Title: A new concept in assessing transmissibility of HIV-1 drug resistance mutations from treated and among drug naïve individuals

Short title: Transmissibility of HIV-1 drug resistance mutations

Raf WINAND [1,2], Kristof THEYS[3], Mónica EUSÉBIO[4,5*], Jan AERTS[1,2], Ricardo J. CAMACHO[3,4,6], Perpetua GOMES[4,6,7], Marc A. SUCHARD[8,9,10], Anne-Mieke VANDAMME[3,4,11], Ana B. ABECASIS[4,5] on behalf of the Portuguese HIV-1 Resistance Study Group

Funding: This project was funded through: Research Council KU Leuven: CoE PFV/10/016 SymBioSys; FWO grant G069214N and KAN 1.5.249.12; iMinds Medical Information Technologies SBO 2014, ICON b-SLIM; Collaborative HIV and Anti-HIV Drug Resistance Network (CHAIN, grant Health-F3-2009-223131, European Community's Seventh Framework Programme FP7/2007-2013); BEST HOPE: Bio-Molecular and Epidemiological Surveillance of HIV Transmitted Drug Resistance, Hepatitis Co-Infections and Ongoing Transmission Patterns in Europe (project funded through HIVERA: Harmonizing Integrating Vitalizing European Research on HIV/Aids, grant 249697); US National Science Foundation: DMS 1264153; L'Oréal Portugal Medals of Honor for Women in Science 2012 (financed through L'Oréal Portugal, Comissão Nacional da Unesco and Fundação para a Ciência e Tecnologia). ABA received a post-doc fellowship by the Fundação de Ciência e Tecnologia (SFRH/BPD/65605/2009). KT receives a FWO postdoctoral fellowship.The funding sources had no role in the writing of the manuscript or on the decision to submit this manuscript for publication.

The corresponding author had full access to all the data in the study and had final responsibility for the decision to submit for publication.

Word count: 3010

Corresponding author:

Ana Abecasis

ana.abecasis@ihmt.unl.pt

Public International Health and Biostatistics

Global Health and Tropical Medicine

Instituto de Higiene e Medicina Tropical

Universidade NOVA de Lisboa

Rua da Junqueira n.º 100 | 1349-008 Lisboa

+351 213 652 600 / ext. 269

Affiliations:

1. KU Leuven, Department of Electrical Engineering (ESAT), STADIUS Center for Dynamical Systems, Signal Processing and Data Analytics, Kasteelpark Arenberg 10 B-3001 Leuven, Belgium

2. iMinds Medical IT, Kasteelpark Arenberg 10, B-3001 Leuven, Belgium

3. KU Leuven - University of Leuven, Department of Microbiology and Immunology, Rega Institute for Medical Research, Clinical and Epidemiological Virology, B-3000 Leuven, Belgium

4. Centro de Malária e outras Doenças Tropicais, Instituto de Higiene e Medicina Tropical, Universidade Nova de Lisboa, Rua da Junqueira, nº 100, 1349-008 Lisboa, Portugal

5. Unidade de Saúde Pública Internacional e Bioestatística, Instituto de Higiene e Medicina Tropical, Universidade Nova de Lisboa, Rua da Junqueira, nº 100, 1349-008 Lisboa, Portugal

6. Laboratório de Microbiologia Clínica e Biologia Molecular, Serviço de Patologia Clínica, Hospital de Egas Moniz, Centro Hospitalar de Lisboa Ocidental, Rua da Junqueira, nº 126, 1349-009 Lisbon, Portugal

7. Centro de Investigação Interdisciplinar Egas Moniz (CiiEM), Instituto Superior de Ciências da Saúde Sul, Campus Universitário, Quinta da Granja, Monte de Caparica, 2829 - 511 Caparica, Portugal

8. Department of Biomathematics, David Geffen School of Medicine at UCLA, University of California, Los Angeles, CA 90095, USA

9. Department of Human Genetics, David Geffen School of Medicine at UCLA, University of California, Los Angeles, CA 90095, USA

