Use of whole genome sequencing in the Dutch Acute HCV in HIV study. A focus on transmitted antiviral resistance.

Mette T. Christiansen[1*#], Sebastiaan J. Hullegie[2*], Martin Schutten[3], Katja Einer-Jensen[4], Helena J. Tutill[1], Judith Breuer[1] and Bart J.A. Rijnders[2]

* Both authors contributed equally to the content of the paper

1. Division of infection and immunity, University College London (UCL), London, United Kingdom
2. Department of Internal Medicine and Infectious Diseases, Erasmus MC, Rotterdam, The Netherlands
3. Department of Virology, Erasmus MC, Rotterdam, The Netherlands
4. QIAGEN-AAR, Arhus, Denmark

New address: Statens Serum Institut, Microbiology and Infection control, Copenhagen, Denmark

(trial registration number NCT01912495*)*

Keywords: Acute Hepatitis C, protease inhibitors, treatment, resistance, HIV

Manuscript word count: 1052

Abstract word count: 182

Funding

This project has received funding from the European Union's Seventh Program for research, technological development and demonstration under grant agreement No 304875. The DAHH-study was supported by MSD. MSD had no role in the conduct or analysis of the data. JB receives funding from the NIHR UCL/UCLH Biomedical Research Centre.

Conflict of interest

MC: none; SH: travel grant from Gilead and MSD; MS: reports other from Viroclinics Bioscience B.V., outside the submitted work; KE: reports grant No 304875 (PathSeek) from EU FP7, during the conduct of the study; HT: none; JB: none; BR: grants from MSD, during the conduct of the study; grants from Gilead, other from MSD, Jansen-Cilag, BMS, Gilead, Pfizer, ViiV, outside the submitted work

Corresponding author contact information

S.J.Hullegie

's Gravendijkwal 230

3015CE, Rotterdam

The Netherlands

Tel: +31107032482

Fax: +31107030040

b.hullegie@erasmusmc.nl

Abstract

Objectives; Within HIV positive men having sex with men the epidemic of hepatitis C (HCV) is ongoing. Transmission of resistant variants of HCV after failure of treatment with directly acting antivirals (DAA) could be a major threat to the effectivity of therapy. We determined whether HCV resistant variants to DAAs were prevalent amongst patients with an acute HCV infection diagnosed in 2013 and 2014 in the Netherlands.

Methods: Target enrichment for viral nucleic acid separation and deep sequencing were used to recover whole HCV genomes of 50 patients with an acute HCV infection. The genomes were assembled by *de novo* assembly and analysed for known DAA resistance mutations.

Results: In acute HCV infected treatment-naïve patients the relevant resistance associated substitutions were Q80K (40%) in NS3/4a, M28V (24%) and Q30H combined with Y93H (2%) in NS5A and M414T (2%) or S556G (2%) in NS5b. Patients who failed on boceprevir, peginterferon and ribavirin therapy developed mutations in NS3 at position T54A and R155K.

Conclusion: Target enrichment and whole genome sequencing was successfully applied directly on clinical samples from patients with an acute HCV infection.

Introduction

Since very effective and well-tolerated direct-acting antivirals (DAAs) have become available, hepatitis C (HCV) has progressed from a difficult to treat to an almost universally curable infection. Currently, there is an urgent need for reliable sequencing techniques to determine whether resistant variants influence treatment efficacy. Whole genome sequencing (WGS) directly from clinical samples can contribute significantly to the understanding of clinical population structures, including minority populations, mixed infections and resistance mutations(1). In this study, we apply the target enrichment approach directly to clinical specimens of acute HCV infected patients treated in the Dutch Acute HCV in HIV Study (DAHHS) (2).

Materials and Methods

Baseline samples (initiation of therapy) of acute HCV genotype 1 infected patients were available from 50 patients participating in the DAHH-Study (supplementary table S1). Two patients with a poor response to therapy at week 4 (above 1000 IU/ml) and three patients who relapsed after treatment were analyzed for acquired resistance. The total number of samples analyzed was 55.

