

Article

Validation of a Genotype-Independent Hepatitis C Virus Near-Whole Genome Sequencing Assay

Hope R. Lapointe ^{1,2}, Weiyan Dong ², Winnie W. Y. Dong ², Don Kirkby ², Conan Woods ², Art F. Y. Poon ³, Anita Y. M. Howe ⁴ , P. Richard Harrigan ¹ and Chanson J. Brumme ^{2,5,*} 

¹ Department of Medicine, Division of Social Medicine, University of British Columbia, Vancouver, BC V6T 1Z4, Canada; hope.lapointe@ubc.ca (H.R.L.); richard.harrigan@ubc.ca (P.R.H.)

² BC Centre for Excellence in HIV/AIDS, Vancouver, BC V6Z 1Y6, Canada; edong@bccfe.ca (W.D.); wdong@bccfe.ca (W.W.Y.D.); dkirkby@bccfe.ca (D.K.); email@woodsc.ca (C.W.)

³ Department of Pathology and Laboratory Medicine, Western University, London, ON N6A 3K7, Canada; apoon42@uwo.ca

⁴ British Columbia Centre for Disease Control, Vancouver, BC V5Z 4R4, Canada; anitahowe@live.com

⁵ Department of Medicine, Division of Infectious Diseases, University of British Columbia, Vancouver, BC V6T 1Z4, Canada

* Correspondence: cbrumme@bccfe.ca



Citation: Lapointe, H.R.; Dong, W.; Dong, W.W.Y.; Kirkby, D.; Woods, C.; Poon, A.F.Y.; Howe, A.Y.M.; Harrigan, P.R.; Brumme, C.J. Validation of a Genotype-Independent Hepatitis C Virus Near-Whole Genome Sequencing Assay. *Viruses* **2021**, *13*, 1721. <https://doi.org/10.3390/v13091721>

Academic Editors: Federico Garcia and Francesca Ceccherini-Silberstein

Received: 18 June 2021

Accepted: 26 August 2021

Published: 30 August 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Abstract: Despite the effectiveness of direct-acting antiviral agents in treating hepatitis C virus (HCV), cases of treatment failure have been associated with the emergence of resistance-associated substitutions. To better guide clinical decision-making, we developed and validated a near-whole-genome HCV genotype-independent next-generation sequencing strategy. HCV genotype 1–6 samples from direct-acting antiviral agent treatment-naïve and -treated HCV-infected individuals were included. Viral RNA was extracted using a NucliSens easyMAG and amplified using nested reverse transcription-polymerase chain reaction. Libraries were prepared using Nextera XT and sequenced on the Illumina MiSeq sequencing platform. Data were processed by an in-house pipeline (MiCall). Nucleotide consensus sequences were aligned to reference strain sequences for resistance-associated substitution identification and compared to NS3, NS5a, and NS5b sequence data obtained from a validated in-house assay optimized for HCV genotype 1. Sequencing success rates (defined as achieving >100-fold read coverage) approaching 90% were observed for most genotypes in samples with a viral load >5 log₁₀ IU/mL. This genotype-independent sequencing method resulted in >99.8% nucleotide concordance with the genotype 1-optimized method, and 100% agreement in genotype assignment with paired line probe assay-based genotypes. The assay demonstrated high intra-run repeatability and inter-run reproducibility at detecting substitutions above 2% prevalence. This study highlights the performance of a freely available laboratory and bioinformatic approach for reliable HCV genotyping and resistance-associated substitution detection regardless of genotype.

Keywords: HCV; direct-acting antiviral agent; resistance-associated substitutions; whole-genome sequencing; genotype-independent

1. Introduction

Hepatitis C virus (HCV) infection is a major public health concern in Canada and worldwide. In 2015, the World Health Organization (WHO) estimated that 71 million people globally had chronic HCV infection [1]. Nearly 400,000 die annually from HCV infection, largely due to liver cirrhosis and hepatocellular carcinoma. Despite an estimated 843,000 individuals being cured in 2015, these estimates are in sharp contrast to the number of new HCV infections globally (1.75 million). Without appropriate treatment scale-up and therapeutic monitoring, the global HCV epidemic may continue to expand in severity.

The therapeutic approach towards HCV chronic infection has shifted in recent years with the advent of oral direct-acting antiviral (DAA) agents. Previously, first-line treat-

ment options were limited to pegylated interferon-, ribavirin- and first-generation protease inhibitor-containing regimens [2]. Interferon- and ribavirin-based therapies were non-specific to HCV and caused moderate to severe side effects universally, while first-generation protease inhibitors also had moderate to high potential for drug–drug interactions [3]. Newer therapies against specific viral targets use direct-acting antiviral agents (DAA), small-molecule inhibitors that can be categorized on the basis of their viral targets. These drug targets are important for viral replication, notably non-structural proteins NS3/4a, NS5a, and NS5b.

Some DAA-based regimens are genotype-specific, requiring personalized therapy to ensure genotype coverage by antiviral agents. The recent introduction of pan-genotypic options (sofosbuvir/velpatasvir, glecaprevir/pibrentasvir) circumvent the need for genotype identification [4,5]. These pan-genotypic regimens are now recommended by the WHO as first-line therapy for all adults with HCV infection regardless of genotype [6,7]. With this highly effective repertoire of DAA medications, overall sustained virologic response rates have demonstrated >90–95% regardless of HCV genotype, baseline HCV RNA viral loads, race, HIV coinfection and hepatic fibrosis [8–11].

However, cases of DAA treatment failure have been associated with the emergence of drug-resistance-associated substitutions (RAS) in all HCV genotypes, detected in both DAA treatment-naïve and treatment-experienced people [12–15]. The failure to achieve sustained virologic response through the emergence of RAS is thought to be the result of the selective pressure of antiviral drugs during treatment or the genetic variation inherent in the virus itself. As such, the identification of RAS can enable clinicians to personalize DAA treatment for a given patient, typically through the addition of ribavirin or by extending therapy duration, thereby enhancing the likelihood of treatment success.

Currently, the use of drug resistance testing in clinical management of HCV differs slightly between national and international societies. The Canadian Association for the Study of the Liver (CASL), the European Association for the Study of the Liver (EASL) and the American Association for the Study of Liver Diseases (AASLD) recommend resistance testing in select clinical scenarios [16–18]. As most HCV-positive individuals achieve sustained virologic response, baseline resistance testing is not universally recommended for all HCV infections [16,17]. RAS identification may be recommended in DAA-treatment-naïve individuals, though almost solely for genotype 1a and 3 infections, dependent on treatment status (naïve vs. experienced), the selected treatment regimen and cirrhosis status. Resistance testing is generally not recommended in patients with other HCV genotypes (1b, 2, 4, 5 and 6). Drug resistance testing can be considered in cases of treatment failure and retreatment, though limited data are available on the clinical utility of this approach. HCV resistance testing nevertheless provides an opportunity to identify RAS, notably in the case of genotype 1a or 3 infections as well as previously treated patients who have failed an initial DAA regimen. Such testing can also be used in the context of epidemiological surveillance of HCV drug resistance to provide valuable insight into treatment and prevention strategies [19]. Drug resistance testing therefore may continue to provide decisional support for certain clinical scenarios, such as optimizing second-line therapy [20].

To our knowledge, there is no commercially available kit to perform HCV drug resistance testing for all marketed DAAs. For testing, clinicians and/or public health labs must submit samples to one of the few select labs with in-house developed assays for clinical HCV RAS identification (e.g., BC Centre for Excellence in HIV/AIDS, Quest Diagnostics, Labcorp). For instance, HCV clinical specimens in Canada are first submitted to a provincial public health laboratory, which subsequently refers the sample to the BC Centre for Excellence for HIV/AIDS for HCV drug resistance testing. Currently, there are a small number of whole-genome sequencing approaches published for HCV, though these have widely variable performance and uses, may not have been fully validated with clinical isolates and may be technically incompatible with certain HCV genotypes or impractical for clinical use [21–30]. Bioinformatic challenges also represent a notable hurdle in this field,

which currently lacks a standardized approach to analyzing next-generation sequencing data. Few open-source software for genome sequencing and drug resistance testing are available, with alternatives typically presenting with technical and financial barriers to their use.

We therefore developed and validated a near-whole genome, HCV genotype-independent sequencing assay using the MiSeq platform (Illumina) to identify RAS that may be associated with treatment failure. This assay also allows for HCV genotype classification using genomic data in lieu of conventional probe-based assays. The results produced by this assay can help guide clinicians and patients when making treatment decisions in specific clinical scenarios, as outlined by treatment guidelines. Other public health laboratories globally can adopt this HCV drug resistance assay, for which an associated bioinformatic pipeline is available specifically for generating HCV drug resistance reports.

2. Materials and Methods

2.1. Study Population

Sample sets for each assay performance parameter are outlined in Table 1. Briefly, HCV-positive samples used to develop and validate this near whole genome, HCV genotype-independent assay were sourced from Merck from the C-WORTHY trial (NCT01717326/Protocol PN035) and the Vancouver Injection Drug Users Study (VIDUS) [31,32]. Samples included specimens from treatment-naïve HCV-infected individuals, as well as DAA-treated persons who did not achieve sustained virologic response at 12 weeks post-treatment. All HCV-positive samples were tested for genotype (GT) 1–6 by sample providers themselves using the VERSANT HCV Genotype 2.0 line probe assay (LiPA; Siemens, Erlangen, Germany).

Table 1. Specimen characteristics for assessing individual assay validation parameters.

