Conference report

A systematic, deep sequencing-based methodology for identification of mixed-genotype hepatitis C virus infections

Andrea D. Olmstead[a,*], Vincent Montoya[a], Celia K. Chui[a], Winnie Dong[a], Jeffrey B. Joy[a,b], Vera Tai[a,1], Art F.Y. Poon[a,2], Thuy Nguyen[a], Chanson J. Brumme[a], Marianne Martinello[c], Gail V. Matthews[c], P. Richard Harrigan[b], Gregory J. Dore[c], Tanya L. Applegate[c], Jason Grebely[c], Anita Y.M. Howe[a]

[a] BC Centre for Excellence in HIV/AIDS, Vancouver, BC, Canada
[b] Faculty of Medicine, Department of Medicine, Division of AIDS, University of British Columbia, Vancouver, BC, Canada
[c] UNSW Sydney, The Kirby Institute, Sydney, NSW, Australia

ARTICLE INFO

ABSTRACT

Hepatitis C virus (HCV) mixed genotype infections can affect treatment outcomes and may have implications for vaccine design and disease progression. Previous studies demonstrate 0–39% of high-risk, HCV-infected individuals harbor mixed genotypes however standardized, sensitive methods of detection are lacking. This study compared PCR amplicon, random primer (RP), and probe enrichment (PE)-based deep sequencing methods coupled with a custom sequence analysis pipeline to detect multiple HCV genotypes. Mixed infection cutoff values, based on HCV read depth and coverage, were identified using receiver operating characteristic curve analysis. The methodology was validated using artificially mixed genotype samples and then applied to two clinical trials of HCV treatment in high-risk individuals (ACTIVATE, 114 samples from 90 individuals; DARE-C II, 26 samples from 18 individuals) and a cohort of HIV/HCV co-infected individuals (Canadian Coinfection Cohort (CCC), 3 samples from 2 individuals with suspected mixed genotype infections). Amplification bias of genotype (G)1b, G2, G3 and G5 was observed in artificially mixed samples using the PCR method while no genotype bias was observed using RP and PE. RP and PE sequencing of 140 ACTIVATE and DARE-C II samples identified the following primary genotypes: 15% ($n = 21$) G1a, 76% ($n = 106$) G3, and 9% ($n = 13$) G2. Sequencing of ACTIVATE and DARE-C II demonstrated, on average, 2% and 1% of HCV reads mapping to a second genotype using RP and PE, respectively, however none passed the mixed infection cutoff criteria and phylogenetics confirmed no mixed infections. From CCC, one mixed infection was confirmed while the other was determined to be a recombinant genotype. This study underlines the risk for false identification of mixed HCV infections and stresses the need for standardized methods to improve prevalence estimates and to understand the impact of mixed infections for management and elimination of HCV.

1. Introduction

Globally, it is estimated that 70 million people are infected with Hepatitis C virus (HCV) (Polaris Observatory HCV Collaborators, 2017). HCV is an extremely diverse pathogen; the virus's polymerase is highly error prone causing a swarm of HCV variants to accumulate over time within each infected individual. This has contributed to global diversification of HCV into at least seven distinct genotypes (G1-7) with varied geographical distributions (Smith et al., 2014).

The advent of highly-effective, short duration direct acting antivirals is greatly improving the outcome and prognosis of individuals chronically infected with HCV (Dore and Feld, 2015). These antiviral regimens are simple, tolerable and effective in > 95% of those treated (Falade-Nwulia et al., 2017). Although pan-genotype HCV therapies

have emerged, many regimens are genotype specific and knowledge of a patient's HCV genotype is required to prescribe the most effective therapy.

As HCV infection does not elicit a protective immune response, individuals repeatedly exposed to HCV can become infected with more than one genotype (Abdel-Hakeem and Shoukry, 2014; Grebely et al., 2012b; Osburn et al., 2010). Mixed-genotype infections have been observed in up to 39% of high-risk populations, such as in people who inject drugs (PWID) and people with acute HCV infection (Cunningham et al., 2015; Grebely et al., 2012a; Herring et al., 2004; McNaughton et al., 2017; Pham et al., 2010; van de Laar et al., 2009). Mixed infections pose a challenge to genotype-specific therapies and have implications for vaccine design-strategies (Abdelrahman et al., 2015; McNaughton et al., 2014). In addition, the prevalence of mixed infections and their impact on disease progression are poorly understood.

HCV genotyping for treatment purposes is frequently performed using commercial assays based on real-time PCR or line probe hybridization (Gao, 2012; Yang et al., 2014). Although these assays can detect mixed HCV infections, they are not specifically validated for this purpose and false positive and negative results occur (Minosse et al., 2016). Research studies examining mixed infections have employed a variety of methods with varying sensitivities and limitations (Cunningham et al., 2015). Methods that rely on PCR require primers and probes that can both capture and differentiate the high amount of existing HCV sequence variation. Deep sequencing methods offer several advantages in that they can theoretically detect any HCV genome irrespective of the diversity and the sequences can be used for phylogenetic and drug resistance analysis. Identifying HCV genotypes and subtypes from a sample using deep sequencing is relatively simple when the most abundant and variable sequences are used, however identifying mixed infections is complicated by the presence of low abundance sequences (i.e. < 2%) and sequences mapping to conserved genomic regions.

Here, a PCR-independent deep sequencing method was developed that uses randomly primed cDNA synthesis with or without probe capture of HCV sequences for detection of HCV mixed genotype infections. A bioinformatics pipeline was developed to detect mixed infections and to rule out false positive results from sequencing reads mapping to incorrect genotype references. The methodology for detecting mixed infections was applied to samples from three high-risk cohort studies of HCV infections.

2. Methods

2.1. Study design

For initial development and validation, de-identified clinical samples of known HCV genotypes were artificially mixed (*n* = 72) in ratios of 98:2, 95:5, 90:10, 75:25, 50:50 and four samples contained only one genotype (Supplementary Tables 1 and 2).

For further evaluation, plasma or serum samples from one cohort and two clinical trials were used. The Canadian Co-infection Cohort (CCC) follows HIV/HCV co-infected individuals in Canada (Klein et al., 2010). ACTIVATE was an international multi-center trial evaluating the efficacy of response-guided treatment with pegylated interferon alpha-2b and ribavirin for 12 or 24 weeks among people with recent injecting drug use (IDU) or receiving opioid substitution therapy with chronic HCV G2 or G3 (Grebely et al., 2017). Ninety-six percent of ACTIVATE participants had a history of IDU, 73% within 6 months and 59% within one month of enrollment. Participants with HIV infection were excluded. DARE-C II was a prospective, open label multi-center study evaluating the efficacy of sofosbuvir and ribavirin for 6 weeks among people with recent or acute HCV infection (Martinello et al., 2016). Eighty-four percent of participants had a history of IDU; 58% reported within 6 months of screening. HIV co-infection was documented in 74% of participants.

All ACTIVATE and DARE-C II participants with available serum and/or plasma samples at screening (SCR), or treatment baseline (BSL), with viral loads > 1000 IU/ml were assessed for mixed infection. Among participants with persistent HCV infection (treatment non-response or recurrence following sustained virological response (SVR)) HCV RNA detectable samples at treatment week 4 (WK4), end of treatment (ETR), and 4, 12 or 24 weeks post-treatment (SVR4, SVR12, SVR24) were assessed for mixed infection.

Ethical approval was granted by the St. Vincent's Hospital Human Research and Ethics Committee in Sydney, Australia.

