. College Station, TX: StataCorp LP) or Microsoft Excel, with a p value <0.05 regarded as significant.

RESULTS

Characterisation of the study population. From July 2011 to December 2013, 58.1% (4,119 out of 7,093) of newly HIV-1 diagnosed MSM in England, Wales and Northern Ireland were tested using RITA by the national reference laboratory in England (Table 1). Of these, 26.7% (1,101) were identified as likely recent infections, of which 442 had sufficient residual volume and were successfully amplified by PCR and subjected to NGS. This represents 43.1% of all RITA positive samples during that period. Samples included 68 from 2011, 145 from 2012 and 229 from 2013. The median age of the study population was 32 years [26-40; IQR]. More than half of the newly diagnosed (51.4%) were from the London region; however, the proportions that were RITA tested and found to be recently infected and then sequenced was similar across the geographic regions ranging from 41.6 to 46.5%. In contrast, a higher proportion (51.2%) of samples from the recently infected aged over 50 years old were sequenced compared to 31.2% from those aged between 15 to 24 years old (Table 1).

Distribution of circulating HIV-1 subtypes among recently infected MSM in the UK. The 442 sequences were subtyped using four web-based subtyping tools as described in the methods section and the assigned subtypes shown in Table 2. For 402 sequences (91%), the results from at least three of the subtyping tools were concordant and the results for 425 sequences (96.2%) were concordant by at least two subtyping tools. The remaining sequences, where the subtyping tools disagreed or returned no particular assignment, were mostly complex unique recombinant forms (URFs). As expected, subtype B was the predominant subtype making up 75.6%. Of note, the most common non-B subtype group consisted of rare CRFs and URFs at 6.6%. The majority (27/29) of the rare CRFs/URFs were composed of a subtype B and another subtype(s) or CRFs (Supplementary Table 1). The subtypes of the remaining samples in order of abundance were subtype F1 (4.5%; n=20), CRF02_AG (4.3%; n=19), A1 (3.4%; n=15), C (2.3%; n=10), CRF01_AE (1.4%; n=6), CRF06_cpx (1.4%; n=6) G (0.5%; n=2) and CRF07_BC (0.2%; n=1).

The proportion of TDR at 20% variant frequency threshold. We determined the proportion of TDR mutations among recently infected MSM at a variant frequency threshold of >20% which is equivalent to that used for Sanger capillary sequencing. At this threshold, TDR mutations were detected in 32 out of the 442 sequences, 7.2% [5.0 - 10.1%; 95% CI]. By drug class, the overall TDR proportion was as follows: protease inhibitors (PIs: 2.7%, n=12), nucleoside reverse transcriptase inhibitors (NRTIs: 3.2%, n=14) and non-nucleoside reverse transcriptase inhibitors (NNRTIs: 1.6%, n=7). The TDR mutations detected in the 32 sequences are shown in Table 3. One sample had dual class resistance with three mutations, two NRTI mutations (M41L and M184V) and one NNRTI mutation (V106A). The remainder had single class resistance with one TDR mutation each except for one sample that had two NNRTI mutations (K103N and Y188L). The most common TDR mutations detected in each class were: L90M (n=7) for PIs, T215rev (n=9) for NRTIs and finally K103N (n=5) for NNRTIs (Table 3).

The detection of low frequency TDR mutations between 2 and 20% mutation frequency thresholds. To establish the threshold for detection of low-level variants in our assay we determined the reproducibility of detection of variants in clinical samples. We compared the frequency of each codon in the *pol* gene amplicon of a particular sample with its corresponding codon in a replicate which had been processed independently from nucleic acid extraction to sequencing. This showed that some codons detected at low frequencies did not have the same percentage occurrence in the two independent runs and this was seen more often with variants detected at < 2% (Figure 1a). Furthermore, the analysis showed that if we included frequencies in the second replicate at plus or minus 50% of the value of the first replicate, the threshold of low frequency variant detection approaches 100% only at cut-off values >2% (Figure

1b). The median depth of coverage for each replicate run was similar and high at 15,782 [11,426-19,502; IQR] and 19,393 [15,375-22,454] for Run 1 and 2, respectively (Figure 1c). Thus, the threshold for low frequency variant detection in our assay was set at 2%.

At the 2% variant frequency threshold, an additional 38 samples were identified to have TDR mutations, representing a significant increase in the overall TDR proportion at 15.8% [13.4 – 20.6%] (McNemar's chi-squared $p<0.0001$). The depth of coverage at sites where low frequency variants were identified was very high and ranged from 7,139 to 47,752 reads (Supplementary Table 2). By drug class, the overall TDR proportions when low frequency variants were included increased by 3.2-fold for PIs at 8.6%, by 1.7-fold for NRTIs at 5.4% and by 1.9-fold for NNRTIs at 2.9%. Low frequency variants were detected in 4 samples that had TDR mutations at a frequency >20%, these being: PI V82A + NNRTI K103N (9.7%), NNRTI K103N and Y188L + PI M46L (2.1%), NRTI T215S + PI M46L (11%) and PI L90M + PI D30N (5.9%). This changed the classification of the first 3 samples from single- to dual-class resistance. The majority of the identified low frequency variants were PI mutations that were detected in 26 out of 42 (62%) samples (Table 3 and Supplementary Table 2).