10. Department of Biostatistics, UCLA Fielding School of Public Health, University of California, Los Angeles, CA 90095, USA

11. Unidade de Microbiologia, Instituto de Higiene e Medicina Tropical, Universidade Nova de Lisboa, Rua da Junqueira, nº 100, 1349-008 Lisboa, Portugal

* Now at: Portuguese Society of Rheumatology, Lisbon, Portugal

Abstract

Objectives

Surveillance drug resistance mutations (SDRMs) in drug naïve patients (DN) are typically used to survey HIV-1 transmitted drug resistance (TDR). We test here how SDRMs in patients failing treatment (TR), the original source of TDR, contribute to assessing TDR, transmissibility and transmission source of SDRMs.

Design

Retrospective observational study analyzing a Portuguese database of HIV-1 infected patients.

Methods

The prevalence of SDRMs to PIs, NRTIs and NNRTIs in DN and TR patients was measured for 3554 HIV-1 subtype B patients. Transmission ratio (prevalence in DN/prevalence in TR), average viral loads and robust linear regression with outlier detection (prevalence in DN versus in TR) were analyzed and used to interpret transmissibility.

Results

Prevalence of SDRMs in DN and TR were linearly correlated, but some SDRMs were classified as outliers – above (PRO: D30N, N88D/S, L90M, RT: G190A/S/E) or below (RT: M184I/V) expectations. The normalized regression slope was 0.073 for PI, 0.084 for NRTI, and 0.116 for NNRTI. Differences between SDRMs transmission ratios were not associated with differences in viral loads.

Conclusions

The significant linear correlation between prevalence of SDRMs in DN and in TR indicates that the later can be useful to predict levels of TDR. The slope is a

cohort-dependent estimate of rate of TDR per drug class and outlier detection reveals comparative persistence of SDRMs. Outlier SDRMs with higher transmissibility are more persistent and more likely to have been acquired from DNs. Those with lower transmissibility have faster reversion dynamics after transmission and are associated with acquisition from TR.

Key words:

HIV drug resistance

Protease Inhibitors

Reverse transcriptase inhibitors

Mutation

Transmission

Antiretroviral therapy

Introduction

Studies in Europe and US show that antiretroviral drug resistance can be found in approximately 10% of drug-naïve HIV-1 infected individuals (DN). In Europe, the prevalence of transmitted drug resistance (TDR) was 8.4% in patients newly diagnosed between 2002 and 2005 [1,2], and 7.7% between 2003 and 2009 [3].

In Sub-Saharan Africa it is up to 5.7% while in the US numbers ranging between 8% and 24.5% are reported [3–6]. Most of these studies count the number of transmitted drug resistance mutations (DRM) in newly diagnosed DN using a list of surveillance DRM (SDRM) that excludes known polymorphisms [7].

These aggregated numbers offer no further insights into transmissibility and transmission source of these SDRMs. Differences in transmissibility of wild-type and SDRM-containing viruses have been reported. For instance, one study showed that strains with SDRMs are found in DN less often than expected from the population of patients failing treatment [8]. Others deduced from mathematical models that the relative fitness of the strain and the treatment-induced reduction in viral load have an impact on transmission of SDRMs [9]. de Mendoza et al. showed some SDRMs to be transmitted more efficiently [10]. SDRMs in DN are mostly singletons, which are rare in patients failing treatment (TR). It is currently assumed this is the result of reversion of some SDRMs to wild type after transmission, since such mutations have a fitness cost for the virus in absence of drugs. In such cases, minor variants still carrying all transmitted SDRMs are to be expected. When after reversal, the major variant with fewer SDRMs is transmitted further among DNs, minor variants with more SDRMs are less likely to exist and the frequency of singletons among DNs will increase further. The distinction between acquisition of TDR from TR or from other DN is important for first line treatment strategies. Current guidelines assume that DN patients with TDR are likely to harbor undetected SDRMs as minor variants and therefore suggest avoiding low genetic barrier regimens, even if mutations against such regimens are not observed [11,12]. Evidence of onward

transmission between DN could therefore influence treatment decisions. To investigate the mode of transmission of TDR, directly from TR vs onwards between DN, most studies rely on transmission chain investigations [13–15], which are complicated and for which the methodology is still under development.

