

**TRANSMITTED HIV-1 DRUG RESISTANCE IN THE START STUDY USING  
NEXT GENERATION SEQUENCING**

John D Baxter<sup>1</sup>, David Dunn<sup>2</sup>, Anna Tostevin<sup>2</sup>, Rasmus L Marvig<sup>3</sup>, Marc Bennedbæk<sup>8</sup>,  
Alessandro Cozzi-Lepri<sup>2</sup>, Shweta Sharma<sup>4</sup>, Michael J Kozal<sup>5</sup>, Mark Gompels<sup>6</sup>, Angie N  
Pinto<sup>7</sup>, and Jens Lundgren<sup>8</sup> for the INSIGHT START Study Group

<sup>1</sup>Cooper University Hospital/Cooper Medical School of Rowan University, Camden, NJ,  
USA, <sup>2</sup>Institute for Global Health, UCL, London, UK, <sup>3</sup>Center for Genomic Medicine,  
Rigshospitalet, Copenhagen University Hospital, Denmark, <sup>4</sup>Division of Biostatistics, School  
of Public Health, University of Minnesota, Minneapolis, MN, USA, <sup>5</sup>Yale University School  
of Medicine, New Haven, CT, USA, <sup>6</sup>North Bristol NHS Trust, UK, <sup>7</sup>The Kirby Institute,  
University of New South Wales, Sydney, Australia, <sup>8</sup>Copenhagen HIV Programme,  
Rigshospitalet, University of Copenhagen, Copenhagen, Denmark.

**Key Words:** Human Immunodeficiency Virus (HIV); HIV drug resistance; Antiretroviral  
therapy; Next generation sequencing

**Corresponding Author:**

John D. Baxter, M.D.  
Professor of Medicine  
Division of Infectious Diseases  
Cooper University Hospital  
Cooper Medical School of Rowan University  
401 Haddon Avenue, Room 270  
Camden, NJ 08103  
**Tel:** (856) 757-7767  
**Fax:** (856) 757-7803  
**E-mail:** [baxter@rowan.edu](mailto:baxter@rowan.edu)

Short Title: HIV-1 TDR in START by NGS

## ABSTRACT

**Objectives:** The aim of this analysis was to characterize transmitted drug resistance (TDR) in START study participants by next generation sequencing (NGS), a sensitive assay capable of detecting low-frequency variants.

**Methods:** Stored plasma from participants with entry HIV RNA >1,000 copies/ml were analyzed by NGS (Illumina MiSeq). TDR was based on the WHO 2009 surveillance definition with the addition of reverse transcriptase (RT) mutations T215N and E138K, and integrase strand transfer inhibitor (INSTI) surveillance mutations (Stanford HIVdb). Drug resistance mutations (DRMs) detected at three thresholds are reported: above 2%, 5%, and 20% of the viral population.

**Results:** Between 2009-2013, START enrolled 4,684 ART-naïve individuals in 35 countries. Baseline NGS data at study entry was available for 2,902 participants. Overall prevalence of TDR using a detection threshold of 2%/5%/20% was 9.2%/5.6%/3.2% for NRTIs, 9.2%/6.6%/4.9% for NNRTIs, 11.4%/5.5%/2.4% for PIs, and 3.5%/1.6%/0.1% for INSTI DRMs and varied by geographic region. Using the 2% detection threshold, individual DRMs with the highest prevalence were: PI M46IL (5.5%), RT K103NS (3.5%), RT G190ASE (3.1%), T215ISCDVEN (2.5%), RT M41L (2.2%), RT K219QENR (1.7%) and PI D30N (1.6%). INSTI DRMs were detected almost exclusively below the 20% detection threshold, most commonly Y143H (0.4%), Q148R (0.4%), and T66I (0.4%).

**Conclusions:** Use of NGS in this study population resulted in the detection of a large proportion of low-level variants which would not have been detected by traditional Sanger sequencing. Global surveillance studies utilizing NGS should provide a more comprehensive assessment of TDR prevalence in different regions of the world.

## INTRODUCTION

Transmitted drug resistance (TDR) refers to the presence of one or more HIV-1 drug resistance mutations (DRMs) in individuals with no prior history of antiretroviral drug exposure. TDR has been associated with reduced susceptibility to antiretroviral agents and increases the risk of a suboptimal virologic response to initial antiretroviral therapy.(1, 2) Multiple studies from the U.S. and Europe have described the prevalence of TDR in treatment naïve individuals, which typically ranges from 5-15%.(3-8) Available data from resource limited countries has shown that TDR is becoming an emerging health issue, with increasing prevalence rates being reported particularly for non-nucleoside reverse transcriptase inhibitors (NNRTIs).(3, 9)

Global surveillance studies of TDR have primarily used standard Sanger sequencing looking for reverse transcriptase (RT) and protease inhibitor (PI) DRMs, with prevalence rates varying by geographic region.(3, 9) There are limited data on global TDR using next generation sequencing (NGS), which is a highly sensitive assay capable of detecting low-frequency (minor) variants.(10-12) NGS can identify the prevalence of DRMs associated with TDR in a population of individuals while allowing for the detection of low frequency drug-resistant related variants which may have reduced viral fitness.

