Accuracy testing of deep sequencing for HIV-1 drug resistance testing in the QCMD 2015 ENVA HIV drug resistance typing

Pornpimon Nimitsuntiwong[1], Chorthip Wathiphaba[1], Ekawat Pasomsab[1], Wasun Chantratita[1,2]*

[1]Virology Laboratory, Department of Pathology, Ramathibodi Hospital, Mahidol University, Bangkok 10400, Thailand
[2]Center of Medical Genomics, Faculty of Medicine, Ramathibodi Hospital, Mahidol University, Bangkok 10400, Thailand

**Corresponding author: wasun.cha@mahidol.ac.th*

ABSTRACT

Sanger sequencing which is a gold standard method for genotypic drug resistance testing has limited sensitivity in detecting HIV-1 drug resistance mutations (DRMs) at frequencies below 20% of viral quasispecies. Deep sequencing is an ultrasensitive method, which allows detecting such mutations and mutations detected by Sanger sequencing. A newly developed HIV-1 deep sequencing drug resistance assay has never been tested the accuracy with external quality assessment (EQA) program. The objective of this study was to test the accuracy of deep sequencing in detecting DRMs in HIV-1 protease (PR) and reverse transcriptase (RT) genes with five samples from ENVA 2015 HIV drug resistance typing EQA program. According to the 2015 EQA program report, deep sequencing could generate complete datasets of all five ENVA15 panel samples, which covered both PR and RT genes and comprised all IAS codons. In addition, deep sequencing detected 339 of 340 DRM codons and was awarded 99.86% of the overall sequence concordance. A manual reviewed bam file of ENVA15-08 sample was performed to investigate the incorrect codon and found at position 2,381 (PR-43) had a mixture of A (wild-type) at 83% and G (mutant type) at 16% which was identical to the expected results (R). Therefore, we reanalyzed the sequences of all samples by another pipeline found deep sequencing detected 334 DRM codons identical to expected results which comprised the incorrect codon and 333 DRM codons. In addition, the other 6 codons which comprised mutants at frequencies below 20% of viral quasispecies were partial concordance with the expected results because 2015 ENVA consensus sequences were created by aligning sequences summited by all participants whose almost all datasets were based on Sanger sequencing technology. In conclusion, deep sequencing has accuracy in detecting HIV-1 DRMs and would be adopted as a clinical laboratory routine in the near future.

Keywords: HIV drug resistance; NGS technologies; deep sequencing; QCMD ENVA HIV drug resistance

INTRODUCTION

Combination antiretroviral therapy (cART) plays significant roles in reducing AID-related morbidity and mortality in HIV-1 infected patients as well as preventing HIV-1 transmission (Rodger *et al.*, 2013; Cohen *et al.*, 2016). However, antiretroviral resistance remains one of major obstacles to effective antiretroviral therapy of HIV-1 infected patients, which impairs the response to therapy and causes virological failure (Paredes *et al.*, 2010). Therefore, HIV-1 drug resistance testing is considered as one of important ART monitoring tests for HIV patient management (Günthard *et al.*, 2014) because this test provides drug resistance mutation (DRM) profile for clinicians using the information to select and construct an antiretroviral drug (ARV) regimen which is more likely to achieve and maintain viral suppression. Nowadays, HIV treatment guidelines have recommended that drug resistance testing should be performed in treatment-experienced patients with virological failure, as well as in treatment-naïve patients for baseline resistance testing and prior to ART initiation (Vandamme *et al.*, 2011; Günthard *et al.*, 2014). A gold standard method for HIV-1 DRM detection based on Sanger sequencing or first-generation sequencing is sequencing bulk RT-PCR products of HIV-1 protease (PR) and reverse transcriptase (RT) genes by dideoxynucleotide triphosphates (ddNTPs). These products are detected by capillary electrophoresis followed by analyzing sequence reads with computer software to generate a consensus sequence. Sanger sequencing such as FDA-approved commercial kits and in-house assays has been widely used in a clinical setting. However, this method has limited sensitivity in detecting low-frequency nucleotide variants at frequencies