To guide strategies to avoid TDR and improve treatment strategies for DN with TDR, a deeper understanding of transmissibility and reversion dynamics of transmitted SDRMs, and the extent of acquisition of TDR from TR versus from other DN is needed. Here we describe an innovative and simple approach of correlating prevalence of SDRM in DN and TR, while using outlier detection to measure transmissibility and source of SDRMs. High or low transmissibility is defined as higher or lower prevalence in DN compared to what would be expected from the regression model. We suggest that high transmissibility is associated with persistence of SDRMs (slower reversion dynamics) and higher rates of onward transmission among DNs, and low transmissibility is associated with faster reversion dynamics and acquisition of TDR from TR. This approach does no longer require a dense sample of patients to study transmission chains of TDR since general epidemiological data for the regression analysis can be extracted from a sparser sample.

Methods

We analysed a clinical database containing HIV-1 genetic sequences sampled in Portugal between April 2001 and March 2013 in the context of antiretroviral resistance testing, either before starting treatment (DN) or at treatment failure

(TR). In order to ensure that both populations were part of the same epidemic, the analysis was limited to subtype B, as identified by Rega subtyping tool v2.0 using protease and reverse transcriptase (RT) sequences. Of a total of 3606 subtype B sequences, obtained from 3354 patients, 1685 sequences were from DN (counting only the first DN sequence per patient) and 1921 from TR (counting only the last viral sequence per patient). The median age of the patient population was 45 years (IQR = 15), 75.0% were male.

A mutation was considered a TDR mutation when it was included in the SDRM list [7]. When reporting prevalence of SDRMs, DN were patients naïve for all drug classes while TR (e.g. PI-TR) were patients with experience with the drug class of interest. Prevalence in TR is presented per drug class or, for comparison between drug classes, normalized for the complete treated population (Table 1). We used robust linear regression to estimate a regression model correlating prevalence of SDRMs in TR and DN for each drug class, allowing for outlier detection.

For each SDRM, we evaluated two measures of transmissibility: a) its transmission ratio (Table 1): prevalence in DN divided by prevalence in TR experienced with the respective drug class; b) whether or not it was an outlier with respect to the regression model. Mutations positioned above or below the regression line have, respectively, higher or lower prevalence in DN than would be expected from the prevalence in TR, and are therefore interpreted as having higher or lower transmissibility. To evaluate the effect of viral load (VL) as a

confounding factor, we correlated VL of DN, TR or DN+TR patients with specific SDRMs and the transmission ratio for those same SDRMs.

Data analyses and manipulation were performed using in-house Python scripts. For statistical analyses, SAS, R and the Rpy2 interface for Python were used [16,17]. To identify outliers in the regression model, we performed a robust linear regression assuming the prevalence in TR to be fixed and independent. Robust linear regression and outlier detection were performed using *ROBUSTREG*, as implemented in SAS.

Results

In this dataset of subtype B infected patients, the overall prevalence of TDR in DN was 10.0% [95% CI, 8.6–11.5], while 67.7% [95% CI, 65.5–69.8] of patients failing treatment carried SDRMs. Most patients with TDR showed single class resistance resulting in a prevalence of 7.4% [95% CI, 6.2–8.8], the majority (5.4%) being singletons. Double class resistance was found in 2.2% [95% CI, 1.6–3.0] and triple class resistance in 0.4% [95% CI, 0.1–0.8] of DN. In TR failing any drug class, 18.6% [95% CI, 16.9–20.4] of individuals showed single class resistance, a minority (1.9%) being singletons; double class resistance was found in 38.4% [95% CI, 36.2–40.6] and triple class in 10.7% [95% CI, 9.3–12.1]. The SDRMs in DN with the highest prevalence were for PI L90M with 1.4%, for NRTI T215Y/F+Rev (Rev = C,D,E,I,S,V) with 2.6% and for NNRTI K103N/S with 2.3%. In TR experienced with the respective drug class we find that the SDRMs with the

highest prevalence were for PI L90M with 19.7%, for NRTI M184I/V with 44.0%, for NNRTI K103N/S with 32.3%. Table 1 lists all SDRMs.

