Newly developed automated-massively parallel sequencing presenting the accuracy of hepatitis C virus genotyping and application on resistance-associated variants detection

Narathon Manee[1], Wasun Chantratita[1, 2], Ekawat Pasomsub[1, 2]*

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

**Corresponding author: ekawat.pas@mahidol.ac.th*

ABSTRACT

Most hepatitis C virus (HCV) infection becomes chronic diseases and predictably develops into liver cirrhosis or hepatocellular carcinoma. HCV can be classified into 7 genotypes (GT) and multiple subtypes (ST). HCV treatment guidelines focus on determining the accuracy of HCV genotyping and subtyping, before starting treatment in all HCV-infected individuals. Efficient HCV genotyping and subtyping have become essential tools for indicating the optimal treatment. Current routine assays for HCV genotyping are easy and quick turnaround, but apparently subtyping is limited to identification. This study evaluated the accuracy of HCV genotyping in a newly developed automated-massively parallel sequencing (MPS) technique of previously identified HCV-infected specimens. The results indicated that by using the reference method, MPS and a bioinformatic pipeline, the genotype and subtype had identical accuracy of 95.65%. Interestingly, MPS could detect the remaining 4.35% as mixed-HCV type infected specimens, while the reference method could not. This result indicated that the MPS technique was highly accurate in identifying the HCV genotype and mixed-type infection. In addition, MPS can detect resistance-associated variants (RAVs) simultaneously with the HCV genotyping, especially in HCV GT1-infected specimens. Since HCV GT1 has a highly frequent presence of pre-existing RAVs, with resistance associated to direct-acting antivirals (DAAs), this study found that up to 60% of HCV GT1 sample were RAVs. This discovery of important RAVs in the patient infected with HCV GT1 could be beneficial in guiding the therapy decisions while avoiding the use of some DAAs. Thereby, the opportunity for patients to access the best treatment is enhanced.

Keywords: HCV genotyping; mixed-HCV infection; massively parallel sequencing; Sanger sequencing

INTRODUCTION

Hepatitis C virus (HCV) is a positive single-stranded RNA virus in the *Hepacivirus* genus of the *Flaviviridae* family (Chevaliez *et al*., 2006; Li *et al*., 2015). Multiple strains, 7 genotypes (GT) and approximately 67 subtypes (ST) of HCV have been distributed differently in various regions worldwide (Smith *et al.*, 2014). No vaccine is currently available to prevent HCV infection, it can be cured with the antiviral drug, which result in cure rates approximately 90%. The treatment goal is sustained virological response (SVR; undetectable HCV RNA three or six months post therapy) (WHO, 2016). The initial treatment of chronic hepatitis C was pegylated interferon (Peg-IFN) combined with ribavirin (RBV). Later, direct-acting antivirals (DAAs in four groups: NS3 inhibitors, NS5A inhibitors, NS5B Nucleoside and Non-Nucleoside Polymerase Inhibitors) was developed and utilized with/without pegylated interferon and Ribavirin (WHO, 2016). The treatment of HCV-infected patients is different by HCV genotyping (GTs 2-6). Moreover, treatment guidance of HCV genotype 1 was specific to a and b subtypes (Chung *et al*., 2015; EASL, 2015; Thai Association for the Study of the Liver, 2016). Thereby, it is necessary to determine the strains of HCV (HCV genotyping) in all patients prior starting the treatment, which benefits the treatment plan that includes duration time to treat, dose of drugs, suitable regimens and prediction of treatment response (Chung *et al*., 2015; EASL, 2015; WHO, 2016). Besides, a single patient may be infected by more than one HCV strain (mixed-type infection). Mixed-type infection was observed frequently in about

14–39 percent of persons sharing a hypodermic needle (Cunningham *et al*., 2015). Methods with low sensitivity are unable to determine all strains that infect an individual patient, as a different HCV-minority genome will be undetected and in turn, causes an inappropriate treatment plan by reducing the response rate to therapy, or treatment failure from undetected, non-sensitive minority strains (Cunningham *et al.*, 2015).