below 20% of viral quasispecies because each nucleotide position of the consensus sequence presents predominant nucleotide variant generally at frequencies above 20% (Palmer *et al.*, 2005; Church *et al.*, 2006; Halvas *et al.*, 2006). On the contrary, Next-Generation sequencing (NGS) or deep sequencing also referred to second generation sequencing technologies namely pyrosequencing (454/Roche), reversible dye termination (Illumina) and semiconductor sequencing (Ion Torrent) is quite different from the Sanger sequencing in term of a massively parallel approach, an ultrasensitive method, a high-throughput and reduced cost (Liu *et al.*, 2012). NGS performs the thousands of millions sequencing reactions of amplified DNA fragments on the solid phase simultaneously and collects the sequencing data in the same time. The generated sequence reads from deep sequencing represent nucleotide variants at frequencies above 20% and those below 20% of viral quasispecies. Deep sequencing therefore has revolutionized HIV-1 drug resistance testing. Previous studies employed deep sequencing to evaluate and to compare results with Sanger sequencing showed that deep sequencing detected all DRM detected by Sanger sequencing and low-frequency DRMs were undetectable by Sanger sequencing (Stelzl *et al.*, 2011; Avidor *et al.*, 2013; Garcia-Diaz *et al.*, 2013; Gibson *et al.*, 2014). Furthermore, deep sequencing used to study the clinical significances of low-frequency DRMs in treatment-naïve patients demonstrated that pre-existing low-frequency NNRTI mutations were associated with the increased risk of virological failure (Simen *et al.*, 2009; Li *et al.*, 2011). A newly developed HIV-1 deep sequencing drug resistance assay based on the semiconductor sequencing technology has never been tested the accuracy with the EQA program. The aim of the study was to test accuracy of this assay in detecting DRMs in HIV-1 PR and reverse RT genes with ENVA panel from the Quality Control for Molecular Diagnostics (QCMD) ENVA 2015 HIV drug resistance typing EQA program.

MATERIALS AND METHODS

Samples

The ENVA15 panel consisted of five plasma samples based on characterized and cultured HIV virions diluted in HIV, HBV and HCV negative plasma in the Quality Control for Molecular Diagnostics (QCMD) ENVA 2015 HIV drug resistance typing EQA program.

A newly developed HIV-1 deep sequencing drug resistance assay

Sentosa® SQ HIV-1 Genotyping Assay (Vela Diagnostics Pte. Ltd.) can perform simultaneously 15 clinical samples per run which are paralleled with an HIV System Control (HIV-SC) serving as a positive control and no template control in order to control the entire workflow and ensure the absence of contamination in the workflow, respectively.

First, HIV-1 RNA was automatically extracted from 730 µl of each ENVA15 sample by a robotic liquid handling system based on the emotion 5075 system (Eppendorf, Germany). Second, master mix of reverse transcriptase polymerase chain reaction (RT-PCR), the extracted RNA of ENVA samples and HIV-SC were prepared in a reaction plate by the robotic liquid handling system. Third, RT-PCR amplification was performed by the Veriti™ Dx 96-well Thermal Cycler. Fourth, library preparation was performed automatically by the robotic liquid handling system which the amplified PR and RT amplicons were sheared to approximately 200 bp DNA fragments and then were normalized and ligated with adaptors followed by pooling together the individual libraries in 1.5 mL DNA LoBind® tube. Fifth, the pooled DNA libraries were used to perform emulsion PCR and enrichment of template-positive IPSs by using Ion Torrent-based deep sequencing. Sixth, template-positive IPSs were loaded on the 318 chip and sequenced by Ion torrent PGM based on semiconductor sequencing. After a 5-h run of sequencing, the generated raw sequence reads were used to perform primary analysis by using the Sentosa® SQ Suite software and then were analyzed by using bioinformatics tools (in-house pipeline) to generate nucleotide sequences which were send to QCMD.

QCMD ENVA 2015 HIV drug resistance typing EQA program

The 2015 ENVA consensus sequences

The sequences summited by all participants in the program were aligned to generate the 2015 ENVA consensus sequence and gave the sequences for each codon. The consensus sequences for each codon were observed in >60 % of the sequences (the majority results), which was used for comparing the results of participants. The 2015 ENVA consensus sequences were in substantial agreement with the consensus sequences calculated from the independent testing results.

ENVA scoring

According to the scoring system as shown in Table 1, the sequences of participants were compared to the consensus sequences for each codon. If the sequence of participants was identical to the codon genotype of the consensus sequence, they gained 1 point. On the other hand, if those was not matched to the consensus sequence, they gained 0 points. For codons containing a mixture of 2 or 3 nucleotides at a certain position, they received 1 point if the mixture was reported or if the correct mutation was reported.

With IAS-USA drug resistance table 2014, each sample comprised the number of codons in the PR gene associated with resistance to protease inhibitors for 36 codons as well as those in the RT gene associated with resistance to reverse transcriptase inhibitors for 32 codons. Therefore, in each sample the maximum performance score could be achieved was 68 points. An overall maximum score achievable is 340 points.