We then looked at the transmission ratio of each SDRM separately (Table 1). The highest transmission ratios were observed for PI N83D (0.165), N88D/S (0.134), and D30N (0.130); for NRTI T215Rev (0.727) (while T215Y/F transmission ratio was much lower, 0.029) and V75A/M/S/T (0.120); for NNRTI K101E/P (0.092) and G190A/E/S (0.091).

By plotting the prevalence of mutations found in DN versus in TR experienced with the respective drug class, a significant linear correlation indicated that the prevalence of SDRMs in DN can be predicted from its prevalence in TR patients belonging to the same epidemic. In Figure 1 (drug class specific) and S1 (normalized) the robust regression lines are drawn together with a 95% confidence interval (CI). The R-squared values range between 0.52 (for PIs) and 0.68 (for NNRTIs). A number of SDRMs are classified as outliers indicating that their transmissibility is either higher or lower compared to the other mutations in the same drug class. For PIs D30N, N88D/S, and L90M have a higher transmissibility; for NRTIs M184I/V has a lower transmissibility; for NNRTIs G190A/E/S has a higher transmissibility (Table S1).

To test the confounding effect of viral load (VL), we plotted the transmission ratio of specific SDRMs against the median VL of individuals carrying such SDRMs in samples taken at most 30 days before or after the resistance test. The correlation was not significant, neither when plotting the median VL in TR carrying the

mutation (Fig. S2), nor when plotting the median VL in DN carrying the mutation or both DN and TR together (data not shown). In addition we performed a Wilcoxon rank sum test on the VL of patients failing therapy for different drug classes and found no significant difference between the VL.

Discussion

We propose to capture differences in transmissibility of drug resistance mutations (DRM) by comparing the prevalence of each surveillance DRM (SDRM) in DN with its prevalence in TR. Such an analysis could convey whether an SDRM is found more often than could be expected from its prevalence in TR, thereby suggesting it may have increased its prevalence in DN by onward transmission among DN. This simple and innovative way to analyze TDR is however dependent on access to the treated population of the same epidemic, as prevalence of SDRM in TR becomes an important parameter to analyze. As far as we know, such an approach has never been suggested.

We tested this in Portuguese subtype B infected patients, representing the bulk of the native infections (thus DN and TR belong to the same epidemic) [18]. Patients were counted among TR (e.g. NNRTI-TR) only when having experience with the drug class of interest, thereby avoiding counting e.g. NNRTI SDRMs in a NNRTI naïve patient albeit experienced with other drug classes that had acquired this NNRTI SDRM through historical TDR. This approach requires knowledge of the treatment regimen. If such information is not available counting among all TR

still gives reliably results, however in our hands this was less sensitive (results not shown).

SDRM transmission ratios (prevalence in DN/prevalence in TR) were high for SDRMs with known high transmissibility (e.g. 0.727 for T215Rev, Table 1), however the precision was low for SDRMs with very low prevalence in TR. For instance, the N83D mutation has a high transmission ratio (0.165) but was only found in 1 DN and 5 TR, consequently the CI is very large. Thus, transmission ratios can only give a reliable indication of transmissibility for mutations with high prevalence in TR.

We then built a robust linear model and detected outliers. We found that the prevalence of SDRMs in DN is linearly correlated with its prevalence in TR (Fig. 1), indicating that the prevalence in TR can be used for surveillance purposes and predicts levels of TDR. Interpretation of the slope is difficult. While faster reversion dynamics will result in a reduced slope, and increased onward transmission among DN will result in a steeper slope, the slope will also vary according to the treatment strategies encountered in the cohort, such that slopes from different drug classes cannot be compared with each other or across cohorts. We therefore built an additional robust linear model normalizing the prevalence in TR to the total population failing treatment (Fig. S1). If SDRMs would not affect fitness in absence of drugs, the normalized regression slope should be similar between drug classes. The differences can be therefore interpreted as differences in reversion dynamics among SDRMs affecting different drug classes, which will consequently impact the rates of onward

transmission of such mutations. Remarkably, the NNRTI normalized slope (0.116) is higher than the PI (0.073) or NRTI (0.084) normalized slope, consistent with the higher persistence of NNRTI mutations [19]. Consequently, NNRTI SDRMs are more likely to be part of DN transmission chains.