2.2. Extraction, PCR, random-primer and probe capture-based deep sequencing

HCV RNA extraction was performed using NucliSENS easyMAG system (bioMérieux) according to the manufacturer's instructions. Three deep sequencing library preparation methods were used (Fig. 1). The first method employed PCR amplification of a portion of the HCV NS5B region as previously described (Koletzki et al., 2010) (See Supplementary Table 3 for complete primer sequences). The second method used rRNA depletion, followed by randomly primed cDNA synthesis, fragmentation, adaptor ligation and individual barcoding of each sample as described in 'NEBNext® Ultra™RNA Library Prep Kit for Illumina Instruction Manual – Chapter 2'. Depletion of rRNA has been previously shown to substantially increase the sensitivity of RNA virus detection (Manso et al., 2017; Matranga et al., 2014). The third method used capture probes to enrich for HCV sequences from the random primer libraries, which were further amplified using low cycle PCR. This was performed using custom designed probes from Illumina targeting HCV Non-Structural (NS) 3, NS5A and NS5B genes and three
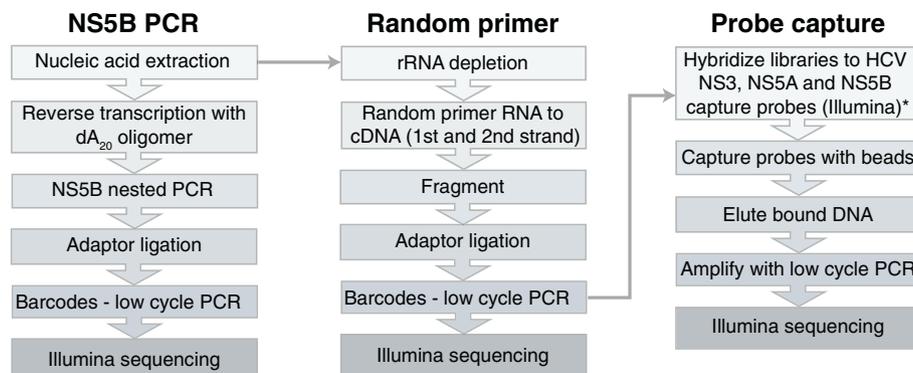

Fig. 1. Workflow summary for PCR, Random Primer and Probe Capture library preparation methods for deep sequencing of mixed infection samples. Capture probes were custom designed by Illumina.

different capture reagent methods were tested (SeqCap EZ Library from Roche, TruSeq RNA Access Library Prep Kit from Illumina, and Hybridization and Wash Kit for xGen® Lockdown® Reagents from IDT). Further details are provided in the supplementary methods.

2.3. Sequence alignment and classification

A bioinformatics pipeline implemented in Python was designed to detect and quantify multiple HCV genotypes within a sample. MiSeq sequencing reads from paired-end FASTQ files, de-multiplexed by Illumina MiSeq Reporter Software, were quality filtered using Cutadapt (Martin, 2011). The nucleotide sequences were subsequently mapped to a set of 571 HCV reference sequences using bowtie2 (Langmead and Salzberg, 2012). The reference set was generated by querying the NCBI GenBank nucleotide database for all full-length HCV sequence records and merging them with the Los Alamos National Laboratory (LANL) HCV Database reference set (Kuiken et al., 2005), and removing phylogenetically similar sequences. The reference set includes the hg38 human genome assembly to exclude human RNA derived reads that could spuriously map to HCV references. Bowtie2 is executed with –local alignment option that tolerates partial reference matches. The alignment results (SAM format) are redirected to the standard output stream, where they are captured by the Python script that parses out the HCV genotype of the reference sequence for each successfully mapped read. By default, only the first read of a pair is used to generate genotype counts. Mapped reads are filtered by first reads whose mate (second read) mapped to a different genotype and by length of the matched region exceeding 100 nucleotides, based on the total number of matched nucleotides indicated by the CIGAR string. The primary output reports the number of reads mapping to each HCV genotype for each sample. Additionally, the number of reads mapping to each amino acid at each position per reference sequence is used to generate consensus sequences for each HCV genotype or subtype of interest.

2.4. Mixed infection assessment

Frequently, reads from a single sample map to more than one HCV genotype or subtype. The *primary genotype* is that having the greatest number of mapped reads, while *secondary genotype*s are those with fewer mapped reads than that of the primary genotype. The relative proportions of primary to secondary genotype reads in a sample varies widely, thus a method was devised for differentiating true mixed infections from artifacts of spurious read mapping due to partial or incorrect reference matches. This method was calibrated using 64 artificially mixed genotype samples (Supplementary Table 1; eight replicate samples with comparable results to their corresponding replicates were not included). Following sequencing of artificially mixed samples, the percent coverage, percent depth and average depth for HCV NS3, NS5A and NS5B were calculated for each genotype in each sample. Percent coverage is the percentage of the total amino acids in the HCV NS3, NS5A and NS5B references with at least 1 mapped read. Average depth is the sum of the read depths for NS3, NS5A and NS5B divided by the total number of positions for the same respective genes. Percent depth is calculated as the average depth for the genotype of interest divided by the total HCV depth in the sample. Receiver operator characteristic (ROC) curves for percent coverage, percent depth and average depth were plotted using the PlotROC function in R (R Core Team, 2013). Youden's index was applied to ROC curves to determine the optimal thresholds for identifying a true mixed HCV genotype infection in clinical samples.

2.5. Phylogenetic analysis

Consensus amino acid sequences were generated for primary and secondary genotype analysis. Any samples with < 500 HCV reads (5 samples) were excluded from phylogenetic analyses. The sequences

were aligned to references from the LANL database (Kuiken et al., 2005) using MUSCLE. Phylogenetic trees were inferred using FastTree2 (Price et al., 2010) and rooted on G6 reference sequences as the outgroup.

3. Results

3.1. Deep amplicon sequencing of artificially mixed HCV genotype samples

PCR amplification of HCV is an efficient method for generating sequencing libraries (Thomson et al., 2016; Trémeaux et al., 2016), however amplification bias towards a subset of HCV variants may occur (Aird et al., 2011; Kanagawa, 2003). To demonstrate this, 40 artificially mixed genotype samples (Supplementary Table 2) were deep sequenced following PCR amplification of a portion of HCV NS5B using primers targeting regions conserved across HCV genotypes. Deep sequencing of the NS5B amplicon demonstrated a lack of correlation between the expected input ratios of various genotypes and what was observed (correlation coefficient: 0.24) (Fig. 2). Examining the average observed versus expected ratio of reads for each genotype, PCR amplification was biased towards HCV G1b, G2, G3 and G5 (i.e. overall average of the observed percentages of HCV reads for each genotype were much higher than what was expected). Conversely, HCV G1a, G4, and G6 consistently yielded lower observed than expected percentages of HCV reads.

3.2. Random primer sequencing of artificially mixed HCV genotype samples

The random primer sequencing method (Fig. 1) was applied to 72 artificially mixed samples (including 40 samples sequenced with NS5B PCR amplification) (Supplementary Table 1). Using this method, there was a high correlation between the expected input ratios of various genotypes and what was observed (correlation coefficient = 0.97) and there did not appear to be a bias towards any of the HCV genotypes or subtypes examined (Fig. 2). The average coverage for individual HCV genotypes was 99% (range 84% - 100%); variable coverage values were observed for genotypes with lower input ratios (i.e. 2–10% input). The average coverage for individual genes was 100% for NS3, 99% for NS5A and 98% for NS5B. The average depth was 919 reads/position (range 10–3702 reads/position) (Supplementary Fig. 1A).

3.3. Probe capture sequencing of artificially mixed HCV genotype samples

A subset of artificially mixed samples (*n* = 36) was sequenced using the probe capture method where there was a high correlation between the expected input ratios of various genotypes and what was observed (correlation coefficient = 0.95) (Fig. 2). Overall, the coverage following probe capture was lower compared to random primer sequencing (average 95%; range 54% - 100%) (Fig. 3A). Poorer HCV coverage occurred primarily in the low input (5% and 10%) sample genotypes from a sequencing run consisting of 24 samples. This was in contrast to a run consisting of only 12 samples, which had higher coverage for low input ratio genotypes (Supplementary Fig. 2). Two different reagents used (Illumina versus IDT reagents) and HCV genotype did not correlate with poorer coverage. The average coverage for individual genes was 96% for NS3, 94% for NS5A and 93% for NS5B.

