Journal of Clinical Virology 73 (2015) 95–100

Contents lists available at ScienceDirect

Journal of Clinical Virology

journal homepage: www.elsevier.com/locate/jcv

Identification of minority resistance mutations in the HIV-1 *integrase* coding region using next generation sequencing

Jannik Fonager[a,*], Jonas T. Larsson[b], Christian Hussing[a], Frederik Neess Engsig[c], Claus Nielsen[a], Thea Kølsen Fischer[a]

[a] Department of Microbiological Diagnostics and Virology, Statens Serum Institut, Copenhagen, Denmark
[b] Department of Microbiology and Infection Control, Statens Serum Institut, Copenhagen, Denmark
[c] Department of Infectious Diseases, Copenhagen University Hospital, Hvidovre, Denmark

ARTICLE INFO

Article history:
Received 18 August 2015
Received in revised form 31 October 2015
Accepted 3 November 2015

Keywords:
Hiv-1
Integrase
Illumina MiSeq
Deep sequencing
Minority variants
Resistance

ABSTRACT

Background: The current widely applied standard method to screen for HIV-1 genotypic resistance is based on Sanger population sequencing (Sseq), which does not allow for the identification of minority variants (MVs) below the limit of detection for the Sseq-method in patients receiving integrase strand-transfer inhibitors (INSTI). Next generation sequencing (NGS) has facilitated the detection of MVs at a much deeper level than Sseq.
Objectives: Here, we compared Illumina MiSeq and Sseq approaches to evaluate the detection of MVs involved in resistance to the three commonly used INSTI: raltegravir (RAL), elvitegravir (EVG) and dolutegravir (DTG).
Study design: NGS and Sseq were used to analyze RT-PCR products of the HIV-1 *integrase* coding region from six patients and in serial samples from two patients. NGS sequences were assembled and analyzed using the low frequency variant detection (LFVDT) tool in CLC genomic workbench.
Results: Sseq detected INSTI resistance and accessory mutations in three of the patients (called INSTI Res+), while no resistance or accessory mutations were detected in the remaining three patients (called INSTI Res-). Additional INSTI resistance and/or accessory mutations were detected by NGS analysis of integrase sequences from all three INSTI Res+ and one INSTI Res- patient.
Conclusion: Our observations suggested that NGS demonstrated a higher sensitivity than sSEQ in the identification of INSTI relevant MVs both in patients at treatment baseline and in patients receiving INSTI therapy. Thus NGS can be a valuable tool in monitoring of antiretroviral minority resistance in patients receiving INSTI therapy.

1. Background

Combination antiretroviral therapy against infection with human immunodeficiency virus (HIV) type 1 is highly efficient in the suppression of HIV-1 replication within infected hosts and can significantly delay or prevent both HIV-1 transmission [1] and progression to AIDS [2]. Nevertheless, transmitted and acquired resistance mutations remain a global problem [3–5], despite decreased occurrence of resistance in some settings [6]. Genotypic resistance testing based on Sseq is primarily performed to identify the most suitable treatment regime. However, Sseq is unable to reliably detect resistance-encoding MVs below ~20% [7–10]. Some studies have found that the presence of resistance MVs could retrospectively predict the clinical outcome [11,12]; however, the impact of detection of MVs on the chosen treatment strategies for HIV-1 infected patients needs further investigation. NGS technology has facilitated the detection of mutants occurring at extremely low levels in HIV-1 infected patients for both *Protease*-, *Reverse transcriptase*- and *Integrase*-genes [13–17]. However, no consensus currently exists as to which NGS technological platform or methodology should be implemented in laboratories performing genotypic resistance testing. It was recently shown that the Illumina HiSeq 2000 platform had a superior ability to detect MVs and also had low false positive rates compared to the 454 platform [14], commonly used in HIV MV studies [18–21]. Other methods, such as allele specific PCR have also been used to detect MVs [22] and produce concordant results when compared with deep sequencing

* Corresponding author at: Department of Microbiological Diagnostics and Virology, Statens Serum Institut, Artillerivej 5, 2300S Copenhagen, Denmark.
E-mail address: FON@ssi.dk (J. Fonager).

http://dx.doi.org/10.1016/j.jcv.2015.11.009
1386-6532/

techniques [23]. However, these methods are sensitive to unknown variations in the primer-binding site and require updating for each newly discovered mutation.

INSTI's block the viral enzyme integrase and INSTI's have shown to be highly efficient in suppressing HIV infection and are recommended as an option in initial treatment of treatment naïve HIV-1 infected patients and as recommended by the European AIDS Clinical Society [22–24]. In Denmark, INSTI's were not in use before 2006. From 2007 to 2011, the use of INSTI was mainly restricted to RAL, initially as part of a salvage regime in patients with virological failure and triple class resistance but since 2009 also on treatment-naïve HIV patients in combination with NRTI's [25].

A low genetic barrier exists for especially RAL and EVG [26–28] and several cross-resistance mutations can confer simultaneous resistance to two or more INSTI [29,30]. Since these resistance mutations might be present as MVs, it is important to investigate how frequently they occur both at baseline and during treatment in order to assess their clinical significance.