Double stranded cDNA synthesis

Nucleic acid was extracted from EDTA plasma using the DSP Virus/Pathogen kit on the QiaSymphony with the complex200 protocol and nucleic acid was eluted in 60ul. RNA was concentrated to ~11 µl and the complete sample was used for first strand cDNA synthesis (Superscript III Reverse Transcriptase kit, Life Technologies). Second strand cDNA synthesis used 20 µl from the first stand synthesis (Second Strand cDNA synthesis kit, NEB). All of the ds cDNA was purified (Genomic DNA clean and concentrator ™ kit, Zymo Research) eluting in 30 µl of ultrapure nuclease free H_2O.

SureSelect^{XT} Target Enrichment: Library preparation, hybridisation and enrichment

The purified ds cDNA were quantified using the Qubit (dsDNA HS assay kit, Life Technologies) and 200-500 ng of ds cDNA was sheared for 150 seconds, using a Covaris E220 focused ultra-sonication system (PIP 175, Duty factor 5, Cycles per Burst 200). End-repair, adapter ligation, hybridisation, PCR (12 cycles pre-capture and 18 or 22 cycles post capture) and all post- reaction clean-up steps were performed according to the SureSelect^{XT} Automated Target Enrichment for Illumina Paired-End Multiplexed Sequencing 200 ng protocol (version F.2) on the Bravo platform WorkStation B from Agilent Technologies. The 120-mer RNA baits spanning the length of 953 GenBank HCV partial and complete reference genomes were designed using an in-house PERL script developed by the PATHSEEK consortium and synthesised by Agilent Technologies. The samples were sequenced in multiple runs on an Illumina MiSeq sequencing platform with 500 bp v2 reagent sets.

Sequence data analysis

Genome mapping, assembly and finishing was performed using CLC Genomics Workbench (version 7.5/7.5.1) including the CLC Microbial Genome Finishing Module (version 1.4) from Qiagen. Trimmed reads were pre-filtered against a GenBank reference list containing 953 partial and complete HCV genomes and a subset of 50,000-100,000 HCV specific reads were subsampled and used for *de novo* assembly. Reference based mapping, using the best matching GenBank HCV reference and high stringency affinity gap cost was performed for samples failing *de novo* assembly.

Positions of interest for DAA resistance (see tables S2-S4 in supplementary material) were identified in the GenBank reference HCV strain H77 (accession no. AF011751). All trimmed reads were mapped against the genes of interest (NS3/NS4a, NS5a and NS5b) using the default affine gap cost parameters followed by removal of duplicated mapped reads. Low frequency variant detection was called with a min. coverage filter of 20x and a min. count of 5 independent reads. A relative read direction filter was

used and variants with a read direction distribution significantly different from the expected were removed.

Results

Result of whole genome sequencing

In this study we performed WGS directly on clinical specimens from 55 HCV positive samples. Using *de novo* assembly, we generated complete genomes (>90-100% recovery of the HCV genome) from 51 samples, partial genomes (>80-90% recovery of the HCV genome) from 3 samples and 1 low titre sample (3070 IU/ml) failed to generate a HCV genome (<50% recovery of the HCV genome) (table S5 in supplementary material). Overall, the sequence data showed strong correlation between the number of HCV copies in the diagnostic sample and the mean read depth across the genome (Pearson correlation coefficient 0.39 P=0.005). From this data-set, the lower limit of detection for WGS of HCV directly from clinical samples was estimated to a diagnostic value of ~3500 IU/ml (figure 1).

Four samples (Ol12, Na01, Ad04, Ef06) failed *de novo* assembly despite having >90% on-target reads (reads mapping to HCV). Reference based mapping showed, from the highly variable region of the HCV *E2* region (H77 numbering 872-1968), an inter-subtype (GT1A plus GT1A) dual infection in patient Ol12(data not shown). Due to lack of sample material this could however not be confirmed using classical techniques for detecting dual infection. A second consensus sequence could not be recovered from patient Na01 and Ad04 indicating that these patients did not have a dual infection. The two consensus sequences recovered from patients Ef06 rooted very close to each other in a phylogenetic tree, suggesting strain variation rather than dual infection within the patient.