The Strategic Timing of AntiRetroviral Treatment (START) study is an international trial of immediate versus deferred ART initiation among treatment-naïve HIV-positive individuals with CD4 counts >500 cells/ $\mu$ L. Participants were enrolled from 215 sites in 35 countries, representing both resource-rich and resource-limited regions of the world and included North America, Europe, Australia, Latin America, Africa, and Asia. A prior study of TDR in a subset of START study participants was previously described using locally performed Sanger

sequencing available almost exclusively in resource-rich regions.(13) In this analysis, we applied a more sensitive assay for detection of drug resistance using NGS to determine TDR prevalence rates from stored specimens for nucleoside reverse transcriptase inhibitor (NRTI), NNRTI, PI, and integrase strand transfer inhibitor (INSTI) DRMs at study entry in the entire START study population.

## METHODS

### *Study population*

The START trial, conducted by the International Network for Strategic Initiatives in Global HIV Trials (INSIGHT), enrolled participants between April 2009 and December 2013. The study design and data collection plan has previously been reported.(14) A plasma sample, taken up to 60 days prior to enrolment, was obtained from participants consenting to the storage of specimens. Samples were stored centrally at the INSIGHT laboratory repository in Cinnaminson, New Jersey, prior to shipment to the Center for Genomic Medicine, Rigshospitalet, Copenhagen for NGS analysis. NGS was attempted on all samples with HIV RNA >1,000 copies/ml at study entry.

### *Sample preparation, amplification of viral RNA, and sequencing*

The plasma samples were thawed from -80C freezers at room temperature. 500 µL plasma was transferred to new RNase-free tubes and centrifuged at 2000xG for 15 minutes. The supernatant was extracted and centrifuged at 21.000xG for 75 minutes and 360 µL of the top supernatant was discarded. Viral RNA was extracted using QIAamp viral RNA extraction kit (Qiagen) on a QIAcube robot using the manufacturer's guidelines. Reverse transcription polymerase chain reaction (RT-PCR) was used to amplify two amplicons from the viral RNA. The primer sequences (available in Appendix, Table 5) were designed by Gall *et*

*al.*(15) The reverse-transcription and amplification was performed using SuperScript III One-Step RT-PCR System with Platinum Taq High Fidelity (Thermo Fisher Scientific) according to manufacturer's instructions with 10 µL viral RNA used for each amplicon. The PCR products were purified using Ampure XP (Agencourt) PCR purification according to the manufacturer's instructions.

The two amplicons were pooled for each sample prior to library preparation. Libraries were prepared using Nextera XT (Illumina) according to manufacturer's protocol except that 1.5X of the library normalization beads was used in the final normalization step. DNA libraries were sequenced on an Illumina MiSeq machine using a MiSeq 150-cycle V3 reagent kit (Illumina).

#### *HIV-1 subtyping*

Subtypes were assigned to samples for which the consensus sequence information was available for at least 90% of either of the two amplicons, or at least 50% of both. The consensus sequence was analyzed with REGA HIV-1 Subtyping Tool version 3.(16) Output was manually inspected to check for the presence of subtype specific sequences within the given consensus sequence. Samples were assigned to be either a pure subtype (A-D, F, G) or a recombinant subtype in cases where the genome showed presence of sequences specific to >1 pure subtype.

Deleted: (17)

#### *Identification of drug resistance mutations with Virvarseq*

Sequence reads (FASTQ files) were analysed with VirVarSeq version 20140929, which calls variants at the codon level.(17) VirVarSeq was run with HIV-1 HXB2 as reference and with default settings, except that soft-clipping as defined by the aligner were ignored and without

the mixture model step (as recommended by Huber *et al.*(18) From the output, we extracted amino acid frequencies in *pol* gene from amino acid position 1 to 935 where positions 1-99 encode protease (PR) protein, positions 100-659 encode reverse transcriptase (RT) protein, and positions 660-935 partially encode integrase (IN) protein (our amplicon did not cover position 936-947).

#### *Definition of transmitted drug resistance and phenotypic drug susceptibility*

As in a previous paper from START reporting the results of locally performed Sanger sequencing, TDR was based on the WHO 2009 surveillance list with the addition of RT mutations T215N and E138K.(13, 19) INSTI mutations, which are not included on this list, were defined as those on the Stanford HIVdb surveillance drug resistance mutation list, namely: T66AIK, E92Q, F121Y, G140ACS, Y143CHR, S147G, Q148HKR, N155HS.(20) Interpretation of phenotypic drug susceptibility was standardized using the Stanford HIVdb algorithm v8.6 which defines drug resistance as: none, potential low level, low level, intermediate, or high.(21) To achieve consistency with WHO resistance reports, predicted potential low level is not reported. It is noted that the Stanford HIVdb algorithm considers a much wider range of mutations than considered by the WHO 2009 surveillance list (including the integrase gene) and these additional TDR DRMs detected by NGS were included for predicted phenotypic drug susceptibility.