Test	No. of Samples	Replicates per Sample	HCV pVL log ₁₀ (IU/mL), Median (Range)	Genotypes (LiPA Assay)
Accuracy	93	2	6.8 (3.0–7.6)	1a (n = 78) 1b (n = 10) 1e (n = 3) 2b (n = 2)
Repeatability	4	12	6.5 (6.1–7.3)	1a (n = 3) 1b (n = 1)
Reproducibility	12	5	6.5 (6.0–7.7)	1a (n = 10) 1b (n = 1) 3 (n = 1)
Sensitivity (viral load) & Genotype coverage	148	1	6.1 (3.0–7.6)	1a (n = 78) 1b (n = 10) 1e (n = 3) 2b (n = 2) 4 (n = 8) 5 (n = 5) 6 (n = 5) Unknown (n = 37)
Sensitivity (minor species)	95	2	NA	Unknown (n = 95)
Specificity-Negatives	5	3	NA	NA

HCV, Hepatitis C Virus; pVL, plasma viral load; IU, international units; LiPA, line probe assay; NA, not available. Dashes indicate no data were available for the associated parameter.

2.2. PCR Amplification

Viral RNA was extracted from 500 µL of plasma on a NucliSENS easyMag RNA/DNA Extractor (bioMérieux, Montréal, QC, Canada) according to manufacturer's protocol. Extracted nucleic acids were eluted in 60 µL of elution buffer and stored at −20 °C until

reverse-transcription-polymerase chain reaction (RT-PCR). A negative control (DEPC-treated water) was included in each run.

Nucleic acids underwent RT-PCR using SuperScript III Reverse Transcriptase (18080-044, Invitrogen, Carlsbad, CA, USA), followed by a nested PCR using Klentaq LA (110, DNA Polymerase Technology Inc., St. Louis, MO, USA) to amplify the HCV genome in two fragments (Figure 1).

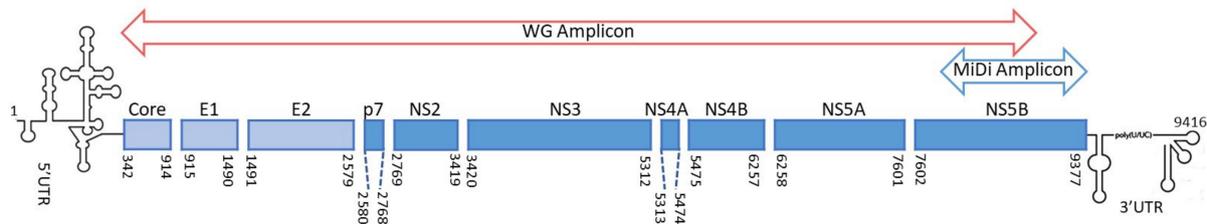


Figure 1. HCV genome map. A schematic representation of the HCV genome depicting coverage of the genotype-independent, near-whole genome sequencing method wherein the genome is amplified in two fragments then sequenced in parallel.

Firstly, one near-whole-genome amplicon (“WG amplicon”) was generated spanning the HCV Core to NS5b codon 336 (GT1a_H77 genome coordinates 342 to 8610). To cover the remaining NS5b portion, a smaller, partially-overlapping amplicon (“MiDi amplicon”) was designed spanning NS5b codons 228 to the 3’ end of NS5b (H77 genome coordinates 7829 to 9377). Nested PCR products were visualized by agarose gel electrophoresis to confirm successful amplification prior to library generation. Amplicons were normalized and purified using AMPure XP magnetic beads (A63880, Beckman Coulter, Mississauga, ON, Canada). Purified amplicon concentrations were quantified using the Invitrogen Quant-iT Picogreen dsDNA assay (P7589, Invitrogen, Carlsbad, CA, USA) and diluted to 1 ng/μL. Note that the amplicons were processed separately until the pooling of the amplicon libraries in preparation for sequencing. Primers for reverse transcription and amplification steps are included in the Supplementary Materials (Laboratory Protocol, Table S1).

2.3. Library Preparation

Individual libraries for each amplicon (MiDi and WG amplicons) were prepared using the Nextera XT DNA Library Preparation Kit (FC-131-1024, Illumina) and Nextera XT Index Kits (FC-131-1002, Illumina) for amplicon tagmentation and dual-index barcoding, respectively. Libraries were prepared according to manufacturer protocols, with a modified tagmentation step. Here, 40% working volumes were used for the tagmentation reagents with 2 ng of DNA as input (see Supplementary Materials–Methods). Indexed amplicons for each target were pooled and purified using AMPure XP beads, and the pooled library was diluted to 1.3 ng/μL. The WG and MiDi pooled libraries were then combined at a roughly 7:1 ratio, based on the approximate amplicon length. Note that the WG and MiDi amplicons from the same samples are not tagged with the same indices; unique index tag combinations are used for each amplicon, allowing for the assessment of sequence concordance in the overlapping regions. The Illumina PhiX control library was then added at 10% ratio to the sample DNA libraries and sequenced on the Illumina MiSeq platform using the 2 × 250 bp MiSeq Reagent Kit v2 (MS-102-2003, Illumina, San Diego, CA, USA). A complete standard operating procedure outlining the laboratory protocol for this assay is available in the Supplementary Materials.

2.4. Data Processing

Raw short-read MiSeq data were processed by an in-house pipeline designed for rapidly evolving RNA viruses (MiCall v6.8, available at <https://github.com/cfe-lab/MiCall> and on Illumina BaseSpace Sequence Hub). Briefly, the MiCall pipeline uses

bowtie2 to construct consensus sequences and estimate nucleotide frequencies following several steps: (1) initial read mapping to a set of HCV reference genome “seed” sequences, (2) sample-specific reference construction, remapping and reconstruction, (3) an iterative mapping process using sample-specific references, and (4) consensus sequence generation [33]. Given the extreme intra- and inter-genotype sequence diversity displayed by HCV, multiple reference sequences are used in the pipeline to maximize the number of individual MiSeq reads that are successfully mapped to a reference.

Firstly, paired-end reads are mapped to 57 HCV genotype/subtype reference sequences curated by the Los Alamos National Laboratory HCV Sequence Database [34]. After this preliminary mapping stage, HCV genotype references having ≥ 10 mapped reads are retained. Multiple different reference genotypes (e.g., GT1a, GT1b, GT2b) may be selected during this preliminary mapping; however, if more than one “seed” reference from the same genotype (e.g., GT1a, GT1b) is retained, only the seed with the most mapped reads is retained for that genotype.

Secondly, for each retained reference sequence, mapped reads are collapsed into a sample-specific reference that retains the genotype identity of the original “seed” Los Alamos reference. All reads are then remapped to each constructed sample-specific reference, which are then reconstructed using newly mapped reads.

Thirdly, reference-building and re-mapping described in the last step occurs iteratively until $\geq 95\%$ of reads map to the sample-specific reference, or no additional reads are mapped in additional rounds of re-mapping. As this process continues and additional reads are mapped, it is expected that individual sample-specific references may diverge away from their initial “seed” reference sequence. For example, it is possible for a GT2b sample-specific reference to sufficiently change to more closely resemble a GT1a reference as re-mapping occurs, and new reads are iteratively incorporated. To address this, each iteration includes a genotype check for each sample-specific reference. Briefly, a genetic distance metric (Levenshtein distance) is computed between sample-specific references and all Los Alamos reference sequences. In instances where sample-specific references have drifted to more closely resemble another genotype than the “seed” genotype, the sample-specific reference is excluded. For instance, the GT2b sample-specific reference could be excluded here, and a more suitable GT1a sample-specific reference is retained for reference-building and re-mapping. In instances where the sample-specific references do not drift away from the expected genotype, they are all retained through to the end of the pipeline. Should this occur, the pipeline outputs multiple consensus sequences from a single sample, one per sample-specific reference. Paired-end reads are merged, and error correction rules are applied; corrections included discarding nucleotides with a quality score < 15 and resolving mismatching base calls in overlapping reads by retaining the higher quality base.

Lastly, all paired-end reads are used to generate nucleotide consensus sequences at various nucleotide prevalence thresholds (1%, 2%, 5%, 10%, 20%, 25%); nucleotide mixtures were called when two or more bases were present above the prevalence cutoff. Amino acid frequencies and sequences were translated directly from paired-end reads and used for RAS analysis. Finally, the MiCall pipeline merges sequence data from the two amplicons WG and MiDi to generate single, near-full genome sequences. The software resolves the overlapping region between the WG and MiDi amplicons by retaining only the data from the amplicon with the greatest coverage. The genotype of the final assembled consensus is inferred again by calculating the shortest Levenshtein distance to the Los Alamos reference set. The resulting dataset covers the near whole-genome of HCV. To demonstrate that both amplicons generate concordant sequence data, we examined nucleotide agreement and RAS quantification within the overlapping segment of both products.

2.5. Quality Control Criteria

MiCall incorporates a final quality control step that verifies sequence coverage check at key resistance-associated positions. Samples with > 1000 -fold coverage at all RAS positions

are automatically approved. Sequences with lower read coverage (100 to 999 reads) at ≥ 1 RAS position are flagged for manual review. Any sequences with fewer than 100 reads at any RAS position are considered to have failed and are excluded from downstream data analyses. Additionally, at least one of the sample's gene regions (e.g., NS3, NS5a, or NS5b) must have at least 100-fold coverage at all RAS positions or will be deemed a failure.

2.6. Genotype Inference & RAS

For analytical purposes, sample genotypes were inferred from one of two sources, MiSeq genomic sequence data and LiPA assay results. Where sequencing was successful, genotypes were inferred according to the data processing outlined above. In instances where no MiSeq sequence was generated during this validation and HCV genotypes could not be inferred, viral classification was based on the LiPA assay genotype performed by the sample provider. Note that all genotypes listed in Table 1 are LiPA-based results.