The average read depth was on average 3609 reads/position (range 62–14,572 reads/position) (Supplementary Fig. 1B), which was a 4-fold increase relative to random primer sequencing alone (Fig. 3A). Probe capture sequencing also yielded a 2.6-fold increase in the total number of HCV reads and a 4-fold increase in the percentage of reads mapping to HCV compared to random primer sequencing (Fig. 3). It is noteworthy that the probe capture sequencing method requires an additional two days of laboratory work per run to complete compared to random primer sequencing.
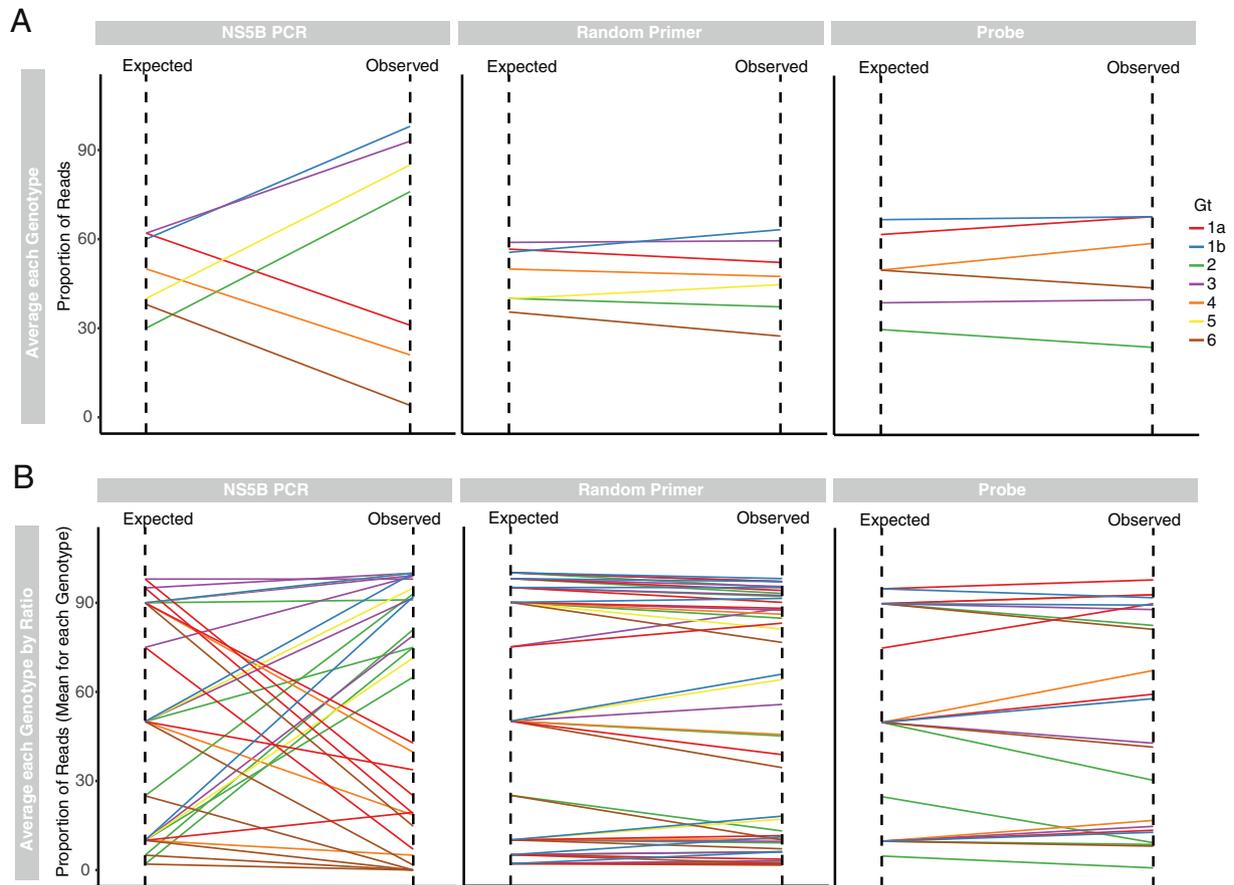
Fig. 2. Mean proportion of observed versus expected Hepatitis C virus (HCV) reads for various HCV genotypes in artificially mixed samples sequenced following Non-Structural (NS)5B PCR, randomly primed cDNA synthesis and probe capture. Note that observed reads are those that map to any position on the HCV genome. A) Means were calculated based on the average of all input ratios, or B) individual input ratios, for each genotype in artificially mixed samples (see Supplementary Tables 1 and 2).
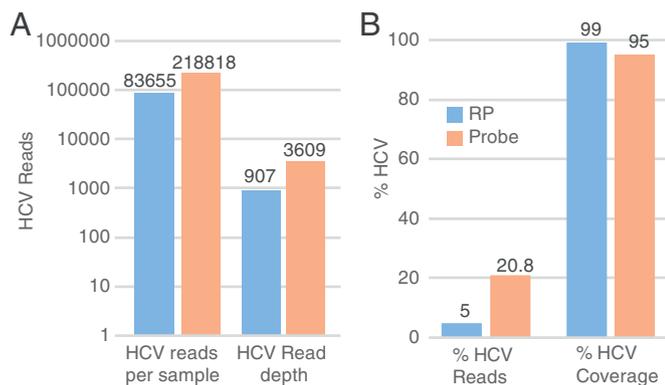

Fig. 3. Comparison of random primer versus capture probe sequencing in artificially mixed genotype Hepatitis C virus (HCV) samples. A) HCV reads/sample and Read Depth and B) Percent HCV reads out of total reads and percent coverage of NS3, NS5A and NS5B.

3.4. Criteria for identifying mixed genotype HCV infections

Although HCV mixed infections may be confidently identified when there is a high depth and coverage of two or more genotypes in a sample, detection of genotypes with low read counts are difficult to interpret. We performed an ROC curve analysis on average read depth, percent depth and percent coverage of 64 artificially mixed samples sequenced with the random primer method (Supplementary Table 1)

(Fig. 4) in order to identify cutoffs for classifying mixed infections. The area under the ROC curves were slightly less than one (0.990–0.998) due to the fact that 3 of 5 sample genotypes with 2% input ratios (the lowest mixed genotype ratio examined) had lower coverages/depths than some genotypes that were falsely identified in other samples. In order to optimize both sensitivity and specificity for identifying HCV mixed infections, it was determined that each genotype should have a minimum percent depth of 1.8%, percent coverage of 98.2% and 39.2 reads per position (average depth) to be confidently identified as a mixed genotype infection.

3.5. Phylogenetic analysis of artificially mixed samples

Phylogenetic analysis was performed using the consensus sequences of primary and secondary genotypes identified in artificially mixed genotype samples sequenced with the random primer method. Both primary and secondary consensus sequences were confirmed to cluster correctly with their respective genotype reference as visualized by clustering of reference and cohort sequences (both primary and secondary genotypes) labeled by similar colors in the phylogenetic tree (Supplementary Fig. 3).