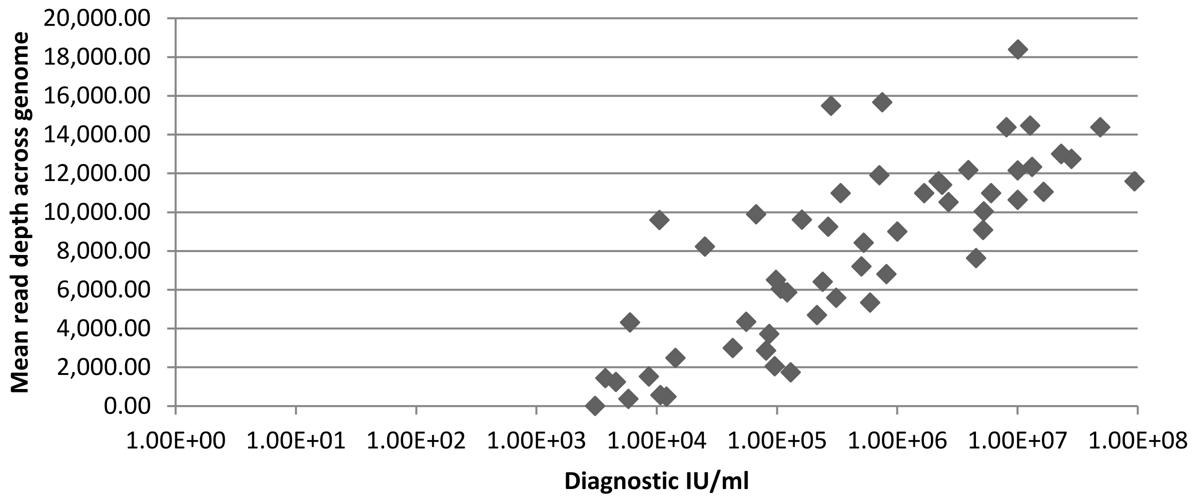
Directly Acting Antivirals (DAAs) resistance analysis

In the DAHHS cohort DAA resistant associated substitutions (RASs) were identified in 31 of the 50 baseline samples. In 38.7% (12/31) of the samples a single DAA RASs was identified and in 61.3% (19/31) multiple RASs were detected. The most prominent RAS in NS3 was Q80K in 40% (20/50) of the samples. In the NS5a region, M28V 24% (12/50) and H58P 26% (13/50) were found as dominant. The A421V 22% (11/50) and I585V 12% (6/50) were seen as most occurring RASs in the NS5b region (Figures 2 a/b/c). Remarkably, in patient Me05, an uncommon combination of Q30H and Y93H in NS5a occurred. This combination is associated with 93.136 fold resistance in EC50 against daclatasvir (3). Additional results of treatment failures are discussed in the supplements.

Discussion

In this study we used the combination of whole-genome enrichment and deep sequencing to recover close to complete genomes for 98% of the sample set. We used the sequence data to assess the prevalence of known DAA resistance mutations within the acute HCV genotype 1 infected patients and fortunately, between 2013 and 2014 no convincing evidence of transmitted drug resistance was found. However, in high-risk groups with ongoing transmission, HCV treating physicians should remain aware of circulating DAA resistant variants when treatment failure is observed.

Figure 1 Relationship between mean read depth across the whole genome and the diagnostic viral load



Legend figure 1. Figure 1 shows the relationship between the HCV copy number identified in the diagnostic sample (IU/ml x 2.7 = copies/ml) and the mean read depth obtained across the whole HCV genome.