#### *Sequencing depth and thresholds for calling DRMs*

Sequence read coverage depth varied markedly across the sequenced amplicons (highest in protease, intermediate in reverse transcriptase, lowest in integrase). We stipulated a minimum read depth of 200 across the region spanning all relevant mutations *within* each gene. For WHO surveillance mutations this was codons 23-90 of protease, codons 41-230 of

RT, and codons 66-155 of integrase; for Stanford predicted phenotypic drug susceptibility this was codons 10-90 of protease, codons 41-348 of RT and codons 51-263 of integrase. This resulted in different denominators for different drug classes, which were therefore analysed separately. DRMs detected at three thresholds by NGS are reported: above 2%, 5%, and 20% of the viral population (the latter comparable to the detection threshold for Sanger sequencing).(11, 18)

#### *Statistical Methods*

If two or more mutations were present in the analysis of drug class specific TDR by detection threshold (Figure 1), the highest frequency was used in the analysis. Fisher's exact test (two-sided) was used to test the association between geographical region and whether TDR variants were observed at 2-5%, 5-20%, or greater than 20%. Logistic regression analysis was used to examine predictors of drug class specific TDR. Odds ratios were adjusted, a priori, for the effects of geographical region, calendar year of enrolment, and age. Subtype was not included in these models as independent effects of subtype and geographical region could not be estimated due to the very strong association between these two variables. All statistical analyses were conducted using STATA, version 15 (StataCorp, Houston, TX). A p-value of < 0.05 was deemed significant.

## RESULTS

### *Study participant baseline characteristics*

START enrolled 4,684 ART-naïve individuals from 35 countries between April 2009 and December 2013.<sup>(22)</sup> Europe had the highest number of participants (1,539), followed by Latin America (1,174), Africa (999), USA (507), Asia (356), and Australia (109). Of the 4,072 individuals with viral load >1,000 copies/ml (and thus eligible for this study), 3,785 (93%) had a baseline specimen analysed by NGS. Baseline characteristics of these individuals (see Table 1) showed a median CD4 count of 643 cells/μL, median HIV RNA of 18,105 copies/mL, and median time since diagnosis was 0.96 years. Using the criteria described in Methods, 2,901 (77%) specimens produced an evaluable result for the assessment of protease TDR, 2,180 (58%) for NRTI/NNTRI TDR, and 1,338 (35%) for integrase TDR. Median (IQR) read depth was 13,134 (5,644-26,671) in protease, 6,891 (3,078-14,194) in RT, and 2,240 (987-5,213) in integrase. Subtype B was the most common subtype, followed by subtype C (Table 1).

### *Prevalence of transmitted drug resistance*

Figure 1 shows the prevalence of TDR, overall and by geographical region, at different detection thresholds. Overall, at the 2% threshold, TDR prevalence was 9.2% for NRTI DRMs, 9.2% for NNRTI DRMs, 11.4% for PI DRMs, and 3.5% for INSTI DRMs. Using a 5% threshold the respective values were 5.6%, 6.6%, 5.5%, and 1.6%; using a 20% threshold the respective values were 3.2%, 4.9%, 2.4%, and 0.1%. Comparing the 2% and 20% thresholds of detection for TDR DRMs by drug class there was a 4.8 (11.4%/2.4%) and 35-fold (3.5% vs 0.1%) greater proportion of viruses with PI and INSTI [minor](#) variants, respectively, whereas this ratio was 2.9 and 1.9 for TDR DRMs in the NRTI and NNRTI drug classes.

Deleted: minor



A comparatively high prevalence of NNRTI DRMs was observed in the USA (mainly mutations above the 20% threshold), as was a comparatively high prevalence of NRTI DRMs in Australia (mainly mutations above the 5% threshold). Otherwise, there was no clear geographical variation in the prevalence of DRMs for any drug class, with the apparent variation in integrase TDR likely due to the small number of cases. Nevertheless, there was some evidence of geographical variability in the *level* of DRM variants, conditional on being above the 2% threshold:  $P=0.05$  for NRTI TDR,  $P=0.007$  for NNRTI,  $P=0.11$  for PI TDR. The NRTI effect was mainly driven by a relative deficit of variants at 5-20% in Europe, and the NNRTI effect by a relative excess of variants at 2-5% in Asia and in Europe and a relative excess at 5-20% in Africa. There also was some evidence that TDR prevalence differed by HIV-1 subtype (data not shown), but these were reflective of the geographical distribution of different subtypes.