Based on the inferred genotype, nucleotide consensus sequences were aligned to appropriate United States Food and Drug Administration (FDA) reference strain sequences for downstream identification of RAS and indels (Supplementary Materials Table S2). RAS positions were restricted to positions and substitutions listed in Supplementary Materials Table S3, adapted from known resistance-associated mutations at the time of analysis [35].

2.7. Assessment of Assay Performance

Performance metrics similar to our previous validation were used to assess assay quality between replicates and/or laboratory methods: (1) the concordance of nucleotide base calls, (2) the concordance of amino acid residue calls, (3) the quantification of RAS prevalence [36] and (4) sequencing success.

Concordance is defined as the proportion of nucleotide or amino acid sequence agreement between all nucleotides or amino acids sequenced across all samples. In cases of compatible discordances, (i.e., where one sequence observed a nucleotide or amino acid mixture whereas the other sequence observed a single component thereof, e.g., nucleotides Y vs. C), they were weighed identically as complete discordances. These concordance metrics were assessed using consensus sequences generated from specific nucleotide prevalence cutoffs.

Sequencing success was calculated as the proportion of sequences having met MiCall quality control criteria outlined above, requiring a minimum 100-fold coverage at all RAS-associated positions in NS3, NS5a and NS5b.

Unless otherwise stated, analyses were performed individually for NS3, NS5a and NS5b gene segments and used a 20% nucleotide or amino acid prevalence threshold for mixture calling, where applicable. A 20% cutoff was selected to mimic Sanger sequencing detection thresholds and is well above the expected minimum detected level [37]. In analyses where sequence data were generated from multiple sample replicates, and no gold-standard sample reference sequence was available, a consensus sequence was constructed from all available replicates for use as a comparator. Nucleotides and amino acids appearing in $\geq 20\%$ of replicate sequences were included as mixtures in the consensus. In addition to the metrics described above, we also assessed concordance in the detection of all nucleotide substitutions across the genome at their respective prevalence.

2.7.1. Accuracy

Accuracy, the level of agreement between a new test and a reference test, was assessed by comparing newly generated sequences to data obtained from a previously validated in-house assay optimized for HCV GT1 on the same Illumina MiSeq platform [36]. Duplicates of 93 samples were used here, consisting mostly of GT1 samples.

2.7.2. Precision

Precision refers to the ability of an assay to reproduce its result, whether within or between tests. Repeatability (intra-run variability) was assessed using 12 replicates of

four GT1 HCV-positive samples, processed together through amplification and sequenced in a single MiSeq run. Reproducibility (inter-assay variability) was assessed using five replicates of 12 samples ($n = 11$ GT1, $n = 1$ GT3), with each set of replicates processed separately in five PCR batches and sequenced on five separate days. To assess sequencing precision, sequences obtained from individual replicates were compared for concordance against a consensus sequence constructed from all available replicate sequences.

2.7.3. Sensitivity

Sensitivity refers to the ability to reliably detect an analyte with acceptable precision. We assessed the HCV viral load limit of detection with two sample sets: (1) the same GT1 samples used for accuracy measurement ($n = 93$) and (2) a set of largely GT3-6 samples ($n = 55$).

The ability to reliably detect low prevalence substitutions or minority variants was assessed using two replicates of 95 clinical samples of multiple genotypes.

2.7.4. Specificity

Analytical specificity refers to the ability of an assay to detect only the intended target and the lack of interference from other analytes. Specificity was assessed in two ways: (1) HCV-negative samples were assayed separately in triplicate and (2) a negative control (DEPC-treated water) was included in all runs for this validation and during routine clinical runs for months following this validation. We did not study interfering substances (e.g., hemoglobin, bilirubin, triglycerides, etc.) since our method of RNA extraction has been shown not to be affected by these substances [38]. In addition, we did not assess substances that may otherwise inhibit or interfere with RT-PCR amplification as these were previously assessed during the validation of an HCV whole-genome NGS method designed for GT1 infections [36,39]. It is reasonable to assume that the prior results would apply to the genotype-independent method as well given the minimal differences between the assays.

2.7.5. Genotype Spectrum

The genotype spectrum of our assay was evaluated using the same GT1 ($n = 93$) and GT3-6 samples ($n = 55$) described above for accuracy and sensitivity. Laboratory staff were blinded to the expected, previously determined probe-based genotype for 37 (67%) of the GT3-6 samples.

3. Results

3.1. Accuracy

A total of 93 samples (predominantly GT1a) were processed in duplicate by the genotype-independent and GT1-optimized methods and used to assess accuracy, using the latter as the reference comparator. The amplification and sequencing of all three genes (NS3, NS5a and NS5b) in at least one replicate was successful for 88 (94.6%) and 85 (91.4%) samples by the genotype-independent and GT1-optimized methods, respectively. For method comparisons, 85 samples (89.2%) had available data at all three genes by at least one replicate for both assays. In instances wherein a method generated sequence data by both replicates, only the first replicate's data was analyzed for that respective method.

As the GT1-optimized assay was designed to amplify and sequence GT1a and GT1b specimens, sequence comparisons between methods were restricted to the successfully sequenced samples of the same genotype, leaving 83 samples for comparison of all three genes. The genotype-independent HCV method showed 99.85%, 99.86% and 99.90% nucleotide concordance with the HCV GT1-optimized NGS method in NS3, NS5a and NS5b, respectively, using a 20% mixed base-calling threshold. Sequence discordances represented 0.14% of the total bases compared ($n = 489/352,418$) and were entirely due to nucleotide mixtures identified by one method and not the other; no incompatible nucleotide discordances were observed. Importantly, neither method was systematically biased towards identifying minority variants. While the genotype-independent method

identified a marginally larger proportion of the discrepant mixtures ($n = 288$; 58.9%), overall agreement in mixture calling was very high (Cohen's $\kappa = 0.94$). Nucleotide concordance matrices between the two NGS methods, as well as between replicates of the genotype-independent method are presented in the Supplementary Materials (Figures S1 and S2). Amino acid sequences were nearly identical for each gene, achieving $\geq 99.92\%$ concordance.

RAS quantification was assessed for samples with >1000-fold coverage for both replicates in both assays. Samples with no detectable RAS at >2% prevalence by either NGS method were excluded. Overall we observed a strong linear correlation between the RAS population prevalence observed by both methods ($R^2 > 0.99$) for NS3 and NS5a (Figure 2A). A systematic bias in amplifying specific RAS was not observed (Figure 2B). Linear regression could not be performed reliably for NS5b, but amino acid substitutions at positions 159, 282 and 320 were highly conserved (near 100% prevalence for all residues).

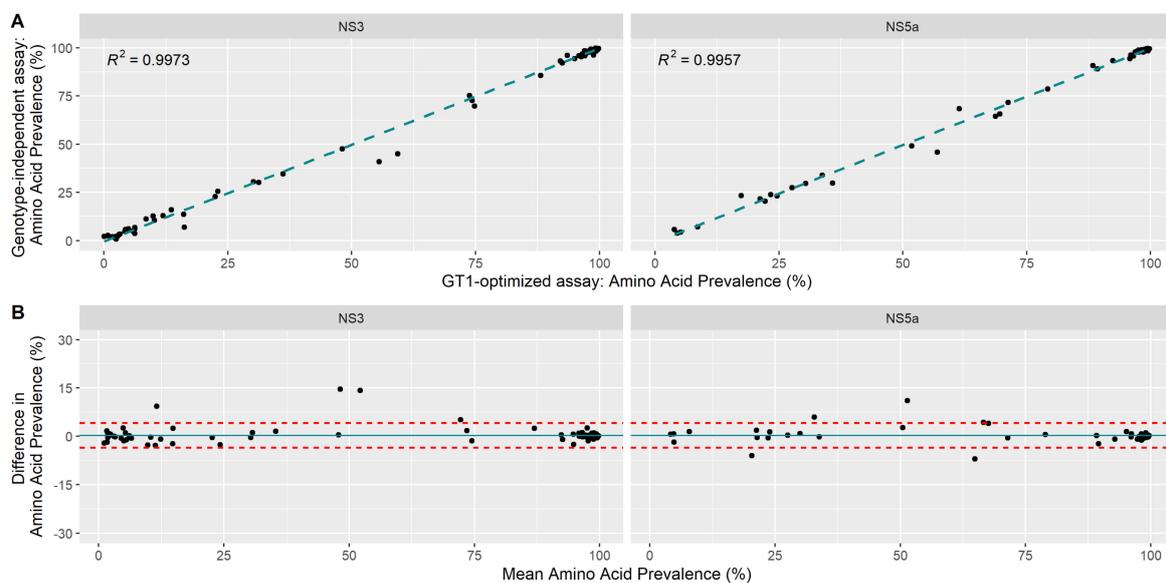


Figure 2. Accuracy of RAS quantification using the genotype-independent sequencing method compared to a genotype 1-optimized assay. Accuracy for NS3 and NS5a was assessed by the comparison of amino acid substitution frequencies at predefined, resistance-associated position as determined by our previously validated genotype 1-optimized and genotype-independent assays. Each point represents one possible RAS; note that positions with multiple possible amino acid residues are represented as separate data points. (A) Amino acid prevalence is compared between both sequencing methods for the same samples. (B) The difference in RAS frequencies between both sequencing methods as a function of mean RAS prevalence. The two red dotted lines represent the 95% limits of agreement.

3.2. Precision

Both the reproducibility and repeatability of all possible amino acid substitutions across NS3, NS5b and NS5a were evaluated in the same manner. Precision metrics assessed the genotype-independent assay's ability to detect nucleotide and amino acid substitutions consistently and quantitatively between replicates of a sample. For sequence concordance analyses, individual replicate sequences were compared to a sample consensus sequence constructed from the replicate sequences, using a 20% cutoff for mixture-calling.