Table 1 The ENVA scoring.

Expected result / Participants	Wild Type	Mutant Type	Mixed Type
Wild Type	1	0	0
Mutant Type	0	1	1
Mixed Type	0	1	1

RESULTS

The newly developed HIV-1 deep sequencing drug resistance assay generated an average read length of 166 bp and an average target region coverage of PR more than 1,000X similar to an average target region coverage of RT genes. Deep sequencing was successful in generating sequences of all five samples which belonged to HIV-1 subtype B, C and D and covered both HIV-1 PR and RT genes.

In 2015 EQA program report, the number of nucleotide sequence datasets submitted were 94 datasets which were obtained from 38 commercial techniques and 56 in-house techniques from 93 respondents out of 107 participants from 39 countries. Almost all of the datasets were based on Sanger sequencing technology, but there were 3 datasets based on Next-Generation Sequencing (NGS) categorized into commercial techniques. Of the total of 94 datasets, 65 (69.1%) were complete datasets (covering both the PR and RT genes) of all five samples. The ENVA15-08 sample has the number of complete datasets (68/94) lower than other samples because 26 datasets were not submitted as sequences covering both genes or were submitted as sequence covering only one gene. This assay, one of three NGS in this EQA program, was also able to generate complete datasets of all samples, and each dataset contained 68 DRM codons according to IAS-USA drug resistance table 2014. Deep sequencing was given the performance score for 339 out of 340 points (99.71%) (Table 2). In other words, deep sequencing could detect 339 DRM codons identical to the expected results which were composed of 3 mixture codon genotypes, 95 mutant codon genotypes or 241 wild type codon genotypes in both PR and RT genes. An incorrect codon belonged to the ENVA15-08 sample. Furthermore, the overall sequence concordance percentage (at nucleotide level) was 99.86% of the consensus sequences for all five ENVA15 panel samples.

DISCUSSION

HIV-1 drug resistance information plays a critical role in selecting and constructing optimal antiretroviral regimens for HIV-1 infected individuals. Recently, deep sequencing was used for HIV-1 drug resistance testing in a clinical laboratory (Avidor *et al.*, 2013; Garcia-Diaz *et al.*, 2013). The newly developed HIV-1 deep sequencing drug resistance assay has never been tested the accuracy with the EQA program.

The newly developed HIV-1 deep sequencing drug resistance assay was successful in generating sequences belonged to HIV-1 subtype B, C, and D of all five ENVA15 panel samples. Although the ENVA15 panel did not cover HIV-1 subtype which is predominant or typically found in Thailand (CRF01-AE), this assay was also successful in generating sequences belonged to HIV-1 subtype CRF01-AE by evaluating this assay with samples previously genotyped with Sanger sequencing (data not shown). Non-B subtypes have virologic and immunologic responses to antiretroviral drugs similar to those of subtype B (Geretti *et al.*,2009; Touloumi *et al.*,2013).

Nevertheless, there were partial differences of resistance mutations among subtypes. For example, reverse transcriptase inhibitor resistance mutations. K65R emerges more frequently and more quickly in subtype C viruses than in subtype B, as well as the V106M has been commonly found in subtype C and CRF01_AE but not in subtype B (Martinez-Cajas *et al.*, 2009).

Table 2 Results of accuracy testing in detecting DRMs in HIV-1 PR and RT genes of ENVA15 panel samples by deep sequencing.

Sample	HIV-1 subtype	Viral load (Log10 Copies/ml)	Performance score/maximum score
ENVA15-07	C	4.37	68/68 (100%)
ENVA15-08	C	4.17	67/68 (98.50%)
ENVA15-09	C	3.53	68/68 (100%)
ENVA15-10	B	4.46	68/68 (100%)
ENVA15-11	D	4.90	68/68 (100%)
Overall			339/340 (99.71%)

