Sequence analysis

The calibrated population resistance tool: standardized genotypic

Robert J. Gifford[1,*], Tommy F. Liu[1], Soo-Yon Rhee[1], Mark Kiuchi[1], Stephane Hue[2], Deenan Pillay[2,3] and Robert W. Shafer[1]

[1]Division of Infectious Diseases, Department of Medicine, Stanford University, Stanford, CA, USA, [2]Department of Infection, University College London, London and [3]Centres for Infection, Health Protection Agency, Colindale, UK

ABSTRACT

Summary: The calibrated population resistance (CPR) tool is a web-accessible program for performing standardized genotypic estimation of transmitted HIV-1 drug resistance. The program is linked to the Stanford HIV drug resistance database and can additionally perform viral genotyping and algorithmic estimation of resistance to specific antiretroviral drugs.

Availability: http://cpr.stanford.edu/cpr/index.html

Contact: robjgiff@gmail.com

1 INTRODUCTION

Antiretroviral (ARV) therapy has greatly advanced the management of human immunodeficiency virus type 1 (HIV-1) infection. When used in combination, ARV drugs targeting the viral reverse transcriptase (RT) and protease (PR) activities can suppress HIV-1 replication to undetectable levels, leading to significant clinical benefit. However, a number of factors can lead to the emergence of drug-resistant virus strains. Once drug-resistant strains have emerged, they are 'archived' in resting white blood cells and can rapidly re-emerge if therapeutic regimens using drugs to which they are resistant are restarted. Expert panels recommend that, where possible, selection of ARV drug regimens should be guided by 'genotypic' screening in which viral drug resistance mutations (DRMs) are identified by population sequencing of the dominant HIV-1 strain in plasma. A number of software and web resources have been developed to support this procedure (Liu and Shafer, 2006).

Drug-resistant viruses selected by treatment can be transmitted, potentially compromising options for first line therapy in untreated individuals (Kuritzkes *et al.*, 2008). Surveillance of HIV-1 drug resistance (HIVDR) is therefore crucial to maintain the success of HIV-1 prevention efforts. However, variations in the methodologies used for surveillance of transmitted HIVDR, such as the specific DRMs taken as indicating transmitted resistance, have so far limited the potential to draw general conclusions from these studies. There is a widely recognized requirement for standardized protocols in this area, so that trends in HIVDR can be investigated through comparison between studies performed in distinct geographic regions and over time (Pillay, 2004; van de Vijver *et al.*, 2007).

We recently published a list of standard surveillance DRMs (SDRMs), endorsed by the World Health Organization (WHO) for epidemiological surveillance of transmitted HIVDR (Bennett *et al.*, 2008b; Shafer *et al.*, 2007, 2008). Here we describe an online program, the calibrated population resistance (CPR) tool, providing a standardized framework for estimating transmitted HIVDR from population-sampled HIV-1 PR and RT sequence sets.

2 FUNCTIONALITY

The CPR program accepts FASTA-formatted HIV-1 PR and/or RT sequence data. Options to carry out genotyping (subtyping) and to estimate genotypic resistance to specific ARV drugs are provided. A profile alignment of the submitted sequence set is created by aligning each nucleotide sequence to a polypeptide reference sequence for the region of the HIV-1 genome encoding PR and RT (by default, a subtype B consensus sequence, available from http://hivdb.stanford.edu/). Mutations, deletions and insertions (defined as changes relative to the reference sequence) are recorded for each submitted sequence. The prevalence of individual mutations is calculated by dividing mutation frequency by the number of valid codons at the corresponding position in the alignment. CPR implements a standard approach to handling contingencies such as missing data (i.e. incomplete sequences) and the nucleotide ambiguities common in HIV-1 sequence data obtained through population sequencing of viral RNA. These procedures are described in the program release notes.

A list of DRMs (by default the most recent version of the SDRM list) is used to compute the prevalence of resistance to each of the three main classes of ARV drug: protease inhibitors (PIs), nucleoside RT inhibitors (NRTIs) and non-nucleoside RT inhibitors (NNRTIs). The prevalence of transmitted HIVDR to each drug class is estimated as the number of sequences containing any DRM specific to that drug class relative to the number of times the target gene is represented in the alignment.