Figure 2 Baseline resistance

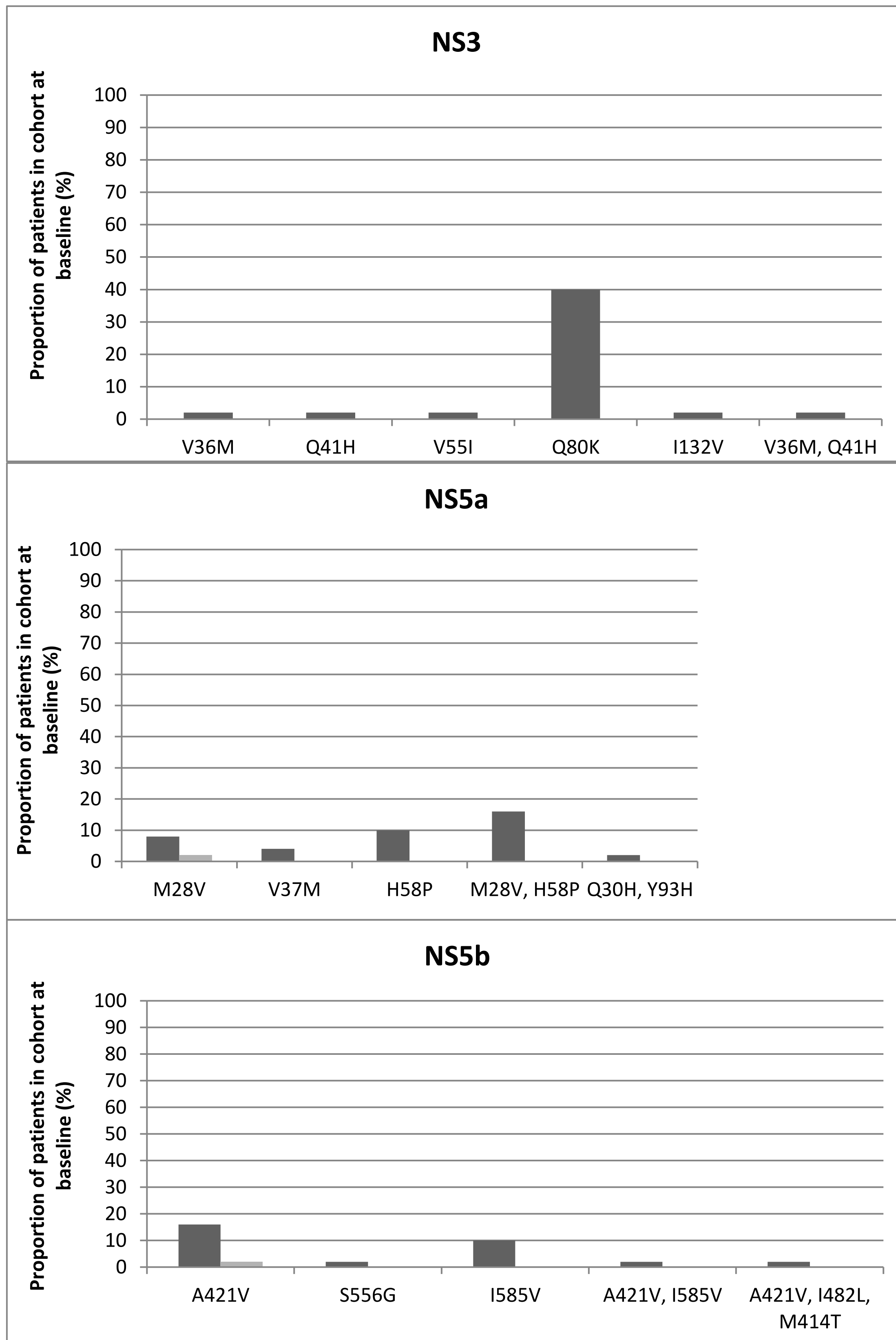


Figure 2 illustrates the known DAA resistance associated mutation identified at baseline (n=50) in the acute HCV cohort. The vertical axis shows the proportion of patients (in percentage) in which the different mutations were identified. a) Known DAA resistance associated mutations in NS3/NS4a, b) Known DAA resistance associated mutations in NS5a and c) Known DAA resistance associated mutations in NS5b.

Footnotes

Acknowledgements: Suzan D. Pas, Gertine W. van Oord, Andre Boonstra. We acknowledge all partners within the PATHSEEK consortium (University College London, Erasmus MC, QIAGEN AAR, and Oxford Gene Technology). Collaborators in the DAHH-Study: Mark A.A. Claassen, Guido E.L. van den Berk, Jan T.M. van der Meer, Dirk Posthouwer, Fanny N. Lauw, Eliane M.S. Leyten, Peter P. Koopmans, Clemens Richter, Arne van Eeden, Wouter F.W. Bierman, Joop E. Arends. We acknowledge the infrastructure support of the MRC Centre for Medical Molecular Virology

References

1. Thomson E, Ip CL, Badhan A, Christiansen MT, Adamson W, Ansari MA, et al. Comparison of next generation sequencing technologies for the comprehensive assessment of full-length hepatitis C viral genomes. J Clin Microbiol. 2016.
2. Hullegie SJ, Claassen MA, van den Berk GE, van der Meer JT, Posthouwer D, Lauw FN, et al. Boceprevir, peginterferon and ribavirin for acute hepatitis C in HIV infected patients. J Hepatol. 2016;64(4):807-12.
3. Fridell RA, Wang C, Sun JH, O'Boyle DR, 2nd, Nower P, Valera L, et al. Genotypic and phenotypic analysis of variants resistant to hepatitis C virus nonstructural protein 5A replication complex inhibitor BMS-790052 in humans: in vitro and in vivo correlations. Hepatology. 2011;54(6):1924-35.

