HAL
open science

Drug resistance mutations in HIV: new bioinformatics approaches and challenges

Luc Blassel, Anna Zhukova, Christian J Villabona-Arenas, Katherine E Atkins, Stéphane Hué, Olivier Gascuel

To cite this version:

Luc Blassel, Anna Zhukova, Christian J Villabona-Arenas, Katherine E Atkins, Stéphane Hué, et al.. Drug resistance mutations in HIV: new bioinformatics approaches and challenges. Current Opinion in Virology, 2021, 51, pp.56 - 64. 10.1016/j.coviro.2021.09.009 . hal-03369954

HAL Id: hal-03369954

https://hal.sorbonne-universite.fr/hal-03369954

Submitted on 7 Oct 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License

Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Virology

Drug resistance mutations in HIV: new bioinformatics approaches and challenges

Luc Blassel[1,2,8], Anna Zhukova[1,3,8], Christian J Villabona-Arenas[4,5], Katherine E Atkins[4,5,6], Stéphane Hué[4,5] and Olivier Gascuel[7]

Drug resistance mutations appear in HIV under treatment pressure. Resistant variants can be transmitted to treatment-naive individuals, which can lead to rapid virological failure and can limit treatment options. Consequently, quantifying the prevalence, emergence and transmission of drug resistance is critical to effectively treating patients and to shape health policies. We review recent bioinformatics developments and in particular describe: (1) the machine learning approaches intended to predict and explain the level of resistance of HIV variants from their sequence data; (2) the phylogenetic methods used to survey the emergence and dynamics of resistant HIV transmission clusters; (3) the impact of deep sequencing in studying within-host and between-host genetic diversity of HIV variants, notably regarding minority resistant variants.

Addresses
[1] Unité Bioinformatique Evolutive, Institut Pasteur, Paris, France
[2] Sorbonne Université, Collège Doctoral, Paris, France
[3] Hub de Bioinformatique et Biostatistique, Institut Pasteur, Paris, France
[4] Centre for the Mathematical Modelling of Infectious Diseases (CMMID), London School of Hygiene & Tropical Medicine, London, UK
[5] Department of Infectious Disease Epidemiology, London School of Hygiene & Tropical Medicine, London, UK
[6] Usher Institute, University of Edinburgh, Edinburgh, UK
[7] Institut de Systématique, Evolution, Biodiversité (ISYEB, UMR 7205 - CNRS, Muséum National d'Histoire Naturelle, EPHE, SU, UA), Paris, France

Corresponding author: Gascuel, Olivier (olivier.gascuel@mnhn.fr)
[8] Co-first authors.

Current Opinion in Virology 2021, **51**:xx–yy

This review comes from a themed issue on **Virus bioinformatics**

Edited by **Alexander Gorbalenya** and **Maria Anisimova**

https://doi.org/10.1016/j.coviro.2021.09.009

1879-6257/© 2021 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Introduction

Drug resistance mutations (DRMs) arise in Human Immunodeficiency Virus-1 (HIV-1) due to antiretroviral treatment pressure, leading to viral rebound and treatment failure [1,2]. Furthermore, drug-resistant HIV variants can be transmitted to treatment-naive individuals and further spread throughout the population over time [3–5]. These transmitted drug-resistant (TDR) HIV variants limit treatment options and have clinical and public health implications worldwide. The scale of TDR varies globally; in the US and Europe, the prevalence of TDR has decreased or stabilized at between 5% and 15% [6–11]. However, in resource-limited countries, the prevalence of TDR is becoming a pressing health issue [11,12], with many regions reporting an exponential increase in prevalence and many surpassing 10% prevalence [13]. Indeed, WHO have suggested that if the prevalence of TDR exceeds 10% in a country, then first-line regimens should be reconsidered [14]. Because of this, a number of countries in Africa and Asia have revised their national treatment guidelines [12].

There are five main classes of HIV-1 antiretroviral therapies, which target different virus proteins: (i) nucleoside/nucleotide reverse transcriptase inhibitors (NRTIs), (ii) non-nucleoside reverse transcriptase inhibitor (NNRTIs), (iii) protease inhibitors (PIs), (iv) integrase inhibitors (INIs or INSTIs) and (v) entry inhibitors. The reverse transcriptase, protease and integrase proteins are encoded in the *pol* gene, while the entry inhibitors induce DRMs in the *env* gene. NRTIs and NNRTIs date back to the 80s and are currently the most commonly used drugs. PIs and INIs appeared more recently, in the mid-90s, and are still in development [15,16]. PIs and INIs are associated with lower levels of resistance compared to reverse transcriptase-based therapy. INIs are increasingly used in first-line regimens in the presence of NNRTI resistance at the population level [12]. In total, there are ~25 available drugs, all of which are associated with known DRMs. A list of DRMs is regularly updated [17•] by a consortium of international experts, who select and classify the DRMs to be surveilled (~175 in total) based on genotype analyses, phenotypic resistance tests and clinical outcome in patients on antiretroviral therapies. Primary DRMs directly confer resistance to treatments, but some mutations have an accessory role, increasing drug resistance when appearing alongside primary DRMs, while others seem to have a compensatory role and reduce the fitness cost for primary DRMs. All this, combined with the development of new antiretroviral drugs [16,18•] and the use of antiretroviral treatments in high-risk populations by pre-exposure prophylaxis [19,20], makes it particularly important to further our understanding of

HIV adaptation, detect new mutations associated with drug resistance, and survey the emergence of resistant-HIV transmission clusters in infected populations, especially in low-income countries.

For all these endeavors to advance, bioinformatics methods and large well-curated sequence databases are essential. The Stanford HIV Drug Resistance Database (https://hivdb.stanford.edu/) is the largest public repository and most widely used online resource for HIV drug resistance. It currently comprises: (1) ~450 000 sequences (reverse transcriptase, protease or integrase) from ~200 000 patients with treatment status, from all around the word; (2) ~60 000 results of drug susceptibility assays from HIV-1 virus isolates; (3) clinical outcome data from 15 clinical trials; (4) many software programs and web services to query this data. Several countries and regions have set up national databases of HIV sequences generated through routine resistance genotyping. These repositories link genotypic data with anonymized clinical and demographic information, and are regularly updated, making these national databases an attractive resource to study and monitor drug resistance. However, due to the sensitive nature of patient-derived information, the content of these national databases is non-public and only available on request. The main national/regional HIV drug resistance repositories include: (i) The UK HIV Drug Resistance Database (https://www.hivrdb.org.uk/), which is the central repository for resistance tests performed as part of routine clinical care throughout the UK since 2001. It currently comprises over 165 000 test results, most in the form of annotated *pol* gene sequences and includes over 60% of the newly diagnosed patients in the UK, with linked clinical data available for the majority of patients. (ii) The Swiss HIV Cohort Study Drug Resistance Database (http://www.shcs.ch/) that includes data and meta-data from over 80% of new diagnoses in Switzerland. (iii) The PANGEA database [21•] with data from sub-Saharan Africa, a radically different region where the pandemic started and is of great concern, which holds over 12 000 nearly complete HIV-1 genomes, with basic-to-extensive associated epidemiological metadata.

In the following, we describe the main approaches to decipher this data, and the potential of Next Generation Sequencing (NGS) to better understand and survey the emergence of DRMs and their transmission.