Factors associated with transmitted drug resistance. Univariable analyses revealed that TDR was significantly associated with subtype B than non-B infections (odds ratio for TDR in non-B subtype infections of 0.41; 95%CI=0.19-0.85; $p=0.017$). This was mostly due to a reduced likelihood of TDR in non-B subtype infections at >20% variant frequency (OR=0.30; 95%CI=0.09-1.01; $p=0.051$ compared to OR=0.59; 95%CI=0.26-1.37; $p=0.223$ at 2-20% variant frequency). Multivariable analyses confirmed subtype as an independent factor associated with TDR (OR for TDR in non-B subtype infections of 0.38; 95%CI=0.17-0.88; $p=0.024$). The only other factor significantly associated with TDR in multivariable analyses was infections that were probably acquired outside the UK (OR=2.64; 95%CI=1.03-6.78; $p=0.044$). However, this effect was slightly attenuated in univariable analyses (OR=2.12; 95%CI=0.93-4.83; $p=0.073$) and was not strongly linked with a particular variant frequency threshold. There was no significant association with age, geographic region, ethnicity and the country of birth in both univariable and multivariable analyses (Table 4).

Predicted susceptibility of samples harbouring low frequency variants to ARVs recommended for first-line therapy in the UK. We investigated the predicted effect of the low frequency TDR mutations detected on susceptibility to the ARVs that are currently recommended for first-line treatment or as alternatives in the UK.[41] The susceptibility of the samples was analysed using the Stanford HIV drug resistance database genotypic interpretation algorithm which assigns five different levels of drug susceptibility, these being: susceptible, potential resistance, low-level resistance, intermediate resistance and high-level resistance. Most of the TDR mutations resulted in low-level resistance but intermediate to high-level resistance was often associated with NNRTIs (Figure 2).

At the >20% variant frequency threshold the drugs most affected were as follows: the NNRTIs nevirapine and efavirenz with 1.9%, the NRTI zidovudine with 2.5% and the PIs atazanavir and lopinavir with 1.7% of the samples showing low- to high-level resistance (Figure 2). The only drugs not associated with any resistance were the PI darunavir and the NRTI tenofovir. When the low-level frequency mutations were included we observed an increase in the proportion of samples with reduced susceptibility to all drugs including resistance to darunavir and tenofovir (Figure 2). Resistance to the NNRTIs nevirapine and efavirenz increased 2-fold to 4% and 3.8%, respectively, whereas that to the PIs lopinavir and atazanavir increased 2.3- and 2.8-fold to 4.6% and 3.8%, respectively. Of note, a significant proportion of the samples showed resistance to PIs and NRTIs that are no longer used in the UK with 2.7% of the samples showing resistance to the older NRTIs and PIs at >20% variant frequency threshold increasing up to 8% at >2% variant frequency threshold.

DISCUSSION

The data show that the proportion of TDR among recently infected MSM doubles when low frequency variants are taken into account from 7.2% to 15.8% at >20% and >2% variant frequency thresholds, respectively. This is in agreement with other studies that have used highly sensitive genotyping methods where the proportion of TDR among treatment-naïve individuals has ranged between 17-30% worldwide.[20, 23, 26, 42-44] A majority (62%) of the low frequency variants were associated with resistance against PIs despite PI-associated drug resistance mutations rarely being observed among treatment-experienced patients failing therapy in the UK at 3.5% in 2013 compared to

16.5% and 23.2% for NRTI and NNRTI mutations (UK HIVDRDb). It is expected that following transmission, drug resistance-associated variants would steadily decline and disappear with time in the absence of drug selective pressure. Thus, these data suggest either a transmission and sustained persistence of low frequency variants or a stochastic *de novo* generation of these mutations in the infected patients. For the latter, the mutations would be expected to be randomly distributed; however, the data show a predominance of particular types of low frequency variants in each drug class i.e. M46IL for PIs, D67GN for NRTIs and G190E for NNRTIs which are different from the most common drug resistance-associated mutations observed at >20% variant frequency threshold: L90M, T215rev and K103N, respectively. Alternatively, this could reflect the impact on replication fitness of individual mutations with those significantly detrimental to viral replication most likely to decrease rapidly in frequency in the absence of drug selection or it could be dependent on differences in the frequency of a given codon change resulting in an amino acid substitution at a particular site.

Several studies have investigated the transmission of low frequency drug resistance variants.[45-47] One study used ultradeep sequencing on samples from 32 recently infected individuals concluded that the bulk of low frequency drug resistance variants were either due to sequencing or *de novo* viral replicative errors.[45] In contrast, a study using allele-specific PCR on samples from recently and chronically infected patients showed direct evidence that low frequency variants can be transmitted.[46] It is possible that the contradictory outcomes could be a result of different experimental methodologies. Thus, the origins and source of these low frequency variants need further investigation using large well-characterized cohorts, as it has been hypothesized that transmitted variants are more likely to persist and establish a latent infection than *de novo* generated variants.

