Published in final edited form as:

J Virol Methods. 2013 April ; 189(1): 232–234. doi:10.1016/j.jviromet.2013.01.019.

Rapid deep sequencing of patient-derived HIV with ion semiconductor technology

Max W. Chang^a, Glenn Oliveira^b, Jinyun Yuan^{a,*}, Jason F. Okulicz^{c,d}, Samuel Levy^{b,†}, and Bruce E. Torbett^{a,#}

^aDepartment of Molecular and Experimental Medicine, The Scripps Research Institute, La Jolla, CA

^bScripps Genomic Medicine, Scripps Translational Science Institute, La Jolla, CA

^cInfectious Disease Clinical Research Program, Uniformed Services University of the Health Sciences, Bethesda, MD

^dInfectious Disease Service, San Antonio Military Medical Center, San Antonio TX

Abstract

The development of next-generation sequencing technologies has facilitated the study of HIV drug resistance evolution. However, the high capacity and per-run cost of many sequencers is not ideal for viral sequencing unless many samples are analyzed simultaneously. Ion semiconductor sequencing has recently emerged as a flexible, lower-cost alternative with short runtime. This paper describes the use of Ion Torrent devices for deep sequencing of drug resistant HIV samples. High levels of sequencing coverage were obtained in HIV Gag and protease, allowing the detection of mutations at low frequencies.

Keywords

HIV-1 drug resistance; deep sequencing; HIV protease; HIV Gag; Ion Torrent

The treatment of HIV/AIDS has made tremendous strides during the last 25 years, with the development of drugs effective against multiple viral targets (De Clercq, 2009). However, the frequent evolution of drug resistance has been a significant factor in the continued persistence of the HIV/AIDS epidemic (Johnson et al., 2011; Kutilek et al., 2003; Shafer et al., 2007). Continued study of drug resistance mechanisms allows for more informed treatment strategies for clinicians and has also led to advances in drug design.

The rapid replication and high mutation rate of HIV ensure that the viral population remains highly heterogeneous, with numerous mutations present at low frequencies. Monitoring the proliferation of various mutations can be difficult employing traditional capillary electrophoresis involving Sanger chemistry, which has difficulty detecting variants occurring with less than 35% frequency (Palmer et al., 2005). The development of next-

© 2013 Elsevier B.V. All rights reserved.

#betorbet@scripps.edu.

*Present address: Department of Molecular Microbiology and Immunology, St. Louis University School of Medicine, St. Louis, MO

†Present address: Genomic Health, Inc., Redwood City, CA

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

generation sequencing technologies, though, has allowed the measurement of mutations occurring with 1% frequency or less (Wang et al., 2007), as dictated by the depth of sequence redundancy obtained.

Consequently, studies utilizing deep sequencing have become more common, predominantly involving the Roche 454 FLX platform. Although the Roche 454 platform generates relatively long reads, its per-base and per-run costs are relatively high among next-generation sequencing platforms that are employed commonly (Glenn, 2011). High capacity sequencers from Illumina, such as the HiSeq 1000, have a low per-base cost, but long runtimes and high per-run costs make these instruments feasible only when a very large number of viral samples are multiplexed. Beginning in 2010, smaller-scale benchtop instruments have been released (Loman et al., 2012), which are more suitable for viral sequencing. Among these was the Ion Torrent PGM (Life Technologies, Guilford, CT, USA), a novel sequencing system employing ion semiconductor technology. It is characterized by a low per-run cost which affords a high level of flexibility in planning experiments, and its ability to measure low frequency mutations in drug resistant HIV samples was evaluated.

Specifically, Ion Torrent devices were used to analyze longitudinal samples from a single HIV-infected patient from the U.S. Military HIV Natural History Study whose treatment with various regimens had failed to maintain long-term suppression of viral replication below the limit of detection (50-400 copies/mL, depending on the time of testing). The patient's initial treatment included a non-nucleoside reverse transcriptase inhibitor (NNRTI)-based regimen of efavirenz paired with nucleoside reverse transcriptase inhibitors (NRTIs) zidovudine/lamivudine. Subsequent regimens were protease-inhibitor based, the first containing lopinavir/ritonavir paired with the NRTIs tenofovir/lamivudine that was later changed to atazanavir/ritonavir combined with didanosine/lamivudine. Viral samples from multiple time points were sequenced: two serum samples before protease inhibitor treatment, two samples during lopinavir/ritonavir treatment, and one sample during atazanavir/ritonavir treatment. Analysis of these samples with the Ion Torrent PGM and existing software tools was able to generate high sequencing coverage and identify relevant drug resistance mutations in HIV Gag and protease.