Machine learning approaches to study and predict resistance

The presence of DRMs before the start of an antiretroviral therapy regimen is a strong predictor of the success or failure of that regimen. Resistance testing using DNA sequencing is performed routinely in upper-income countries, and with increasing frequency in low-income and middle-income countries. To this end, computer programs are used to analyze the virus sequence of the patient (i.e. the virus genotype) and predict the level of resistance of this sequence to available drugs (i.e. the resistance phenotype of the virus). Computer programs can also be used to optimize the combination of multiple drugs [22].

The standard approach to predict the level of resistance of HIV sequences (either in the reverse transcriptase, protease or integrase proteins) is to rely on known resistance mutations to various antiretroviral therapies. HIVdb [23] uses expert rules to combine mutations (primary or accessory) observed in the studied sequences, while WebPSSM [24] uses position-specific scoring matrices. However, machine learning methods are increasingly used for this purpose, often via web services [25,26]. These methods first learn a statistical model from a set of training examples, that is, virus sequences and their resistance level measured experimentally using PhenoSense assays [27], and then assess the accuracy of the learned model using an independent set of testing examples or a cross-validation procedure. We distinguish classification methods, which predict the effectiveness of a given antiretroviral therapy [28•], and regression approaches, which predict the fold resistance ratio of the given sequence compared to the wild type [29]. Initial approaches were based on decision trees [30], support vector machines [25], logistic regression [31] and neural networks [29]. The latter showed higher accuracy (on average ~85%) than the rule-based methods used, for instance, by HIVdb (~70%) [29].

Deep learning models (i.e. neural networks with complex architectures and a large number of hidden neurons [32]) are a major focus in current machine learning research and have been successfully applied to many biological problems [33]. Moreover, recent methods make it possible to map model outputs back to subsets of the most influential input features [34]. This approach was explored by Steiner *et al.* [28•], who evaluated the performance of three deep learning architectures (multilayer perceptron, bidirectional recurrent neural network, and convolutional neural network) for drug resistance prediction using genotype-phenotype data available from HIVdb, as training and testing data (via cross-validation). The resistance to 18 antiretroviral therapies was learned from ~2100 sequences associated with PI susceptibility, ~1800 sequences associated with NNRTI susceptibility, and 2100 sequences associated with NRTI susceptibility (measured by PhenoSense assays [27], as for PI and NNRTI data). The accuracy of convolutional neural networks ranged from 86% to 96% and a large number of known DRMs were among the most influential input features. Authors suggest that other influential mutations could also be associated with resistance. These findings underscore the gain in accuracy brought by machine learning approaches, compared to rule-based methods (e.g. HIVdb). However, the main limitation is the low

number of available sequences with drug susceptibility measurement given that deep learning is commonly used with much larger data sets (>10 000 and frequently >100 000 training examples).

Another approach was explored in Ref. [35•] to study resistance patterns, epistasis and discover new DRMs using: (i) A much larger reverse transcriptase sequence dataset (~55 000 sequences) for training; (ii) A classification task to discriminate treatment-naïve from treatment-experienced sequences; (iii) Simpler machine learning models, such as random forest and logistic regression; (iv) Testing on a very different African dataset with subtypes not seen in the training data to improve robustness and to limit the impact of phylogenetic confounding factors. These choices were made with one goal in mind: interpretability, because it allows the easy extraction of mutations associated with resistance from important (influential) classifier features. To summarize, more DRMs are expected among treated patients than among naïve ones, even if we expect some DRMs to be present among naïve patients due to TDRs. To extract DRMs we can then perform tests (e.g. exact Fisher tests [36]) or use more advanced, interpretable machine learning methods [35•]. To confirm and further explore the nature of newly discovered resistance associated mutations, the training process was repeated after removing features and sequences corresponding to known DRMs (Figure 1a). This approach allowed the finding of six new potential accessory mutations. Two of these are L228H and L228R (i.e. mutations from L to H and L to R, respectively, at reference position 228 of reverse transcriptase), which are spatially very close to both the active and regulatory sites of reverse transcriptase (Figure 1b), and are overrepresented in sequences containing known DRMs (Figure 1c.1 and c.2).

Phylogenetic methods to decipher the spread of resistance

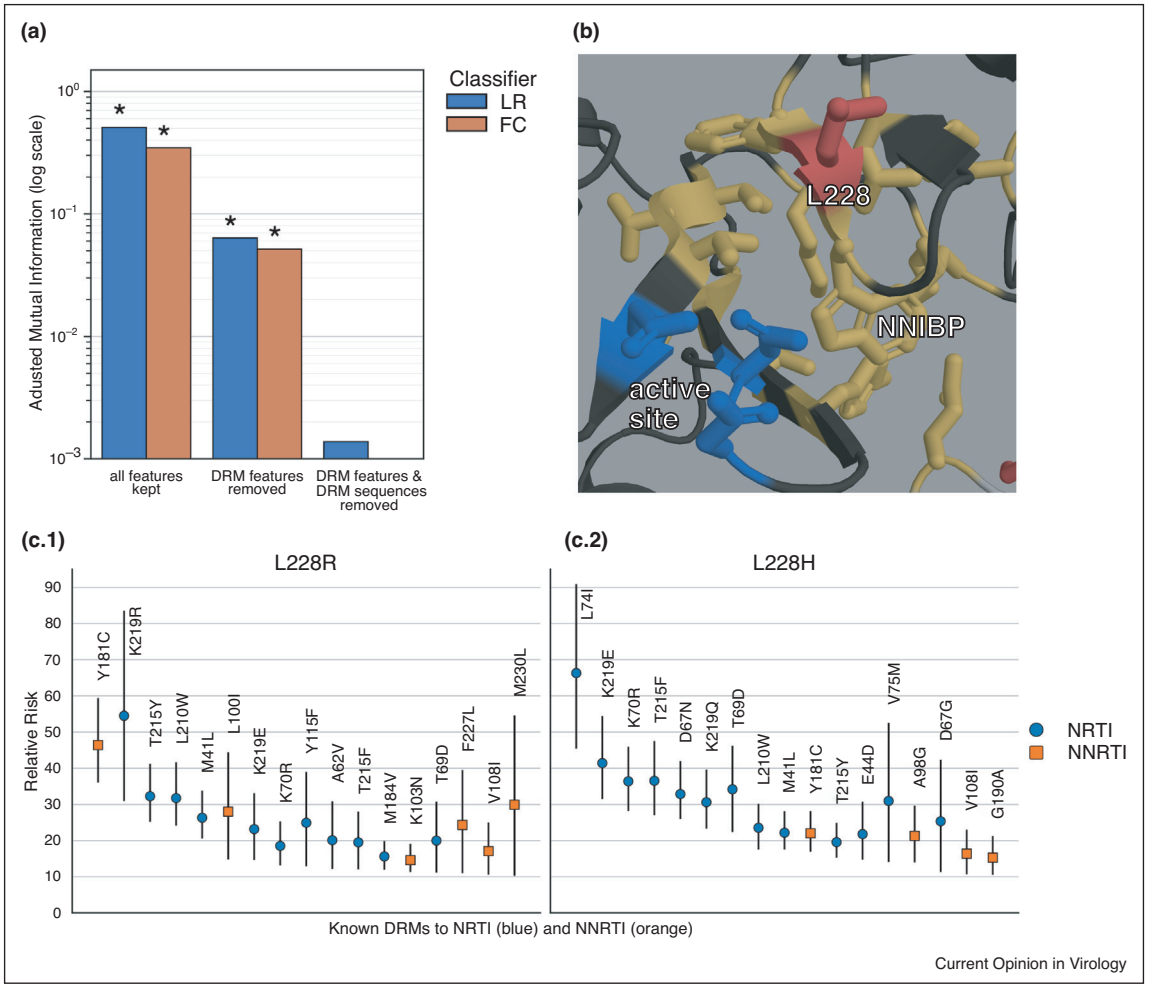
Following acquisition under treatment pressure, DRMs and resistance-associated mutations can be transmitted to treatment naïve patients. We distinguish acquired and transmitted drug resistances (TDRs). TDRs can be further separated into those corresponding to treated-to-naïve versus naïve-to-naïve transmissions. The latter are particularly problematic, as they can cause the emergence of resistance clusters in the naïve population. On the other hand, DRMs have some fitness cost and in the absence of treatment they tend to be reverted to the wild type amino acid. Some DRMs have been shown to revert rapidly (e.g. M184V in reverse transcriptase, associated to NRTIs [37,38]), while others have a low fitness cost (e.g. L90M in protease [39]) and tend to induce large resistance clusters ([40]; Figure 2).