Similar to previous studies of TDR prevalence in the UK the most common TDR mutations we identified at >20% variant frequency threshold confer resistance to drugs no longer used for treatment of HIV-1 infection i.e. PI L90M and NRTI T215rev.[9, 48] These mutations are likely to have been initially transmitted from ARV-experienced individuals further back in the transmission chain and despite absence of drug pressure have persisted in the population.[48, 49] Interestingly, TDR mutations especially those present at a variant frequency greater than 20% were observed to significantly occur more frequently in subtype B than non-B subtypes in keeping with the notion that the resistance is mostly historical due to ART having been in use for longer in subtype B than non-B infections.

To date, studies describing the impact of low frequency variants detected at baseline on treatment outcome have linked NNRTI-resistant mutations with a two-fold increase in the risk of virologic failure.[21] One factor determined to be associated with this increased risk is the mutational load which is a product of the frequency of the variant and viral load.[23, 50] Viral load data were incomplete for this study but are likely to be relatively high for acute infections. However, we observed that the frequency of NNRTI low frequency variants was often higher (between 3.1% and 15.4%, and thus likely to represent a higher mutational load) compared to PI and NRTI variants that were mostly between 2% and 3% (Supplementary Table 2). Further large case-control or cohort studies are required to determine the impact of specific low frequency variants on treatment outcome.

It has been reported that the proportion of non-B and non-C subtypes among the treatment-naïve MSM population in the UK has increased significantly from 5.7% to 13.6% between 2002 and 2010.[51] In this study this proportion was 23.2%, a further increase on the 2010 figures and in keeping with the upward trend in the proportion of non-B and non-C subtypes among MSM. We also show that rare CRFs/URFs were the most frequent non-B subtypes observed comprising ~7% of the samples. This proportion is likely higher than reported here as only 15% of the genome was sequenced and recombination could have occurred in the non-sequenced portions of the genome. The increase in inter-subtype recombinants could be due to increased migration from Africa and Eastern Europe, where they are more common, but could also reflect the emergence of novel recombinant forms due to an increased probability of inter-subtype co-infections among MSM. The latter is supported by the fact that the majority of the rare recombinants were composed of a subtype B and a non-B subtype or CRFs.

A limitation of this study is the threshold for detection of low frequency variants. As described earlier, the low frequency variants detected in a sample could have several sources including real transmitted variants, variants introduced during *de novo* viral replication *in vivo* or laboratory artefacts introduced during RT-PCR amplification and/or sampling bias. Sampling bias occurs at several steps during the process: at RNA extraction, at RT-PCR and at DNA library preparation, all of which result in bottleneck effects. Laboratory artefacts and *de novo* viral replication errors have been shown to result in as high as 2% variant frequency using clinical samples from pre-ART era.[52] By themselves RT-PCR and sequencing errors on Illumina machines have been shown to account for less than 0.5 to 1% of observed errors.[53, 54] Our experiments using repeat independent amplification and sequencing of the same clinical samples showed results that are consistent with these previous observations with most discrepancies in variant calls observed at frequencies below 2%. Therefore, the 2% threshold chosen for our assay probably results in the ruling out of most if not all false positive variants i.e. high specificity, but it is likely to result in under calling of true variants i.e. less sensitivity.

In summary, this study shows that the use of NGS can provide detailed and enhanced genomic information on TDR and subtype distribution in newly diagnosed HIV-1 patients as part of a national surveillance program. These data gathered in real time together with demographic data and in tandem to determination of recent infection are a useful extension to public health surveillance of HIV to better inform individual clinical prescribing practice, population-based prevention strategies and would also be useful for the validation of current diagnostic tools.

Acknowledgements. We thank Kieren Lythgow for additional bioinformatics support. We also wish to thank Samuel Moses, Gkikas Magiorkinis and the NIHR HPRU in Blood Borne and Sexually Transmitted Infections Steering Committee: Caroline Sabin (Director), Anthony Nardone (PHE Lead), Catherine Mercer, Gwenda Hughes, Jackie Cassell, Greta Rait, Samreen Ijaz, Tim Rhodes, Kholoud Porter, William Rosenberg.

Funding. This research was funded by Public Health England (PHE) and by the National Institute for Health Research and undertaken by the National Institute for Health Research Health Protection Research Unit (NIHR HPRU) in Blood Borne and Sexually Transmitted Infections at UCL in partnership with Public Health England (PHE) and in collaboration with the London School of Hygiene and Tropical Medicine. The views expressed in this publication are those of the author(s) and not necessarily those of the NHS, the National Institute for Health Research, the Department of Health or Public Health England.

Transparency Declarations. All authors declare no potential conflicts of interest.

Author contributions. J.L.M., O.N.G., P.A.C. and V.D. conceived the idea for the study and analysis. E.C. and YTC performed the sequencing experiments and initial data analysis. A.A., G.M. and J.T. collected the metadata and co-ordinating RITA testing. D.F.B. and R.M. performed the bioinformatics analyses. R.J.H performed statistical analyses. A.A., D.F.B., N.F., G.M. and J.T. provided valuable input regarding the analyses. E.C. and J.L.M. drafted the manuscript. All authors provided critical reading that shaped the manuscript.

Reference List

1. Wittkop L, Gunthard HF, de WF, et al. Effect of transmitted drug resistance on virological and immunological response to initial combination antiretroviral therapy for HIV (EuroCoord-CHAIN joint project): a European multicohort study. *Lancet Infect Dis* 2011; **11**: 363-71.

