Journal of Virological Methods 205 (2014) 7–16

Contents lists available at ScienceDirect

Journal of Virological Methods

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

Targeted deep sequencing of HIV-1 using the IonTorrentPGM platform

Gustavo H. Kijak[a,*], Eric Sanders-Buell[a], Elizabeth A. Harbolick[a], Phuc Pham[a], Agnes L. Chenine[a], Leigh Anne Eller[a], Kathleen Rono[b], Merlin L. Robb[a], Nelson L. Michael[c], Jerome H. Kim[c], Sodsai Tovanabutra[a]

[a] U.S. Military HIV Research Program, Henry M. Jackson Foundation for the Advancement of Military Medicine, Bethesda, MD, USA
[b] Kenya Medical Research Institute/Walter Reed Project, Kericho, Kenya
[c] U.S. Military HIV Research Program, Walter Reed Army Institute of Research, Silver Spring, MD, USA

Article history:
Received 9 January 2014
Received in revised form 11 April 2014
Accepted 25 April 2014
Available online 4 May 2014

Keywords:
Targeted deep sequencing
Next-generation sequencing
HIV-1
Molecular evolution
IonTorrent

A B S T R A C T

The characterization of mixed HIV-1 populations is a key question in clinical and basic research settings. This can be achieved through targeted deep sequencing (TDS), where next-generation sequencing is used to examine in depth a sub-genomic region of interest. This study explores the suitability of IonTorrent PGM(LifeTechnologies) for the TDS-based analysis of HIV-1 evolution. Using laboratory reagents and primary specimens sampled at pre-peak viremia the error rates from misincorporation and in vitro recombination were <0.5%. The sequencing error rate was 2- to 3-fold higher in/around homopolymeric tracts, and could be discerned from true polymorphism using bidirectional sequencing. The limit of detection of complex variants was further lowered by using haplotyping. The application of this system was illustrated on primary samples from an individual infected with HIV-1 followed from pre-peak viremia through six months post-acquisition. TDS provided an augmented view of the extent of genetic diversity, the covariation among polymorphisms, the evolutionary pathways, and the boundaries of the mutational space explored by the viral swarm. Based on its performance, the system can be applied for the characterization of minor viral variants in support of studies of viral evolution, which can inform the rational design of the next generation of vaccines and therapeutics.

1. Introduction

As of the end of 2012, more than 35 million people were living worldwide with HIV/AIDS, and hundreds of millions were at risk of acquiring the infection (UNAIDS, 2013). Despite colossal research efforts in the past three decades, the quest for a safe and effective preventive vaccine and a cure for the disease are still ongoing and constitute main public health priorities.

The extensive genetic variation of the virus, which results from the combination of high rates of mutation, recombination and replication, is responsible for the fast emergence of viral strains that can evade host immune responses and resist the effects of antiretroviral agents (Coffin and Swanstrom, 2013). During most of the course of the infection, viral populations within an individual are heterogeneous, with varying levels of genetic diversity (Salazar-Gonzalez et al., 2009). Being able to characterize mixed viral populations is a key question both in the clinical and in the basic research settings. For instance, the study of the ontogeny of viral escape from the host adaptive immune responses, such as antibodies and cytotoxic T lymphocytes (CTLs), can support the rational design of preventive vaccine (Goonetilleke et al., 2009). Similarly, the detection of viral quasispecies that are resistant to antiretroviral drugs can inform the selection of regimens with better prospects of success; the risk of treatment failure may be significantly increased even when the drug-resistant viruses are present at 1% (Li et al., 2011).

Traditionally, viral populations have been sampled using cloning in bacterial systems or single genome amplification (SGA) (Palmer et al., 2005), followed by Sanger sequencing; these methods tend to be costly, tedious, low-throughput, and are not prone to extensive automation (McCutchan et al., 2005; Shankarappa et al., 1999). Furthermore, power calculations show that the detection of viral subpopulations carrying multiple mutations and present

Abbreviations: CAT, chloramphenicol acetyltransferase; CTL, cytotoxic T lymphocyte; DEPC, diethylpyrocarbonate; ePCR, emulsion polymerase chain reaction; NAT, nucleic acid testing; GS, next-generation sequencing; PASS, parallel allele-specific sequencing; pVL, plasma viral load; RT, reverse transcriptase; SGA, single genome amplification; SNP, single nucleotide polymorphism; T/F, transmitted/founder; TDS, targeted deep sequencing; vRNA, viral RNA.