In an era where DAAs are used increasingly, variants associated with drug resistance (resistance-associated variants: RAVs) were found and frequently presented in patients infected with genotype 1. RAVs were detected on three DAA-targeted genes of the HCV genome, and observed frequently in the NS3 and NS5A regions, but were rarely found in the NS5B region. RAVs were identified in patients before starting therapy with an estimated 8.6 and 1.4 percent in subtypes 1a and1b, respectively. Moreover, HCV RAVs have been found in most patients infected with genotype 1 after treatment failure with a first generation NS3/NS4A protease inhibitor (Ahmed and Felmlee, 2015). The Q80K polymorphism is a variant that is common and occurs naturally at *NS3* gene position of the HCV subtype 1a (19−48%) (Poveda *et al*., 2014), which significantly reduces the response rate of Simeprevir treatment (second-generation NS3/4A protease inhibitor), when compared to patients infected with HCV subtype 1a that does not appear with Q80K (Wyles, 2013; Wyles and Gutierrez, 2014; Ahmed and Felmlee, 2015; Lenz *et al*., 2015; Sarrazin *et al*., 2015). The HCV treatment guide from the World Health Organization (WHO), European Association for the Study of the liver (EASL), America American Association for the Study of Liver Diseases and the Infectious Diseases Society of America (AASLD/IDSA) advised against the use of Simeprevir when Q80K is presented in naive patients infected with HCV subtype 1a, as well as those re-starting treatment after previous treatment failure (retreatment) (Au *et al*., 2015; EASL, 2015; WHO, 2016). Approximately 10−15 percent of patients infected with HCV genotype 1 was detected with variants related to drug resistance to Daclatasvir, Ledipasvir, Elbasvir or Ombitasvir (NS5A inhibitors) (AASLD-IDSA, 2016a; 2016b). These RAVs are important, found commonly in patients infected with HCV subtype 1a and occur at the M28, Q30, L31, and Y93 positions in the HCV *NS5A* gene. The effects of these variants reduced the response rate of NS5A inhibitors treatment by more than 5 times when compared with patients without these variants (Wyles and Gutierrez, 2014; AASLD-IDSA, 2016a; 2016b). The EASL and AASLD/IDSA recommend avoiding the use of some NS5A inhibitors in patients infected with genotype 1 if they appear to have RAVs. The AASLD/IDSA recommends detecting these RAVs before re-starting treatment in patients experiencing treatment failure (AASLD-IDSA, 2016b).

The 5'UTR (untranslated region) and the core region are common for HCV genotyping. However, there are limitations in the accuracy of results that they do not cover the identification of all subtypes (Avo *et al*., 2013; Liu *et al*., 2015; Quer *et al*., 2015). The *NS5B* gene is capable of identifying all strains with high precision, and used commonly in HCV epidemiology. It is therefore considered as the gold standard for HCV genotyping (Enache and Enache, 2008; Gryadunov *et al.*, 2010). The VERSANT HCV Genotype 2.0 Line probe assay (LiPA) and the Abbott m2000 RealTime HCV Genotype II are popular methods used for HCV genotyping in clinical diagnostic laboratories. However, both approaches have the disadvantage of inaccuracy in genotype separation when compared to the reference method, and they also are weak in identifying the HCV subtype level (Avo *et al*., 2013; Liu *et al*., 2015; Quer *et al*., 2015).

The first-generation sequencing or Sanger sequencing is a reference method for HCV genotyping, performing by DNA synthetic HCV RNA at the *NS5B* gene position. However, Sanger sequencing has the disadvantage of sequencing only one sample per run and is therefore inappropriate for use in a clinical laboratory. Second-generation sequencing; also known as massively parallel sequencing (MPS) technology, is a newly developed technology with high throughput and high sensitivity (Cunningham *et al.*, 2015). MPS was developed recently as a new model that worked under the automatic system, in contrast to the older version. Furthermore, this newly developed automated-MPS added the ability of detecting HCV RAVs, which can be performed simultaneously with HCV genotyping. This study evaluated the performance of the newly developed automated-MPS in determining accurate GTs and STs including evaluation of an additional application performance for identifying RAVs often found in the HCV GT1.