Table 2 shows the results of a logistic regression analysis of selected predictors for drug-class specific DRMs (2% detection threshold). This confirmed the visual impression of regional variability in the prevalence NRTI DRMs ( $P=0.002$ ) and NNRTI DRMs ( $P=0.02$ ) that was observed in Figure 1. The prevalence of PI DRMs was strongly associated with age (although lacking a clear trend) and weakly associated with gender; however, these associations were not observed in sensitivity analyses limited to mutant variants detected above the 5% or 20% threshold (results not shown). There was a weak suggestion of an overall increase in the prevalence of NNRTI DRMs (but not NRTI DRMs or PI DRMs) over the 4½ year recruitment period to START. Calendar time trends stratified by geographical region were also examined but revealed no notable findings.

Specific DRMs observed in at least 0.5% of samples (at the 2% detection threshold) are shown in Table 3. The most common NRTI DRMs were T215 revertants (2.5%), M41L (2.2%), and K219QENR (1.7%). The T215 revertants and, to a lesser extent M41L, were mostly high level variants representing the majority of the quasispecies (i.e. occurring in >80% of the viral population) whereas a wider range of variant frequency was observed for K219QENR. Notably, the K65R mutation was not observed in any sample, even at the 2% threshold. M184V and M184I were detected in 7 and 18 samples respectively, the latter generally as a low level variant (2-5%). The most common NNRTI DRMs – K103NS (3.5%), G190ASE (3.1%), and E138K (1.6%) – showed diverse patterns. K103NS variants were mainly observed (60/77, 78%) in more than 80% of the viral quasispecies, E138K mainly (20/34, 59%) in less than 5%, and G190ASE displayed a more uninform spread. Estimates of PI TDR are strongly influenced by the M46IL mutation, which was observed in 5.5% of all samples, mostly as low-level variants (Table 3). Excluding M46IL variants would reduce estimated PI TDR (at 2% threshold) from 11.4% to 6.6%. The D30N mutation was mainly (41/46, 89%) present as a low level variant below 20%, whereas most (12/18, 67%) L90M were detected in more than 80% of the viral quasispecies. All individual INSTI DRMs were observed below the 20% threshold, with the exception of G140S (detected at the 25% level, sample from Spain collected in 2012) and G140A (21%, sample from Peru in 2011).

[The majority of DRMs detected at each threshold occurred as solitary mutations in a single drug class for individual participant samples. There were a relatively small number of samples which had more than one DRM within a drug class or mutations present in multiple drug classes. For individual samples where TDR DRMs were detected at the 2% threshold, multiple within class DRMs occurred in 18.5% of those with NRTI DRMs, 13.4% with](#)

[NNRTI DRMs, 13.5% with PI DRMs, and 4.3% with INSTI DRMs. Of those with TDR DRMs detected at the 2% threshold in both RT and protease, multi-class resistance with at least one NRTI, NNRTI, and PI DRM occurred in only three participant samples.](#)

We further examined (but did not formally analyse) specific DRMs by geographical region, which indicated marked variability (Appendix, Table 4). For example, K103NS was the most common NNRTI mutation in USA and Latin America whereas G190ASE was the most common NNRTI mutation in Europe, Africa, and Asia. This would appear to be the main explanation for the geographical variability in the level of DRM variants which was observed in Figure 1.

#### *Predicted phenotypic drug susceptibility*

Analysis of predicted phenotypic resistance (low level, intermediate, and high level) for selected antiretroviral agents according to threshold of detection (Figure 2) resulted in several important findings. First, for certain drugs, particularly nelfinavir and raltegravir, using a 2% rather than a 20% variant threshold dramatically increases estimates of predicted phenotypic resistance (much more than drug class specific TDR estimates). This is due to the wider range of mutations incorporated in the Stanford HIVdb algorithm than in the WHO or Stanford drug mutation surveillance lists. For example, the integrase mutations 92G, 138K, and 163R, which were generally observed at 2-5% variant level by NGS, individually predict low-level resistance to raltegravir, but are not counted on the Stanford drug mutation surveillance list. Conversely, resistance to efavirenz is mainly predicted by the K103NS mutations. As these mainly occur at high variant levels (Table 3), the frequency of predicted resistance is largely unaffected by the variant threshold that is used. Secondly, there was minimal predicted resistance to darunavir and dolutegravir regardless of the detection threshold.

Finally, the frequency of tenofovir resistance was estimated to be 1.3% at 20% threshold, 1.6% at 5% threshold, and 2.4% at 2% threshold, despite the complete absence of the K65R mutation. This is explained by the contribution of thymidine analogue mutations to predicted tenofovir resistance.