Traditionally, for routine resistance genotyping a unique consensus virus sequence per patient is used to characterize the variants infecting a given individual, and phylogenetic trees are inferred from these consensus sequences to study the emergence, transmission and reversion of DRMs at the population level. In these trees, sequences that cluster together represent transmission clusters and each of the internal tree nodes corresponds to a transmission event. However, with one sequence per patient one cannot infer the direction of transmission, that is, distinguish the transmitter and recipient partners corresponding to a given tree node. With multiple sequences per patient, as obtained from NGS, phylogenetic methods help to infer the most likely transmission history [41]. However, reliably identifying the direction of transmissions remains challenging [42] and depends on, among other factors, the genetic diversity captured in the virus sequences of the individuals [43]. To summarize, the genetic diversity of the virus is expected to be significantly higher for the transmitter than for the recipient, but both can be similar, for example, when the infection dates are close. Moreover, one can never rule out the possibility of an intermediate, unsampled individual. Despite these limitations, phylogenetic inference has proved a promising tool for the population-level analysis of HIV resistance transmission. For example, phylogenetic tools are key in the PANGEA project [21•] to analyze the source-sink dynamics in several Sub-Sahara African settings, aiming to find generalizable characteristics of transmitters and transmission events, and guide recommendations for HIV treatment and prevention policies.

To decipher DRM transmissions, the most likely transmission clusters are extracted from the phylogeny. Genetic clusters correspond to well-supported subtrees that contain sequences closely related to each other and distant from the rest of the tree based on user-defined genetic thresholds [45]. A genetic cluster can be interpreted as representing a recent outbreak, for example, when a virus acquires a DRM and the patient starts transmitting the resistant virus. If most of the individuals in this cluster contain the same DRM, they form a resistance cluster, from which the number of within-cluster naïve-to-naïve TDRs can be estimated. This approach was used to study TDRs in Switzerland [46,47], Denmark [48], Ethiopia [49] and the USA [50].

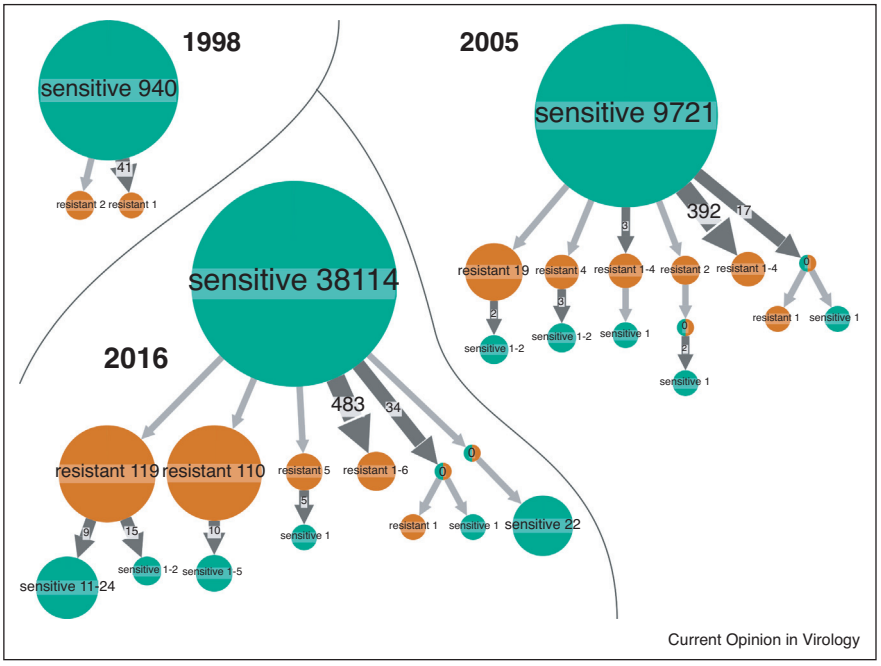
The second approach refines the previous one by using ancestral state reconstruction of a binary character describing the presence/absence of the studied DRM. Tree tips are annotated using the presence or absence of mutations, and the internal node states are inferred using parsimony [4] or maximum-likelihood [51•] methods. The clusters are defined by subsets of tips and nodes, all of which have the same resistance status and descend from a unique node corresponding to the first within-cluster transmission. Isolated, resistant tips with treatment-experienced status are interpreted as acquired drug

Figure 1



Detecting resistance associated mutations using machine learning.
This figure is adapted from Ref. [35•]. The authors used a large UK dataset (n = 55 539) of reverse transcriptase sequences from HIV patients, who have received treatment or not [44]. Sequences from this dataset were encoded as binary vectors where each feature corresponded to a specific mutation. These vectors were used to train classifiers, logistic regression (LR) and a Fisher-test-based classifier (FC), to discriminate between treatment-experienced and treatment-naïve sequences. These classifiers were then evaluated on a smaller and phylogenetically very different African dataset (n = 3990) using the adjusted mutual information (panel **(a)**). This criterion measures how well the classifiers discriminate the two types of sequences (0: random classifier; 10° = 1: perfect classifier) and can be used to compute a p-value to see if the results are truly different to ones from a random classifier (an asterisk denotes a p-value $\leq$ 0.05). In panel (a), when all features are kept, both classifiers have high (close to 1), highly significant discriminatory power. In order to check where this power comes from, the authors did the same procedure, but this time removing features corresponding to known DRMs from the encoded vectors. The adjusted mutual information is lower than when using the full vectors, but still significantly better than random. Finally, the authors repeated the procedure after removing all sequences that had at least one known DRM from the training set. This time the adjusted mutual information indicates that the classifiers are no better than random. This shows that even after removing known DRMs from the data, there remains some resistance-associated information in the sequences, which differentiates treatment-naïve and treatment-experienced sequences. Furthermore, this information seems to be in the sequences that already contained DRMs, meaning that it most likely corresponds to accessory mutations that appear alongside known DRMs. By examining the LR and FC classifiers, we can extract the most important mutations in their decision-making; L228H and L228R of reverse transcriptase are two such mutations studied in Ref. [35•]. Site 228 in panel **(b)** is positioned right between the active site (where NRTIs act) and the regulatory 'NNIBP' site (where NNRTIs act). To check the accessory nature of these mutations, the authors computed relative risk between L228R/H and known DRMs. For a given known DRM, the relative risk corresponds to the prevalence of L228R/H in sequence that have that DRM, divided by the prevalence of L228R/H in sequences that do not have that DRM. In panels **(c.1)** and **(c.2)**, relative risks for L228R and L228H are shown with their 95% confidence intervals. These relative risks show that L228R and L228H are highly overrepresented both in sequences that contain DRMs to NRTI and NNRTI. This, as well as the physical proximity of site 228 to the sites where both classes of drugs operate, point to a potential role as accessory mutations to known DRMs.