Analysis generates a report that summarizes the input dataset in terms of drug resistance, genetic diversity and sequence quality. The CPR report includes a graphical overview of DRMs and resistance-associated mutations present in the input dataset, and a plot showing coverage across the target region (i.e. the PR and RT genes). If the option to perform genotypic estimation of resistance is selected, resistance scores [ranging from 1 (susceptible) to 5 (highly

*To whom correspondence should be addressed.

resistant)] to specific PI, NRTI and NNRTI drugs are shown for each sequence. Once the report is generated the submitted sequences are deleted.

The HIV drug resistance database (HIVDB) is used to develop a list of PR and RT amino acid variants that have been observed at a prevalence of $\geqslant 0.1\%$ in a database containing sequences from about 20 000 individuals (Rhee *et al.*, 2003). Atypical mutations that have been reported less frequently and that are not known drug-resistant variants are highlighted in the report. APOBEC3G-mediated sequence editing [see, Holmes *et al.* (2007) for review] is detected using a subset of atypical mutations that typically occur in edited sequences (Gifford *et al.*, 2008). Other sequence quality indicators, such as stop codons and frameshifts are also identified and listed in a quality analysis section of the report.

A number of mutations have been described that are marginal with respect to their inclusion on the SDRM list, and an option is provided to highlight these 'borderline/suspicious' mutations in the report in addition to SDRMs. The CPR report shows the prevalence of individual mutations in the query dataset alongside their corresponding prevalence (stratified by subtype) in sequences from untreated patients in HIVDB. This allows investigators to rapidly identify sequence polymorphisms that are disproportionally represented in query datasets, and to discriminate between subtype-specific polymorphisms, sequence quality problems and mutational markers of prior-drug selection pressure. Mutation lists used within the program are standardized and version-tracked, as it is expected that changes may occur as new information about drug resistance and viral polymorphism becomes available.

The CPR tool is written in PERL and can readily be installed on computers running UNIX or LINUX operating systems. Alignments are constructed using LAP (Huang and Zhang, 1996). Viral subtypes are assigned using STAR (Myers *et al.*, 2005). Genotypic estimation of resistance is performed using the Stanford SIERRA web service.

3 DISCUSSION

The CPR tool aims to promote consistency between epidemiological studies by providing investigators worldwide with ready access to a simple, standard protocol for genotypic estimation of transmitted HIVDR. Because the CPR tool is closely linked to HIVDB, it allows investigators to leverage the power of large quantities of published HIV-1 sequence data within their analyses. Additionally, by standardizing protocols for genotypic estimation of transmitted HIVDR, the CPR program can facilitate comparison between sequence datasets that cannot be shared due to legal or proprietary constraints. These include datasets collated by some of the largest national and international surveillance programs (Little *et al.*, 2002; SPREAD programme, 2008; UK Collaborative Group on HIV Drug Resistance, 2007; Yerly *et al.*, 2007).

In regions of the world with minimal health infrastructure and large numbers of HIV-1 infected individuals, management of ART is necessarily based on simplified, standard treatment protocols (Bennett *et al.*, 2008a; Gilks *et al.*, 2006). The WHO has developed a minimum-resource approach for surveillance of transmitted HIVDR to accompany the expansion of access to ART in these regions, based on routine genotypic screening in a representative subset of the HIV-infected, untreated population (Bennett *et al.*, 2008b). Due to resource constraints, the number of individuals surveyed is likely to be small ($\leqslant 47$), and surveillance is likely to rely partly on archived and convenience samples (Bertagnolio *et al.*, 2007). Since the threshold for implementing changes in ART policy is low [$>5\%$ prevalence of transmitted HIVDR (Bennett *et al.*, 2008b)] it is crucial that protocols deal accurately and consistently with the contingencies of sequence analysis. The quality control measures implemented in CPR will help investigators to identify artifacts in sequence datasets collected for surveillance purposes [e.g. spurious 'drug resistance mutations' introduced by APOBEC-mediated sequence editing (Gifford *et al.*, 2008)] so that expensive and unnecessary changes in health policy may be avoided.