3.6. Random primer and probe capture sequencing of cohort samples

The methodology for identifying mixed infections was applied to two clinical trials of high-risk individuals undergoing HCV treatment, ACTIVATE and DARE-C II. The ACTIVATE study recruited 93 individuals with recent IDU or on opioid substitution therapy from clinics
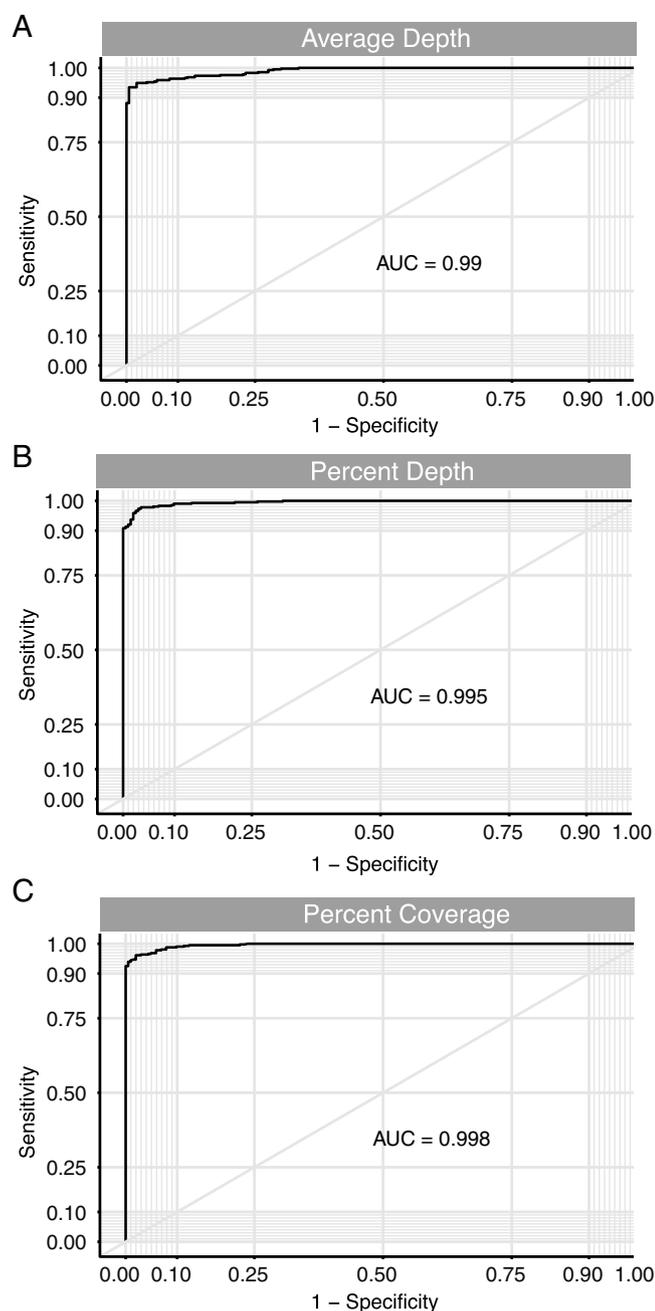
Fig. 4. Receiver Operating Characteristic (ROC) curve analysis of 64 artificially mixed genotype Hepatitis C virus (HCV) samples. Plots are sensitivity versus 1–specificity of average depth, percent depth (of total HCV depth), and percent coverage of various HCV genotypes present in artificially mixed samples. Area under the curve (AUC) provides a measure of how well the parameters can be used to differentiate mixed from non-mixed genotype infections.

in Canada, Australia and five European countries and treated them with pegylated interferon-alpha-2b and ribavirin for 12 or 24 weeks. In this study, a total of 114 samples from 90 individuals were evaluated (three individual's samples were not sequenced due to low viral loads). Of these, 77 individuals had one sample, six had two samples, three had three samples and four had four samples collected from various time points before, during or after treatment. The DARE-C II study recruited 19 individuals with recent HCV infection from tertiary hospitals in Australia and New Zealand for 6 weeks of sofosbuvir and ribavirin therapy. In this study, 26 samples from 18 individuals were evaluated (one individual's samples were not sequenced due to low viral load). Of these, 10 individuals had one sample and eight had two samples.

All of the ACTIVATE and DARE-C II samples were sequenced using the random primer approach as this method proved sufficient to identify mixed HCV genotypes in all artificially mixed samples, and because of the increased hands on time associated with probe capture sequencing. A random subset of ACTIVATE samples ($n = 16$) were also sequenced with two different probe capture reagents. Samples that had low read counts with random primer sequencing or potential mixed infections were also sequenced using the probe capture method (ACTIVATE, $n = 10$; DARE-C II, $n = 12$).

Random primer sequencing of these samples yielded on average 1.5×10^5 total HCV reads/sample with an average 17% of reads mapping to HCV. Probe capture increased the number to an average of 1.4×10^6 reads/sample with 60% of the reads mapping to HCV. Correlation between the total number of HCV reads/sample and the pre-determined viral load for both sequencing methods was observed (Fig. 5A). The majority of the samples had complete or near complete coverage while a wide range of depths were observed (Fig. 5B).

The distribution of major genotypes in the samples was: 15% ($n = 21$) G1a, 76% ($n = 106$) G3, and 9% ($n = 13$) G2. The primary genotypes were represented on average by 1.4×10^5 reads for random primer sequencing and 1.4×10^6 reads using probe capture sequencing. On average 95% of the HCV reads mapped to the primary genotype and 2% mapped to a second genotype using random primer sequencing, while 98% versus 2% of the reads mapped to primary and secondary genotypes, respectively, using probe capture sequencing.

Initially, several samples appeared to contain mixed genotype infections. For example, six random primer samples had 5–25% of the total HCV reads mapping to a secondary genotype; however, these samples had very low read counts (61 to 933 HCV reads/sample). Eight additional samples appeared to contain mixed infections with > 4000 reads (up to 12,233 reads) mapped to secondary genotypes but these were represented by only 1–2% of the total HCV reads.

When the cutoffs for percent coverage, percent depth and average depth were applied to the ACTIVATE and DARE-C II samples sequenced using random primer or probe capture sequencing, only one primary HCV genotype was confirmed in every sample (i.e. no mixed genotype infections were identified).

Two of the samples had secondary genotypes that passed the cutoff for average depth only and would have been classified as mixed G3 and G6 if they had passed all three cutoffs. Also, four samples had secondary genotypes with depth and coverage values falling just below the cutoffs and would be classified as mixed G1a and G1b if the secondary genotypes had passed the thresholds.

3.7. Phylogenetic analysis of clinical HCV infections

To rule out the possibility of the cutoffs being too strict, a phylogenetic analysis was performed on the consensus sequences for primary and secondary genotypes identified in the above-mentioned samples that had values close to the cutoffs for mixed infections, plus one additional sample with the highest number of reads mapping to a second genotype (seven samples total). Phylogenetic analysis demonstrated that none of these samples represented true mixed infections as secondary genotypes clustered more closely to their respective primary genotype than to the predicted reference sequence (Fig. 6 and Supplementary Figs. 4 and 5). This clustering pattern was confirmed using the probe capture derived consensus sequences (data not shown). The coverage of NS3, NS5A and NS5B in putative secondary genotypes ranged from 57 to 95% (mean 79%).