Figure 2



Emergence and transmission of resistance in protease (mutation L90M, subtype B, UK).
Ancestral state reconstruction of the presence/absence of DRM L90M over time was performed and visualized by PastML [51•] on a phylogenetic tree inferred from 39 224 UK subtype B *pol* gene sequences [44] with RAxML-NG [52] and dated with LSD2 [60]. Resistance status was detected with sierrapy [23]. A sensitive resistance status for all tree nodes and tips before 1995 (year of acceptance of Saquinavir, the first antiretroviral therapy that can provoke L90M DRM) was imposed as in Ref. [61]. Circles denote clusters of samples with the same L90M state (green when the mutation is absent, orange for resistant strains); the sample sizes of clusters are indicated in the labels, for example, the circle 'resistant 119' represents the largest resistance cluster in 2016 (119 patients). Clusters with a '0' and two colors indicate internal tree nodes for which both resistant and sensitive states had similar marginal probabilities. Arrows between two circles denote transmissions from the top to the bottom cluster (i.e. acquired drug resistances correspond to the sensitive-to-resistant transmissions, while reversions correspond to the resistant-to-sensitive ones). The size and the number on top of the arrows indicate that the arrows represent multiple transmission events leading to clusters of similar sizes (e.g. the arrow of size 483 represents 483 acquired drug resistances). Overall, we see both a large number of independent acquisitions of drug resistance (arrows from green to orange circles), and the emergence of resistance clusters (orange circles of size >1). As expected, we do not see any resistance cluster in 1998, and small ones in 2005 (≤ 19 patients). We also see a substantial amount of reversions (e.g. 9 + 15 from the largest 2016 resistance cluster).

resistances (~83% in average, in UK subtype B [4]), while in resistance clusters we mostly observe naïve-to-naïve TDRs (~70% in average, in UK subtype B [4]). Reversions correspond to non-resistant tips and clusters descending from a resistance cluster. This approach is illustrated in Figure 2, where we used maximum-likelihood [52] to build a large tree containing 39 224 subtype B sequences from the UK, and infer [51•] the resistance status of all tree nodes for the L90M protease DRM. This mutation has a low fitness cost (see above), which likely explains its high frequency and high probability of transmission between treatment-naïve individuals, resulting in large resistance clusters and low reversion rate [4].

Next-generation sequencing, resistant minority variants

Standard population-based Sanger sequencing provides the genotypes of the predominant variants in a patient, but fails to detect resistant minority variants present in less than ~20% of the total viral population [53]. By contrast, next generation sequencing (NGS)-based pipelines not only lower sequencing costs, but also enable reliable and specific detection of resistant variants accounting for ~2% of the viral population [54,55•]. NGS is thus becoming the new standard for genotypic drug resistance testing for HIV [56,57•,58•,59•].

Resistant minority variants are suspected to cause virological failures that are difficult to predict using Sanger sequencing when their frequency is below 20%. In fact, the clinical impact of resistant minority variants is not uniform across drug classes and depends on the genetic barrier to resistance to specific drugs. NNRTIs in particular have a low genetic barrier (a single DNA mutation can drastically affect drug susceptibility) and many studies [62] have shown that resistant minority variants may

adversely affect the response to NNRTIs. Moreover, there is increasing evidence showing that resistant minority variants increase the risk of treatment switches and DRM accumulation [63]. All this, combined with the fact that NGS enables the quantification of DRM frequencies (and not solely their detection, as with Sanger sequencing), led to the development of many software pipelines to extract and quantify resistant minority variants from NGS data [55•]. For example, Hivmmer [64] is an alignment and variant-calling pipeline for Illumina HIV deep sequences, based on the probabilistic aligner Hmmer [65]. While the main pipelines are able to detect and quantify DRM frequencies [55•], there is still a need for standardization and quality assurance [57•]. Moreover, to our knowledge, no tool to predict resistance to major drugs of a representative sample of variants hosted by a patient exists for NGS data.

Resistant minority variants are also suspected to play a part in the transmission of DRMs. The study of a large international cohort of naïve patients using NGS resulted in the detection of a large fraction of DRMs corresponding to minority variants, which would not have been detected by traditional Sanger sequencing [59•]. Phylogenetic analyses [58•,66] indicate that some of these rare variants likely result from transmissions. However, careful analysis of resistance clusters favors the hypothesis that most resistant minority variants found in naïve patients are likely generated *de novo* as a result of replication errors [66].

Finally, new tools specifically designed for parsing the large volume of information contained within NGS datasets have recently begun to gain traction. For example, by simultaneously analyzing within-host and between-host pathogen sequences, phyloscanner [41] provides unprecedented resolution into the transmission process, allowing inference of the direction of transmission, the identification of TDRs and the detection of multiply infected individuals. Such an approach combined with rich NGS data and metadata should be of great help in phylodynamic studies [67•].

Perspectives

HIV drug resistance surveillance is essential to track TDR trends and shape first-line regimen recommendations, especially in low-income countries where DRMs are frequent, often multiple, and tend to increase [12,14,36,68]. We are at a crossroads where NGS should occupy a major place in HIV resistance surveillance and clinical care, thanks to its decreasing costs and ability to reveal resistant minority variants and study their impact. However, adoption of NGS-based HIV resistance genotyping poses pressing challenges [56,57•], especially for low-income countries, where they are most needed [58•,69]. In particular, there is a need for standardized analyses, validated pipelines, and public large-scale databases providing not only the within-host diversity of the virus at different time points, but also rich patient metadata (e.g. treatment history). In this context, machine learning and phylogenetic approaches are expected to play a major role, as they have already done with Sanger sequencing. Moreover, the use of modeling should increase to develop and monitor first-line and second-line treatment regimens [70•], and to characterize the impact of DRMs [71]. Lastly, the analysis of transmission networks [39,72–74] should help us gain further insight in HIV drug resistance surveillance.

Funding

This work was supported by: Horizon 2020 framework program (Award number 634650, VIROGENESIS project, OG recipient, LB master salary); Agence Nationale de la Recherche (Award number ANR-19-P3IA-0001, PIA3 programme, PRAIRIE project, OG recipient, LB PhD salary); European Research Council (ERC) Starting Grant (Number 757688, awarded to Katherine E. Atkins, KEA and CJVA).

Conflict of interest statement

Nothing declared.

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- • of special interest

1. Larder BA, Kemp SD: **Multiple mutations in HIV-1 reverse transcriptase confer high-level resistance to zidovudine (AZT)**. *Science* 1989, **246**:1155-1158 http://dx.doi.org/10.1126/science.2479983.

2. Lepri AC, Sabin CA, Staszewski S, Hertogs K, Müller A, Rabenau H, Phillips AN, Miller V: **Resistance profiles in patients with viral rebound on potent antiretroviral therapy**. *J Infect Dis* 2000, **181**:1143-1147 http://dx.doi.org/10.1086/315301.