Patient serum samples were provided in 1 mL aliquots, and the viral particles present were concentrated by centrifugation at $18,000 \times g$ for 1 hour. After removing 860 μL of supernatant, the remainder was processed using a viral RNA extraction kit (QIAamp Viral RNA Mini, Qiagen, Valencia, CA, USA). The resulting RNA eluate was used immediately, without freezing, as a template for RT-PCR (OneStep RT-PCR, Qiagen). Each 50 μL RT-PCR reaction contained 15 μL of the RNA solution. Two ~1 kb amplicons spanning Gag and protease were amplified with 40 PCR cycles. Primer design was based on conserved regions of the HIV-1 genome (sequences shown in Table 2). Successful amplification of the targets was verified by agarose gel electrophoresis.

The amplified cDNA was prepared for sequencing using the Ion Fragment Library kit (Life Technologies, Foster City, CA, USA). Specifically, the two amplicons were pooled in equimolar amounts, then 1 μg of this mixture was sheared to an average size of 175 bp via mechanical shearing (S2 instrument, Covaris, Woburn, MA, USA). Sequence-specific adapters were ligated to the inserts and the resulting fragments were separated electrophoretically and size-selected at the 200 bp mark. The fragments were then amplified using six cycles of PCR. After validating the libraries using the Bioanalyzer (Agilent, Santa Clara, CA, USA) and Qubit system (Life Technologies), the libraries were amplified clonally and affixed to Ion Spheres using the Ion Torrent OneTouch template kit (Life Technologies). The Ion Spheres from each sample were loaded onto their own 314 chip. The

five 314 chips were run separately on the Personal Genome Machine (Life Technologies) to yield an average of roughly 13 Mb (Q20) bases per sample.

The resulting sequencing reads were mapped to an HIV consensus B Gag-Pol reference sequence (Kuiken et al., 2010) using ReadClean454 (RC454) (Henn et al., 2012). RC454, written originally for 454 sequencing reads, performs read mapping using MOSAIK and carries out additional correction, such as identifying homopolymer indels. As the Ion Torrent platform has comparable difficulty to 454 in sequencing homopolymeric regions, similar analysis and error correction techniques are applicable. Alternatively, BWA and the native Ion Suite mapper TMAP were capable of mapping the reads, but did not readily fit into an analysis pipeline suitable for pooled viral sequencing. RC454 and its companion programs V-Phaser and V-Profiler (described below), on the other hand, were written specifically to analyze viral populations. Other programs for this purpose include Segminator II (Archer et al., 2012) and ShoRAH (Zagordi et al., 2011), which may be useful for those requiring a graphical user interface or haplotype reconstruction, respectively.

The pattern of mapped reads indicated a high level of sequencing coverage for all five samples, with 13,700-fold coverage averaged over Gag and protease (Figure 1, Table 1). The sample with the lowest level of coverage showed an average depth of 5,300 and a minimum depth of 368. In multiple studies involving 454 pyrosequencing, this level of coverage was sufficient to identify mutations present in less than 1% of the population (Henn et al., 2012; Zagordi et al., 2010). Isolated dips in sequencing coverage were related to homopolymeric regions. A drop in coverage near position 500, for example, occurs in a sequence of five consecutive guanine nucleotides. Decreased coverage near the amplicon ends and in the overlap between amplicons resulted from the use of amine-modified primers which mitigate over-sampling in these regions (Harismendy and Frazer, 2009).