2. Little SJ, Holte S, Routy JP, et al. Antiretroviral-drug resistance among patients recently infected with HIV. *N Engl J Med* 2002; **347**: 385-94.

3. Pillay D, Cane PA, Shirley J, et al. Detection of drug resistance associated mutations in HIV primary infection within the UK. *AIDS* 2000; **14**: 906-8.

4. UK Collaborative Group on Monitoring the Transmission of HIV Drug Resistance. Analysis of prevalence of HIV-1 drug resistance in primary infections in the United Kingdom. *BMJ* 2001; **322**: 1087-8.

5. UK Collaborative Group on HIV Drug Resistance, UK Collaborative HIV Cohort Study, UK Register of HIV Seroconverters. Evidence of a decline in transmitted HIV-1 drug resistance in the United Kingdom. *AIDS* 2007; **21**: 1035-9.

6. Masquelier B, Bhaskaran K, Pillay D, et al. Prevalence of transmitted HIV-1 drug resistance and the role of resistance algorithms: data from seroconverters in the CASCADE collaboration from 1987 to 2003. *J Acquir Immune Defic Syndr* 2005; **40**: 505-11.

7. Cane P, Chrystie I, Dunn D, et al. Time trends in primary resistance to HIV drugs in the United Kingdom: multicentre observational study. *BMJ* 2005; **331**: 1368.

8. Aghaizu A, Brown AE, Nardone A, et al. HIV in the United Kingdom 2013 Report: data to end 2012. In: Public Health England, London., 2013.

9. Dolling D, Sabin C, Delpech V, et al. Time trends in drug resistant HIV-1 infections in the United Kingdom up to 2009: multicentre observational study. *BMJ* 2012; **345**: e5253.

10. Vercauteren J, Wensing AM, van de Vijver DA, et al. Transmission of drug-resistant HIV-1 is stabilizing in Europe. *J Infect Dis* 2009; **200**: 1503-8.

11. Yin Z, Brown AE, Hughes G, et al. HIV in the United Kingdom 2014 Report: data to end 2013. In: Public Health England, 2014.

12. Pingen M, Nijhuis M, de Bruijn JA, et al. Evolutionary pathways of transmitted drug-resistant HIV-1. *J Antimicrob Chemother* 2011; **66**: 1467-80.

13. Barbour JD, Hecht FM, Wrin T, et al. Persistence of primary drug resistance among recently HIV-1 infected adults. *AIDS* 2004; **18**: 1683-9.

14. Parisi SG, Mazzi R, Boldrin C, et al. Drug-resistance mutations can be archived very early in HIV primary infection. *AIDS* 2006; **20**: 1337-8.

15. Little SJ, Frost SD, Wong JK, et al. Persistence of transmitted drug resistance among subjects with primary human immunodeficiency virus infection. *J Virol* 2008; **82**: 5510-8.

16. Jain V, Sucupira MC, Bacchetti P, et al. Differential persistence of transmitted HIV-1 drug resistance mutation classes. *J Infect Dis* 2011; **203**: 1174-81.

17. Pao D, Andrady U, Clarke J, et al. Long-term persistence of primary genotypic resistance after HIV-1 seroconversion. *J Acquir Immune Defic Syndr* 2004; **37**: 1570-3.

18. Bezemer D, de RA, Prins M, et al. Evolution of transmitted HIV-1 with drug-resistance mutations in the absence of therapy: effects on CD4+ T-cell count and HIV-1 RNA load. *Antivir Ther* 2006; **11**: 173-8.

19. Castro H, Pillay D, Cane P, et al. Persistence of HIV-1 transmitted drug resistance mutations. *J Infect Dis* 2013; **208**: 1459-63.

20. Johnson JA, Li JF, Wei X, et al. Minority HIV-1 drug resistance mutations are present in antiretroviral treatment-naive populations and associate with reduced treatment efficacy. *PLoS Med* 2008; **5**: e158.

21. Li JZ, Paredes R, Ribaudo HJ, et al. Low-frequency HIV-1 drug resistance mutations and risk of NNRTI-based antiretroviral treatment failure: a systematic review and pooled analysis. *JAMA* 2011; **305**: 1327-35.

22. Cozzi-Lepri A, Noguera-Julian M, Di GF, et al. Low-frequency drug-resistant HIV-1 and risk of virological failure to first-line NNRTI-based ART: a multicohort European case-control study using centralized ultrasensitive 454 pyrosequencing. *J Antimicrob Chemother* 2015; **70**: 930-40.

23. Gega A, Kozal MJ. New Technology to detect low-level drug-resistant HIV variants. *Future Virology* 2011; **6**: 17-26.

24. Li JZ, Kuritzkes DR. Clinical implications of HIV-1 minority variants. *Clin Infect Dis* 2013; **56**: 1667-74.

25. Geretti AM, Paredes R, Kozal MJ. Transmission of HIV drug resistance: lessons from sensitive screening assays. *Curr Opin Infect Dis* 2015; **28**: 23-30.

26. Lataillade M, Chiarella J, Yang R, et al. Prevalence and clinical significance of HIV drug resistance mutations by ultra-deep sequencing in antiretroviral-naive subjects in the CASTLE study. *PLoS One* 2010; **5**: e10952.