MATERIALS AND METHODS

Samples

This retrospective study was conducted in 23 HCV-infected patients, who were previously identified as HCV genotype by using the Versant-Line Probe Assay (HCV genotyping LiPA2.0; Siemens Healthcare Diagnostics, Eragny, France). All of the HCV-infected patients were collected randomly between March and

August 2016 without patient history-associated demographic and clinical information. The inclusion criteria comprised tested samples with HCV viral load of more than 1,000 copies/mL, with sufficient plasma volume for using the newly developed automated-MPS, and Sanger sequencing to confirming the HCV genotyping result. This study was approved by the Committee on Human Rights Related to Research Involving Human Subjects, Faculty of Medicine, Ramathibodi Hospital, Mahidol University, Bangkok, Thailand and performed based on the principles of the Declaration of Helsinki (ID 09-58-43).

The newly developed automated-massively parallel sequencing

The MPS-based *Sentosa*® SQ HCV Genotyping Assay was performed on the Ion Torrent PGM platform (Vela Operations Singapore Pte Ltd, The Kendall, 50 Science Park Road, Singapore). It worked under an automatic system as follows: HCV RNA extraction from 530 µl of samples and PCR set-up were performed on the *Sentosa*® robotic SX101 instrument with primers specific to the *NS5B*, *NS5A* and *NS3* genes. The off-board RT-PCR was administered on the Veriti® Dx 96-Well Thermal Cycler. After that, the reaction plate was transferred back onto the *Sentosa*® SX101 worktable in order to continue the library preparation, which generated 200-nucleotide fragments of a DNA amplicon, followed by adaptors ligated to the fragments. The final step of the library preparation was presented with a DNA library pool containing 15 samples and one HCV System Control. The DNA library pool was carried out on the *Sentosa*® ST401 instrument by clonal amplification in a scheme of PCR emulsion. Sequencing was executed on a *Sentosa*® SQ301 Sequencer based on the measurement changes of pH from release of hydrogen ions, due to nucleotide incorporation, before converting into an electrical voltage signal. *Sentosa*® SQ Suite software performed the primary analysis continuously on the raw sequencing data generated by the *Sentosa*® SQ301 Server. Then, the data was transferred to a *Sentosa*® SQ Reporter for the secondary analysis and a generation report (comprising GTs calling concomitantly with RAVs calling, which is found often in GT1). The report contained a full audit trail relating to the actions within the run, and also released output files representing reads (uBAM), mapped reads (BAM), and variant calls (VCF).

Data Analysis

Bioinformatic programs were used to analyze a very large number of sequenced reads from sequencing. All of the sequenced reads were mapped to all complete HCV genomes from NCBI and LANL database by using Ion Torrent TMAP 4.0.5. A sequenced read which could specify a map to the NS5B region of HCV genomes (mapped NS5B specific read) was analyzed for HCV genotyping, while all of the sequenced reads that could specify a map to *NS3*, *NS5A* and *NS5B* regions of complete HCV genomes were analyzed for the HCV resistance-associated variants occurring in the sample. A MIRA Assembler 3.4.1 was used to assemble mapped NS5B specific reads for HCV genotyping into NS5B contigs, which were then aligned to all *NS5B* sequences from the *NS5B* database by using BLAST. Top hit strain was chosen if the *NS5B* alignment score showed ≥ 80% similarity, and subsequent hits were chosen if their alignment score was ≥ 0.98 of the top hit. A phylogenetic Tree was used to determine the genotype.

When detecting HCV RAVs, all mapped specific reads with *NS3*, *NS5A* and *NS5B* regions of complete HCV genomes, which only indicated HCV genotype 1, were assembled to NS3, NS5A and NS5B contigs by using the MIRA Assembler 3.4.1. This step was carried out only if the read inferred genotype 1a and 1b, and then the assembled contigs aligned to reference genome of genotype 1a and 1b with the NCBI and LANL database, respectively, by using BLAST. Moreover, the mapped reads were aligned to the assembled contigs in order to obtain the variants (mutation) frequency, as in some cases, different mutations may be assembled into separate contigs. Finally, the mutations were called from reads rather than contigs. Variants were called per Codon and output in terms of AA change. However, in partially detecting RAVs, the phenotype of resistance-association drugs and therapy guidance decisions had not been discussed. Therefore, the geno2pheno web service (http://hcv.geno2pheno.org/index.php) was used to determine the resistance-association variant for each drug (Geno2pheno hcv, 2011).