Following the read mapping, further analyses were able to detect relevant mutations in protease and Gag at low frequencies. This variant calling and additional correction were carried out using V-Phaser and V-Profiler (Macalalad et al., 2012). Overall, primary and accessory mutations in protease tended to be fixed or nearly absent, while certain resistance-associated mutations in Gag were found at low to moderate frequencies (Figure 2). In protease, marked changes in the viral population were observed only after switching to an atazanavir-containing regimen. The most notable event was the fixation of the N88S drug resistance mutation, which was accompanied by the L10I accessory mutation. The Y132F mutation in the MA-CA Gag cleavage site may have provided compensation for the N88S protease mutation (Fun et al., 2012). Mutations in the p1-p6 cleavage site, I437V and S440F, were also found, although only at relatively low frequency. These changes, as well as the resistance-associated G62R mutation in Gag, would have likely been missed using traditional Sanger sequencing unless a large number of clones were analyzed.

As sequencing technology continues to improve, massively parallel sequencing strategies are likely to become the primary method for the genetic analysis of viral populations. This study focused on a handful of samples taken from a single patient, but scaling up to a larger number of samples is relatively straightforward using higher capacity chips and barcoded multiplexed samples. Benchtop alternatives to the Ion Torrent include the Illumina MiSeq, which does not suffer from homopolymer errors but has a longer run time and less flexibility in run cost (Loman et al., 2012). The 454 GS Junior can generate long reads, though its overall throughput is relatively low. While the choice of platform involves certain trade-offs, the continued adoption of benchtop sequencing instruments will facilitate the study of viral evolution.

Acknowledgments

This work was supported by NIH/NCRR UL1 RR025774, NIH P01 GM083658, and 1P50 GM103368. MWC was supported by NIH T32AI007354. Patient specimens were obtained from the U.S. Military HIV Natural History Study (IDCRP-000-03) as part of the Infectious Disease Clinical Research Program (IDCRP) funded in part by the National Institute of Allergy and Infectious Diseases, NIH under Inter-Agency Agreement Y1-AI-5072. This is manuscript 21964 from The Scripps Research Institute.

References

- Archer J, Baillie G, Watson SJ, Kellam P, Rambaut A, Robertson DL. Analysis of high-depth sequence data for studying viral diversity: a comparison of next generation sequencing platforms using Segminator II. *BMC Bioinformatics*. 2012; 13:47. [PubMed: 22443413]
- De Clercq E. Anti-HIV drugs: 25 compounds approved within 25 years after the discovery of HIV. *Int. J. Antimicrob. Agents*. 2009; 33:307–20. [PubMed: 19108994]
- Fun A, Wensing AM, Verheyen J, Nijhuis M. Human Immunodeficiency Virus gag and protease: partners in resistance. *Retrovirology*. 2012; 9:63. [PubMed: 22867298]
- Glenn TC. Field guide to next-generation DNA sequencers. *Mol. Ecol. Resour.* 2011; 11:759–69. [PubMed: 21592312]
- Harismendy O, Frazer K. Method for improving sequence coverage uniformity of targeted genomic intervals amplified by LR-PCR using Illumina GA sequencing-by-synthesis technology. *BioTechniques*. 2009; 46:229–31. [PubMed: 19317667]
- Henn MR, Boutwell CL, Charlebois P, Lennon NJ, Power KA, Macalalad AR, Berlin AM, Malboeuf CM, Ryan EM, Gnerre S, Zody MC, Erlich RL, Green LM, Berical A, Wang Y, Casali M, Streeck H, Bloom AK, Dudek T, Tully D, Newman R, Axten KL, Gladden AD, Battis L, Kemper M, Zeng Q, Shea TP, Gujja S, Zedlack C, Gasser O, Brander C, Hess C, Gunthard HF, Brumme ZL, Brumme CJ, Bazner S, Rychert J, Tinsley JP, Mayer KH, Rosenberg E, Pereyra F, Levin JZ, Young SK, Jessen H, Altfeld M, Birren BW, Walker BD, Allen TM. Whole genome deep sequencing of HIV-1 reveals the impact of early minor variants upon immune recognition during acute infection. *PLoS Path.* 2012; 8:e1002529.
- Johnson VA, Calvez V, Gunthard HF, Paredes R, Pillay D, Shafer R, Wensing AM, Richman DD. 2011 update of the drug resistance mutations in HIV-1. *Top. Antivir. Med.* 2011; 19:156–64. [PubMed: 22156218]
- Kuiken, C.; Foley, B.; Leitner, T.; Apetrei, C.; Hahn, B.; Mizrahi, I.; Mullins, J.; Rambaut, A.; Wolinsky, S.; Korber, B., editors. *HIV Sequence Compendium 2010*. Theoretical Biology and Biophysics Group, Los Alamos National Laboratory; New Mexico: 2010.
- Kutilek VD, Sheeter DA, Elder JH, Torbett BE. Is resistance futile? *Curr. Drug. Targets Infect. Disord.* 2003; 3:295–309. [PubMed: 14754431]
- Loman NJ, Misra RV, Dallman TJ, Constantinidou C, Gharbia SE, Wain J, Pallen MJ. Performance comparison of benchtop high-throughput sequencing platforms. *Nat. Biotechnol.* 2012; 30:434–9. [PubMed: 22522955]
- Macalalad AR, Zody MC, Charlebois P, Lennon NJ, Newman RM, Malboeuf CM, Ryan EM, Boutwell CL, Power KA, Brackney DE, Pesko KN, Levin JZ, Ebel GD, Allen TM, Birren BW, Henn MR. Highly sensitive and specific detection of rare variants in mixed viral populations from massively parallel sequence data. *PLoS Comp. Biol.* 2012; 8:e1002417.
- Palmer S, Kearney M, Maldarelli F, Halvas EK, Bixby CJ, Bazmi H, Rock D, Falloon J, Davey RT Jr, Dewar RL, Metcalf JA, Hammer S, Mellors JW, Coffin JM. Multiple, linked human immunodeficiency virus type 1 drug resistance mutations in treatment-experienced patients are missed by standard genotype analysis. *J. Clin. Microbiol.* 2005; 43:406–13. [PubMed: 15635002]
- Shafer RW, Rhee SY, Pillay D, Miller V, Sandstrom P, Schapiro JM, Kuritzkes DR, Bennett D. HIV-1 protease and reverse transcriptase mutations for drug resistance surveillance. *AIDS*. 2007; 21:215–23. [PubMed: 17197813]
- Wang C, Mitsuya Y, Gharizadeh B, Ronaghi M, Shafer RW. Characterization of mutation spectra with ultra-deep pyrosequencing: application to HIV-1 drug resistance. *Genome Res.* 2007; 17:1195–201. [PubMed: 17600086]