3. Hué S, Gifford RJ, Dunn D, Fernhill E, Pillay D: **Resistance on B of the UCG on HD: demonstration of sustained drug-resistant human immunodeficiency virus type 1 lineages circulating among treatment-naïve individuals**. *J Virol* 2009, **83**:2645-2654 http://dx.doi.org/10.1128/JVI.01556-08.

4. Mourad R, Chevennet F, Dunn D, Fearnhill E, Delpech V, Asboe D, Gascuel O, Hue S: **A phylotype-based analysis highlights the role of drug-naive HIV-positive individuals in the transmission of antiretroviral resistance in the UK**. *AIDS* 2015, **29**:1917-1925 http://dx.doi.org/10.1097/QAD.0000000000000768.

5. Zhukova A, Cutino-Moguel T, Gascuel O, Pillay D: **The role of phylogenetics as a tool to predict the spread of resistance**. *J Infect Dis* 2017, **216**:S820-S823 http://dx.doi.org/10.1093/infdis/jix411.

6. Novak RM, Chen L, MacArthur RD, Baxter JD, Hullsiek KH, Peng G, Xiang Y, Henely C, Schmetter B, Uy J *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**. *Clin Infect Dis* 2005, **40**:468-474 http://dx.doi.org/10.1086/427212.

7. Geretti AM: **Epidemiology of antiretroviral drug resistance in drug-naïve persons**. *Curr Opin Infect Dis* 2007, **20**:22-32 http://dx.doi.org/10.1097/QCO.0b013e328013caff.

8. Ross L, Lim ML, Liao Q, Wine B, Rodriguez AE, Weinberg W, Shaefer M: **Prevalence of antiretroviral drug resistance and**

resistance-associated mutations in antiretroviral therapy-naïve HIV-infected individuals from 40 United States cities. *HIV Clin Trials* 2007, **8**:1-8 http://dx.doi.org/10.1310/hct0801-1.

9. Wheeler WH, Ziebell RA, Zabina H, Pieniazek D, Prejean J, Bodnar UR, Mahle KC, Heneine W, Johnson JA, Hall HI *et al.*: **Prevalence of transmitted drug resistance associated mutations and HIV-1 subtypes in new HIV-1 diagnoses, U.S.–2006**. *AIDS* 2010, **24**:1203-1212 http://dx.doi.org/10.1097/QAD.0b013e3283388742.

10. Frentz D, Boucher CAB, van de Vijver DAMC: **Temporal changes in the epidemiology of transmission of drug-resistant HIV-1 across the world**. *AIDS Rev* 2012, **14**:17-27 https://pubmed.ncbi.nlm.nih.gov/22297501/.

11. Günthard HF, Calvez V, Paredes R, Pillay D, Shafer RW, Wensing AM, Jacobsen DM, Richman DD: **Human immunodeficiency virus drug resistance: 2018 recommendations of the international antiviral society–USA panel**. *Clin Infect Dis* 2019, **68**:177-187 http://dx.doi.org/10.1093/cid/ciy463.

12. World Health Organization: *HIV Drug Resistance Report 2019*. World Health Organization; 2019 http://www.who.int/hiv/pub/drugresistance/hivdr-report-2019/en/.

13. Gupta RK, Gregson J, Parkin N, Haile-Selassie H, Tanuri A, Andrade Forero L, Kaleebu P, Watera C, Aghokeng A, Mutenda N *et al.*: **HIV-1 drug resistance before initiation or re-initiation of first-line antiretroviral therapy in low-income and middle-income countries: a systematic review and meta-regression analysis**. *Lancet Infect Dis* 2018, **18**:346-355 http://dx.doi.org/10.1016/S1473-3099(17)30702-8.

14. World Health Organization: *Consolidated Guidelines on the Use of Antiretroviral Drugs for Treating and Preventing HIV Infection*. World Health Organization; 2016 http://www.who.int/hiv/pub/arv/arv-2016/en/.

15. Wensing AMJ, van Maarseveen NM, Nijhuis M: **Fifteen years of HIV Protease Inhibitors: raising the barrier to resistance**. *Antiviral Res* 2010, **85**:59-74 http://dx.doi.org/10.1016/j.antiviral.2009.10.003.

16. Trivedi J, Mahajan D, Jaffe RJ, Acharya A, Mitra D, Byrareddy SN: **Recent advances in the development of integrase inhibitors for HIV treatment**. *Curr HIV/AIDS Rep* 2020, **17**:63-75 http://dx.doi.org/10.1007/s11904-019-00480-3.

17. • Wensing AM, Calvez V, Ceccherini-Silberstein F, Charpentier C, Günthard HF, Paredes R, Shafer RW, Richman DD: **2019 update of the drug resistance mutations in HIV-1**. *Top Antivir Med* 2019, **27**:111-121 https://pubmed.ncbi.nlm.nih.gov/31634862/

The 2019 edition of the IAS-USA drug resistance mutations list and Figure. The mutations listed are those that have been identified by specific criteria for evidence and drugs described. The Figure is designed to assist practitioners in identifying key mutations associated with resistance to antiretroviral drugs. This regularly updated material is useful for anyone working on HIVDR.

18. • Tzou PL, Rhee S-Y, Descamps D, Clutter DS, Hare B, Mor O, Grude M, Parkin N, Jordan MR, Bertagnolio S *et al.*: **Integrase strand transfer inhibitor (INSTI)-resistance mutations for the surveillance of transmitted HIV-1 drug resistance**. *J Antimicrob Chemother* 2020, **75**:170-182 http://dx.doi.org/10.1093/jac/dkz417

Recent classification of the INSTI-resistance mutations for transmitted HIV-1 drug resistance (TDR) surveillance. Criteria include: presence on published expert lists, conservation in INSTI-naive persons, frequency in INSTI-treated persons and contribution to reduce *in vitro* susceptibility. Importantly, selected DRMs are non-polymorphic (i.e. commonly found in wild variants). A set of 24 DRMs is selected, as being likely to be useful for quantifying INSTI-associated TDR.

19. Grant RM, Lama JR, Anderson PL, McMahan V, Liu AY, Vargas L, Goicochea P, Casapía M, Guanira-Carranza JV, Ramirez-Cardich ME *et al.*: **Preexposure chemoprophylaxis for HIV prevention in men who have sex with men**. *N Engl J Med* 2010, **363**:2587-2599 http://dx.doi.org/10.1056/NEJMoa1011205.

20. McCormack S, Dunn DT, Desai M, Dolling DI, Gafos M, Gilson R, Sullivan AK, Clarke A, Reeves I, Schembri G *et al.*: **Pre-exposure prophylaxis to prevent the acquisition of HIV-1 infection (PROUD): effectiveness results from the pilot phase of a pragmatic open-label randomised trial**. *Lancet* 2016, **387**:53-60 http://dx.doi.org/10.1016/S0140-6736(15)00056-2.

21. • Abeler-Dörner L, Grabowski MK, Rambaut A, Pillay D, Fraser C, on behalf of the PANGEA consortium: **PANGEA-HIV 2: phylogenetics and networks for generalised epidemics in Africa**. *Curr Opin HIV AIDS* 2019, **14**:173-180 http://dx.doi.org/10.1097/COH.0000000000000542

The PANGEA (Phylogenetics and Networks for Generalized Epidemics in Africa) consortium has generated over 18 000 NGS HIV sequences from five countries in Eastern and Southern Africa: Botswana, Kenya, South Africa, Uganda and Zambia, sampled over 2014–2018. Combining phylogenetics, phylodynamics and epidemiology PANGEA-II will highlight where prevention efforts should be focused in sub-Saharan Africa to reduce the HIV epidemic most effectively. PANGEA offers accreditation to and welcomes project proposals and data contributions from external researchers who share their aims.