Within a drug class, some SDRMs were found to be outliers below or above the regression line. Outlier detection also yielded significant results for some mutations with low prevalence in TR. It could be argued that the higher the viral load of a patient, the more often the virus will be transmitted, and this may confound our interpretation of transmissibility of specific SDRMs [20]. We tested this hypothesis and this was not the case, suggesting it was rather related to the intrinsic transmission efficiency and/or reversion dynamics of variants carrying those mutations.

If the prevalence in DN of an SDRM is below the regression line, this SDRM is transmitted less often and/or reverses more quickly after transmission due to a higher fitness cost in absence of the drug. In both cases it means that the transmission fitness of that mutation is lower than for other SDRMs of the same drug class. If the prevalence in DN is above the regression line, then the mutation is more frequently transmitted, has a lower rate of reversion after transmission to DN due to a lower fitness cost, and/or has a longer persistence in transmission chains of DN patients. Therefore outlier detection can inform about transmission fitness. Such an interpretation is consistent with Yang et al. [21], who stated that it is likely that SDRMs with a high fitness cost are transmitted mainly from patients failing therapy and SDRMs with a low fitness cost may more likely be

transmitted by DN patients. However some differences may also be due to actual differences in transmission frequency as was suggested by Chateau et al. who observed a possible reduced vaginal transmission efficiency for K65R [22]. Both effects together can be expressed as transmissibility.

Our results are consistent with other knowledge in the field. We show the NRTI SDRM M184I/V to be an outlier with lower transmissibility. This is consistent with Castro et al. who showed the same mutation to be less persistent after transmission to DN patients [23]. In the same study G190A was found to have a higher persistence and this mutation is here identified as an outlier with higher transmissibility. Castro et al. did not present results on all PI SDRMs, while our approach suggests the outliers D30N, N88D/S, and L90M to have high transmissibility. Although L90M shows a decreased replicative fitness, the polymorphism L63P has been shown to increase the fitness of the virus carrying L90M up to wild-type levels [24]. In our dataset we found that 82.6% of DN with L90M also carry L63P explaining the higher transmissibility of outlier L90M despite the replicative cost of the mutation in absence of therapy.

In addition to these robust outliers we find a number of additional SDRMs above or below the 95% confidence interval of the regression line. For instance, the NRTI SDRMs M41L, T215Rev and K219E/N/Q/R are above and K70E/R, L74I/V, and T215Y/F below the CI of the regression line. For the NNRTI SDRMs, K101E/P is above and Y181C/I/V below the CI of the regression line. The PI SDRMs I84V and I54A/L/M/S/T/V are below the CI of the regression line with Castro et al. showing a lower persistence for I84V and I54V. While these findings

are consistent with Castro's results, statistical significance needs to be confirmed using models built from larger datasets.

We pooled T215Y/F and T215 revertants as they can all be traced back to development of T215Y/F under NRTI selective pressure in TR. While the combined mutations can be found within the CI of the regression line, T215Rev and T215Y/F mutations are located above and below the regression line respectively. These findings support our model, as T215Y/F has a fitness cost in absence of drug and will revert rather quickly after transmission into one of these other non-wild-type SDRMs (together called T215Rev) and will show less onward transmission between DN. Any T215Y/F in a DN is therefore a testimony of a relatively recent acquisition from TR. Conversely, T215Rev located well above the regression line, is confirming that such outliers are more persistent and an indicator of more onward transmission between DNs. The higher persistence and consequent higher rates of onward transmission of T215 revertants has recently been described by Dolling et al. [25].