* Corresponding author. Tel.: +1 301 319 9746.
E-mail address: gkijak@hivresearch.org (G.H. Kijak).

http://dx.doi.org/10.1016/j.jviromet.2014.04.017
0166-0934/

at frequencies below 5% requires acquiring >100 sequences (Keele et al., 2008), which renders this approach impractical and cost-prohibitive for routine detection of such minor viral variants. Real-time PCR-based methods represent cost-effective methods for the sensitive and selective detection of single nucleotide polymorphisms (SNPs) at frequencies <1% (Wei et al., 2011) but they are not designed to capture linkage among multiple polymorphisms. The parallel allele-specific sequencing (PASS) assay can detect low-frequency variants and can determine linkage among polymorphisms (Cai et al., 2007) but its use is not widespread because no commercial kit is available.

The advent of next-generation sequencing (NGS) technologies brings the promise of simplicity, increased throughput, low cost, and automation, which can translate into affordable higher sequencing depth. At the same time, the error rate associated with NGS is recognized to be higher than that of Sanger-based sequencing (Chang, 2009). When applied to HIV-1, the measured noise of NGS results from the composite of: (1) reverse transcription (when using viral RNA (vRNA) as initial template), (2) PCR for library preparation, (3) emulsion PCR (ePCR) or cluster generation, and (4) sequencing. The introduction of noise in the first two processes can be limited using high-fidelity enzymes (Di Giallonardo et al., 2013; Larsen et al., 2013), whereas the end-user has little capacity to influence the conditions of processes 3 and 4. In the study of HIV-1, the capacity to accurately discern true genetic signals from background noise is paramount; especially considering that SNPs can drastically impact susceptibility to antiretroviral drugs (Richman et al., 1994) or immune effectors (Leslie et al., 2004).

The main experimental approaches of HIV-1 NGS have been whole genome sequencing, whole gene sequencing, and targeted deep sequencing (TDS). The first two approaches involve the amplification, fragmentation, sequencing, and alignment of partially overlapping reads. TDS examines a defined sub-genomic region of interest at great sampling (i.e., number of input genomes) and sequencing (i.e., number of reads) depth to determine the frequency of the different variants. One of the main advantages of TDS compared to the other approaches derives from the fact that the capacity to obtain longer reads has increased substantially over the recent years. As in TDS the obtained reads are, for the most part, fully overlapping, it is now possible to directly determine, rather than just infer (Zagordi et al., 2011), the linkage among measured polymorphisms (i.e., haplotyping). This is particularly important in HIV-1, where the presence of multiple mutations in the same genome can result in viruses with increased levels of antiretroviral drug resistance (Larder and Kemp, 1989)or can compensate for the loss of fitness that may have resulted from particular CTL-escape mutations (Brockman et al., 2007). Additionally, as PCR efficiency is inversely associated with amplicon length (Piyamongkol et al., 2003), targeting shorter regions results in more efficient library preparation, larger pools of sequenceable genomes, and lower chances of allelic dropout. Targeting shorter regions also results in lower in vitro recombination rates, and this is particularly important in HIV-1, where linkage can be created or disrupted by recombination events occurring during HIV-1 reverse transcription (Clavel et al., 1989), but it can also be an in vitro artifact generated in NGS library preparation (Shao et al., 2013).