22. Zazzi M, Cozzi-Lepri A, Prosperi MCF: **Computer-aided optimization of combined anti-retroviral therapy for HIV: new drugs, new drug targets and drug resistance**. *Curr HIV Res* 2016, **14**:101-109.

23. Liu TF, Shafer RW: **Web resources for HIV type 1 genotypic-resistance test interpretation**. *Clin Infect Dis* 2006, **42**:1608-1618 http://dx.doi.org/10.1086/503914.

24. Jensen MA, Coetzer M, van't Wout AB, Morris L, Mullins JI: **A reliable phenotype predictor for human immunodeficiency virus type 1 subtype C based on envelope V3 sequences**. *J Virol* 2006, **80**:4698-4704 http://dx.doi.org/10.1128/JVI.80.10.4698-4704.2006.

25. Beerenwinkel N, Däumer M, Oette M, Korn K, Hoffmann D, Kaiser R, Lengauer T, Selbig J, Walter H: **Geno2pheno: estimating phenotypic drug resistance from HIV-1 genotypes**. *Nucleic Acids Res* 2003, **31**:3850-3855 http://dx.doi.org/10.1093/nar/gkg575.

26. Riemenschneider M, Hummel T, Heider D: **SHIVA - a web application for drug resistance and tropism testing in HIV**. *BMC Bioinformatics* 2016, **17**:314 http://dx.doi.org/10.1186/s12859-016-1179-2.

27. Zhang J, Rhee S-Y, Taylor J, Shafer RW: **Comparison of the precision and sensitivity of the Antivirogram and PhenoSense HIV drug susceptibility assays**. *J Acquir Immune Defic Syndr* 2005, **38**:439-444 http://dx.doi.org/10.1097/01.qai.0000147526.64863.53.

28. • Steiner MC, Gibson KM, Crandall KA: **Drug resistance prediction using deep learning techniques on HIV-1 sequence data**. *Viruses* 2020, **12**:560 http://dx.doi.org/10.3390/v12050560

This article presents recent advances in the application of deep learning to predict the resistance (phenotype) of HIV variants from their sequence (genotype). It utilizes publicly available HIV-1 sequence data and drug resistance assay results for 18 antiretroviral therapies to evaluate the performance of three architectures (multilayer perceptron, bidirectional recurrent neural network, and convolutional neural network). It identifies convolutional neural networks as the best performing architecture, with accuracy clearly superior to standard rule-based approaches, and distinguishes several sequence sites that are not known DRM locations and seem to be associated to resistance.

29. Sheik Amamuddy O, Bishop NT, Tastan Bishop Ö: **Improving fold resistance prediction of HIV-1 against protease and reverse transcriptase inhibitors using artificial neural networks**. *BMC Bioinformatics* 2017, **18**:369 http://dx.doi.org/10.1186/s12859-017-1782-x.

30. Beerenwinkel N, Schmidt B, Walter H, Kaiser R, Lengauer T, Hoffmann D, Korn K, Selbig J: **Diversity and complexity of HIV-1 drug resistance: a bioinformatics approach to predicting phenotype from genotype**. *Proc Natl Acad Sci U S A* 2002, **99**:8271-8276 http://dx.doi.org/10.1073/pnas.112177799.

31. Heider D, Senge R, Cheng W, Hüllermeier E: **Multilabel classification for exploiting cross-resistance information in HIV-1 drug resistance prediction**. *Bioinformatics* 2013, **29**:1946-1952 http://dx.doi.org/10.1093/bioinformatics/btt331.

32. Goodfellow I, Bengio Y, Courville A: *Deep Learning*. MIT Press; 2016 http://www.deeplearningbook.org.

33. Angermueller C, Pärnamaa T, Parts L, Stegle O: **Deep learning for computational biology**. *Mol Syst Biol* 2016, **12**:878 http://dx.doi.org/10.15252/msb.20156651.

34. Molnar C: **iml: an R package for interpretable machine learning**. *J Open Source Softw* 2018, **3**:786 http://dx.doi.org/10.21105/joss.00786.

35. • Blassel L, Tostevin A, Villabona-Arenas CJ, Peeters M, Hué S, Gascuel O: **Using machine learning and big data to explore the drug resistance landscape in HIV**. *PLoS Comput Biol* 2021, **17**: e1008873 http://dx.doi.org/10.1371/journal.pcbi.1008873

This article uses machine learning to discover new DRMs and study potential epistasis effects. It applies this approach to a very large UK dataset comprising ≈55 000 reverse transcriptase sequences. Results robustness is checked on different UK and African datasets. Six new mutations associated to resistance are found. All six have a low genetic barrier and show high correlations with known DRMs. Importantly, the results indicate that epistasis seems to be limited to the classical scheme where primary DRMs confer resistance and associated mutations modulate the strength of the resistance and/or compensate for the fitness cost induced by DRMs.

36. Villabona-Arenas CJ, Vidal N, Guichet E, Serrano L, Delaporte E, Gascuel O, Peeters M: **In-depth analysis of HIV-1 drug resistance mutations in HIV-infected individuals failing first-line regimens in West and Central Africa**. *AIDS* 2016, **30**:2577-2589 http://dx.doi.org/10.1097/QAD.0000000000001233.

37. Paredes R, Sagar M, Marconi VC, Hoh R, Martin JN, Parkin NT, Petropoulos CJ, Deeks SG, Kuritzkes DR: **In vivo fitness cost of the M184V mutation in multidrug-resistant human immunodeficiency virus type 1 in the absence of lamivudine**. *J Virol* 2009, **83**:2038-2043 http://dx.doi.org/10.1128/JVI.02154-08.

38. Clutter DS, Jordan MR, Bertagnolio S, Shafer RW: **HIV-1 drug resistance and resistance testing**. *Infect Genet Evol* 2016, **46**:292-307 http://dx.doi.org/10.1016/j.meegid.2016.08.031.

39. Wertheim JO, Oster AM, Johnson JA, Switzer WM, Saduvala N, Hernandez AL, Hall HI, Heneine W: **Transmission fitness of drug-resistant HIV revealed in a surveillance system transmission network**. *Virus Evol* 2017, **3** http://dx.doi.org/10.1093/ve/vex008.

40. Turner D, Amit S, Chalom S, Penn O, Pupko T, Katchman E, Matus N, Tellio H, Katzir M, Avidor B: **Emergence of an HIV-1 cluster harbouring the major protease L90M mutation among treatment-naïve patients in Tel Aviv, Israel**. *HIV Med* 2012, **13**:202-206 http://dx.doi.org/10.1111/j.1468-1293.2011.00960.x.

41. Wymant C, Hall M, Ratmann O, Bonsall D, Golubchik T, de Cesare M, Gall A, Cornelissen M, Fraser C, STOP-HCV Consortium TMPC and The BEEHIVE Collaboration: **PHYLOSCANNER: inferring transmission from within- and between-host pathogen genetic diversity**. *Mol Biol Evol* 2018, **35**:719-733 http://dx.doi.org/10.1093/molbev/msx304.