As far as we know, this is the first time the correlation of SDRM prevalence in TR and DN is used to make claims on transmissibility and acquisition of TDR from TR or from DN. While other studies have investigated transmission ratios, they only considered patients failing treatment as source of SDRMs [8,10]. Our approach enables the identification of SDRMs for which DNs are a likely additional source. Future efforts should include the development of a more sophisticated statistical model to address the uncertainty associated with less frequent SDRMs.

We anticipate that our population of DN is mainly chronically infected, even though we know that we have a significant proportion of acutely infected DN [18]. Both for mutations with high or low persistence, the time since infection will affect its observed prevalence in DN: the longer the time, the higher the chance that such mutation will have reverted and vice-versa. A dataset with more chronically infected patients will therefore result in a more sensitive outlier detection.

Our results support that for 1st line treatment strategies, prevalence of SDRMs in local patients failing treatment should be considered, as such SDRMs will generally be more prevalent in local DN. We deem this more realistic and more practical than only counting SDRMs in DN as is currently done because: 1) the prevalence of SDRM in TR is higher than in DN and therefore the CI is smaller; 2) many DN patients have not been diagnosed and have therefore not been sampled; 3) while most TR patients are tested for drug resistance immediately at treatment failure, in many countries most DN are either not tested or only just before starting treatment when eventual reversion already took place.

In conclusion, correlating prevalence of SDRMs in DN patients with the prevalence in patients failing treatment gives a refined view of the relationship between transmission dynamics, persistence, viral fitness, and source of TDR. In addition robust outlier detection could inform the extent of transmission of particular SDRMs among DN, which could help public health and strategic measures to avoid increase of TDR.

Acknowledgements

Raf Winand: database implementation, data analysis, figures, data interpretation, writing; Kristof Theys: data interpretation, writing; Mónica Eusébio: data analysis; Jan Aerts: data interpretation, writing; Ricardo Camacho: data collection, data curation, database implementation; Perpetua Gomes: data collection, data curation; Marc Suchard: data analysis, writing; Anne-Mieke Vandamme: study design, data interpretation, figures, writing; Ana Abecasis: study design, data analysis, data interpretation, figures, writing; Portuguese HIV-1 Resistance Study Group: data collection.

References

1 Vercauteren J, Wensing AMJ, van de Vijver D a MC, Albert J, Balotta C, Hamouda O, *et al.* Transmission of drug-resistant HIV-1 is stabilizing in Europe. *J Infect Dis* 2009; **200**:1503–8.

2 Wensing AMJ, van de Vijver D a, Angarano G, Asjö B, Balotta C, Boeri E, *et al.* Prevalence of drug-resistant HIV-1 variants in untreated individuals in Europe: implications for clinical management. *J Infect Dis* 2005; **192**:958–66.

3 Frentz D, Boucher C a B, van de Vijver D a MC. Temporal changes in the epidemiology of transmission of drug-resistant HIV-1 across the world. *AIDS Rev* 2012; **14**:17–27.

4 Bennett DE, Myatt M, Bertagnolio S, Sutherland D, Gilks CF. Recommendations for surveillance of transmitted HIV drug resistance in countries scaling up antiretroviral treatment. *Antivir Ther* 2008; **13 Suppl 2**:25–36.

5 Booth CL, Geretti AM. Prevalence and determinants of transmitted antiretroviral drug resistance in HIV-1 infection. *J Antimicrob Chemother* 2007; **59**:1047–56.

6 Stadeli KM, Richman DD. Rates of emergence of HIV drug resistance in resource-limited settings: a systematic review. *Antivir Ther* 2013; **18**:115–23.

7 Bennett DE, Camacho RJ, Otelea D, Kuritzkes DR, Fleury H, Kiuchi M, *et al.* Drug resistance mutations for surveillance of transmitted HIV-1 drug-resistance: 2009 update. *PLoS One* 2009; **4**:e4724.

8 Leigh Brown AJ, Frost SDW, Mathews WC, Dawson K, Hellmann NS, Daar ES, *et al.* Transmission fitness of drug-resistant human immunodeficiency virus and the prevalence of resistance in the antiretroviral-treated population. *J Infect Dis* 2003; **187**:683–6.