We also assessed amino acid substitution detection at various prevalence categories. For each amino acid substitution, the mean prevalence across sequencing replicates was calculated and binned into <1%, 1–2%, 2–5%, 5–10%, 10–20% and >20% groups. The variant detection rate was then defined as the percent of sequenced replicates in which the amino acid was detected with a frequency above one of two minimum frequency thresholds: a 2% prevalence threshold was used to assess the potential for NGS sequencing to detect minority variants, while a 20% threshold was used to mimic Sanger sequencing detection limits.

3.3. Repeatability

Repeatability (intra-run variability) was assessed using 12 replicates of four GT1 HCV-positive samples. Nearly all replicates (42/48, 87.5%) were successfully sequenced by the genotype-independent method at all three genes, NS3, NS5a, NS5b. Note that five of the observed replicate sequencing failures were in NS5b by a single sample. Nucleotide and amino acid concordance relative to the consensus sequences was high for all samples; median nucleotide and amino acid concordance was above $\geq 99.90\%$ and $\geq 99.93\%$ respectively, across NS3, NS5a and NS5b.

At a 2% cut-off for mixture calling, the amino acid substitutions with a mean prevalence $>5\%$ were detected in 100% of the replicates in NS3, NS5A and NS5B. (Figure 3A). For substitutions with a mean prevalence between 2% and 5%, the repeatability of detection for NS3 and NS5A was $\sim 80\%$; lower repeatability ($\sim 50\%$) was observed in NS5B. As expected, there was a significant drop in the detection rates at mean prevalence around or below the 2% variant threshold. Substitutions present below 1% mean prevalence were correctly identified as conserved positions with only a few outliers. In the same analysis using a 20% prevalence threshold, substitutions present above 20% mean prevalence were detected in nearly all replicates with less than 0.01% outliers (Supplementary Materials Figure S3A).

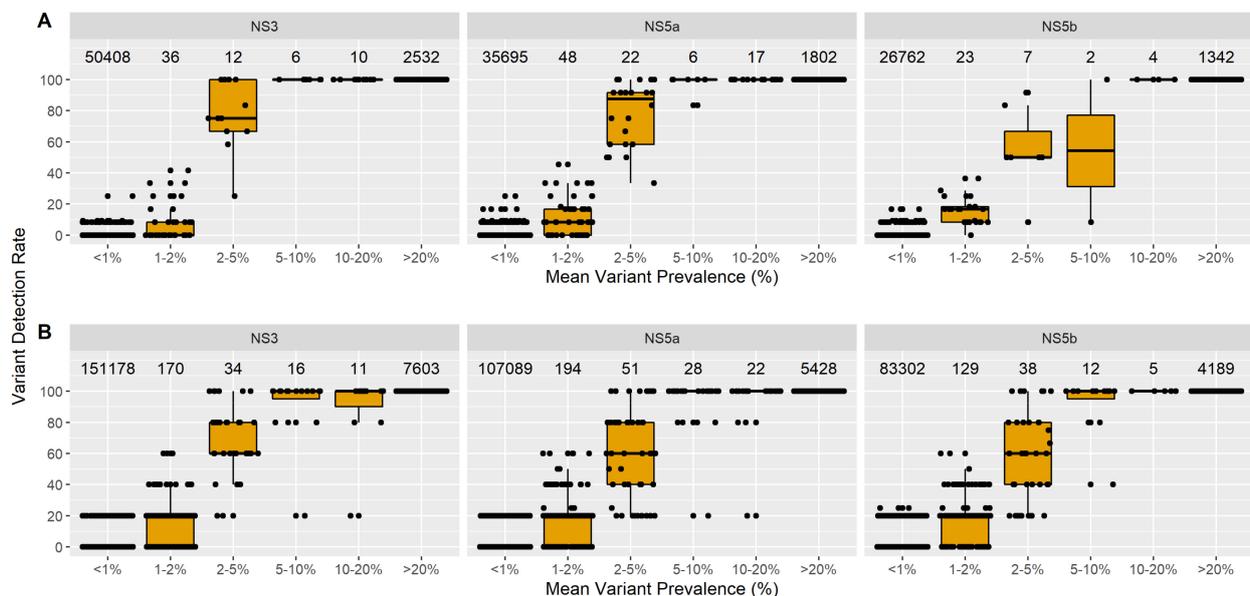


Figure 3. Repeatability and reproducibility of amino acid substitutions in NS3, NS5a and NS5b. Substitutions observed at a prevalence $\geq 2\%$ in a given replicate were considered “detected”. Variant detection rate (defined as the % of replicates per sample, in which a substitution was detected) is categorized by the mean prevalence of a substitution across all replicates of a sample. Replicate testing of each sample began from the same RNA extract; all steps beginning from the RT-PCR were repeated. Numbers indicated on the top of each graph represent the total number of substitutions in each bin. (A) Repeatability of amino acid substitutions was determined using four samples (3 GT1a, 1 GT1b) in 12 replicates processed on a single MiSeq run. (B) The reproducibility of amino acid substitutions was determined using 12 samples (10 GT1a, 1 GT1b, 1 GT3) in 5 replicates processed on separate days on five separate MiSeq runs.

3.4. Reproducibility

Reproducibility (inter-run variability) was assessed using five replicates of 12 samples and displayed similar results as the repeatability metric. With a single exception, all replicates were successfully sequenced at NS3, NS5a and NS5b (59/60, 98.3%); in one instance, NS5b was not sequenced in one replicate while the remaining genes were analyzed successfully. The median nucleotide and amino acid sequence concordance relative to the consensus sequence was $\geq 99.86\%$ and $\geq 99.90\%$, respectively, for NS3, NS5a and NS5b.

Nearly 100 of all amino acid substitutions with mean prevalence >5% were reproducibly identified at the 2% detection threshold (Figure 3B). Reproducibility was lower for substitutions with a mean prevalence between 2 and 5%, 60% in NS3 and NS5A, and 40% in NS5B. The detection of substitutions with mean prevalence at or below 2% was low as expected, given these residues likely have true frequencies below 2%. Conserved amino acids (<1% mean prevalence) were also identified correctly, with few exceptions. As with the repeatability analysis, we repeated this analysis using a 20% prevalence cut-off for mixtures (Supplementary Materials Figure S3B). Substitutions present above 20% prevalence were detected in most inter-assay replicates but with a greater number of outliers compared to intra-assay replicates. Although a 2% cutoff enables the detection of low prevalence substitutions or minority variants (i.e., <20% prevalence), stochastic effects introduced through RNA extraction, the sampling of small volumes or in RT-PCR efficiency result in the reduced reproducibility of detecting amino acid substitutions when present below 5% prevalence.

3.5. Sensitivity–Viral Load

To assess the ability of the genotype-independent assay to reliably amplify and sequence the HCV virus of different genotypes, we evaluated the PCR and sequencing success rate 146 samples consisting of GT 1 to 6 samples. No GT2 samples were available at the time of assessing this validation parameter.

Nearly all GT1 samples with a viral load >5.5 log₁₀ IU/mL were successfully sequenced ($n = 85/86$, 98.8%; Table 2). A sequencing success rate of 20–50% was obtained for samples with a viral load <5.5 log₁₀ IU/mL. However, only a total of seven samples were tested, four of which have a viral load in the 3–4 log range. Similarly, GT3–6 samples with a viral load >5.5 log₁₀ IU/mL were all successfully sequenced (17/17, 100%). In the 23 GT3–6 samples with a viral load between 5.1 and 5.5 log₁₀ IU/mL, NS3 and NS5A sequences were obtained in 20 (87%) of the samples, whereas NS5b sequence was generated in only 15 (65%) samples. The drop in sequencing success rates in NS5B is likely due to decreased coverage within NS5b beyond the WG amplicon, primarily in proximity to the amplification primer site. Only half of GT3–6 samples with viral load <5.0 log₁₀ IU/mL could be amplified and sequenced. Based on these results, the lower limit of detection was determined to be at >5.0 log₁₀ IU/mL.

Table 2. Sequencing success of the genotype-independent NGS assay in GT1 and GT3–6 sample sets stratified by viral load category.

HCV pVL log ₁₀ (IU/mL)	Genotype 1 ($n = 93$)			Genotypes 3–6 ($n = 53$) [^]		
	NS3	NS5A	NS5B	NS3	NS5A	NS5B
>7	23/23 (100%)	23/23 (100%)	23/23 (100%)	2/2 (100%)	2/2 (100%)	2/2 (100%)
6.6–7.0	32/33 (97%)	32/33 (97%)	32/33 (97%)	1/1 (100%)	1/1 (100%)	1/1 (100%)
6.1–6.5	14/14 (100%)	14/14 (100%)	14/14 (100%)	4/4 (100%)	4/4 (100%)	4/4 (100%)
5.6–6.0	16/16 (100%)	16/16 (100%)	16/16 (100%)	10/10 (100%)	10/10 (100%)	10/10 (100%)
5.1–5.5	1/2 (50%)	1/2 (50%)	1/2 (50%)	20/23 (87%)	20/23 (87%)	15/23 (65%)
<5	1/5 (20%)	2/5 (40%)	1/5 (20%)	7/13 (7/54%)	7/13 (54%)	7/13 (54%)

HCV, hepatitis C virus; pVL, plasma viral load; IU, international units. Genotypes were inferred through sequence analysis or in the absence of sequence data from assigned genotypes from LiPA assays. HCV subtype information was obtained using phylogenetic analyses based on MiSeq data. Sequencing success was defined as sequences having a minimum of 100-fold coverage at all RAS-associated positions in each respective gene. [^] Two genotype 3–6 samples from the set of 55 samples were excluded from this table, as no viral load data were available.