Sanger sequencing

HCV RNA was extracted from 400 µL of EDTA-plasma sample. The PCR product was generated by RT-PCR amplification of the HCV RNA *NS5B* region, and sequenced directly by Sanger sequencing (Biosystems® 3500 Series Genetic Analyzer: Integrated DNA Technologies, Singapore). The consensus sequences were submitted to geno2pheno (HCV) web service for HCV genotype and subtype analyses.

Reference sequences

All HCV genome sequences available in the National Center for Biotechnology Information (NCBI) and Los Alamos National Laboratory (LANL) database were used as the reference sequences for the *NS5B*, *NS5A* and *NS3* genes for HCV genotyping. The GenBank accession numbers NC004102 and EU256045 were used as references for the variant calling of ST1a and ST1b, respectively, in partial RAVs detection.

Statistical analysis

Kappa statistics were used to evaluate inter-rater agreement (inter-method) between massively automated parallel sequencing and the reference method. The Kappa value ranged from 0 to 1.00, but generally, a Kappa of more than 0.80 is considered a satisfactory agreement (McHugh, 2012; Tan *et al*., 2015). A *p*-value of < 0.05 was considered statistically significant. Kappa statistics were performed using SPSS Software version 20.

RESULTS AND DISCUSSION

This study classified evaluation of the result into 2 parts; HCV genotyping and partial RAVs detection, both of which were obtained from the newly developed automated-MPS. Detail in the Table 1 shows the HCV genotyping result of MPS compared with Sanger sequencing as a reference method. Both approaches show identical results of HCV GTs and STs as 95.65% (22/23). The remaining 4.35% (1/23) was interesting as MPS could identify mixed-type HCV infection of GT 6n and 3b, while the reference method was unable to detect the GT 6n, which infected as another mixed-type in this sample (Sample no. 23). Sample number 3 interestingly had a low viral load of 219 IU/ml, but MPS as well as the reference method could accurately determine the genotype and subtype. This data indicated that MPS had a high accuracy in determining strains of HCV, which is consistent with the reference method (Quer *et al*., 2015). Furthermore, this data showed that MPS had high sensitivity in detecting infection with multiple HCV strains in a patient (Cunningham *et al*., 2015; Quer *et al*., 2015). Besides, the Kappa coefficient was 1.00 combined with a *p*-value < 0.001, which showed the statistical significance of perfect HCV genotyping agreement between MPS and the reference method. Thus, Kappa statistic indicated that automated-MPS was equivalent to the reference method in being reliable and valid in determining the HCV genotype and subtype.

Table 1 HCV genotype obtained from two methods; newly developed automated- massively parallel sequencing versus the reference method.

Sample No.	Viral load (IU/ml)	Genotyping result	
		NS5B Direct Sanger sequencing (reference method)	*Sentosa*® SQ HCV Genotyping Assay (newly developed-MPS)
1	3,022,691	1a	1a
2	4,393,604	1a	1a
3	219	1b	1b
4	1,075,521	1b	1b
5	2,793,458	1b	1b
6	5,330,000	3a	3a
7	156,973	3b	3b
8	11,051,463	3a	3a
9	11,051	3a	3a
10	1,826,305	3a	3a
11	825,356	6f	6f
12	2,502,365	6i	6i
13	3,795,542	6i	6i
14	13,537	6e	6e
15	6,805,737	6n	6n
16	655,832	6i	6i
17	511,188	6i	6i
18	278,112	6f	6f
19	10,147	3a	3a
20	80,517	6n	6n
21	4,599,512	6v	6v
22	6,998,899	6e	6e
23	676,280	3b	6n + 3b