9 Blower SM, Aschenbach AN, Gershengorn HB, Kahn JO. Predicting the unpredictable: transmission of drug-resistant HIV. *Nat Med* 2001; **7**:1016–20.

10 De Mendoza C, Rodriguez C, Corral A, del Romero J, Gallego O, Soriano V. Evidence for differences in the sexual transmission efficiency of HIV strains with distinct drug resistance genotypes. *Clin Infect Dis* 2004; **39**:1231–8.

11 Vandamme A-M, Camacho RJ, Ceccherini-Silberstein F, de Luca A, Palmisano L, Paraskevis D, *et al.* European recommendations for the clinical use of HIV drug resistance testing: 2011 update. *AIDS Rev* 2011; **13**:77–108.

12 Panel on Antiretroviral Guidelines for Adults and Adolescents. Guidelines for the use of antiretroviral agents in HIV-infected adults and adolescents. 2014; **5**:79–104.

13 Yerly S, Junier T, Gayet-Ageron A, Amari EB El, von Wyl V, Günthard HF, *et al.* The impact of transmission clusters on primary drug resistance in newly diagnosed HIV-1 infection. *AIDS* 2009; **23**:1415–23.

14 Audelin AM, Gerstoft J, Obel N, Mathiesen L, Laursen A, Pedersen C, *et al.* Molecular phylogenetics of transmitted drug resistance in newly diagnosed HIV Type 1 individuals in Denmark: a nation-wide study. *AIDS Res Hum Retroviruses* 2011; **27**:1283–90.

15 Drescher SM, von Wyl V, Yang W-L, Böni J, Yerly S, Shah C, *et al.* Treatment-naive individuals are the major source of transmitted HIV-1 drug resistance in men who have sex with men in the Swiss HIV Cohort Study. *Clin Infect Dis* 2014; **58**:285–94.

16 R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria. 2013.http://www.r-project.org/ (accessed 10 Aug2014).

17 Gautier and Rpy2 contributors. Rpy2 (version 2.4). 2014.http://rpy.sourceforge.net/rpy2/doc-2�4/html/index.html (accessed 10 Aug2014).

18 Palma a C, Araújo F, Duque V, Borges F, Paixão MT, Camacho R. Molecular epidemiology and prevalence of drug resistance-associated mutations in newly diagnosed HIV-1 patients in Portugal. *Infect Genet Evol* 2007; **7**:391–8.

19 De Béthune M-P. Non-nucleoside reverse transcriptase inhibitors (NNRTIs), their discovery, development, and use in the treatment of HIV-1 infection: a review of the last 20 years (1989-2009). *Antiviral Res* 2010; **85**:75–90.

20 Quinn TC, Wawer MJ, Sewankambo N, Serwadda D, Li C, Wabwire-Mangen F, *et al.* Viral load and heterosexual transmission of human immunodeficiency virus type 1. Rakai Project Study Group. *N Engl J Med* 2000; **342**:921–9.

21 Yang W-L, Scherrer A, Yerly S, Böni J, Klimkait T, Aubert V, *et al.* Prevalence of transmitted drug resistance and relation to mean population viral load of treatment failing patients: a 16-year analysis within the Swiss HIV Cohort Study (SHCS). *Antivir Ther* 2013; **18 Suppl 1**:A21.

22 Chateau M, Swanson MD, Garcia JV. Inefficient vaginal transmission of tenofovir-resistant HIV-1. *J Virol* 2013; **87**:1274–7.

23 Castro H, Pillay D, Cane P, Asboe D, Cambiano V, Phillips A, *et al.* Persistence of HIV-1 transmitted drug resistance mutations. *J Infect Dis* 2013; **208**:1459–63.

24 Suñé C, Brennan L, Stover DR, Klimkait T. Effect of polymorphisms on the replicative capacity of protease inhibitor-resistant HIV-1 variants under drug pressure. *Clin Microbiol Infect* 2004; **10**:119–26.