27. Aghaizu A, Murphy G, Tosswill J, et al. Recent infection testing algorithm (RITA) applied to new HIV diagnoses in England, Wales and Northern Ireland, 2009 to 2011. *Euro Surveill* 2014; **19**.

28. Health Protection Agency. HIV in the United Kingdom: 2012 Report. In. London: Health Protection Services, Colindale, 2012.

29. Birrell PJ, Gill ON, Delpech VC, et al. HIV incidence in men who have sex with men in England and Wales 2001-10: a nationwide population study. *Lancet Infect Dis* 2013; **13**: 313-8.

30. Duong YT, Qiu M, De AK, et al. Detection of recent HIV-1 infection using a new limiting-antigen avidity assay: potential for HIV-1 incidence estimates and avidity maturation studies. *PLoS One* 2012; **7**: e33328.

31. Suligoi B, Galli C, Massi M, et al. Precision and accuracy of a procedure for detecting recent human immunodeficiency virus infections by calculating the antibody avidity index by an automated immunoassay-based method. *J Clin Microbiol* 2002; **40**: 4015-20.

32. Kassanjee R, Pilcher CD, Busch MP, et al. Viral load criteria and threshold optimization to improve HIV incidence assay characteristics - a CEPHIA analysis. *AIDS* 2016.

33. Cane P. HIV drug resistance testing. *Methods Mol Biol* 2011; **665**: 123-32.

34. Alcantara LC, Cassol S, Libin P, et al. A standardized framework for accurate, high-throughput genotyping of recombinant and non-recombinant viral sequences. *Nucleic Acids Res* 2009; **37**: W634-W642.

35. de Oliveira T, Deforche K, Cassol S, et al. An automated genotyping system for analysis of HIV-1 and other microbial sequences. *Bioinformatics* 2005; **21**: 3797-800.

36. Kosakovsky Pond SL, Posada D, Stawiski E, et al. An evolutionary model-based algorithm for accurate phylogenetic breakpoint mapping and subtype prediction in HIV-1. *PLoS Comput Biol* 2009; **5**: e1000581.

37. Zhang M, Schultz AK, Calef C, et al. jpHMM at GOBICS: a web server to detect genomic recombinations in HIV-1. *Nucleic Acids Res* 2006; **34**: W463-W465.

38. Struck D, Lawyer G, Ternes AM, et al. COMET: adaptive context-based modeling for ultrafast HIV-1 subtype identification. *Nucleic Acids Res* 2014; **42**: e144.

39. 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.

40. Shafer RW. Rationale and uses of a public HIV drug-resistance database. *J Infect Dis* 2006; **194 Suppl 1**: S51-S58.

41. Williams I, Churchill D, Anderson J, et al. British HIV Association guidelines for the treatment of HIV-1-positive adults with antiretroviral therapy 2012 (Updated

November 2013. All changed text is cast in yellow highlight.). *HIV Med* 2014; **15 Suppl 1**: 1-85.

42. Palmer S, Boltz V, Maldarelli F, et al. Selection and persistence of non-nucleoside reverse transcriptase inhibitor-resistant HIV-1 in patients starting and stopping non-nucleoside therapy. *AIDS* 2006; **20**: 701-10.

43. Simen BB, Simons JF, Hullsiek KH, et al. Low-abundance drug-resistant viral variants in chronically HIV-infected, antiretroviral treatment-naive patients significantly impact treatment outcomes. *J Infect Dis* 2009; **199**: 693-701.

44. Buckton AJ, Prabhu D, Motamed C, et al. Increased detection of the HIV-1 reverse transcriptase M184V mutation using mutation-specific minority assays in a UK surveillance study suggests evidence of unrecognized transmitted drug resistance. *HIV Med* 2011; **12**: 250-4.

45. Gianella S, Delport W, Pacold ME, et al. Detection of minority resistance during early HIV-1 infection: natural variation and spurious detection rather than transmission and evolution of multiple viral variants. *J Virol* 2011; **85**: 8359-67.

46. Metzner KJ, Scherrer AU, Preiswerk B, et al. Origin of minority drug-resistant HIV-1 variants in primary HIV-1 infection. *J Infect Dis* 2013; **208**: 1102-12.

47. Lipscomb JT, Switzer WM, Li JF, et al. HIV reverse-transcriptase drug resistance mutations during early infection reveal greater transmission diversity than in envelope sequences. *J Infect Dis* 2014; **210**: 1827-37.

48. Mbisa JL, Fearnhill E, Dunn DT, et al. Evidence of Self-Sustaining Drug Resistant HIV-1 Lineages Among Untreated Patients in the United Kingdom. *Clin Infect Dis* 2015; **61**: 829-36.

49. Hue S, Gifford RJ, Dunn D, et al. Demonstration of sustained drug-resistant human immunodeficiency virus type 1 lineages circulating among treatment-naive individuals. *J Virol* 2009; **83**: 2645-54.

50. Gupta S, Lataillade M, Kyriakides TC, et al. Low-frequency NNRTI-resistant HIV-1 variants and relationship to mutational load in antiretroviral-naive subjects. *Viruses* 2014; **6**: 3428-37.

51. The UK Collaborative Group on HIV Drug Resistance. The increasing genetic diversity of HIV-1 in the UK, 2002-2010. In: 2014; 773-80.