Lastly, our methodology was applied to three samples from two individuals in the CCC cohort, suspected of having mixed infections based on a large number of reads mapping to two different genotypes in each sample. In one individual's sample (1_CCC), 51% and 48% of the HCV reads mapped to G1a and G3, respectively. This sample passed the cutoff criteria for mixed infection, which was confirmed using phylogenetic analysis (Fig. 6).
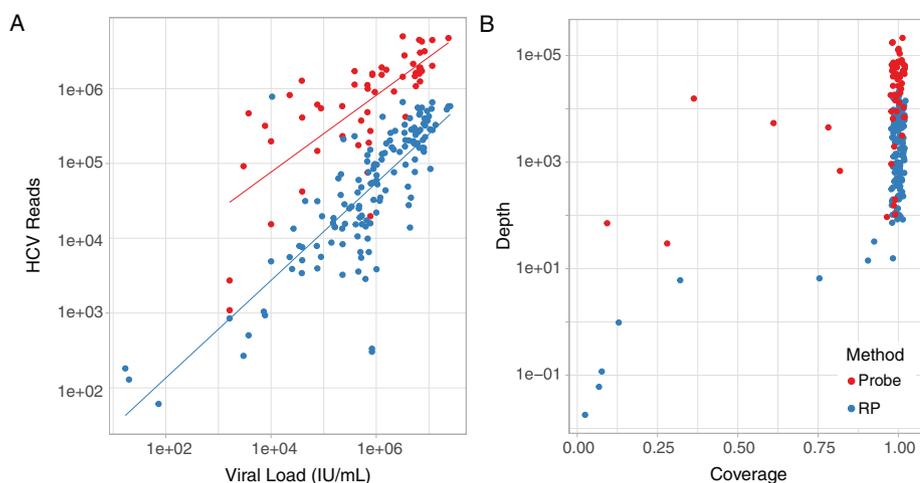
Fig. 5. Random Primer and Probe Capture sequencing of high-risk Hepatitis C virus (HCV) cohorts, ACTIVATE and DARE-C II. A) HCV viral load versus the total number of HCV reads in a sample for random primer (correlation coefficient 0.70) and probe capture (correlation coefficient 0.72) sequencing. B) The average total HCV depth (for Non-Structural (NS)3, NS5A, NS5B) versus coverage for cohort samples sequenced using random primer and probe capture sequencing.

A second individual with two samples (2_CCC, time point 1 and 2) had ~85% and 14% of the HCV reads mapping to G1a and G1b, respectively. Neither sample passed the mixed infection cutoff criteria but phylogenetic analysis yielded interesting results. In NS3 and NS5A, both G1a and G1b consensus sequences clustered with G1a references (Supplementary Figs. 4 and 5). In NS5B, G1a and G1b consensus sequences clustered with G1b references (Fig. 6). Further, we mapped G1a and G1b reads from random primer sequencing across the entire HCV genome and determined that reads map to G1a upstream of nucleotide position 7241 (located within NS5A), while reads downstream of this position map to G1b. This analysis suggests individual 2_CCC is infected with a unique recombinant HCV with the breakpoint occurring in NS5A (Supplementary Fig. 6).

4. Discussion

This study developed a novel methodology for identifying mixed-genotype HCV infections. Three methods were evaluated for generating sequencing libraries. Deep amplicon sequencing was biased towards several HCV genotypes while random primer and probe capture sequencing accurately captured the relative proportions of HCV genotypes in artificially mixed genotype samples. For analysis, a large reference library and relaxed alignment settings allowed detection of multiple HCV genotypes in a sample while cutoff values based on depth and coverage limited false mixed-genotype interpretations. Applying our methodology to 140 samples from two studies of high-risk individuals, surprisingly, did not identify any mixed genotype infections. However, in two other cases, the methodology confirmed one suspected mixed genotype infection and ruled out another as a novel recombinant HCV genotype.

The HCV genotype an individual may be infected with has important implications for treatment, vaccine development, epidemiology and progression of disease. Many rapid, commercial HCV genotyping assays are available but are limited in their ability detect mixed and recombinant genotype infections due in part to their focus on more conserved HCV genomic regions (Avó et al., 2013; Gao, 2012; Larrat et al., 2013; Yang et al., 2014). Laboratory developed real-time PCR assays offer better resolution but require a separate probe for each subtype of interest, which may limit detection of mixed infections involving rare-subtypes or when unexpected sequence variation is present (Davalieva et al., 2014; Lindh and Hannoun, 2005; Nakatani et al., 2010; Olmstead et al., 2017). Sanger sequencing is limited because genotypes that make up < 20% of the viral population cannot be detected unless more laborious strategies are employed (e.g. clonal sequencing). Deep sequencing on the other hand can be used to look at large portions or the entire viral genome, can detect low frequency

variants, and template-independent methods can be used (del Campo et al., 2017; Minosse et al., 2016; Qiu et al., 2015; Quer et al., 2015; Thomson et al., 2016; Wei et al., 2016). Although PCR amplification is an efficient strategy for generating deep sequencing libraries, the primers can be biased towards specific genotypes or subtypes as we demonstrated here. Despite the fact that the primers used in this study for NS5B amplification were previously shown to be equally sensitive in amplifying multiple genotypes in isolation (Koletzki et al., 2010), the amplification bias observed in this study is consistent with challenges in designing primers in general for HCV (Murphy et al., 2007; Olmstead et al., 2017), as well as with results from related studies attempting to amplify multiple species from a mixed population sample (Brooks et al., 2015; Lee et al., 2012).

To avoid the bias of PCR, we optimized two template-independent methodologies for detecting mixed HCV infections. While both methods accurately captured the relative proportion of HCV genotypes in a sample, random primer sequencing was preferred because it requires significantly less hands-on time. Due to the increased HCV read depth and relative proportion of HCV reads provided by probe capture sequencing more samples can be processed per run but, this method requires a large financial investment in probes, which should be considered when calculating overall costs.

Similar laboratory sequencing methods as those described here have been previously reported; however, many of these studies either did not investigate the application to mixed infection or did not develop a standardized strategy for ruling out false positive results (Bagaglio et al., 2015; Qiu et al., 2015; Thomson et al., 2016; Wei et al., 2016). Developing standardized sequencing analysis methods is critical for clinically relevant studies especially considering the potential for sequencing errors, cross-contamination, and bleed-through in many common sequencing platforms (McElroy et al., 2014; Schirmer et al., 2015; Schnell et al., 2015).