42. Volz EM, Frost SDW: **Inferring the source of transmission with phylogenetic data**. *PLoS Comput Biol* 2013, **9**:e1003397 http://dx.doi.org/10.1371/journal.pcbi.1003397.

43. Villabona-Arenas CJ, HueÂ S, Baxter J, Hall M, Lythgoe KA, Bradley J, Atkins KE: ***Using Phylogenetics to Accurately Infer HIV-1 Transmission Direction***. 2021 https://doi.org/10.1101/2021.05.12.21256968.

44. Dunn D, Pillay D: **UK HIV drug resistance database: background and recent outputs**. *J HIV Ther* 2007, **12**:97-98 In: *https://pubmed.ncbi.nlm.nih.gov/18578092/*.

45. Poon AFY: **Impacts and shortcomings of genetic clustering methods for infectious disease outbreaks**. *Virus Evol* 2016, **2** http://dx.doi.org/10.1093/ve/vew031.

46. Yerly S, Junier T, Gayet-Ageron A, Amari EBE, von Wyl V, Günthard HF, Hirschel B, Zdobnov E, Kaiser L: **Study and the SHC: the impact of transmission clusters on primary drug resistance in newly diagnosed HIV-1 infection**. *AIDS* 2009, **23**:1415-1423 http://dx.doi.org/10.1097/QAD.0b013e32832d40ad.

47. Drescher SM, von Wyl V, Yang W-L, Böni J, Yerly S, Shah C, Aubert V, Klimkait T, Taffé P, Furrer H *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**:285-294 http://dx.doi.org/10.1093/cid/cit694.

48. Audelin AM, Lohse N, Obel N, Gerstoft J, Jørgensen LB: **The incidence rate of HIV type-1 drug resistance in patients on antiretroviral therapy: a nationwide population-based Danish cohort study 1999–2005**. *Antivir Ther* 2009, **14**:995-1000 http://dx.doi.org/10.3851/IMP1412.

49. Arimide DA, Abebe A, Kebede Y, Adugna F, Tilahun T, Kassa D, Assefa Y, Balcha TT, Björkman P, Medstrand P: **HIV-genetic diversity and drug resistance transmission clusters in Gondar, Northern Ethiopia, 2003-2013**. *PLoS One* 2018, **13** http://dx.doi.org/10.1371/journal.pone.0205446.

50. Rhee S-Y, Clutter D, Fessel WJ, Klein D, Slome S, Pinsky BA, Marcus JL, Hurley L, Silverberg MJ, Kosakovsky Pond SL *et al.*: **Trends in the molecular epidemiology and genetic mechanisms of transmitted human immunodeficiency virus type 1 drug resistance in a large US clinic population**. *Clin Infect Dis* 2019, **68**:213-221 http://dx.doi.org/10.1093/cid/ciy453.

51. • Ishikawa SA, Zhukova A, Iwasaki W, Gascuel O: **A fast likelihood method to reconstruct and visualize ancestral scenarios**. *Mol Biol Evol* 2019, **36**:2069-2085 http://dx.doi.org/10.1093/molbev/msz131

PastML is a fast maximum likelihood method and tool for ancestral scenario reconstruction and visualization on phylogenetic trees. It uses decision-theory concepts to associate each node in the tree to a set of likely states. Its application to HIV1-C data set inferred many cases of independent emergence of resistance mutations under treatment pressure, and detected few resistance clusters, corresponding to transmissions among untreated patients.

52. Kozlov AM, Darriba D, Flouri T, Morel B, Stamatakis A: **RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference**. *Bioinformatics* 2019, **35**:4453-4455 http://dx.doi.org/10.1093/bioinformatics/btz305.

53. Grant RM, Kuritzkes DR, Johnson VA, Mellors JW, Sullivan JL, Swanstrom R, D'Aquila RT, Gorder MV, Holodniy M, Lloyd Robert M *et al.*: **Accuracy of the TRUGENE HIV-1 genotyping kit**. *J Clin Microbiol* 2003, **41**:1586-1593 http://dx.doi.org/10.1128/JCM.41.4.1586-1593.2003.

54. Fogel JM, Bonsall D, Cummings V, Bowden R, Golubchik T, de Cesare M, Wilson EA, Gamble T, del Rio C, Batey DS *et al.*: **Performance of a high-throughput next-generation sequencing method for analysis of HIV drug resistance and viral load**. *J Antimicrob Chemother* 2020, **75**:3510-3516 http://dx.doi.org/10.1093/jac/dkaa352.

55. • Lee ER, Parkin N, Jennings C, Brumme CJ, Enns E, Casadellà M, Howison M, Coetzer M, Avila-Rios S, Capina R *et al.*: **Performance comparison of next generation sequencing analysis pipelines for HIV-1 drug resistance testing**. *Sci Rep* 2020, **10**:1634 http://dx.doi.org/10.1038/s41598-020-58544-z

Many NGS HIVDR data analysis pipelines have been independently developed, each with variable outputs and data management protocols. This article compares the performance of five NGS HIVDR pipelines (HyDRA, MiCall, PASeq, Hivmmer and DEEPGEN) using proficiency panel samples from NIAID Virology Quality Assurance (VQA) program. All pipelines detect amino acid variants (AAV) at full range of frequencies (1–100%) and demonstrate good linearity as compared to the reference frequency values. However, their specificity dramatically decreases at AAV frequencies <2%, suggesting that 2% threshold may be a more reliable reporting threshold for ensured specificity in AAV calling and reporting. Findings from this study highlight the need for standardized strategies for NGS HIVDR data analysis.

56. Noguera-Julian M, Edgil D, Harrigan PR, Sandstrom P, Godfrey C, Paredes R: **Next-generation human immunodeficiency virus sequencing for patient management and drug resistance surveillance**. *J Infect Dis* 2017, **216**:S829-S833 http://dx.doi.org/10.1093/infdis/jix397.

57. • Ávila-Ríos S, Parkin N, Swanstrom R, Paredes R, Shafer R, Ji H, Kantor R: **Next-generation sequencing for HIV drug resistance testing: laboratory, clinical, and implementation considerations**. *Viruses* 2020, **12**:617 http://dx.doi.org/10.3390/v12060617

Several challenges still exist for the standardization and quality assurance of NGS-based HIVDR genotyping. This article highlight considerations of these challenges as related to laboratory, clinical, and implementation of NGS for HIV drug resistance testing. Several sources of variation and bias

occur in each step of the general NGS workflow, that is, starting material, sample type, PCR amplification, library preparation method, instrument and sequencing chemistry-inherent errors, and data analysis options and limitations. Additionally, adoption of NGS-based HIVDR genotyping, especially for clinical care, poses pressing challenges, especially for resource-poor settings, including infrastructure and equipment requirements and cost, logistic and supply chains, instrument service availability, personnel training, validated laboratory protocols, and standardized analysis outputs.

58. • Bonsall D, Golubchik T, de Cesare M, Limbada M, Kosloff B, MacIntyre-Cockett G, Hall M, Wymant C, Ansari MA, Abeler-Dörner L *et al.*: **A comprehensive genomics solution for HIV surveillance and clinical monitoring in low-income settings**. *J Clin Microbiol* 2020, **58** http://dx.doi.org/10.1128/JCM.00382-20

Despite decreasing costs, next-generation sequencing (NGS) is still prohibitively costly for routine use in generalized HIV epidemics in low-income and middle-income countries. This article present veSEQ-HIV, a high-throughput, cost-effective NGS sequencing method and computational pipeline tailored specifically to HIV, which can be performed using leftover blood drawn for routine CD4 cell count testing. This method overcomes several major technical challenges that have prevented HIV sequencing from being used routinely in public health efforts. The complete veSEQ-HIV pipeline provides viral load estimates and quantitative summaries of drug resistance mutations; using phyloscanner software [41], it also exploits information on within-host viral diversity to construct directed transmission networks.