Abbreviations: PR, protease, RT, reverse transcriptase

The sequences of all samples covered both PR and RT genes and comprised all IAS codons. Comparing these sequences to the consensus sequences for each codon demonstrated that deep sequencing could correctly detect codons for 339 out of 340 DRM codons. The only incorrect codon is PR-43 of ENVA15-08 reported as AAA codon genotype which was not identical to expected results (AAR). This codon was reported as AAA approximately 22 out of 69 laboratories submitting PR sequences. To investigate the incorrect codon, a manual reviewed bam file of ENVA15-08 sample was performed and found at position 2,381 (PR-43) had 3,454 sequence reads and had A (wild-type) at 83% and G (mutant type) at 16%. In other word, deep sequencing could detect a mixture of A and G nucleotides as R, according to IUB code, which was identical to the expected results. The incorrect codon genotype of the codon was likely to associate with data analysis pipeline. Therefore, the generated raw sequence reads of all samples were reanalyzed by Sentosa® SQ reporter software (Vela Diagnostics Pte. Ltd.) with Sentosa® SQ HIV-1 analysis pipeline that can call variants as low as 5%. This pipeline demonstrated that deep sequencing correctly identified 334 DRM codons identical to expected results which were composed of PR-43 of ENVA15-08 sample and 333 codons detected mutant types at frequencies higher than 20%, almost all of which had detected mutants nearly 100%. These codons had the average number of coverages more than 1,000X. In addition, deep sequencing identified 6 DRM codons as IUB codes comprising mutants detected at frequencies lower than 20%, which were partial concordance with ENVA 2015 consensus sequences for each codon. Of the 6 low-frequency DRM codons, 50% were at level 5.95 to 7.95% and 50% were at level 10.65 to 16.58%, and all codons had the number of coverage more than 1,000X. The low-frequency DRMs were partial concordance with ENVA 2015 consensus sequences because nearly all datasets submitted in the EQA program were based on Sanger sequencing technology which is bulk sequencing the viral RNA in plasma and results in each position of nucleotide sequences presenting the predominant variant at frequencies above 20% of viral quasispecies (Palmer *et al.*, 2005; Church *et al.*, 2006; Halvas *et al.*, 2006). Theoretically, the coverage required to detect minor variants at various thresholds, for example, in detecting a low-frequency variant present in $\geq 1\%$ of viral quasispecies is required the number of coverage more than 1,000 to ensure the true variant (Wang *et al.*, 2007). In the EQA program, all DRM codons detected by deep sequencing at frequencies of 5 to 20% and at above 20% had the average number of coverage more than 1,000 reads, so these DRMs was very accurate and reliable. Therefore, NGS parameters and a bioinformatics pipeline for dealing with the enormous amounts of NGS data are important to correctly detect HIV-1 DRMs which play role in selecting an optimal antiretroviral regimen.

In conclusion, deep sequencing has a superior performance in detecting HIV-1 drug resistance mutations in both PR and RT genes in case of generating complete datasets of all five panel samples successfully and, in particular, in case of the

detecting all DRMs with expected results in the EQA program correctly. Therefore, deep sequencing has the accuracy in HIV-1 DRM detection as well as it could be adopted as routine HIV-1 drug resistance testing in clinical laboratories in the near future.

ACKNOWLEDGEMENTS

This work is supported by the 60th Year Supreme Reign of his Majesty King Bhumibol Adulyadej Scholarship, granted by the Faculty of Graduate Studies Academic Year 2014, Mahidol University. Vela Diagnostics Pte. Ltd. provided Sentosa® SQ HIV-1 Genotyping Assay and reanalyze bam files of all five ENVA15 panel samples.

REFERENCES

Avidor B, Girshengorn S, Matus N, Talio H, Achsanov S, Zeldis I, Fratty IS, Katchman E, Brosh-Nissimov T, Hassin D, *et al.* (2013) Evaluation of a benchtop HIV ultradeep pyrosequencing drug resistance assay in the clinical Laboratory. J Clin Microbiol 51: 880–886.

Church JD, Jones D, Flys T, Hoover D, Marlowe N, Chen S, Shi C, Eshleman JR, Guay LA, Jackson JB, *et al.* (2006) Sensitivity of the ViroSeq HIV-1 genotyping system for detection of the K103N resistance mutation in HIV-1 subtypes A, C, and D. J Mol Diagn 8: 430–432.

Cohen MS, Chen YQ, McCauley M, Gamble T, Hosseinipour MC, Kumarasamy N, Hakim JG, Kumwenda J, Grinsztejn B, Pilotto JHS, *et al.* (2016) Antiretroviral therapy for the prevention of HIV-1 transmission. N Engl J Med 375: 830-839.

Garcia-Diaz A, Guerrero-Ramos A, McCormick AL, Macartney M, Conibear T, Johnson MA, Haque T, Webster DP (2013) Evaluation of the Roche prototype 454 HIV-1 ultradeep sequencing drug resistance assay in a routine diagnostic laboratory. J Clin Virol 58: 468–473.

Geretti AM, Harrison L, Green H, Sabin C, Hill T, Fearnhill E, Pillay D, Dunn D, UK Collaborative Group on HIV Drug Resistance (2009) Effect of HIV-1 subtype on virologic and immunologic response to starting highly active antiretroviral therapy. Clin Infect Dis 48: 1296-1305.