52. Johnson JA, Li JF, Wei X, et al. Simple PCR assays improve the sensitivity of HIV-1 subtype B drug resistance testing and allow linking of resistance mutations. *PLoS One* 2007; **2**: e638.

53. Archer J, Baillie G, Watson SJ, et al. Analysis of high-depth sequence data for studying viral diversity: a comparison of next generation sequencing platforms using Segminator II. *BMC Bioinformatics* 2012; **13**: 47.

54. Orton RJ, Wright CF, Morelli MJ, et al. Distinguishing low frequency mutations from RT-PCR and sequence errors in viral deep sequencing data. *BMC Genomics* 2015; **16**: 229.

Table 1. Demographic characteristics of UK patients included in the molecular surveillance of MSM recently infected with HIV, July 2011 - December 2013

		Category				% of rec. infections sequenced
		New diagnoses	RITA tested	Recent infections	Sequenced	
Geographic Region	North	1193	712	149	62	41.6
	Mid/East	914	482	114	53	46.5
	London	3646	2208	657	278	42.3
	South	1039	605	157	71	45.2
	NI/Wales	301	112	24	10	41.7
	Total	7093	4119	1101	474	43.1
Age Group	15-24	1071	659	234	73	31.2
	25-34	2643	1558	459	203	44.3
	35-49	2598	1493	330	158	47.9
	50+	781	409	78	40	51.2
	Total	7093	4119	1101	474	43.1

Table 2. Distribution of HIV-1 subtypes among recently infected MSM in the UK

Subtype	number	%	95% CI
A1	15	3.4	[1.9-5.5]
B	334	75.6	[71.3-79.5]
C	10	2.3	[1.1-4.1]
F1	20	4.5	[2.8-6.9]
G	2	0.5	[0.06-1.6]
CRF01_AE	6	1.4	[0.5-2.9]
CRF02_AG	19	4.3	[2.6-6.6]
CRF06_cpx	6	1.4	[0.5-2.9]
CRF07_BC	1	0.2	[0.01-1.3]
Rare CRF/URF	29	6.6	[4.4-9.3]
Total	442	100	-

Table 3. Specific TDR mutations identified at different variant frequency thresholds

	PI mutations		NRTI mutations		NNRTI mutations	
Variant frequency threshold	>20%	2-20%	>20%	2-20%	>20%	2-20%
Mutations (n)	L90M (7)	M46IL (10)	T215rev (9)	D67GN (4)	K103N (5)	G190E (2)
	M46IL (3)	V32I (3)	K219N (3)	K70RE (2)	V106A	K101E
	V82AL (2)	D30N (3)	M41L	T215rev (2)	Y188L	Y188H
		N83D (2)	K70R	F77L	K101E	K103N
		I47V (2)	M184V	V75A		Y181C
		V82A (2)				
		L90M				
		N88D				
		I54L				
		I50V				
Total	12	26	15[a]	10	8[b]	6

[a] two NRTI mutations present in one sample (M41L and M184V)

[b] two NNRTI mutations present in one sample (K103N and Y188L)

Table 4. Factors associated with transmitted drug resistance

Variant frequency threshold	Parameter	Univariate		Adjusted (multivariate)	
		OR [CI]	*P*-value	OR [CI]	*P*-value
>2% (all TDR)	Infected outside UK	2.12 [0.93-4.83]	0.073	**2.64 [1.03-6.78]**	**0.044**
	Born outside UK	1.26 [0.75-2.11]	0.382	0.75 [0.37-1.55]	0.442
	Non-white ethnicity	1.38 [0.73-2.61]	0.315	1.63 [0.76-3.55]	0.211
	Outside London	1.37 [0.81-2.33]	0.241	1.69 [0.89-3.21]	0.107
	Age (15-34)	1.02 [0.61-1.70]	0.936	1.34 [0.74-2.44]	0.335
	Non-B Subtype	**0.41 [0.19-0.85]**	**0.017**	**0.38 [0.17-0.88]**	**0.024**
>20% (high frequency TDR)	Infected outside UK	1.74 [0.57-5.38]	0.333	3.13 [0.83-11.74]	0.092
	Born outside UK	0.91 [0.43-1.91]	0.802	0.43 [0.15-1.26]	0.124
	Non-white ethnicity	1.12 [0.44-2.83]	0.811	1.61 [0.53-4.87]	0.398
	Outside London	1.07 [0.51-2.22]	0.862	1.67 [0.70-3.95]	0.247
	Age (15-34)	1.47 [0.70-3.08]	0.310	1.66 [0.72-3.83]	0.237
	Non-B Subtype	0.30 [0.09-1.01]	0.051	0.30 [0.08-1.08]	0.065
2-20% (low frequency TDR)	Infected outside UK	2.36 [0.90-6.19]	0.082	2.08 [0.70-6.16]	0.185
	Born outside UK	1.60 [0.84-3.02]	0.150	1.23 [0.52-2.91]	0.640
	Non-white ethnicity	1.81 [0.87-3.79]	0.115	2.02 [0.83-4.90]	0.122
	Outside London	1.51 [0.77-2.95]	0.230	1.38 [0.62-3.09]	0.427
	Age (15-34)	0.94 [0.50-1.77]	0.841	1.45 [0.68-3.06]	0.337
	Non-B Subtype	0.59 [0.26-1.37]	0.223	0.58 [0.22-1.52]	0.265

Figure Legends

Figure 1. Specificity and sensitivity for detection of low frequency variants. (A) Correlation of translated codon frequencies, and (B) concordance of translated amino acid variant frequencies in protease and N-terminal half of RT (up to codon 340) for a clinical sample in two independent experiments. Concordance was considered at two levels, exact frequency (dark gray bars) or the frequency of the repeat experiment being within 50% of the frequency in first experiment (light gray bars). (C) Box-and-whisker plot showing the median, lower and upper quartile depth of coverage for the two independent runs, and the variability outside the lower and upper quartiles.