Applying our methodology to two high-risk cohorts of HCV infected individuals, ACTIVATE and DARE-C II, surprisingly, did not identify any mixed infections. Although on average 1–2% (as high as 25%) of HCV reads matched to a second genotype in random primer samples, none of these passed our cutoff criteria. To ensure that the cutoffs were not too strict we performed phylogenetic analysis on the consensus sequences of the putative secondary genotypes in a subset of samples, which were found to cluster with their respective primary genotype, not the predicted reference, suggesting these reads were incorrectly classified. This indicates that mixed genotypes did not impact treatment failure in these cohorts. Other factors such as poor adherence, cirrhosis and shortened treatment durations were previously suggested to influence lack of treatment response in these cohorts (Grebely et al., 2017; Martinello et al., 2016). It is notable that longitudinal samples from
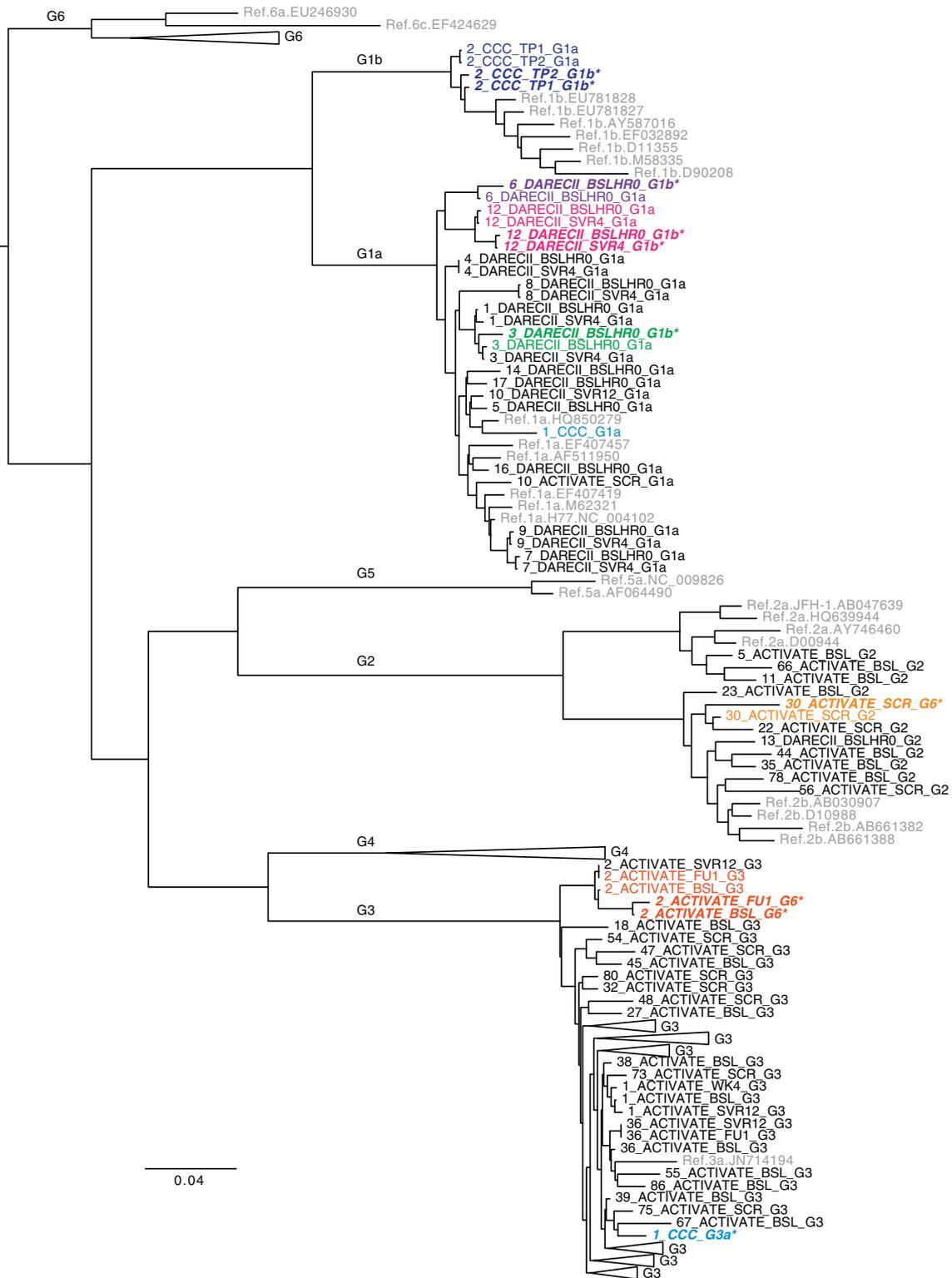
Fig. 6. Phylogenetic analysis of ACTIVATE, DARE-C II and CCC (Non-Structural) NS5B consensus sequences. Samples with putative mixed infections are highlighted with different colors; several individuals are represented by sequences from multiple time points. Potential secondary genotypes are bold-italicized and indicated with an asterisk. Reference sequences are in grey. Parts of the tree have been collapsed to better visualize relevant branches. Naming convention for clinical samples is: PatientID_Cohort_Timepoint_PredictedGenotype. Genotype = G; screening = SCR; baseline = BSL; treatment week 4 = WK4; end of treatment = ETR; 4, 12, 24 weeks post-treatment = SVR4, SVR12, SVR24; follow-up = FU1; time point 1 = TP1; time point 2 = TP2.

these individuals also clustered together indicating that no reinfections or genotype switching occurred.

The absence of mixed infections in these cohorts was surprising as the majority of individuals had a history of IDU, many within 1 to

6 months prior to study enrollment. Rates of mixed infections detected in cohorts of PWID varies considerably from study to study ranging from 14% to as high as 39% when sensitive assays are used (Herring et al., 2004; Pham et al., 2010; van de Laar et al., 2009). The

discrepancy between our study and previous studies may be that the rates of mixed infection in ACTIVATE and DARE-C II are in fact lower than in other IDU cohorts. In ACTIVATE, individuals were recruited from hospital and community-based drug and alcohol clinics, and this active engagement in the health care system may be associated with greater awareness and caution in terms of risks of virus transmission through IDU (Alavi et al., 2015; Bruneau et al., 2014; Kwiatkowski et al., 2002; Vidal-Trécan et al., 2000). An additional possibility is that some studies are overestimating the prevalence of mixed infections due to lack of standardized methods for differentiating true mixed infections from recombinants or false positive results.

Overall, the methodology presented here will be valuable for further investigating the prevalence of mixed infections in populations of interest and the sequences obtained can be used to characterize rare genotypes and recombinants, and to identify resistance associated substitutions. These studies have important implications for HCV treatment, disease progression and future vaccine development strategies.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.meegid.2019.01.016.

References

Abdel-Hakeem, M.S., Shoukry, N.H., 2014. Protective immunity against hepatitis C: many shades of gray. Front. Immunol. 5, 274. https://doi.org/10.3389/fimmu.2014.00274.
Abdelrahman, T., Hughes, J., Main, J., McLauchlan, J., Thursz, M., Thomson, E., 2015. Next-generation sequencing sheds light on the natural history of hepatitis C infection in patients who fail treatment. Hepatology 61, 88–97. https://doi.org/10.1002/hep.27192.
Aird, D., Ross, M.G., Chen, W.-S., Danielsson, M., Fennell, T., Russ, C., Jaffe, D.B., Nusbaum, C., Gnirke, A., 2011. Analyzing and minimizing PCR amplification bias in Illumina sequencing libraries. Genome Biol. 12, R18. https://doi.org/10.1186/gb-2011-12-2-r18.
Alavi, M., Spelman, T., Matthews, G.V., Haber, P.S., Day, C., van Beek, I., Walsh, N., Yeung, B., Bruneau, J., Petoumenos, K., Dolan, K., Kaldor, J.M., Dore, G.J., Hellard, M., Grebely, J., ATAHC Study Group, 2015. Injecting risk behaviours following treatment for hepatitis C virus infection among people who inject drugs: the Australian Trial in Acute Hepatitis C. Int. J. Drug Policy 26, 976–983. https://doi.org/10.1016/j.drugpo.2015.05.003.
Avó, A.P., Água-Doce, I., Andrade, A., Pádua, 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, 815–822. https://doi.org/10.1002/jmv.23545.
Bagaglio, S., Uberti-Foppa, C., Di Serio, C., Trentini, F., Andolina, A., Hasson, H., Messina, E., Merli, M., Porrino, L., Lazzarin, A., Morsica, G., 2015. Dynamic of mixed HCV Infection in Plasma and PBMC of HIV/HCV patients under Treatment with Peg-IFN/Ribavirin. Medicine. https://doi.org/10.1097/MD.0000000000001876. (Baltimore). 94, e1876.
Brooks, J.P., Edwards, D.J., Harwich, M.D., Rivera, M.C., Fettweis, J.M., Serrano, M.G., Reris, R.A., Sheth, N.U., Huang, B., Girerd, P., Strauss, J.F., Jefferson, K.K., Buck, G.A., 2015. The truth about metagenomics: quantifying and counteracting bias in 16S