3.6. Sensitivity–Minority Variants

Given the increased sensitivity of NGS methods in detecting minority variants and the high degree of genetic variation amongst HCV samples, we also assessed the assay's ability to reliably detect these species. This was done by sequencing 95 clinical specimens

of multiple genotypes, each processed in duplicate. Upon successful sequencing, the prevalence of each individual nucleotide (A, T, C, G) at every position in the genome was determined for all sequences. Subsequently, for each pair of replicates, the mean and coefficient of variation (CV) of nucleotide prevalence was calculated for each base/position combination. The median and interquartile range of the coefficient of variation was visualized for all observed mean nucleotide prevalence values (Figure 4).

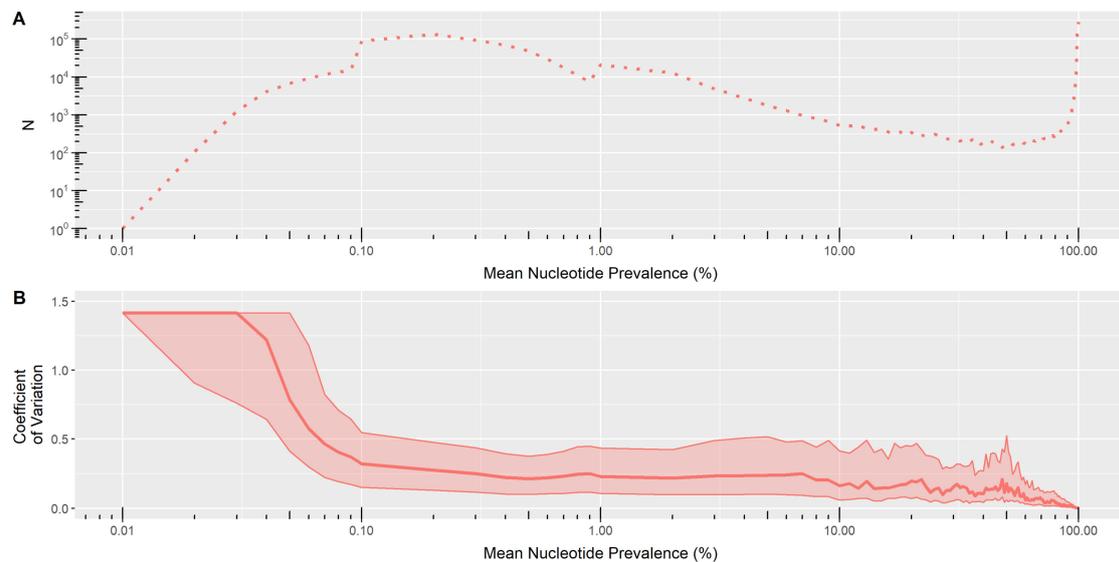


Figure 4. Assay sensitivity in detecting minority variants between sample duplicates at different mean nucleotide frequencies (defined as the mean prevalence of all permutations of nucleotides and positions within a sequence). This parameter was assessed using 95 samples run in duplicate, of which 84 passed sequencing. (A) The total number of nucleotide substitutions (N) detected at a particular mean substitution prevalence. (B) Median coefficient of variation (CV) of minor species with interquartile range (IQR) at different mean substitution prevalence.

Of the 95 samples selected for analysis in this parameter, 84 were successfully sequenced (88.4%). MiCall genotype assignment using sequence data alone was used for genotype assignment, as no prior LiPA-based genotype result was available for these samples. Sequences were assigned the following genotypes: GT1a ($n = 69$), GT1b ($n = 10$), GT1e ($n = 3$), GT2b ($n = 2$). Note that the 11 specimens that failed sequencing were determined to harbor GT1a HCV after analysis of the partial sequences assembled by the MiCall pipeline.

Through inspection of the relationship between CV and mean nucleotide frequency, we determined that nucleotide substitutions $>0.2\%$ prevalence could be reproducibly detected. Minimal variation CV was observed for nucleotide substitutions present in $>0.2\%$ of sequence reads, indicating that substitutions observed at these levels in one replicate are likely to be detected in the other replicate at a similar prevalence. However, nucleotide detection becomes less reliable for rare substitutions with $<0.2\%$ prevalence as shown by the sharp increase in CV at lower prevalence levels. However, a more conservative threshold may be warranted given MiSeq error rates may be higher. In a separate analysis of two HCV clones combined at different ratios (data not shown), we estimated the combined MiSeq random and systematic error rate to be approximately 0.5%. Given the greater error rate, as well as the limitations of PCR, the technical lower limit of minority species detection should be considered 0.5%.

3.7. Specificity

Across all analyses included in this validation, there was no evidence of contamination observed. Neither the HCV-negative sample nor the negative controls (DEPC-treated water) included in runs of this validation, as well routine clinical runs following this study, produced a detectable PCR product by agarose gel electrophoresis. Common interfering

substances (e.g., hemoglobin) are unlikely to affect our assay due to the RNA extraction method.

Analysis for interfering substances was not performed during this study but was previously done during the validation of the GT1-optimized method. In that study, the lack of interference from other common viruses was assessed using HCV-positive plasma was spiked with one of three additional specimens: a clinically derived HIV-1 plasma sample, a HIV-1 molecular clone or a clinically-derived HBV-positive plasma sample. These mixed-infection samples were subsequently processed in triplicate using the GT1-optimized method. No evidence of the off-target amplification of HIV or hepatitis B virus from spiked samples was observed in that study, nor in coinfecting samples using the newly described method since its clinical implementation [36].

3.8. Genotype Spectrum

A total of 146 samples, representing 7 different genotypes and subtypes, were used to evaluate the ability of this assay to classify genotypes against the previously reported genotype. Roughly 90% sequencing success rates were observed for the epidemic strains present in G7 countries, such as GT1a, GT1b, 2b, 3a, and GT6 (Table 3). About 70% success rates were observed in GT4 and GT5. Two of the 55 GT3-6 samples were excluded in the genotype spectrum analysis due to viral loads $<5 \log_{10}$ IU/mL. One other sample from the presumed GT3-6 sample set, which had no available LiPA-based genotype result, was identified as GT1b using sequence data, leaving 52 GT3-6 specimens. In total, NS3 and NS5a were both successfully sequenced in 45 out of 52 GT3-6 samples (87%), and NS5b was successfully sequenced in 39/52 (75%) samples in these genotypes. Small drops in coverage at this position are to be expected, given its proximity to the location of the 2nd round PCR amplification primer position (NS5b codon 336) and the decreased efficiency of Nextera XT processing near the ends of the amplicon. All samples failing NS5b sequencing, but not NS3 and NS5a, had viral loads $>5.1 \log_{10}$ IU/mL.

Table 3. Sequencing success for the HCV NS3, NS5a and NS5b genes using the genotype-independent-HCV assay for genotypes 1–6.

Genotype	Genotype Subtypes *	Samples Attempted ^	Sequenced (%)		
			NS3	NS5a	NS5b
1	1a	78	72 (92%)	72 (92%)	72 (92%)
	1b	11	11 (100%)	11 (100%)	11 (100%)
	1e	3	2 (67%)	2 (67%)	2 (67%)
2	2b	2	2 (100%)	2 (100%)	2 (100%)
3	3a	20	17 (100%)	17 (100%)	17 (100%)
4	4a, 4n	7	5 (71%)	5 (71%)	2 (29%)
5	5a	4	3 (75%)	3 (75%)	3 (75%)
6	6a, 6e, 6h, 6k, 6t	21	20 (95%)	20 (95%)	17 (85%)
Overall	-	146	132 (90%)	132 (90%)	126 (86%)

* Genotypes were inferred through sequence analysis or in the absence of sequence data from assigned genotypes from LiPA assays. HCV subtype information was obtained using phylogenetic analyses based on MiSeq data. Sequencing success was defined as sequences having a minimum of 100-fold coverage at all RAS-associated positions in each respective gene. ^ Only samples with viral loads above the estimated assay limit of detection ($5 \log_{10}$ IU/mL) were included.

We next evaluated the ability of the sequencing assay to accurately identify the HCV genotypes of the 146 samples described above. Genotypes inferred from the sequence data by MiCall were compared with the LiPA-based genotype data provided by the sample providers where available. Genotype comparisons could be assessed for 95 samples, as our study was blinded to the LiPA-based genotypes for 37 samples originating from the C-WORTHY trial. There was a 100% concordance between the genotype calls provided by the LiPA probe-based assay and the genotype calls determined by direct MiSeq sequence

mapping. Subtype information was identical for all genotype 1–3 samples across both genotyping methods. In cases of genotypes 4–6 ($n = 18$), ambiguous subtype information (e.g., genotype 4 or 4a/c/d) was provided by LiPA for all genotype 4 and 6 samples ($n = 13/18$). Direct sequence mapping provided definitive subtype information (e.g., genotype 4a) for all but three of these samples ($n = 10/18$), where the remaining three could not be sequenced.

3.9. Overlapping Region of Amplicons

The MiDi amplicon was designed to provide sequence coverage to the end of NS5b, beyond the WG amplicon that extends from Core to codon 336 of NS5b. However, only two RAS positions are covered by the overlap of WG and MiDi amplicons, and no sequences with known resistance-associated mutations at these positions were observed in this validation (Supplementary Materials Table S4). We therefore assessed the proportion of each nucleotide at all overlapping positions in NS5b (not only RAS positions). There was an excellent correlation between the proportion each nucleotide observed between MiDi and WG ($r^2 > 0.98$). A summary of the mean number of sequence reads, amino acid counts and amino acid prevalence for the two NS5B RAS (S282 and L320) within the overlapping region obtained by WG and MiDi amplicons is shown in Supplementary Materials Table S4.