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

Legends

Table 1: Overview of the identified SDRMs at each position for each drug class with prevalence and transmission ratios. Bold entries indicate an SDRM located above or below the 95% CI in the robust linear regression (Fig. 1). Variants at a single position are pooled (e.g. 46I/L). SDRM – Surveillance Drug Resistance Mutations, DN – Drug Naïve patients, TR – Patients failing treatment, PI – Protease Inhibitors, NRTI – Nucleoside Reverse Transcriptase Inhibitors, NNRTI – Non-Nucleoside Reverse Transcriptase Inhibitors, PI-TR – Patients failing a PI-containing treatment, NRTI-TR – Patients failing a NRTI-containing treatment, NNRTI-TR – Patients failing a NNRTI-containing treatment, CI – confidence interval.

Table S1: Overview of the SDRMs identified as outliers in the robust linear regression model. The Mahalanobis distance and standardized robust residual are given. Cutoff level for outliers is 3.00. PI – Protease Inhibitors, NRTI – Nucleoside Reverse Transcriptase Inhibitors, NNRTI – Non-Nucleoside Reverse Transcriptase Inhibitors.

Figure 1: Robust regression model relating prevalence of SDRMs (codon position) in TR per drug class versus in prevalence of SDRMs in DN. The linear regression line is shown together with the 95% CI, slope, intercept, and R-Square. SDRMs above or below the 95% CI are labeled with codon position, robust outliers (see Table S1) are shown as red circles. **A**. Prevalence of each PI SDRM in PI-TR (i.e. [# PI SDRM in PI-TR]/# PI-TR, see Table 1) versus

prevalence of each PI SDRM in DN i.e. [# PI SDRM in DN]/# DN; **B**. Prevalence of each NRTI SDRM in NRTI-TR versus prevalence of each NRTI SDRM in DN. T215Y/F and T215Rev are shown in green but were taken together (T215) for the linear regression; **C**. Prevalence of each NNRTI SDRM in NNRTI-TR versus prevalence of each NNRTI SDRM in DN. SDRM – Surveillance Drug Resistance Mutations, DN – Drug Naïve individuals, PI – Protease Inhibitors, NRTI – Nucleoside Reverse Transcriptase Inhibitors, NNRTI – Non-Nucleoside Reverse Transcriptase Inhibitors, PI-TR – Patients failing a PI-containing treatment, NRTI-TR – Patients failing a NRTI-containing treatment, NNRTI-TR – Patients failing a NNRTI-containing treatment, CI – Confidence interval

Figure S1: Robust regression model relating normalized prevalence of SDRMs (codon position) in TR versus prevalence of SDRMs in DN. The linear regression line is shown together with the 95% CI, slope, intercept, and R-Square. SDRMs above or below the 95% CI are labeled, robust outliers (see Table S1) are shown as red circles. **A**. Normalized prevalence of each PI SDRM in TR (i.e. [# PI SDRM in PI-TR]/# TR, see Table 1) versus prevalence of each PI SDRM in DN; **B**. Normalized prevalence of each NRTI SDRM in TR versus prevalence of each NRTI SDRM in DN. T215Y/F and T215Rev are shown in green but were taken together (T215) for the linear regression; **C**. Normalized prevalence of each NNRTI SDRM in TR versus prevalence of each NNRTI SDRM in DN. SDRM – Surveillance Drug Resistance Mutations, DN – Drug Naïve individuals, PI – Protease Inhibitors, NRTI – Nucleoside Reverse Transcriptase Inhibitors, NNRTI – Non-Nucleoside Reverse Transcriptase Inhibitors, PI-TR – Patients failing a PI-containing treatment, NRTI-TR – Patients failing a NRTI-containing treatment,

NNRTI-TR – Patients failing a NNRTI-containing treatment, CI – Confidence interval

Figure S2: Transmission ratio of each SDRM versus the log10 of the median viral load found in TR for samples taken at most 30 days before or after the resistance test. SDRM – Surveillance Drug Resistance Mutations, TR – Patients failing treatment, PI – Protease Inhibitors, NRTI – Nucleoside Reverse Transcriptase Inhibitors, NNRTI – Non-Nucleoside Reverse Transcriptase Inhibitors