2. Objectives

Currently, few studies have investigated resistance mutations in the *integrase* coding region using the NGS Illumina methodology [14,31]. Here, the Illumina MiSeq platform and the CLC Genomic workbench were used to sequence and analyze the occurrence of integrase minority mutations compared with traditional Sseq. The main objective was to evaluate if the combination of NGS and CLC analysis could be used to detect INSTI resistance relevant MVs not detected by traditional Sseq.

3. Study design

3.1. Patient samples

Six HIV-1 infected patients (designated 1-HIV to 6-HIV) were included in this study. Four patients (1-HIV, 2-HIV, 3-HIV and 6-HIV) had been infected with HIV-1 for a long time (median: 18 years), while 2 patients (4-HIV and 5-HIV) were infected for a shorter period (median: 5 years). 1-HIV, 2-HIV and 3-HIV (collectively called INSTI Res+) had detectable INSTI resistance mutations by Sseq, while no resistance mutations were detected in sequences from 4-HIV, 5-HIV and 6-HIV (collectively called INSTI Res-). Three historical samples taken from the 2-HIV patient 5–7 years earlier (2-HIV_Hist1, 2-HIV_Hist2 and 2-HIV_Hist3) and one taken from the 3-HIV patient 8 years earlier (3-HIV_Hist1) were also included and analyzed in duplicate (designated by a or b in sample name). The samples had the following reported HIV-1 copies/mL (provided by the departments of infectious diseases in charge of antiviral treatment): 1-HIV: 4200, 2-HIV_Hist1: 6580, 2-HIV_Hist2: 8140, 2-HIV_Hist3: 4710, 2-HIV: 1700, 3-HIV_Hist1:96000, 3-HIV: 6100, 4-HIV: 24221, 5-HIV: 393824, 6-HIV: 4000000.

3.2. Nucleic acid extraction

Nucleic acids were extracted from plasma samples by use of MagNa Pure LC (Roche Diagnostics) and the Viral NA Small Volume Kit (Roche Diagnostics) according to the manufacturer's instructions.

3.3. RT-PCR and Sanger sequence analysis

Amplification, sequencing and analysis of the *integrase* coding region are described in details in Supplementary Fig. SF1.

3.4. Illumina MiSeq sequencing

Sequencing libraries were prepared from RT-PCR products using Illuminaís Nextera XT kit according to the manufacturer's instructions. DNA concentrations in each of the six libraries were measured using Qubit dsDNA BR and ssDNA or the Qubit DNA HS assay kit. Libraries were pair-end sequenced with the MiSeq reagent kit v2, 2×250. The run was successful and after filtering, yielded 96,323,694–297,757,424 nt sequence data from each sample.

3.5. Illumina Miseq deep sequencing analysis

Thirty-seven representative reference sequences were obtained from Los Alamos [32] and imported into CLC. FASTQ files were imported into the CLC Workbench 7.5.1 using the NCBI/Sanger & Illumina 1.8 and later pipeline and failed reads were removed. Mapping was performed to both full-length genome and the *integrase* coding region (amino acid 1–288; nucleotide position 4230–5093) on the appropriate reference sequence. Consensus sequences with MVs were generated from reference assembled reads (using a noise cut off at 0.01 and detecting variants occurring at >1% and covered by at least 10 reads).

Mapping data generated in this study have been deposited in the European Nucleotide Archive under the study accession number PRJEB9552 (http://www.ebi.ac.uk/ena/data/view/PRJEB9552).

3.6. Genotypic resistance profile and subtype analysis

Sequence subtypes, mutation profiles and resistance levels were analyzed by submission of sequences to Stanford HIVDB [33].

3.7. Identification of minority variants

The performance LFVDT was evaluated on an Illumina generated control library [14] using a noise cut off at 0.01 detecting variants occurring at both $\geq$0.1% and $\geq$1% covered by at least 10 reads and using the quality score option. Variants in patient samples were identified $\geq$0.1% for the historical and $\geq$1% for the non-historical samples. The tracks generated by the LFVDT were further analyzed by filtering the marginal variant calls (CLC settings: minimum average base quality = 30, minimum frequency = 1% and minimum forward/reverse balance = 0.05). Variant read percentages were calculated automatically by LFVDT. Both Single nucleotide variants (SNVs) and multiple nucleotide variants (MNVs) were included in the evaluation of minority resistance; however MNVs were excluded from other comparisons. Reference variants were removed after manual inspection.

3.8. Statistical analysis

Statistical analysis was performed in GraphPad Prism version 5.01. Fischerís exact test was used due to limited samples size and categorical variables.

4. Results

4.1. Run characteristics

Run characteristics and PHRED score at or above 30 ($\geq$99.9% accuracy) is shown in Table 1.

4.2. Mapping of reads

Reads from each sample were initially mapped to the *integrase* coding region of all the 37 subtype reference sequences to identify to which reference sequence most reads mapped. After identifying