4. Discussion

We had previously developed HCV genotyping assays that were optimized to detect the NS3 Q80K polymorphism in HCV GT1 specimens only, using both Sanger and NGS (MiSeq) sequencing technologies [36]. Although these assays are suitable to capture the resistance profile of the most prevalent HCV genotype, GT1 represented only 44% of all new HCV infections globally in 2015 [40]. Despite the availability of HCV RAS screening laboratory services in the United States (e.g., Labcorp, Quest Diagnostics), these have only been validated for GT1 infections; no commercially available test to our knowledge has been extensively validated to characterize non-GT1 HCV infections. We have therefore developed and validated an NGS method to infer HCV genotypes and identify RAS for clinical purposes regardless of the associated HCV genotype/subtype classification. This genotype-independent method can be used to screen for all currently known RAS within NS3, NS5a and NS5b and potentially guide therapeutic decision making in treatment-naïve and -experienced individuals.

The genotype-independent sequencing assay described here demonstrated similar performance as the previously validated GT1-optimized NGS method [36]. Overall, high accuracy and precision in all replicates tested was observed (>99.7%). Importantly, samples used to assess both inter- and intra-assay concordance were highly diverse as illustrated by the number of low-prevalence (2–5%) nucleotide mixtures observed in sequences from these samples. Despite this intra-sample diversity, NGS sequencing was capable of accurately and reproducibly characterizing the diversity of HCV species within a sample, regardless of genotype. Note that reproducibility displayed similar results as the repeatability metric, albeit with slightly more variability at both the 2% and 20% prevalence thresholds for amino acid substitution detection.

Substitution prevalence thresholds used in this study focused on two cutoffs: 20% to reflect a Sanger-sequencing limit of detection and 2% to utilize the potential of NGS to detect minority variants. A suitable question stemming from this study is which cutoff is most optimal for the detection of RAS and is sufficient to correlate with clinical outcomes. A separate study found that a 10% cutoff predicted DAA response in GT3 infections [41]. Parameters in this study intended to analyze the assay's capacity to detect substitutions consistently and quantitatively. Results show that when a prevalence cutoff is utilized to make a nucleotide or amino acid call, the most variance in detecting substitutions is centered around the cutoff itself, as one would expect. Using a 20% cutoff, effectively all high-prevalence (i.e., $\geq 20\%$) substitutions are detected consistently, and no substitutions with a true prevalence below 10% are detected. Substitutions whose prevalence is between

10 and 20% are only detected incorrectly after repeat testing of the same sample, on average in 1 of 5 replicates. In contrast, utilizing a 2% cutoff shows a greater spread in its ability to detect substitutions around and above the threshold. Generally, only substitutions with a presumed prevalence above 10% could be reliably detected despite the lower cutoff. However, further studies are needed to elucidate on potential clinical implications of modifying these prevalence cutoffs and the effect of minority variants on treatment outcomes.

In this study, sample providers provided LiPA-based genotype classifications for samples, where available. In rare cases, the LiPA assay can misclassify infections or fail to resolve genotype and/or subtype entirely [42–45]. By using a set of pan-genomic PCR primers and inferring HCV genotype *in silico* from deep-sequencing genomic data, this assay obviates the need to perform HCV genotyping by LiPA prior to sequencing and potentially resolves HCV subtypes with greater accuracy. Laboratories may also opt to perform NS5b sequencing for genotype identification rather than LiPA assays or whole-genome sequencing assays such as the one described here. NS5B sequencing has demonstrated lower rates of genotype misclassification compared to LiPA and other genotyping kits, and may also be more cost-effective [18]. However, this approach restricts genomic data to the NS5b region only and fails to capture RAS located outside this gene.

This whole-genome sequencing assay can be used to better characterize the prevalence of RAS across the HCV genome and within non-GT1 HCV infections. Data generated by this method, in conjunction with phenotype and observational studies, can be used to identify novel RAS within or outside of these genes and possibly correlate with treatment outcomes. Additional validation studies would be required to reassess the performance of this assay for those new mutations.

Having demonstrated similar performance to a previously validated GT1 assay, this genotype-independent HCV sequencing assay further expands the availability of RAS screening to other HCV genotypes. This method has been implemented clinically at the BC Centre for Excellence in HIV/AIDS and has been used to deliver ~4000 resistance test reports to physicians across Canada as of August 2021. The assay protocols and paired software are freely available to interested labs worldwide; software is available at <https://github.com/cfe-lab/MiCall>.

Both the validated GT1-optimized method and new genotype-independent method exhibited similar limitations. In terms of limit of detection, both assays require significantly more input starting material ($\sim 5 \log_{10}$ IU/mL) than their Sanger sequencing counterparts [36]. Despite this, the plasma viral load required for testing is below the expected HCV viral load range for most untreated individuals, including those living with HIV, cirrhosis, advanced liver disease, or those who have previously failed DAA treatment [46–50]. Another limitation of the assay described here is the requirement for two amplicons to cover the entire HCV genome and capture the full spectrum of HCV genotype, except for the 5' and 3' untranslated regions. The generation of the additional MiDi amplicon to cover the remainder of NS5B beyond codon 336 has increased cost and laboratory work implications. However, this additional cost may be warranted due to the potential for RAS beyond codon 336 and the use of NS5b sequence data for phylogenetic cluster identification [51].

A notable consideration of this study is the genotype representativeness within individual validation parameters. With the exception of sensitivity and genotype coverage that were analyzed using a large spectrum of available genotypes, the remaining validation tests were assessed using predominantly GT1 specimens. Additionally, these consisted largely of GT1a samples, with the minimal inclusion of GT1b specimens. Equal representativeness of all genotypes in each validation parameter was not possible in this validation and represents an important limitation. GT1b and non-GT1 samples, particularly rare genotypes, could not be included in the study in large numbers due to a lack of sample availability. These factors may limit our ability to infer assay performance from this study for these uncommon genotypes. Although many genotypes are under-represented in this study, the assay demonstrated acceptable metrics in analyzing the available specimens.

Since utilizing this method clinically at the BC Centre for Excellence in HIV/AIDS, no negative bias has been observed towards these less common genotypes (data not shown).

5. Conclusions

In summary, we have developed and characterized a genotype-independent, whole-genome HCV sequencing workflow on the Illumina MiSeq platform. This assay can provide accurate HCV drug resistance profiles to currently known RAS, notably in cases of DAA-treatment failure or in select treatment-naïve cases, to personalize HCV therapy. We demonstrated positive and consistent performance of the test including accuracy, repeatability, reproducibility and sensitivity. Importantly, the assay successfully amplified and sequenced 90% of samples regardless of genotype. Since the method generates near-full-genome coverage for any HCV genotype, the method could be expanded to screen for other known RAS and phylogenetic cluster identification. The assay and associated bioinformatic software also simultaneously provide genotype classification, which can supplement validated, probe-based genotyping results or be used on a research-use-only basis in instances wherein no genotype assignments are otherwise available. Together, this assay and bioinformatic pipeline can also facilitate the implementation of HCV drug resistance testing by other laboratories globally, with an end-to-end, sample-to-report workflow.

Supplementary Materials: The following are available online at <https://www.mdpi.com/article/10.3390/v13091721/s1>, Laboratory Protocol, Table S1. Primers used in the genotype-independent sequencing protocol. Primers were purified by standard desalting, Table S2. FDA guideline of Reference Strains for Reporting of Amino Acid Sequence Data, Table S3. HCV resistance-associated substitutions (RAS) in genotype 1a and 1b that were analyzed for accuracy. X indicates a variant with any mutation except the wildtype amino acid residue, Table S4. A summary of the mean number of sequence reads and wildtype amino acid counts for the two NS5B RAS within the overlapping region obtained by WG and MiDi amplicons: S282 and L320, Figure S1. Accuracy nucleotide concordance between GT1-optimized and the genotype-independent sequencing assays, Figure S2. Nucleotide concordance between accuracy replicates ($n = 93$ samples), Figure S3. Repeatability and Reproducibility of amino acid substitution detection amino acid substitutions.

Author Contributions: Conceptualization, W.W.Y.D. and P.R.H.; Formal analysis, H.R.L. and C.J.B.; Funding acquisition, P.R.H., A.Y.M.H. and C.J.B.; Investigation, H.R.L., W.D., A.Y.M.H. and C.J.B.; Methodology, H.R.L., W.D., W.W.Y.D., D.K., C.W., A.F.Y.P., A.Y.M.H., P.R.H. and C.J.B.; Software, D.K. and A.F.Y.P.; Supervision, P.R.H.; Validation, H.R.L., W.D., W.W.Y.D., D.K., C.W., A.F.Y.P., A.Y.M.H., P.R.H. and C.J.B.; Writing—original draft, H.R.L. and A.Y.M.H.; Writing—review and editing, H.R.L., D.K., C.W., A.F.Y.P., A.Y.M.H., P.R.H. and C.J.B. All authors have read and agreed to the published version of the manuscript.

Funding: This study was funded in part by Merck and a User Partnership Program grant from Genome British Columbia to P.R.H., A.Y.M.H. and C.J.B. (UPP029).

Institutional Review Board Statement: Ethical approval was granted by the University of British Columbia Research Ethics Board (H16-00586).

Informed Consent Statement: Patient consent for use of specimens for research purposes was acquired as part of the C-WORTHY (NCT01717326/Protocol PN035) and Vancouver Injection Drug Users Study (VIDUS) studies.