rRNA studies. BMC Microbiol. 15, 66. https://doi.org/10.1186/s12866-015-0351-6.
Bruneau, J., Zang, G., Abrahamowicz, M., Jutras-Aswad, D., Daniel, M., Roy, E., 2014. Sustained drug use changes after hepatitis C screening and counseling among recently Infected Persons who inject drugs: a Longitudinal Study. Clin. Infect. Dis. 58, 755–761. https://doi.org/10.1093/cid/cit938.
Cunningham, E.B., Applegate, T.L., Lloyd, A.R., Dore, G.J., Grebely, J., 2015. Mixed HCV infection and reinfection in people who inject drugs—impact on therapy. Nat. Rev. Gastroenterol. Hepatol. 12, 218–230. https://doi.org/10.1038/nrgastro.2015.36.
Davalieva, K., Kiprijanovska, S., Plaseska-Karanfilska, D., 2014. Fast, reliable and low cost user-developed protocol for detection, quantification and genotyping of hepatitis C virus. J. Virol. Methods 196, 104–112. https://doi.org/10.1016/j.jviromet.2013.11.002.
del Campo, J.A., Parra-Sánchez, M., Figueruela, B., García-Rey, S., Quer, J., Gregori, J., Bernal, S., Grande, L., Palomares, J.C., Romero-Gómez, M., 2017. HCV deep-sequencing for sub-genotypes identification of mixed infections: a real life experience. Int. J. Infect. Dis. 67, 114–117. https://doi.org/10.1016/j.ijid.2017.12.016.
Dore, G.J., Feld, J.J., 2015. Hepatitis C virus therapeutic development: in pursuit of Perfectovir. Clin. Infect. Dis. 60, 1829–1836. https://doi.org/10.1093/cid/civ197.
Falade-Nwulia, O., Suarez-Cuervo, C., Nelson, D.R., Fried, M.W., Segal, J.B., Sulkowski, M.S., 2017. Oral direct-acting agent therapy for hepatitis C virus infection. Ann. Intern. Med. 166, 637. https://doi.org/10.7326/M16-2575.
Gao, Z., 2012. Comparison of three different HCV genotyping methods: Core, NS5B sequence analysis and line probe assay. Int. J. Mol. Med. https://doi.org/10.3892/ijmm.2012.1209.
Grebely, J., Pham, S.T., Matthews, G.V., Petoumenos, K., Bull, R.A., Yeung, B., Rawlinson, W., Kaldor, J., Lloyd, A., Hellard, M., Dore, G.J., White, P.A., ATAHC Study Group, 2012a. Hepatitis C virus reinfection and superinfection among treated and untreated participants with recent infection. Hepatology 55, 1058–1069. https://doi.org/10.1002/hep.24754.
Grebely, J., Prins, M., Hellard, M., Cox, A.L., Osburn, W.O., Lauer, G., Page, K., Lloyd, A.R., Dore, G.J., 2012b. Hepatitis C virus clearance, reinfection, and persistence, with insights from studies of injecting drug users: towards a vaccine. Lancet Infect. Dis. 12, 408–414. https://doi.org/10.1016/S1473-3099(12)70010-5.
Grebely, J., Dalgard, O., Cunningham, E.B., Hajarizadeh, B., Foster, G.R., Bruggmann, P., Conway, B., Backmund, M., Robaeys, G., Swan, T., Amin, J., Marks, P.S., Quiene, S., Applegate, T.L., Weltman, M., Shaw, D., Dunlop, A., Hellard, M., Bruneau, J., Midgard, H., Bourgeois, S., Staehelin, C., Dore, G.J., ACTIVATE Study Group, 2017. Efficacy of response-guided directly observed pegylated interferon and self-administered ribavirin for people who inject drugs with hepatitis C virus genotype 2/3 infection: the ACTIVATE study. Int. J. Drug Policy 47, 177–186. https://doi.org/10.1016/j.drugpo.2017.05.020.
Herring, B.L., Page-Shafer, K., Tobler, L.H., Delwart, E.L., 2004. Frequent Hepatitis C virus superinfection in injection drug users. J. Infect. Dis. 190, 1396–1403. https://doi.org/10.1086/424491.
Kanagawa, T., 2003. Bias and artifacts in multitemplate polymerase chain reactions(PCR). J. Biosci. Bioeng. 96, 317–323. https://doi.org/10.1263/jbb.96.317.
Klein, M.B., Saeed, S., Yang, H., Cohen, J., Conway, B., Cooper, C., Cote, P., Cox, J., Gill, J., Haase, D., Haider, S., Montaner, J., Pick, N., Rachlis, A., Rouleau, D., Sandre, R., Tyndall, M., Walmsley, S., 2010. Cohort profile: the Canadian HIV-Hepatitis C Co-infection Cohort Study. Int. J. Epidemiol. 39, 1162–1169. https://doi.org/10.1093/ije/dyp297.
Koletzki, D., Dumont, S., Vermeiren, H., Fevery, B., De Smet, P., Stuyver, L.J., 2010. Development and evaluation of an automated hepatitis C virus NS5B sequence-based subtyping assay. Clin. Chem. Lab. Med. 48, 1095–1102. https://doi.org/10.1515/CCLM.2010.236.
Kuiken, C., Yusim, K., Boykin, L., Richardson, R., 2005. The Los Alamos HCV Sequence Database. Bioinformatics 21, 379–384.
Kwiatkowski, C.F., Fortuin Corsi, K., Booth, R.E., 2002. The association between knowledge of hepatitis C virus status and risk behaviors in injection drug users. Addiction 97, 1289–1294. https://doi.org/10.1046/j.1360-0443.2002.00208.x.
Langmead, B., Salzberg, S.L., 2012. Fast gapped-read alignment with Bowtie 2. Nat. Methods 9, 357–359. https://doi.org/10.1038/nmeth.1923.
Larrat, S., Poveda, J.-D., Coudret, C., Fusillier, K., Magnat, N., Signori-Schmuck, A., Thibault, V., Morand, P., 2013. Sequencing assays for failed genotyping with the versant hepatitis C virus genotype assay (LiPA). J. Clin. Microbiol. 51, 2815–2821. https://doi.org/10.1128/JCM.00586-13.
Lee, C.K., Herbold, C.W., Polson, S.W., Wommack, K.E., Williamson, S.J., McDonald, I.R., Cary, S.C., 2012. Groundtruthing next-gen sequencing for microbial ecology–biases and errors in community structure estimates from PCR amplicon pyrosequencing. PLoS One 7, e44224. https://doi.org/10.1371/journal.pone.0044224.
Lindh, M., Hannoun, C., 2005. Genotyping of hepatitis C virus by Taqman real-time PCR. J. Clin. Virol. 34, 108–114. https://doi.org/10.1016/j.jcv.2005.02.002.
Manso, C.F., Bibby, D.F., Mbisa, J.L., 2017. Efficient and unbiased metagenomic recovery of RNA virus genomes from human plasma samples. Sci. Rep. 7, 4173. https://doi.org/10.1038/s41598-017-02239-5.
Martin, M., 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet J. 17, 10. https://doi.org/10.14806/ej.17.1.200.
Martinello, M., Gane, E., Hellard, M., Sasadeusz, J., Shaw, D., Petoumenos, K., Applegate, T., Grebely, J., Maire, L., Marks, P., Dore, G.J., Matthews, G.V., 2016. Sofosbuvir and ribavirin for 6 weeks is not effective among people with recent hepatitis C virus infection: the DARE-C II study. Hepatology 64, 1911–1921. https://doi.org/10.1002/hep.28844.
Matranga, C.B., Andersen, K.G., Winnicki, S., Busby, M., Gladden, A.D., Tewhey, R., Stremlau, M., Berlin, A., Gire, S.K., England, E., Moses, L.M., Mikkelsen, T.S., Odia, I., Ehiane, P.E., Folarin, O., Goba, A., Kahn, S.H., Grant, D.S., Honko, A., Hensley, L., Happi, C., Garry, R.F., Malboeuf, C.M., Birren, B.W., Gnirke, A., Levin, J.Z., Sabeti,

P.C., 2014. Enhanced methods for unbiased deep sequencing of Lassa and Ebola RNA viruses from clinical and biological samples. Genome Biol. 15, 519. https://doi.org/10.1186/PREACCEPT-1698056557139770.

McElroy, K., Thomas, T., Luciani, F., 2014. Deep sequencing of evolving pathogen populations: applications, errors, and bioinformatic solutions. Microb. Inform. Exp. 4, 1. https://doi.org/10.1186/2042-5783-4-1.

McNaughton, A.L., Thomson, E.C., Templeton, K., Gunson, R.N., Leitch, E.C.M., 2014. Mixed genotype hepatitis C infections and implications for treatment. Hepatology 59, 1209. https://doi.org/10.1002/hep.26544.

McNaughton, A.L., Sreenu, V.B., Wilkie, G., Gunson, R., Templeton, K., Leitch, E.C.M., 2017. Prevalence of mixed genotype hepatitis C virus infections in the UK as determined by genotype-specific PCR and deep sequencing. J. Viral Hepat. https://doi.org/10.1111/jvh.12849.