## DISCUSSION

The START trial represents one of the largest global cohorts with NGS characterization of TDR. Prevalence of TDR overall and by drug class was highest using the 2% detection threshold as a significant number of mutations identified were minor variants. At the 2% threshold the highest TDR prevalence was observed for PI DRMs (11.4%) while at the 20% threshold the highest TDR prevalence occurred with NNRTI DRMs (4.9%). Within drug classes, a greater proportion of participants harbored PI and INSTI ~~minor~~ variants compared to the NRTI and NNRTI associated DRMs. As such, Sanger sequencing, which is comparable to NGS at the 20% threshold of detection, would be expected to markedly underestimate TDR prevalence particularly for the PI and INSTI drug classes.

Deleted: minor

As described in this study, a higher proportion of PI minor variants identified when using NGS for detection of TDR was also reported in a UK surveillance study in recently infected MSMs.(11) Cunningham *et al.* found that the majority of low frequency variants (62%) detected were PI DRMs despite these mutations rarely being observed by traditional Sanger sequencing in treatment-experienced patients failing therapy in the UK.

Unlike most other studies of global TDR, our data also included NGS for surveillance INSTI DRMs. The INSTI DRMs identified were almost exclusively minor variants which occurred mostly between the 2% – 5% thresholds. The most commonly detected surveillance INSTI DRMs were G143CHR, associated with high level resistance to raltegravir, but these were only detected as minor variants. The only INSTI DRMs detected above the 20% threshold were G140S and G140A, which are associated with intermediate resistance to raltegravir and elvitegravir.(21, 23) Given the time period in which these samples were collected there would have been limited prior population exposure to integrase inhibitors, particularly in

regions such as Africa and Asia, therefore the majority of these mutations likely represent naturally occurring low level variants rather than actual DRMs selected under drug pressure and then transmitted. However, their significance remains uncertain and it is unknown if these low level INSTI variants could potentially compromise efficacy of integrase based regimens. INSTI DRMs have been reported to occur as natural polymorphisms and can be detected in samples from individuals obtained prior to the advent of the integrase inhibitors.(24) Further research in understanding their potential clinical significance is warranted given the wide-spread use of integrase-based regimens for initial therapy. In a recent analysis of data from INTEGRATE, a retrospective study from nine European HIV cohorts, baseline INSTI DRMs were found in one of 512 (0.2%) antiretroviral naïve individuals by Sanger sequencing, suggesting that INSTI TDR is uncommon and unlikely to be detected with this methodology.(25)

Variability of TDR by geographic region was observed in the START study population. By drug class, the highest prevalence of NNRTI DRMs was seen in the USA while the highest prevalence of NRTI DRMs occurred in Australia. Prior studies have shown that the prevalence of transmitted NNRTI resistance has historically been higher in North America compared to Europe.(3, 4, 6) The number of samples from participants included in this analysis was more limited from Australia but the relatively high prevalence of NRTI TDR has been observed previously with Sanger sequencing.(4, 13) [This is likely due to extensive prior exposure of viruses to NRTIs in the pool of individuals transmitting HIV-1 in this region.](#)

Geographic variability was also seen in the level of DRM variants detected at different thresholds, particularly for NNRTI TDR. Overall NNRTI TDR was lowest in Africa,

Deleted: N

occurring in 6.6% of study participants, however unlike the USA and South America, the majority of NNRTI DRMs occurred as minor variants (mostly in 5-20% of the quasispecies). Similarly, the prevalence of NNRTI TDR in Asia was 7.5% mostly in 2-5% of the quasispecies. This would suggest that most NNRTI DRMs associated with TDR in these regions may be occurring as minor variants and would not be detected by traditional Sanger sequencing.

[The clinical relevance of minor HIV-1 drug resistant variants remains controversial.](#)

[Transmission of minor variants associated with resistance to different drug classes has been documented in acutely infected individuals, although these are probably rare events.](#)(26) [A recent review of 103 studies of low-abundance drug-resistant variants found that it is difficult to evaluate the clinical impact of minor variants on first-line ART regimens given the heterogeneity of study designs and different laboratory methods used.](#)(27) [However, multiple studies assessing the impact of NNRTI minor variants have shown that the presence of these low level DRMs prior to initiation of treatment may reduce virologic response and increase the likelihood of failure with first-line NNRTI-based regimens.](#)(3, 28-31)

Deleted: (28)

Deleted: S

Deleted:

Deleted: Furthermore, m

[More recent global TDR surveillance data using Sanger sequencing in treatment naïve individuals has demonstrated significantly increasing prevalence rates of NNRTI TDR in Africa and Asia, reaching levels of >10% in some countries.](#)(9) WHO guidelines have suggested that if the prevalence of pretreatment drug resistance (which includes treatment naïve persons and those with prior ART exposure starting first-line therapy) exceeds 10% in a country, then use of non-NNRTI based first-line regimens should be considered.(32) Because of this, a number of countries in Africa and Asia have revised national treatment guidelines to INSTI based regimens as preferred initial therapy.(9) Furthermore, if non-

NNRTI alternatives are unaffordable then pre-treatment drug resistance testing is recommended.(33) Assuming low level NNRTI variants are clinically important, our data would suggest that the prevalence of NNRTI TDR is being underestimated in many regions of the world where previous surveillance efforts have been based on traditional sequencing methods.