Acknowledgments: The authors thank the leadership and laboratory team at the BC Centre for Excellence in HIV/AIDS for their work in developing and validating this protocol. We also thank A.F.Y.P. and D.K. for their contributions in the development and implementation of MiCall for bioinformatic data processing.

Conflicts of Interest: Authors have no conflict of interest to declare. The funders had no role in designing the study; in analyses or interpretation of data; in the writing of the manuscript.

References

1. World Health Organization. *Global Hepatitis Report 2017*; World Health Organization: Geneva, Switzerland, 2017.
2. World Health Organization. *Guidelines for the Screening, Care and Treatment of Persons with Chronic Hepatitis C Infection*; World Health Organization: Geneva, Switzerland, 2014. Available online: <http://www.ip-watch.org/weblog/wp-content/uploads/2014/04/WHO-Hepatitis-C-Guidelines-.pdf> (accessed on 27 April 2014).
3. Pons, S.T.; Lamblin, G.; Boyer, A.; Sautou, V.; Abergel, A. Drug interactions and protease inhibitors used in the treatment of hepatitis C: How to manage? *Eur. J. Clin. Pharmacol.* **2014**, *70*, 775–789. [[CrossRef](#)]
4. Mavyret. AbbVie Inc., North Chicago, IL, USA. 2020. Available online: https://www.rxabbvie.com/pdf/mavyret_pi.pdf (accessed on 23 March 2020).
5. Eplclusa. Gilead Sciences, Inc., Foster City, CA, USA. 2020. Available online: https://www.rxabbvie.com/pdf/mavyret_pi.pdf (accessed on 23 March 2020).
6. World Health Organization. *Guidelines for the Screening, Care and Treatment of Persons with Chronic Hepatitis C Infection*; World Health Organization: Geneva, Switzerland, 2016. Available online: http://apps.who.int/iris/bitstream/handle/10665/205035/9789241549615_eng.pdf?sequence=1 (accessed on 27 April 2016).
7. World Health Organization. *Guidelines for the Care and Treatment of Persons Diagnosed with Chronic Hepatitis C Virus Infection*; World Health Organization: Geneva, Switzerland, 2018. Available online: <https://apps.who.int/iris/bitstream/handle/10665/273174/9789241550345-eng.pdf> (accessed on 27 July 2018).
8. Foster, G.R.; Afdhal, N.; Roberts, S.K.; Br, N.; Gane, E.J.; Pianko, S.; Lawitz, E.; Thompson, A.; Shiffman, M.L.; Cooper, C.; et al. Sofosbuvir and velpatasvir for HCV Genotype 2 and 3 infection. *N. Engl. J. Med.* **2015**, *373*, 2608–2617. [[CrossRef](#)]
9. Feld, J.J.; Jacobson, I.M.; Hode, C.; Asselah, T.; Ruane, P.J.; Gruener, N.; Abergel, A.; Mangia, A.; Lai, C.L.; Chan, H.L.Y.; et al. Sofosbuvir and velpatasvir for hcv genotype 1, 2, 4, 5, and 6 infection. *N. Engl. J. Med.* **2015**, *373*, 2599–2607. [[CrossRef](#)] [[PubMed](#)]
10. Afdhal, N.; Zeuzem, S.; Kwo, P.; Chojkier, M.; Gitlin, N.; Puoti, M.; Romero-Gomez, M.; Zarski, J.P.; Agarwal, K.; Buggisch, P.; et al. Ledipasvir and sofosbuvir for untreated HCV genotype 1 infection. *N. Engl. J. Med.* **2014**, *370*, 1889–1898. [[CrossRef](#)]
11. Arends, J.E.; Kracht, P.A.M.; Hoepelman, A.I.M. Performance of hepatitis C virus (HCV) direct-acting antivirals in clinical trials and daily practice. *Clin. Microbiol. Infect.* **2016**, *22*, 846–852. [[CrossRef](#)]
12. Pawlotsky, J.-M. Hepatitis C virus resistance to direct-acting antiviral drugs in interferon-free regimens. *Gastroenterology* **2016**, *151*, 70–86. [[CrossRef](#)] [[PubMed](#)]
13. Liu, R.; Curry, S.; McMonagle, P.; Yeh, W.W.; Ludmerer, S.W.; Jumes, P.A.; Marshall, W.L.; Kong, S.; Ingravallo, P.; Black, S.; et al. Susceptibilities of genotype 1a, 1b, and 3 hepatitis C virus variants to the NS5A inhibitor elbasvir. *Antimicrob. Agents Chemother.* **2015**, *59*, 6922–6929. [[CrossRef](#)] [[PubMed](#)]
14. Sorbo, M.C.; Cento, V.; di Maio, V.C.; Howe, A.Y.M.; Garcia, F.; Perno, C.F.; Ceccherini-Silberstein, F. Hepatitis C virus drug resistance associated substitutions and their clinical relevance: Update 2018. *Drug Resist. Updates* **2018**, *37*, 17–39. [[CrossRef](#)]
15. Wang, G.P.; Terrault, N.; Reeves, J.D.; Liu, L.; Li, E.; Zhao, L.; Lim, J.K.; Morelli, G.; Kuo, A.; Levitsky, J.; et al. Prevalence and impact of baseline resistance-associated substitutions on the efficacy of ledipasvir/sofosbuvir or simeprevir/sofosbuvir against GT1 HCV infection. *Sci. Rep.* **2018**, *8*, 1–14. [[CrossRef](#)]
16. Panel, A.-I.H.G. Hepatitis C guidance 2018 update: AASLD-IDS recommendations for testing, managing, and treating hepatitis C virus infection. *Clin. Infect. Dis. Off. Publ. Infect. Dis. Soc. Am.* **2018**, *67*, 1477. [[CrossRef](#)]
17. Shah, H.; Bilodeau, M.; Burak, K.W.; Cooper, C.; Klein, M.; Ramji, A.; Smyth, D.; Feld, J.J. The management of chronic hepatitis C: 2018 guideline update from the Canadian Association for the Study of the Liver. *CMAJ* **2018**, *190*, E677–E687. [[CrossRef](#)]
18. Chueca, N.; Rivadulla, I.; Lovatti, R.; Reina, G.; Blanco, A.; Fernandez-Caballero, J.A.; Cardeñoso, L.; Rodriguez-Granjer, J.; Fernandez-Alonso, M.; Aguilera, A.; et al. Using NS5B Sequencing for hepatitis C virus genotyping reveals discordances with commercial platforms. *PLoS ONE* **2016**, *11*, e0153754. [[CrossRef](#)]
19. Howe, A.Y.M.; Ceccherini-Silberstein, F.; Dietz, J.; Popping, S.; Grebely, J.; Rodrigo, C.; Lennerstrand, J.; Douglas, M.W.; Parczewski, M.; Harrigan, P.R.; et al. SHARED: An international collaboration to unravel hepatitis C resistance. *Viruses* **2021**, *13*, 1580. [[CrossRef](#)] [[PubMed](#)]
20. Cento, V.; Chevaliez, S.; Perno, C.F. Resistance to direct-acting antiviral agents. *Curr. Opin. HIV AIDS* **2015**, *10*, 381–389. [[CrossRef](#)] [[PubMed](#)]
21. Thomson, E.; Ip, C.L.C.; Badhan, A.; Christiansen, M.T.; Adamson, W.; Ansari, M.A.; Bibby, D.; Breuer, J.; Brown, A.; Bowden, R.; et al. Comparison of next-generation sequencing technologies for comprehensive assessment of full-length hepatitis C viral genomes. *J. Clin. Microbiol.* **2016**, *54*, 2470–2484. [[CrossRef](#)] [[PubMed](#)]
22. Bull, R.A.; Eltahla, A.A.; Rodrigo, C.; Koekkoek, S.M.; Walker, M.; Pirozyan, M.R.; Betz-Stablein, B.; Toepfer, A.; Laird, M.; Oh, S.; et al. A method for near full-length amplification and sequencing for six hepatitis C virus genotypes. *BMC Genom.* **2016**, *17*, 1–10. [[CrossRef](#)] [[PubMed](#)]
23. Ho, C.K.Y.; Welkers, M.R.A.; Thomas, X.V.; Sullivan, J.C.; Kieffer, T.L.; Reesink, H.W.; Rebers, S.P.H.; de Jong, M.D.; Schinkel, J.; Molenkamp, R. A comparison of 454 sequencing and clonal sequencing for the characterization of hepatitis C virus NS3 variants. *J. Virol. Methods* **2015**, *219*, 28–37. [[CrossRef](#)]
24. Sato, M.; Maekawa, S.; Komatsu, N.; Tatsumi, A.; Miura, M.; Muraoka, M.; Suzuki, Y.; Amemiya, F.; Takano, S.; Fukasawa, M.; et al. Deep sequencing and phylogenetic analysis of variants resistant to interferon-based protease inhibitor therapy in chronic hepatitis induced by genotype 1b hepatitis C virus. *J. Virol.* **2015**, *89*, 6105–6116. [[CrossRef](#)]