- Zagordi O, Bhattacharya A, Eriksson N, Beerenwinkel N. ShoRAH: estimating the genetic diversity of a mixed sample from next-generation sequencing data. *BMC Bioinformatics*. 2011; 12:119. [PubMed: 21521499]
- Zagordi O, Klein R, Daumer M, Beerenwinkel N. Error correction of next-generation sequencing data and reliable estimation of HIV quasispecies. *Nucleic Acids Res*. 2010; 38:7400–9. [PubMed: 20671025]

Figure 1.

High level of sequencing coverage throughout Gag and protease. Intermittent dips in sequencing coverage correspond to homopolymeric repeats. The top section indicates the regions of the HIV genome covered by the RT-PCR amplicons.

Figure 2.

Evolution of selected mutations observed with deep sequencing. The timeline (A) indicates serum sampling points and protease inhibitor regimens. The upper graph (B) shows mutations in the Gag polypeptide: G62R, Y132F, and I437V are associated with drug resistance, while S126K is a likely polymorphism not associated with resistance. Mutations in protease are shown in the bottom graph (C): L10I, A71V, and N88S are primary or accessory resistance mutations, and N37S is a polymorphism not associated with resistance.

Table 1

Viral load of patient samples with summary of sequencing results

Sample Date	Viral load (cp/mL)	Total reads generated	Average read length	Average coverage	Reads aligned (%)
3/1/2004	>750,000	239,386	131	14,680	91.0
8/30/2004	>100,000	167,577	124	10,168	93.1
3/14/2005	21,000	82,848	128	5,264	94.5
9/26/2005	>750,000	435,850	124	27,965	95.9
5/7/2007	241,000	162,977	129	10,559	94.7

Primer Sequence	Orientation	Position (HXB2 reference)
GCGACTGGTGAGTACGCC	Sense	737-754
GGACCAACAAGGTTTCTGTCATCC	Antisense	1759-1736
TCCACCTATCCCAGTAGGAGAA	Sense	1548-1569
TTTGGGCCATCCATTCCTGG	Antisense	2608-2589