In this diverse population of study participants, we found no clear baseline participant demographic predictors for the presence of drug-class specific TDR other than an association between PI DRMs and age. This association appears to have been largely due to the presence of low level PI variants and was not observed above the 5% detection threshold. There was also a weak trend of increasing NNRTI TDR over the time of study enrolment from 2009 – 2013.

When examining individual mutations by variant threshold, different patterns of DRM detection by drug class was observed. Some DRMs tended to occur almost exclusively or predominantly as low level variants and this was observed in each drug class. The most commonly detected NRTI DRMs, K219QENR and M184VI occurred predominately as minor variants while the majority of patients with T215 revertants had the mutation in more than 80% of the quasispecies. Of the most frequent NNRTI DRMs, G190ASE occurred predominately as minor variants while K103NS was dominant in the quasispecies. Within the PI-related DRMs the same picture emerged with M46IL and D30N predominately seen as minor variants, whereas L90M (associated with saquinavir resistance) if present dominated the quasispecies. Although smaller numbers of individuals were included in the Cunningham *et al.* study, a similar pattern was observed with the M46IL and D30N detected



predominately as minor variants while the majority of L90M mutations occurred above the 20% threshold.

The detection of some DRMs predominately at low levels is likely due to impaired viral fitness, which has been described for mutations such as M184VI and D30N.(3, 5, 34-37) The M184VI may also be linked with other DRMs and tends to wane over time due to overgrowth of more replication competent wild type virus.(38) However, mutations like the RT K103NS, the T215 revertants, and the PI L90M appear to have little effect on viral fitness and may persist for prolonged periods in individuals with TDR.(3, 8, 29, 37, 39, 40)

Predicted phenotypic susceptibility was based on an expanded list of mutations which had been detected by NGS and results varied by threshold of detection. Given that certain mutations occurred more frequently as minor variants, some agents had a greater degree of predicted phenotypic resistance, such as nelfinavir and raltegravir, when using the 2% threshold of detection. At the 20% threshold, predicted resistance to these agents was significantly diminished and remained mostly low or intermediate resistance. In contrast, efavirenz and 3TC/FTC had mostly high-level predicted resistance at all thresholds. This is due to single mutations causing high level resistance to these agents, which were detected at all thresholds, although the 184VI was less frequently detected above the 20% threshold. Overall, the first generation NNRTIs efavirenz and nevirapine (which has a similar resistance profile) would be predicted as the least active agents in those initiating therapy in this population. Although there was a moderate amount of reduced susceptibility to rilpivirine observed at all thresholds, this was mostly intermediate and low-level resistance. Rilpivirine is a second generation NNRTI which usually requires multiple mutations to cause high level resistance. Of note, there was very little predicted phenotypic resistance to darunavir and

dolutegravir. These agents would likely be highly active in this population, as well as the newer second generation integrase inhibitor bictegravir.

Potential limitations of this study include the threshold of detection for minor variants, as well as the ability to reliably sequence participant samples. NGS is a highly sensitive assay and there is the possibility of low level variants occurring as laboratory artefacts or sequencing errors, but prior studies have demonstrated that most discrepancies in variant calls occur below the 2% threshold.(11, 18, 41, 42) We were limited by our ability to produce an evaluable result by NGS from all available samples, particularly for reverse transcriptase and integrase DRMs. Another limitation of our study was that we did not assess TDR in those individuals with low level viremia, as NGS was only attempted in those participants with baseline HIV RNA levels >1,000 copies/ml. Furthermore, although the average time from diagnosis of the participants was one year prior to study enrolment and all had early stage HIV disease with CD4 counts >500 cells/ $\mu$ L, this cohort likely represents a mix of those with more recent infection and others with a significantly longer duration of infection.

In summary, using NGS in the START population revealed significant geographic diversity in prevalence of TDR. Our study resulted in detection of a large proportion of low-level variants which would not have been detected by traditional Sanger sequencing, the method commonly used to assess individual pre-treatment drug resistance and for population TDR surveillance. Further studies will be needed to assess the potential clinical impact of transmitted minority variants on treatment response. Given that TDR continues to occur in most regions of the world and prevalence has been increasing to certain agents at an alarming rate in some countries, it will be important to continue surveillance efforts and consider the use of more sensitive drug resistance assays.

## **Acknowledgements**

We wish to thank the study participants and clinical staff of the START trial. See Initiation of Antiretroviral Therapy in Early Asymptomatic Individuals, *New England Journal of Medicine* 2015;373:795-807 for the complete list of START investigators.