The vast majority of the HIV-1 TDS work published to date has been produced using the 454 pyrosequencing platform(Armenia et al., 2012; Di Giallonardo et al., 2013; Fischer et al., 2010; Hedskog et al., 2010, 2013; Messiaen et al., 2012; Mild et al., 2011; Mukherjee et al., 2011; Poon et al., 2010; Redd et al., 2011, 2012; Rozera et al., 2009; Shao et al., 2013; Simen et al., 2009; Swenson et al., 2010, 2011; Tsibris et al., 2009; Vandenbroucke et al., 2010) which was the first commercial NGS system, and there is also one report using the Pacific Bioscience platform (Archer et al., 2012). The use of the Illumina and Ion Torrent PGM systems for HIV-1 genotyping has been reported for whole-gene sequencing (Archer et al., 2012; Chang et al., 2013). Most recently, Gibson et al. (2014) have reported on DEEPGEN™ HIV, a genotyping platform that applies Ion Torrent PGM whole-gene sequencing and TDS, but is tailored toward the clinical management of patients. While other platforms offer higher yields, at this time, the IonTorrent PGM system offers one of the lowest costs per experiment (http://www.molecularecologist.com/next-gen-table-2b/), making this a preferred choice for small laboratories. The aim of the current study was to determine the suitability of TDS implemented in the IonTorrent PGM platform for the analysis of HIV-1 evolution. Using a combination of laboratory reagents and primary clinical specimens the system performed with a lower limit of detection of 0.5% for single nucleotide variants, and with an even lower limit for more complex haplotypes. Based on these specifications, the system can be applied for the characterization of minor viral variants in support of studies of viral evolution, as illustrated here for the analysis of an HIV-1 individual infected followed longitudinally from pre-peak viremia through six months post-infection.

2. Materials and methods

2.1. Samples under study

Plasmid pBCKS(+) plasmid (Agilent Technologies, Santa Clara, CA), encoding the chloramphenicol acetyltransferase (CAT) gene, was expanded in Stbl2 bacterial system (Invitrogen/LifeTechnologies, Carlsbad, CA) in the presence of chloramphenicol, and was extracted using the PureYield Plasmid Maxiprep System (Promega, Madison, WI), according to manufacturer's instructions. The concentration of the extracted plasmid DNA was determined by OD260 spectrophotometry. Three samples from a female commercial sex worker from Kenya enrolled in the RV217/ECHO cohort (henceforth referred to as "participant 20225") were used to assess the assay's performance on primary specimens. The individual was enrolled before HIV-1 infection. HIV-1 acquisition was documented by HIV-1 nucleic acid testing (NAT)(APTIMA HIV-1 RNA Qualitative Assay, GenProbe, San Diego, CA), plasma viral load (pVL) (Abbott RealTime HIV-1, Abbott Laboratories. Abbott Park, IL, lower limit of detection: 40 copies/ml) and serology(Robb et al., manuscript in preparation). Specimens were sampled at day 5 (pVL = 17,782,794 copies/ml), day 36(pVL = 131,826 copies/ml), and day 178 (pVL = 17783) (considering day-4 as the date of the last NAT negative determination, day 0 as the date of first positive HIV-1 NAT, and day 9 as peak viremia) (Supplemental Fig. 1). vRNA was extracted using the QIAamp Viral RNA Mini Kit (QIAGEN,Valencia, CA) (see Supplemental Table 1 for utilized plasma volumes).

2.2. Library preparation

2.2.1. Reverse transcription

For HIV-1 subgenomic regions gp120-V2 and gp120-C3, vRNA was retrotranscribed using the SuperScript III First Strand Synthesis System for RT-PCR (Invitrogen/LifeTechnologies, Carlsbad, CA). All primer sequences are shown in Supplemental Table 2 Master mix 1 contained 2 mM dNTPs, 5 μM of outer reverse primer, extracted vRNA, and diethylpyrocarbonate (DEPC)-treated water in a final volume of 10 μl. The mix was incubated at 65 °C for 5 min then immediately transferred to ice for 4 min. Master mix 2 (10× RT buffer, 0.02 M DTT, 10 mM MgCl2, RnaseOUT, and SuperScript III RT in a final volume of 10 μl) was then added on top of master mix 1. The thermocycler routine was: 50 °C for 3 h, 85 °C for 5 min. Then 1 μl of *Escherichia coli* Rnase H was added and incubated at 37 °C for 20 min.