Figure 2. Predicted drug susceptibility of the samples containing TDR among recently infected MSM. The susceptibility of each sample at ≥20% and >2% mutation frequency to licensed ARV drugs was predicted using the Stanford HIV drug resistance database genotypic interpretation algorithm. The graph shows the proportion of samples in each of the top three drug resistance levels used by the algorithm: low, intermediate and high level. The effect on drugs currently recommended for first-line treatment in the UK are shown individually whereas the effect on older PI and NRTI drugs that are no longer used in first-line therapy (other PI and NRTI) are shown together at the top of the graph. AZT, zidovudine; ABC, abacavir; TDF, tenofovir; 3TC, lamivudine; FTC, emtricitabine, RPV, rilpivirine; NVP, nevirapine; EFV, efavirenz; ETR, etravirine; ATV, atazanavir; LPV, lopinavir; DRV, darunavir; Other PIs, fosamprenavir, indinavir, nelfinavir, saquinavir, tipranavir; Other NRTI, stavudine, didanosine.

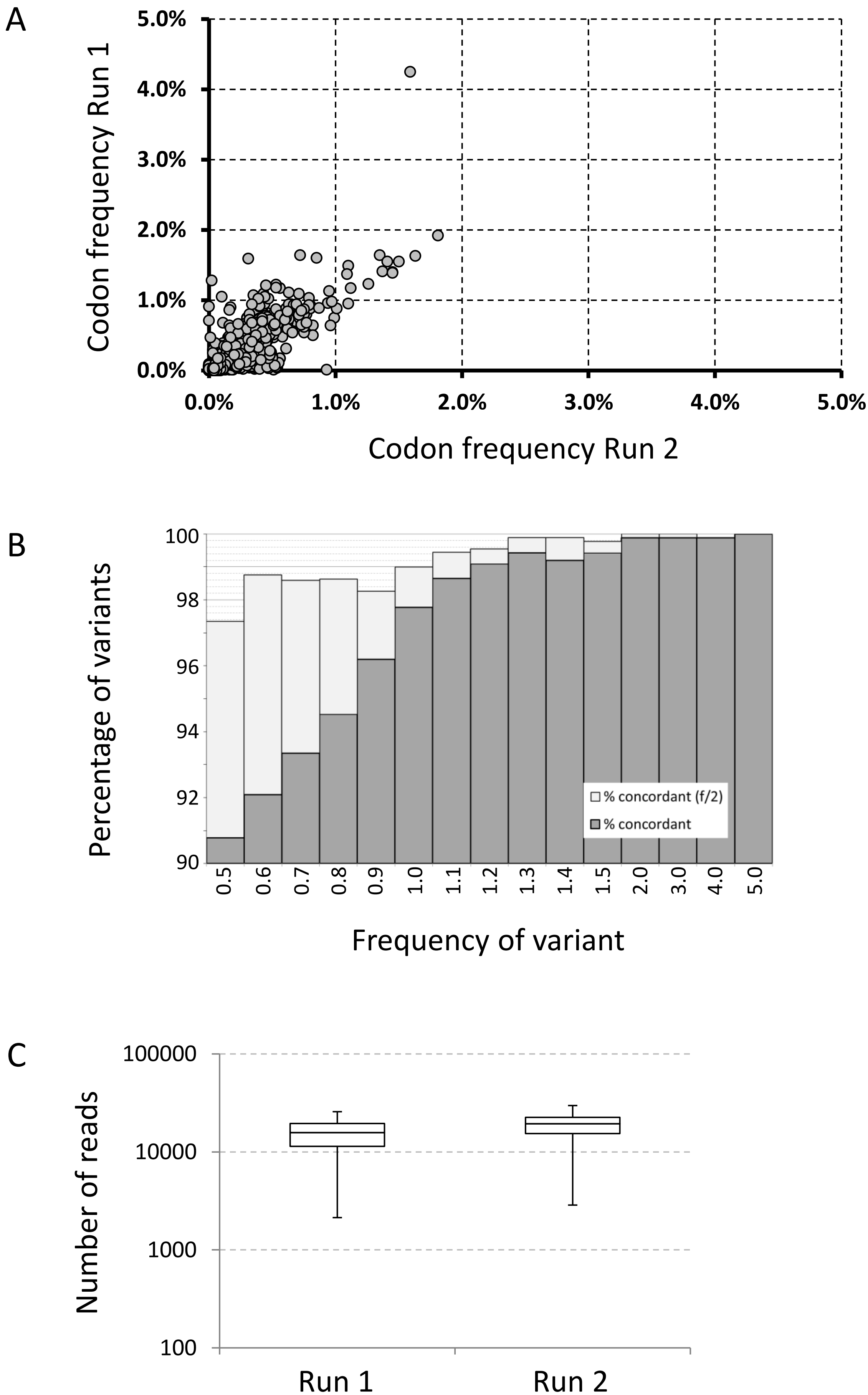
A
Codon frequency Run 1
5.0%
4.0%
3.0%
2.0%
1.0%
0.0%
0.0%
1.0%
2.0%
3.0%
4.0%
5.0%
Codon frequency Run 2
B
Percentage of variants
100
98
96
94
92
90
% concordant (f/2)
% concordant
0.5
0.6
0.7
0.8
0.9
1.0
1.1
1.2
1.3
1.4
1.5
2.0
3.0
4.0
5.0
Frequency of variant
C
Number of reads
100000
10000
1000
100
Run 1
Run 2