The study was supported in part by the National Institute of Allergy and Infectious Diseases (NIAID), National Institute of Health (NIH, grants UM1-AI068641, UM1-AI120197, and 1U01-AI36780), National Institutes of Health Clinical Center, National Cancer Institute, National Heart, Lung, and Blood Institute, Eunice Kennedy Shriver National Institute of Child Health and Human Development, National Institute of Mental Health, National Institute of Neurological Disorders and Stroke, National Institute of Arthritis and Musculoskeletal and Skin Diseases, Agence Nationale de Recherches sur le SIDA et les Hépatites Virales (France), National Health and Medical Research Council (Australia), National Research Foundation (Denmark), Bundesministerium für Bildung und Forschung (Germany), European AIDS Treatment Network, Medical Research Council (United Kingdom), National Institute for Health Research, National Health Service (United Kingdom), and University of Minnesota. Antiretroviral drugs were donated to the central drug repository by AbbVie, Bristol-Myers Squibb, Gilead Sciences, GlaxoSmithKline/ViiV Healthcare, Janssen Scientific Affairs, and Merck.

## Figure Legends

Figure 1. Prevalence of transmitted drug resistance by detection threshold and geographical region.

Footnote. Data also shown in tabular form in Appendix (Table 1).

Figure 2. Predicted phenotypic drug susceptibility by detection threshold. Panel A- above 2% detection threshold, B above 5% threshold, and C above 20% threshold.

Footnote. Predicted susceptibility based on Stanford HIVdb algorithm version 8.6. Data also shown in tabular form in Appendix (Table 2).

## References

1. Pillay D, Bhaskaran K, Jurriaans S, et al. The impact of transmitted drug resistance on the natural history of HIV infection and response to first-line therapy. *Aids*. 2006; 20:21-8.
2. Little SJ, Holte S, Routy JP, et al. Antiretroviral-drug resistance among patients recently infected with HIV. *The New England journal of medicine*. 2002; 347:385-94.
3. Günthard HF, Calvez V, Paredes R, et al. Human immunodeficiency virus drug resistance: 2018 recommendations of the international antiviral society–USA panel. *Clinical Infectious Diseases*. 2018; 68:177-87.
4. Frentz D, Boucher C and Van De Vijver D. Temporal changes in the epidemiology of transmission of drug-resistant HIV-1 across the world. *AIDS reviews*. 2012; 14:17-27.
5. Geretti AM. Epidemiology of antiretroviral drug resistance in drug-naive persons. *Current opinion in infectious diseases*. 2007; 20:22-32.
6. Wheeler WH, Ziebell RA, Zabina H, et al. Prevalence of transmitted drug resistance associated mutations and HIV-1 subtypes in new HIV-1 diagnoses, U.S.-2006. *Aids*. 2010; 24:1203-12.
7. Ross L, Lim ML, Liao Q, et al. Prevalence of antiretroviral drug resistance and resistance-associated mutations in antiretroviral therapy-naive HIV-infected individuals from 40 United States cities. *HIV clinical trials*. 2007; 8:1-8.
8. Novak RM, Chen L, MacArthur RD, et al. Prevalence of Antiretroviral Drug Resistance Mutations in Chronically HIV–Infected, Treatment-Naive Patients: Implications for Routine Resistance Screening before Initiation of Antiretroviral Therapy. *Clinical infectious diseases*. 2005; 40:468-74.
9. Organization WH. HIV Drug Resistance Report 2019. Geneva, Switzerland, 2019.
10. Derache A, Iwuji CC, Baisley K, et al. Impact of Next-generation Sequencing Defined Human Immunodeficiency Virus Pretreatment Drug Resistance on Virological Outcomes in the ANRS 12249 Treatment-as-Prevention Trial. *Clinical Infectious Diseases*. 2018.
11. Cunningham E, Chan Y-T, Aghaizu A, et al. Enhanced surveillance of HIV-1 drug resistance in recently infected MSM in the UK. *Journal of Antimicrobial Chemotherapy*. 2016:dkw404.
12. Moscona R, Ram D, Wax M, et al. Comparison between next-generation and Sanger-based sequencing for the detection of transmitted drug-resistance mutations among recently infected HIV-1 patients in Israel, 2000–2014. *Journal of the International AIDS Society*. 2017; 20:21846.
13. Baxter J, Dunn D, White E, et al. Global HIV-1 transmitted drug resistance in the INSIGHT S strategic Timing of Antiretroviral Treatment (START) trial. *HIV medicine*. 2015; 16:77-87.
14. Group ISS, Lundgren JD, Babiker AG, et al. Initiation of Antiretroviral Therapy in Early Asymptomatic HIV Infection. *The New England journal of medicine*. 2015; 373:795-807.
15. Gall A, Ferns B, Morris C, et al. Universal amplification, next-generation sequencing, and assembly of HIV-1 genomes. *J Clin Microbiol*. 2012; 50:3838-44.
16. Pineda-Peña A-C, Faria NR, Imbrechts S, et al. Automated subtyping of HIV-1 genetic sequences for clinical and surveillance purposes: performance evaluation of the new REGA version 3 and seven other tools. *Infection, genetics and evolution*. 2013; 19:337-48.
17. Verbist B, Thys K, Reumers J, et al. VirVarSeq: a low-frequency virus variant detection pipeline for Illumina sequencing using adaptive base-calling accuracy filtering. *Bioinformatics (Oxford, England)*. 2015; 31:94.