In the partial RAVs detection, Table 2 shows the important mutation list detected in HCV GT1 samples obtained by additional MPS application, and predictions of phenotypic resistance to HCV direct-acting antivirals (DAAs) by the Geno2pheno web-service. This study observed RAVs as 60% (3/5) of the HCV GT1 samples; where there were Q30H combined with Y93H in one of the ST1a infected specimens (sample no.1). Likewise, the Y93H also was presented in a ST1b infected specimen (sample no.5). The Q30H and Y93H played an important clinical role by resisting NS5A inhibitors (Daclatasvir; Ledipasvir; Ombitasvir and Elbasvir). Moreover, the Q80K mutation was found in a ST1a infected specimen (sample no.2), which had clinical importance as resistant to Simeprevir (NS3/NS4A protease inhibitor). These detected RAVs in HCV GT1-infected patients are currently important mutations, and the EASL and AASLD/IDSA suggest to avoid the use of some DAAs contained in treatment regimens. This is of concern because these drugs (Boceprevir; Sofosbuvir; Simeprevir; Daclatasvir; Ledipasvir; Ombitasvir, etc.) are commonly used in Thailand to treat patients infected with hepatitis virus (Thai Association for the Study of the Liver, 2016). Furthermore, it was evident that all samples infected with HCV ST1a showed RAVs, while only one of three samples infected with HCV ST1b showed them, thus suggests that samples infected with HCV ST1a had a greater incidence of RAVs than ST1b, which is consistent with other studies (Poveda *et al*., 2014; Wyles and Gutierrez, 2014). In conclusion, the newly developed automated-MPS promises to be a reliable tool for subtype and genotype determinations, and also has an improvement in sensitivity for detecting mixed HCV infections. However, this study was limited by the small sample size in the mixed HCV-infection risk group. Therefore, large sample sizes may better demonstrate the capabilities of the MPS to identify different HCV strains in mixed-infected in individual patients. In additional, MPS might be unable to detect mixed-type infection when the minor strains contain <1% of the total viral population (Cunningham *et al*., 2015). A significant advantage of using MPS is its ability to determine the HCV subtype and genotype and also identify RAVs simultaneously, which benefits clinical decision making and better treatment plans with DAA-containing regimens, especially for patients infected with the HCV genotype 1. This newly developed method has an automatic workflow that helps to save time, reduce manual steps and save cost in the long term when compared to traditional methods. It also has the advantage of its ability to sequence more than one sample, and more than one gene per run, thus, it is appropriate to use in a clinical diagnostic laboratory.

Table 2. List of important mutations obtained by additional application of newly developed-MPS, including predictions of phenotypic resistance to HCV direct-acting antivirals (DAAs).

Sample No.	Genotyping result	Mutation result	
		RAVs	Resistance associated to
1	1a	Q30H, Y93H	NS5A inhibitors
2	1a	Q80K	NS3/NS4A inhibitor
3	1b	-	
4	1b	-	
5	1b	Y93H	NS5A inhibitors

ACKNOWLEDGEMENTS

This study thanks the virology laboratory and Center for Medical Genomics in the Department of Pathology, Faculty of Medicine Ramathibodi Hospital, Mahidol University, Bangkok, Thailand for valuable advice and materials**.**

REFERENCES

AASLD-IDSA panel (2016a) Initial treatment of HCV infection. Hepatitis C guidance: Recommendations for Testing, Managing, and Treating Hepatitis C, http://www.hcvguidelines.org/full-report/initial-treatment-hcv-infection (November 2016)

AASLD-IDSA panel (2016b) Retreatment of a person with chronic HCV infection in whom prior therapy has failed. Hepatitis C guidance: Recommendations for Testing, Managing, and Treating Hepatitis C, http://www.hcvguidelines.org/full-report/retreatment-persons-whom-prior-therapy-has-failed (November 2016)

Ahmed A, Felmlee DJ (2015) Mechanisms of hepatitis C viral resistance to direct acting antivirals. Viruses 7: 6716–6729.

Au TH, Destache CJ, Vivekanandan R (2015) Hepatitis C therapy: Looking toward interferon-sparing regimens. J Am Pharm Assoc 55: e72–e86.

Avo AP, Agua Doce I, Andrade A, Padua E, (2013) Hepatitis C virus subtyping based on sequencing of the C/E1 and NS5B genomic regions in comparison to a commercially available line probe assay. J Med Virol 85: 815–22.

Chevaliez S, Pawlotsky JM (2006) HCV genome and life cycle. In: Tan SL (ed) Genomes and molecular biology. Hepatitis C Viruses, Horizon Bioscience, pp. 5–47.