Gibson RM, Meyer AM, Winner D, Archer J, Feyertag F, Ruiz-Mateos E, Leal M, Robertson DL, Schmotzer CL, Quiñones-Mateu ME (2014) Sensitive deep-sequencing-based HIV-1 genotyping assay to simultaneously determine susceptibility to protease, reverse transcriptase, integrase, and maturation inhibitors, as well as HIV-1 coreceptor tropism. Antimicrob Agents Chemother 58: 2167–2185.

Günthard HF, Aberg JA, Eron JJ, Hoy JF, Telenti A, Benson CA, Burger DM, Cahn P, Gallant JE, Glesby MJ, *et al.* (2014) Antiretroviral treatment of adult HIV infection: 2014 recommendations of the International Antiviral Society-USA Panel. JAMA 312: 410–425.

Halvas EK, Aldrovandi GM, Balfe P, Beck IA, Boltz VF, Coffin JM, Frenkel LM, Hazelwood JD, Johnson VA, Kearney M, *et al.* (2006) Blinded, multicenter comparison of methods to detect a drug-resistant mutant of human immunodeficiency virus type 1 at low Frequency. J Clin Microbiol 44: 2612–2614.

Li JZ, Paredes R, Ribaudo HJ, Svarovskaia ES, Metzner KJ, Kozal MJ, Hullsiek KH, Balduin M, Jakobsen MR, Geretti AM, *et al.* (2011) Low-frequency HIV-1 drug resistance mutations and risk of NNRTI-based antiretroviral treatment failure: a systematic review and pooled analysis. JAMA 305: 1327–1335.

Liu L, Li Y, Li S, Hu N, He Y, Pong R, Lin D, Lu L, Law M (2012) Comparison of Next-Generation Sequencing Systems. J Biomed Biotechnol 2012: 1-11

Martinez-Cajas JL, Pai NP, Klein MB, Wainberg MA (2009) Differences in resistance mutations among HIV-1 non-subtype B infections: a systematic review of evidence (1996-2008). J Int AIDS Soc 12: 11.

Palmer S, Kearney M, Maldarelli F, Halvas EK, Bixby CJ, Bazmi H, Rock D, Falloon J, Davey RT, Dewar RL, *et al.* (2005) Multiple, linked human immunodeficiency virus type 1 drug resistance mutations in treatment-experienced patients are missed by standard genotype analysis. J Clin Microbiol 43: 406–413.

Paredes R, Clotet B (2010) Clinical management of HIV-1 resistance. Antiviral Res 85: 245-265.

Rodger AJ, Lodwick R, Schechter M, Deeks S, Amin J, Gilson R, Paredes R, Bakowska E, Engsig FN, Phillips A; INSIGHT SMART, ESPRIT Study Groups (2013) Mortality in well controlled HIV in the continuous antiretroviral therapy arms of the SMART and ESPRIT trials compared with the general population. AIDS 27: 973-979.

Simen BB, Simons JF, Hullsiek KH, Novak RM, MacArthur RD, Baxter JD, Huang C, Lubeski C, Turenchalk GS, Braverman MS, *et al.* (2009) Low-abundance drug-resistant viral variants in chronically HIV-infected, antiretroviral treatment-naive patients significantly impact treatment outcomes. J Infect Dis 199: 693–701.

Stelzl E, Pröll J, Bizon B, Niklas N, Danzer M, Hackl C, Stabentheiner S, Gabriel C, Kessler HH (2011) Human immunodeficiency virus type 1 drug resistance testing: Evaluation of a new ultra-deep sequencing-based protocol and comparison with the TRUGENE HIV-1 Genotyping Kit. J Virol Methods 178: 94–97.

Touloumi G, Pantazis N, Chaix M-L, Bucher HC, Zangerle R, Kran A-MB, Thiebaut R, Masquelier B, Kucherer C, Monforte Ad, *et al.* (2013) Virologic and immunologic response to cART by HIV-1 subtype in the CASCADE Collaboration. PLoS ONE 8: e71174.

Vandamme AM, Camacho RJ, Ceccherini-Silberstein F, de Luca A, Palmisano L, Paraskevis D, Paredes R, Poljak M, Schmit JC, Soriano V, *et al.* (2011) European recommendations for the clinical use of HIV drug resistance testing: 2011 update. AIDS Rev 13: 77–108.

Wang C, Mitsuya Y, Gharizadeh B, Ronaghi M, Shafer RW (2007) Characterization of mutation spectra with ultra-deep pyrosequencing: application to HIV-1 drug resistance. Genome Res 17: 1195-1201.