Figure 1

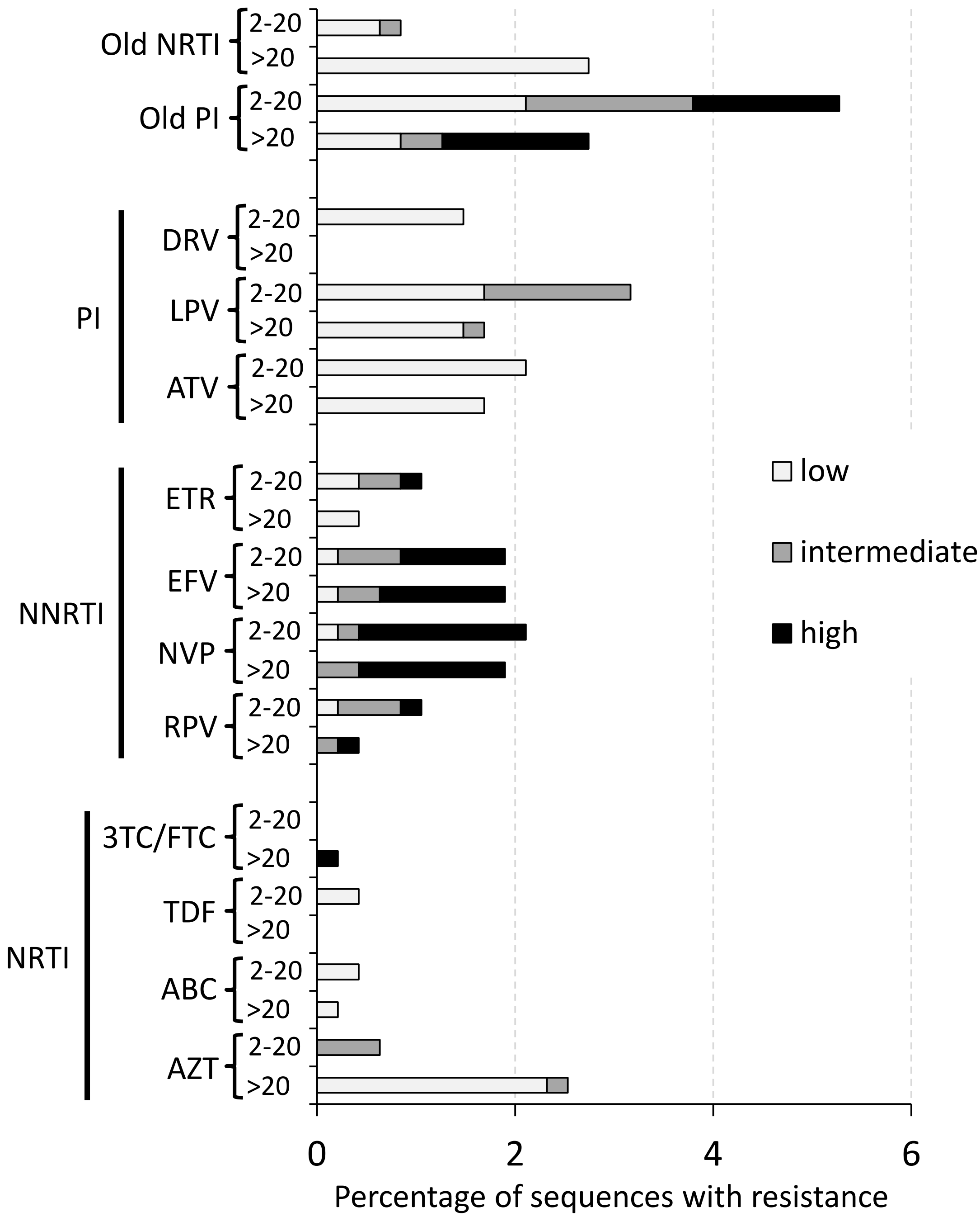
Old NRTI
2-20
>20
Old PI
2-20
>20
PI
DRV
2-20
>20
LPV
2-20
>20
ATV
2-20
>20
NNRTI
ETR
2-20
>20
EFV
2-20
>20
NVP
2-20
>20
RPV
2-20
>20
NRTI
3TC/FTC
2-20
>20
TDF
2-20
>20
ABC
2-20
>20
AZT
2-20
>20
low
intermediate
high
0
2
4
6
Percentage of sequences with resistance

Figure 2