Although designed specifically for surveillance of HIVDR, we propose that the framework implemented in the CPR program represents a prototype for other areas of molecular epidemiology—in particular studies of microbial drug resistance—in which the primary unit of analysis is a population-sampled set of sequences.

Conflict of Interest: none declared.

REFERENCES

Bennett,D.E. *et al.* (2008a) The World Health Organization's global strategy for prevention and assessment of HIV drug resistance. *Antivir. Ther.*, **13**(Suppl. 2), 1–13.

Bennett,D.E. *et al.* (2008b) Recommendations for surveillance of transmitted HIV drug resistance in countries scaling up antiretroviral treatment. *Antivir. Ther.*, **13** (Suppl. 2), 25–36.

Bertagnolio,S. *et al.* (2007) HIV-1 drug resistance surveillance using dried whole blood spots. *Antivir. Ther.*, **12**, 107–113.

Gifford,R.J. *et al.* (2008) Sequence editing by Apolipoprotein B RNA-editing catalytic component and epidemiological surveillance of transmitted HIV-1 drug resistance. *AIDS*, **22**, 717–725.

Gilks,C.F. *et al.* (2006) The WHO public-health approach to antiretroviral treatment against HIV in resource-limited settings. *Lancet*, **368**, 505–510.

Holmes,R.K. *et al.* (2007) APOBEC-mediated viral restriction: not simply editing? *Trends Biochem. Sci.*, **32**, 118–128.

Huang,X. and Zhang,J. (1996) Methods for comparing a DNA sequence with a protein sequence. *Comput. Appl. Biosci.*, **12**, 497–506.

Kuritzkes,D.R. *et al.* (2008) Preexisting resistance to nonnucleoside reverse-transcriptase inhibitors predicts virologic failure of an efavirenz-based regimen in treatment-naive HIV-1-infected subjects. *J. Infect. Dis.*, **197**, 867–870.

Little,S.J. *et al.* (2002) Antiretroviral-drug resistance among patients recently infected with HIV. *N. Engl. J. Med.*, **347**, 385–394.

Liu,T.F. and Shafer,R.W. (2006) Web resources for HIV type 1 genotypic-resistance test interpretation. *Clin. Infect. Dis.*, **42**, 1608–1618.

Myers, *et al.* (2005) A statistical model for HIV-1 sequence classification using the subtype analyser (STAR). *Bioinformatics*, **21**, 3535–3540.

Pillay,D. (2004) Current patterns in the epidemiology of primary HIV drug resistance in North America and Europe. *Antivir. Ther.*, **9**, 695–702.

Rhee,S.Y. *et al.* (2003) Human immunodeficiency virus reverse transcriptase and protease sequence database. *Nucleic Acids Res.*, **31**, 298–303.

Shafer,R.W. *et al.* (2007) HIV-1 protease and reverse transcriptase mutations for drug resistance surveillance. *AIDS*, **21**, 215–223.

Shafer,R.W. *et al.* (2008) Consensus drug resistance mutations for epidemiological surveillance: basic principles and potential controversies. *Antivir. Ther.*, **13**(Suppl. 2), 59–68.

SPREAD Programme (2008) Transmission of drug-resistant HIV-1 in Europe remains limited to single classes. *AIDS*, **22**, 625–635.

UK Collaborative Group on HIV Drug Resistance (2007) Evidence of a decline in transmitted HIV-1 drug resistance in the United Kingdom. *AIDS*, **21**, 1035–1039.

van de Vijver,D.A.M.C *et al.* (2007) The epidemiology of transmission of drug resistant HIV-1. In Leitner,T. *et al.* (eds) *HIV Sequence Compendium 2007*. Theoretical Biology and Biophysics Group, Los Alamos National Laboratory, Los Alamos, NM.

Yerly,S. *et al.* (2007) Transmission of HIV-1 drug resistance in Switzerland: a 10-year molecular epidemiology survey. *AIDS*, **21**, 2223–2229.