59. • Baxter JD, Dunn D, Tostevin A, Marvig RL, Bennedbæk M, Cozzi-Lepri A, Sharma S, Kozal MJ, Gompels M, Pinto AN *et al.*: **Transmitted HIV-1 drug resistance in a large international cohort using next-generation sequencing: results from the Strategic Timing of Antiretroviral Treatment (START) study**. *HIV Med* 2021, **22**:360-371 http://dx.doi.org/10.1111/hiv.13038

This analysis characterized TDR in Strategic Timing of Antiretroviral Treatment (START) study participants by NGS. START enrolled ~5000 antiretroviral therapy naïve individuals in 35 countries between 2009 and 2013, out of which for ~3000 participants baseline NGS data at study entry were available. The study used statistical methods to analyze the WHO 2009 surveillance DRMS, as well as reverse transcriptase mutations T215N and E138K, and INSTI surveillance mutations from Stanford HIVdb, using three thresholds: > 2%, > 5% and >20% of the viral population. A large proportion of low-level variants was detected, which would not have been possible with traditional Sanger sequencing.

60. To T-H, Jung M, Lycett S, Gascuel O: **Fast dating using least-squares criteria and algorithms**. *Syst Biol* 2016, **65**:82-97 http://dx.doi.org/10.1093/sysbio/syv068.

61. Zhukova A, Voznica J, Felipe MD, To T-H, Pérez L, Martínez Y, Pintos Y, Méndez M, Gascuel O, Kouri V: **Cuban history of CRF19 recombinant subtype of HIV-1**. *PLoS Pathog* 2021, **17**:e1009786 http://dx.doi.org/10.1371/journal.ppat.1009786.

62. Stella-Ascariz N, Arribas JR, Paredes R, Li JZ: **The role of HIV-1 drug-resistant minority variants in treatment failure**. *J Infect Dis* 2017, **216**:S847-S850 http://dx.doi.org/10.1093/infdis/jix430.

63. Vandenhende M-A, Bellecave P, Recordon-Pinson P, Reigadas S, Bidet Y, Bruyand M, Bonnet F, Lazaro E, Neau D, Fleury H *et al.*: **Prevalence and evolution of low frequency HIV drug resistance mutations detected by ultra deep sequencing in patients experiencing first line antiretroviral therapy failure**. *PLoS One* 2014, **9**:e86771 http://dx.doi.org/10.1371/journal.pone.0086771.

64. Howison M, Coetzer M, Kantor R: **Measurement error and variant-calling in deep Illumina sequencing of HIV**. *Bioinformatics* 2019, **35**:2029-2035 http://dx.doi.org/10.1093/bioinformatics/bty919.

65. Eddy SR: **Accelerated profile HMM searches**. *PLoS Comput Biol* 2011, **7**:e1002195 http://dx.doi.org/10.1371/journal.pcbi.1002195.

66. Mbisa JL, Kirwan P, Tostevin A, Ledesma J, Bibby DF, Brown A, Myers R, Hassan AS, Murphy G, Asboe D *et al.*: **Determining the origins of human immunodeficiency virus type 1 drug-resistant minority variants in people who are recently infected using phylogenetic reconstruction**. *Clin Infect Dis* 2019, **69**:1136-1143 http://dx.doi.org/10.1093/cid/ciy1048.

67. • Ratmann O, Grabowski MK, Hall M, Golubchik T, Wymant C, Abeler-Dörner L, Bonsall D, Hoppe A, Brown AL, de Oliveira T *et al.*: **Inferring HIV-1 transmission networks and sources of epidemic spread in Africa with deep-sequence phylogenetic analysis**. *Nat Commun* 2019, **10**:1411 http://dx.doi.org/10.1038/s41467-019-09139-4

The study uses phyloscanner software [41] on deep-sequencing data from a large population-based sample of HIV-infected individuals in Rakai District, Uganda [21] to investigate the sources of epidemic spread by reconstructing partial transmission networks, and inferring the direction of transmission.

68. Magambo B, Nazziwa J, Bbosa N, Gupta RK, Kaleebu P, Parry CM: **The arrival of untreatable multidrug-resistant HIV-1 in sub-Saharan Africa**. *AIDS* 2014, **28**:1373-1374 http://dx.doi.org/10.1097/QAD.0000000000000216.

69. Inzaule SC, Tessema SK, Kebede Y, Ogwell Ouma AE, Nkengasong JN: **Genomic-informed pathogen surveillance in Africa: opportunities and challenges**. *Lancet Infect Dis* 2021, **21**:e281-e289 http://dx.doi.org/10.1016/S1473-3099(20)30939-7.

70. • Hauser A, Kusejko K, Johnson LF, Wandeler G, Riou J, Goldstein F, Egger M, Kouyos RD: **Bridging the gap between HIV epidemiology and antiretroviral resistance evolution: modelling the spread of resistance in South Africa**. *PLoS Comput Biol* 2019, **15**:e1007083 http://dx.doi.org/10.1371/journal.pcbi.1007083

The MARISA model presented in this article aims at investigating the time trends and factors driving NNRTI resistance in South Africa. MARISA is a compartmental model that includes the key aspects of the local HIV epidemic: continuum of care, disease progression, and gender. The dynamics of NNRTI resistance emergence and transmission are then added to this framework. Using this novel approach of triangulating clinical and resistance data from various sources, MARISA reproduces the time trends of HIV in South Africa in 2005–2016, with a decrease in new infections, undiagnosed individuals, and AIDS-related deaths. MARISA also captures the dynamics of the spread of NNRTI resistance: high levels of acquired drug resistance, and increasing transmitted drug resistance. MARISA shows that rapid antiretroviral therapy scale-up and inadequate switching to second-line treatment were the key drivers of the spread of NNRTI resistance in South Africa. These results highlight the need of alternative first-line regimens in this region.

71. Pečerska J, Kühnert D, Meehan CJ, Coscollá M, de Jong BC, Gagneux S, Stadler T: **Quantifying transmission fitness costs of multi-drug resistant tuberculosis**. *Epidemics* 2021, **36**:100471 http://dx.doi.org/10.1016/j.epidem.2021.100471.

72. Lewis F, Hughes GJ, Rambaut A, Pozniak A, Brown AJL: **Episodic sexual transmission of HIV revealed by molecular phylodynamics**. *PLoS Med* 2008, **5**:e50 http://dx.doi.org/10.1371/journal.pmed.0050050.

73. Pines HA, Wertheim JO, Liu L, Garfein RS, Little SJ, Karris MY: **Concurrency and HIV transmission network characteristics among men who have sex with men with recent HIV infection**. *AIDS* 2016, **30**:2875-2883 http://dx.doi.org/10.1097/QAD.0000000000001256.

74. Ragonnet-Cronin M, Hu YW, Morris SR, Sheng Z, Poortinga K, Wertheim JO: **HIV transmission networks among transgender women in Los Angeles County: network analysis of surveillance data**. *Lancet HIV* 2019, **6**:e164-e172 http://dx.doi.org/10.1016/S2352-3018(18)30359-X.