25. Donaldson, E.F.; Harrington, P.R.; O'Rear, J.J.; Naeger, L.K. Clinical evidence and bioinformatics characterization of potential hepatitis C virus resistance pathways for sofosbuvir. *Hepatology* **2015**, *61*, 56–65. [[CrossRef](#)] [[PubMed](#)]
26. Wang, G.P.; Sherrill-Mix, S.A.; Chang, K.M.; Quince, C.; Bushman, F.D. Hepatitis C virus transmission bottlenecks analyzed by deep sequencing. *J. Virol.* **2010**, *84*, 6218–6228. [[CrossRef](#)]
27. Verbinnen, T.; van Marck, H.; Vandenbroucke, I.; Vijgen, L.; Claes, M.; Lin, T.; Simmen, K.; Neyts, J.; Fanning, G.; Lenz, O. Tracking the evolution of multiple in vitro hepatitis C virus replicon variants under protease inhibitor selection pressure by 454 deep sequencing. *J. Virol.* **2010**, *84*, 11124–11133. [[CrossRef](#)]
28. Hiraga, N.; Imamura, M.; Abe, H.; Hayes, C.N.; Kono, T.; Onishi, M.; Tsuge, M.; Takahashi, S.; Ochi, H.; Iwao, E.; et al. Rapid emergence of telaprevir resistant hepatitis C virus strain from wildtype clone in vivo. *Hepatology* **2011**, *54*, 781–788. [[CrossRef](#)]
29. Chevaliez, S.; Rodriguez, C.; Pawlotsky, J.M. New virologic tools for management of chronic hepatitis B and C. *Gastroenterology* **2012**, *142*, 1303–1313. [[CrossRef](#)] [[PubMed](#)]
30. Fourati, S.; Pawlotsky, J.-M. Virologic tools for HCV drug resistance testing. *Viruses* **2015**, *7*, 6346. [[CrossRef](#)] [[PubMed](#)]
31. Sulkowski, M.S.; Hezode, C.; Gerstoft, J.; Vierling, J.M.; Mallolas, J.; Pol, S.; Kugelmas, M.; Murillo, A.; Weis, N.; Nahass, R. Efficacy and safety of MK-5172+ MK-8742 ± ribavirin in HCV mono-infected and HIV/HCV co-infected treatment-naïve, non-cirrhotic patients with hepatitis C virus genotype 1 infection: The C-WORTHY study (Final results, Parts A and B). *Hepatology* **2014**, *60*, 318A–319A.
32. Lawitz, E.; Gane, E.; Pearlman, B.; Tam, E.; Ghesquiere, W.; Guyader, D.; Alric, L.; Bronowicki, J.P.; Lester, L.; Sievert, W.; et al. Efficacy and safety of 12 weeks versus 18 weeks of treatment with grazoprevir (MK-5172) and elbasvir (MK-8742) with or without ribavirin for hepatitis C virus genotype 1 infection in previously untreated patients with cirrhosis and patients with previous null response with or without cirrhosis (C-WORTHY): A randomised, open-label phase 2 trial. *Lancet* **2015**, *385*, 1075–1086. [[CrossRef](#)] [[PubMed](#)]
33. Langmead, B.; Salzberg, S.L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **2012**, *9*, 357–359. [[CrossRef](#)]
34. Kuiken, C.; Yusim, K.; Boykin, L.; Richardson, R. The Los Alamos hepatitis C sequence database. *Bioinformatics* **2005**, *21*, 379–384. [[CrossRef](#)]
35. Lontok, E.; Harrington, P.; Howe, A.; Kieffer, T.; Lennerstrand, J.; Lenz, O.; McPhee, F.; Mo, H.; Parkin, N.; Pilot-Matias, T.; et al. Hepatitis C virus drug resistance-associated substitutions: State of the art summary. *Hepatology* **2015**, *62*, 1623–1632. [[CrossRef](#)] [[PubMed](#)]
36. Chui, C.K.S.; Dong, W.W.Y.; Joy, J.B.; Poon, A.F.Y.; Dong, W.Y.; Mo, T.; Woods, C.K.; Beatty, C.; Hew, H.; Harrigan, P.R.; et al. Development and validation of two screening assays for the hepatitis C virus NS3 Q80K polymorphism associated with reduced response to combination treatment regimens containing simeprevir. *J. Clin. Microbiol.* **2015**, *53*, 2942–2950. [[CrossRef](#)]
37. Larder, B.A.; Kohli, A.; Kellam, P.; Kemp, S.D.; Kronick, M.; Henfrey, R.D. Quantitative detection of HIV-1 drug resistance mutations by automated DNA sequencing. *Nature* **1993**, *365*, 671–673. [[CrossRef](#)]
38. Boom, R.; Sol, C.J.; Salimans, M.M.; Jansen, C.L.; Wertheim-van Dillen, P.M.; van der Noordaa, J. Rapid and simple method for purification of nucleic acids. *J. Clin. Microbiol.* **1990**, *28*, 495–503. [[CrossRef](#)] [[PubMed](#)]
39. Zhang, E.Z.; Bartels, D.J.; Frantz, J.D.; Seepersaud, S.; Lippke, J.A.; Shames, B.; Zhou, Y.; Lin, C.; Kwong, A.; Kieffer, T.L. Development of a sensitive RT-PCR method for amplifying and sequencing near full-length HCV genotype 1 RNA from patient samples. *Viol. J.* **2013**, *10*, 53. [[CrossRef](#)]
40. Polaris Observatory HCV Collaborators. Global prevalence and genotype distribution of hepatitis C virus infection in 2015: A modelling study. *Lancet Gastroenterol. Hepatol.* **2017**, *2*, 161–176. [[CrossRef](#)]
41. McPhee, F.; Hernandez, D.; Zhou, N. Effect of minor populations of NS5A and NS5B resistance-associated variants on HCV genotype-3 response to daclatasvir plus sofosbuvir, with or without ribavirin. *Antivir. Ther.* **2017**, *22*, 237–246. [[CrossRef](#)]
42. Verbeeck, J.; Stanley, M.J.; Shieh, J.; Celis, L.; Huyck, E.; Wollants, E.; Morimoto, J.; Farrior, A.; Sablon, E.; Jankowski-Hennig, M.; et al. Evaluation of versant hepatitis C virus genotype assay (LiPA) 2.0. *J. Clin. Microbiol.* **2008**, *46*, 1901–1906. [[CrossRef](#)] [[PubMed](#)]
43. Bouchardeau, F.; Cantaloube, J.F.; Chevaliez, S.; Portal, C.; Razer, A.; Lefrère, J.J.; Pawlotsky, J.M.; de Micco, P.; Laperche, S. Improvement of hepatitis C virus (HCV) genotype determination with the new version of the INNO-LiPA HCV assay. *J. Clin. Microbiol.* **2007**, *45*, 1140–1145. [[CrossRef](#)]
44. Guelfo, J.R.; Macias, J.; Neukam, K.; di Lello, F.A.; Mira, J.A.; Merchante, N.; Mancebo, M.; Nuñez-Torres, R.; Pineda, J.A.; Real, L.M. Reassessment of genotype 1 hepatitis C virus subtype misclassification by LiPA 2.0: Implications for direct-acting antiviral treatment. *J. Clin. Microbiol.* **2014**, *52*, 4027–4029. [[CrossRef](#)]
45. Howe, A.; de Salazar, A.; di Maio, V.; Popping, S.; Fourat, S.; Tay, E.; Rodrigo, C.; Cunningham, E.; Kjellin, M.; Fay, F.; et al. Genotype misclassification and its impact on treatment choices, outcomes and drug resistance. *Can. Liver Meet. Abstr.* **2020**. [[CrossRef](#)]
46. Younossi, Z.M.; Stepanova, M.; Feld, J.; Zeuzem, S.; Jacobson, I.; Agarwal, K.; Hezode, C.; Nader, F.; Henry, L.; Hunt, S. Sofosbuvir/velpatasvir improves patient-reported outcomes in HCV patients: Results from ASTRAL-1 placebo-controlled trial. *J. Hepatol.* **2016**, *65*, 33–39. [[CrossRef](#)]
47. Wyles, D.L.; Ruane, P.J.; Sulkowski, M.S.; Dieterich, D.; Luetkemeyer, A.; Morgan, T.R.; Sherman, K.E.; Dretler, R.; Fishbein, D.; Gathe, J.C.; et al. Daclatasvir plus sofosbuvir for HCV in patients coinfecting with HIV-1. *N. Engl. J. Med.* **2015**, *373*, 714–725. [[CrossRef](#)] [[PubMed](#)]

48. Poordad, F.; Schiff, E.R.; Vierling, J.M.; Landis, C.; Fontana, R.J.; Yang, R.; McPhee, F.; Hughes, E.A.; Noviello, S.; Swenson, E.S. Daclatasvir with sofosbuvir and ribavirin for hepatitis C virus infection with advanced cirrhosis or post-liver transplantation recurrence. *Hepatology* **2016**, *63*, 1493–1505. [[CrossRef](#)] [[PubMed](#)]
49. Asselah, T.; Hezode, C.; Qaqish, R.B.; ElKhashab, M.; Hassanein, T.; Papatheodoridis, G.; Feld, J.; Alqahtani, S.; Pilot-Matias, T.; Yu, Y.; et al. High SVR rates in patients with genotype 4 chronic hepatitis C infection and compensated cirrhosis with ombitasvir/paritaprevir/ritonavir co-administered with ribavirin (AGATE-I). *J. Hepatol.* **2016**, *64*, S827. [[CrossRef](#)]
50. Gane, E.J.; Shiffman, M.L.; Etzkorn, K.; Morelli, G.; Stedman, C.A.M.; Davis, M.N.; Hineostrova, F.; Dvory-Sobol, H.; Huang, K.C.; Osinusi, A.; et al. Sofosbuvir-velpatasvir with ribavirin for 24 weeks in hepatitis C virus patients previously treated with a direct-acting antiviral regimen. *Hepatology* **2017**, *66*, 1083–1089. [[CrossRef](#)]
51. Olmstead, A.D.; Joy, J.B.; Montoya, V.; Luo, I.; Poon, A.F.Y.; Jacka, B.; Lamoury, F.; Applegate, T.; Montaner, J.; Khudyakov, Y.; et al. A molecular phylogenetics-based approach for identifying recent hepatitis C virus transmission events. *Infect. Genet. Evol.* **2015**, *33*, 101–109. [[CrossRef](#)]