Chung RT, Davis GL, Jensen DM, Masur H, Saag MS, Thomas DL (2015) Hepatitis C guidance: AASLD-IDSA recommendations for testing, managing, and treating adults infected with hepatitis C virus. Hepatology 62: 932–954.

Cunningham EB, Applegate TL, Lloyd AR, Dore GJ, Grebely J (2015) Mixed HCV infection and reinfection in people who inject drugs-impact on therapy. Nat Rev Gastroenterol Hepatol 12: 218–30.

EASL (2015) EASL Recommendations on Treatment of Hepatitis C 2015. J Hepatol 63: 199–236.

Enache EL, Enache LS (2008) Versant HCV genotype 2.0 assay (LiPA) in hepatitis C virus genotype determination. Rev Romana Med Lab 12: 47–53.

Geno2pheno hcv [Internet]. Germany: GENAFOR; 2011 [updated 2017 Feb 24; cited 2017 Mar 13]. Available from: http://hcv.geno2pheno.org/index.php.

Gryadunov D, Nicot F, Dubois M, Mikhailovich V, Zasedatelev A, Izopet J (2010) Hepatitis C virus genotyping using an oligonucleotide microarray based on the NS5B sequence. J Clin Microbiol 48: 3910–3917.

Lenz O, Verbinnen T, Fevery B, Tambuyzer L, Vijgen L, Peeters M, Buelens A, Ceulemans H, Beumont M, Picchio G, *et al.* (2015) Virology analyses of HCV isolates from genotype 1-infected patients treated with simeprevir plus peginterferon/ribavirin in Phase IIb/III studies. J Hepatol 62: 1008–1014.

Li Y, Yamane D, Masaki T, Lemon SM (2015) The yin and yang of hepatitis c: Synthesis and decay of hepatitis c virus RNA. Nat Rev Microbiol 13: 544–558.

Liu CH, Liang CC, Liu CJ, Lin CL, Su TH, Yang HC, Chen PJ, Chen DS, Kao JH (2015) Comparison of Abbott realtime HCV genotype II with Versant line probe assay 2.0 for hepatitis C virus genotyping. J Clin Microbiol 53: 1754–1757.

McHugh ML (2012) Interrater reliability: the kappa statistic. Biochem Med 22: 276–282, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3900052/?report=printable (November 2016)

Poveda E, Wyles DL, Mena Á, Pedreira JD, Castro-Iglesias Á, Cachay E (2014) Update on hepatitis C virus resistance to direct-acting antiviral agents. Antiviral Res 108:181–191.

Quer J, Gregori J, Rodriguez F, Buti M, Madejon A, Perez del Pulgar S, Garcia Cehic D, Casillas R, Blasi M, Homs M, *et al.* (2015) High-resolution hepatitis C virus subtyping using NS5B deep sequencing and phylogeny, an alternative to current methods. J Clin Microbiol 53: 219–26.

Sarrazin C, Lathouwers E, Peeters M, Daems B, Buelens A, Witek J, Wyckmans Y, Fevery B, Verbinnen T, Ghys A, *et al.* (2015) Prevalence of the hepatitis C virus NS3 polymorphism Q80K in genotype 1 patients in the European region. Antiviral Res 116: 10–16.

Smith DB, Bukh J, Kuiken C, Muerhoff AS, Rice CM, Stapleton JT, Simmonds P (2014) Expanded classification of hepatitis C virus into 7 genotypes and 67 subtypes: updated criteria and genotype assignment web resource. Hepatology 59: 318–27.

Tan CL, Hassali MA, Saleem F, Shafie AA, Aljadhey H, Gan VB (2015) Development, test-retest reliability and validity of the Pharmacy Value-Added Services Questionnaire (PVASQ). Pharm Pract 13: 598

Thai Association for the Study of the Liver (2016) Thailand Practice Guideline for Management of Chronic Hepatitis C 2016. Thai Association for the Study of Liver, pp. 7–32.

World Health Organization (2014) Guidelines for the screening, sare and treatment of persons with hepatitis C infection.

Wyles DL (2013) Antiviral resistance and the future landscape of hepatitis C virus infection therapy. J Infect Dis 207: S33–S39.

Wyles DL, Gutierrez JA (2014) Importance of HCV genotype 1 subtypes for drug resistance and response to therapy. J Viral Hepat 21: 229–40.