Minosse, C., Giombini, E., Bartolini, B., Capobianchi, M.R., Garbuglia, A.R., 2016. Ultradeep sequencing characterization of HCV samples with equivocal typing results determined with a commercial assay. Int. J. Mol. Sci. 17. https://doi.org/10.3390/ijms17101679.

Murphy, D.G., Willems, B., Deschenes, M., Hilzenrat, N., Mousseau, R., Sabbah, S., 2007. Use of sequence analysis of the NS5B region for routine genotyping of hepatitis C virus with reference to C/E1 and 5' untranslated region sequences. J. Clin. Microbiol. 45, 1102–1112. https://doi.org/10.1128/JCM.02366-06.

Nakatani, S.M., Santos, C.A., Riediger, I.N., Krieger, M.A., Duarte, C.A.B., Lacerda, M.A., Biondo, A.W., Carrilho, F.J., Carilho, F.J., Ono-Nita, S.K., 2010. Development of hepatitis C virus genotyping by real-time PCR based on the NS5B region. PLoS One 5, e10150. https://doi.org/10.1371/journal.pone.0010150.

Olmstead, A.D., Lee, T.D., Chow, R., Gunadasa, K., Auk, B., Krajden, M., Jassem, A.N., 2017. Development and validation of a real-time, reverse transcription PCR assay for rapid and low-cost genotyping of hepatitis C virus genotypes 1a, 1b, 2, and 3a. J. Virol. Methods 244, 17–22. https://doi.org/10.1016/j.jviromet.2017.02.009.

Osburn, W.O., Fisher, B.E., Dowd, K.A., Urban, G., Liu, L., Ray, S.C., Thomas, D.L., Cox, A.L., 2010. Spontaneous control of primary hepatitis C virus infection and immunity against persistent reinfection. Gastroenterology 138, 315–324. https://doi.org/10.1053/j.gastro.2009.09.017.

Pham, S.T., Bull, R.A., Bennett, J.M., Rawlinson, W.D., Dore, G.J., Lloyd, A.R., White, P.A., 2010. Frequent multiple hepatitis C virus infections among injection drug users in a prison setting. Hepatology 52, 1564–1572. https://doi.org/10.1002/hep.23885.

Polaris Observatory HCV Collaborators, 2017. Global prevalence and genotype distribution of hepatitis C virus infection in 2015: a modelling study lancet. Gastroenterol. Hepatol. 2, 161–176. https://doi.org/10.1016/S2468-1253(16)30181-9.

Price, M.N., Dehal, P.S., Arkin, A.P., 2010. FastTree 2–approximately maximum-likelihood trees for large alignments. PLoS One 5, e9490. https://doi.org/10.1371/journal.pone.0009490.

Qiu, P., Stevens, R., Wei, B., Lahser, F., Howe, A.Y.M., Klappenbach, J.A., Marton, M.J., 2015. HCV genotyping from NGS short reads and its application in genotype detection from HCV mixed infected plasma. PLoS One 10, e0122082. https://doi.org/10.1371/journal.pone.0122082.

Quer, J., Gregori, J., Rodríguez-Frias, F., Buti, M., Madejon, A., Perez-del-Pulgar, S., Garcia-Cehic, D., Casillas, R., Blasi, M., Homs, M., Tabernero, D., Alvarez-Tejado, M., Muñoz, J.M., Cubero, M., Caballero, A., delCampo, J.A., Domingo, E., Belmonte, I., Nieto, L., Lens, S., Muñoz-de-Rueda, P., Sanz-Cameno, P., Sauleda, S., Bes, M., Gomez, J., Briones, C., Perales, C., Sheldon, J., Castells, L., Viladomiu, L., Salmeron, J., Ruiz-Extremera, A., Quiles-Pérez, R., Moreno-Otero, R., López-Rodríguez, R., Allende, H., Romero-Gómez, M., Guardia, J., Esteban, R., Garcia-Samaniego, J., Forns, X., Esteban, J.I., 2015. High-resolution hepatitis C virus subtyping using NS5B deep sequencing and phylogeny, an alternative to current methods. J. Clin. Microbiol. 53, 219–226. https://doi.org/10.1128/JCM.02093-14.

R Core Team, 2013. R : A Language and Environment for Statistical Computing. Vienna, Austria.

Schirmer, M., Ijaz, U.Z., D'Amore, R., Hall, N., Sloan, W.T., Quince, C., 2015. Insight into biases and sequencing errors for amplicon sequencing with the Illumina MiSeq platform. Nucleic Acids Res. 43, e37. https://doi.org/10.1093/nar/gku1341.

Schnell, I.B., Bohmann, K., Gilbert, M.T.P., 2015. Tag jumps illuminated - reducing sequence-to-sample misidentifications in metabarcoding studies. Mol. Ecol. Resour. 15, 1289–1303. https://doi.org/10.1111/1755-0998.12402.

Smith, D.B., Bukh, J., Kuiken, C., Muerhoff, A.S., Rice, C.M., Stapleton, J.T., Simmonds, P., 2014. Expanded classification of hepatitis C virus into 7 genotypes and 67 subtypes: updated criteria and genotype assignment web resource. Hepatology 59, 318–327. https://doi.org/10.1002/hep.26744.

Thomson, E., Ip, C.L.C., Badhan, A., Christiansen, M.T., Adamson, W., Ansari, M.A., Bibby, D., Breuer, J., Brown, A., Bowden, R., Bryant, J., Bonsall, D., Da Silva Filipe, A., Hinds, C., Hudson, E., Klenerman, P., Lythgow, K., Mbisa, J.L., McLauchlan, J., Myers, R., Piazza, P., Roy, S., Trebes, A., Sreenu, V.B., Witteveldt, J., STOP-HCV Consortium, S.-H., Barnes, E., Simmonds, P., 2016. Comparison of next-generation sequencing technologies for comprehensive assessment of full-length hepatitis C viral genomes. J. Clin. Microbiol. 54, 2470–2484. https://doi.org/10.1128/JCM.00330-16.

Trémeaux, P., Caporossi, A., Thélu, M.-A., Blum, M., Leroy, V., Morand, P., Larrat, S., 2016. Hepatitis C virus whole genome sequencing: current methods/issues and future challenges. Crit. Rev. Clin. Lab. Sci. 53, 341–351. https://doi.org/10.3109/10408363.2016.1163663.

van de Laar, T.J.W., Molenkamp, R., van den Berg, C., Schinkel, J., Beld, M.G.H.M., Prins, M., Coutinho, R.A., Bruisten, S.M., 2009. Frequent HCV reinfection and superinfection in a cohort of injecting drug users in Amsterdam. J. Hepatol. 51, 667–674. https://doi.org/10.1016/j.jhep.2009.05.027.

Vidal-Trécan, G., Coste, J., Varescon-Pousson, I., Christoforov, B., Boissonnas, A., 2000. HCV status knowledge and risk behaviours amongst intravenous drug users. Eur. J. Epidemiol. 16, 439–445. https://doi.org/10.1023/A:1007622831518.

Wei, B., Kang, J., Kibukawa, M., Chen, L., Qiu, P., Lahser, F., Marton, M., Levitan, D., 2016. Development and validation of a template-independent next-generation sequencing assay for detecting low-level resistance-associated variants of hepatitis C virus. J. Mol. Diagnostics 18, 643–656. https://doi.org/10.1016/J.JMOLDX.2016.04.001.

Yang, R., Cong, X., Du, S., Fei, R., Rao, H., Wei, L., 2014. Performance comparison of the versant HCV genotype 2.0 assay (LiPA) and the Abbott realtime HCV genotype II assay for detecting hepatitis C virus genotype 6. J. Clin. Microbiol. 52, 3685–3692. https://doi.org/10.1128/JCM.00882-14.