18. Huber M, Metzner KJ, Geissberger FD, et al. MinVar: A rapid and versatile tool for HIV-1 drug resistance genotyping by deep sequencing. *Journal of virological methods*. 2017; 240:7-13.
19. 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.
20. Database SUHDR. 2016 Surveillance Drug Resistance Mutation (SDRM) Worksheet: INIs.
21. Liu TF and Shafer RW. Web resources for HIV type 1 genotypic-resistance test interpretation. *Clinical infectious diseases*. 2006; 42:1608-18.
22. Group ISS. Initiation of antiretroviral therapy in early asymptomatic HIV infection. *New England Journal of Medicine*. 2015; 373:795-807.
23. Wensing AM, Calvez V, Ceccherini-Silberstein F, et al. 2019 update of the drug resistance mutations in HIV-1. *Top Antivir Med*. 2019; 27:111-21.
24. Lataillade M, Chiarella J and Kozal M. Natural polymorphism of the HIV-1 integrase gene and mutations associated with integrase inhibitor resistance. *ANTIVIRAL THERAPY*. Vol 11, INT MEDICAL PRESS LTD 2-4 IDOL LANE, LONDON EC3R 5DD, ENGLAND, 2006: S28-S.
25. Rossetti B, Fabbiani M, Di Carlo D, et al. Prevalence of InSTI resistance and effectiveness of InSTI-based regimens in HIV-infected patients: results from a European cohort study. *HIV MEDICINE*. Vol 20, WILEY 111 RIVER ST, HOBOKEN 07030-5774, NJ USA, 2019: 16-7.
26. Metzner KJ, Scherrer AU, Preiswerk B, et al. Origin of minority drug-resistant HIV-1 variants in primary HIV-1 infection. *The Journal of infectious diseases*. 2013; 208:1102-12.
27. Mbunkah HA, Bertagnolio S, Hamers RL, et al. Low-Abundance Drug-Resistant HIV-1 Variants in Antiretroviral Drug-Naive Individuals: A Systematic Review of Detection Methods, Prevalence, and Clinical Impact. *The Journal of infectious diseases*. 2020; 221:1584-97.
28. Li JZ, Paredes R, Ribaldo 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.
29. 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. *The Journal of infectious diseases*. 2009; 199:693-701.
30. 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 medicine*. 2008; 5:e158.
31. Cozzi-Lepri A, Noguera-Julian M, Di Giallonardo F, 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. *Journal of antimicrobial chemotherapy*. 2014; 70:930-40.
32. Organization WH. Consolidated guidelines on the use of antiretroviral drugs for treating and preventing HIV infection: recommendations for a public health approach. World Health Organization; 2016.
33. Organization WH. Guidelines on the public health response to pretreatment HIV drug resistance: July 2017. 2017.
34. Gupta S, Lataillade M, Kyriakides T, et al. Low-frequency NNRTI-resistant HIV-1 variants and relationship to mutational load in antiretroviral-naive subjects. *Viruses*. 2014; 6:3428-37.

35. Yanik EL, Napravnik S, Christopher B, et al. Prevalence of transmitted antiretroviral drug resistance differs between acutely and chronically HIV-infected patients. *Journal of acquired immune deficiency syndromes (1999)*. 2012; 61:258.
36. Cong M-e, Heneine W and García-Lerma JG. The fitness cost of mutations associated with human immunodeficiency virus type 1 drug resistance is modulated by mutational interactions. *Journal of virology*. 2007; 81:3037-41.
37. Wensing AM, van Maarseveen NM and Nijhuis M. Fifteen years of HIV Protease Inhibitors: raising the barrier to resistance. *Antiviral research*. 2010; 85:59-74.
38. Wainberg MA, Moisi D, Oliveira M, Toni TDA and Brenner BG. Transmission dynamics of the M184V drug resistance mutation in primary HIV infection. *Journal of antimicrobial chemotherapy*. 2011; 66:2346-9.
39. Castro H, Pillay D, Cane P, et al. Persistence of HIV-1 transmitted drug resistance mutations. *The Journal of infectious diseases*. 2013; 208:1459-63.
40. Little SJ, Frost SD, Wong JK, et al. Persistence of transmitted drug resistance among subjects with primary human immunodeficiency virus infection. *Journal of virology*. 2008; 82:5510-8.
41. Archer J, Baillie G, Watson SJ, Kellam P, Rambaut A and Robertson DL. 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.
42. 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.