Supplements

Development of RASs during therapy

Insufficient response at week 4 of therapy

Patient Ad04 was diagnosed with genotype 1a and had a baseline HCV viral load of $16.9x10^6$ IU/ml. At week 4 the patient presented with a viral load of $23.1x10^6$ IU/ml and reported that he had discontinued therapy one week before. BOC resistance associated mutations were not identified in patient Ad04 at the two sampling points. Surprisingly, based on the sequence results this patient turned out to have a genotype 4 infection at both time points instead of the genotype 1.

Patient On14 was diagnosed with genotype 1a with a baseline viral load of $16.4x10^6$ IU/ml. At week four of treatment he had a viral load of 5790 IU/ml and reported to be treatment adherent. No baseline DAA RASs were detected, but after four weeks of therapy 2 BOC associated mutations (T54A and R155K) were identified at 53.3% and 14% variant frequencies, that result in a 10-fold decrease in BOC susceptibility (Tong et al, Biochemistry).

Relapse 12 weeks after treatment

Patient Ec03 was treated for 12 weeks and had an undetectable HCV RNA at week 4, but became detectable at week 12 after treatment. The baseline and week 12, post-treatment samples showed two different HCV genotype 1a strains. Second best reference approach showed that the genotype present at failure was not detected at baseline (data not shown). Therefore, WGS proved that this patient did not have a relapse but a reinfection.

Patient Md04 (genotype 1b) and Oe05 (genotype 1a) were successfully treated for 12 weeks. Both had a detectable viral load at week 4 of treatment but became negative at week 8. Both patients became detectable at week 12 after therapy and strains in both cases were identical to the baseline strains. At the time of relapse, patient Oe05 showed a R155K dominant (92%) mutation in the NS3 region, which is associated with major BOC resistance. In contrast, Md04 had no known acquired BOC mutations at relapse but was positive for the I132V RASs in the NS3 region at baseline.

Supplementary table S1. Baseline characteristics

		n=50	
male		100%	
age	years	40	(33-47)
genotype 1	a	98%	
	b	2%	
HCV viral load	IU/ml	439.500	(40.450-2.362.500)
Cd4	E6/mL	650	(450-833)
Il 28B genotype			
CC		38%	
non-CC		54%	
unknown		8%	

Table S1 Legend. Characteristics of patients analysed at initiation of treatment. All results are presented as median with interquartile range or as percentage.

Table S2. Mutations known to be associated with resistance in NS3/NS4a

Amino acid position (per gene)	Sensitive aa	Amino acid in ref. H77 1a AF011751	3-letter code in H77	Resistant aa
36	V	V	Val	M/L
41	Q	Q	Gln	R/N/L/K/H
43	F	F	Phe	V/S/L/I
54	T	T	Thr	S/A
55	V	V	Val	I/A
56	Y	Y	Tyr	H
80	Q	Q	Gln	R/N/L/K/H
95	T	T	Thr	S
109	R	R	Arg	K
122	S	S	Ser	S/R/N/G/A
132	I	I	Ile	V
155	R	R	Arg	W/T/S/Q/M/K/G
156	A	A	Ala	V/T/S/N/G/F
168	D	D	Asp	Y/N/I/H/G/E/A
170	V	I	Ile	A

Legend table S2. Table S2 lists the known DAA resistance associated mutations in NS3/NS4a which was assessed in this study.

Table S3. Mutations known to be associated with resistance inNS5a

Amino acid position (per gene)	**Sensitive aa**	**Amino acid in ref. H77 1a AF011751**	**3-letter code in H77**	**Resistant aa**
28	M/L	M	Met	V/T
30	R/Q	Q	Gln	E/H/Q/R
31	L	L	Leu	F/M/V
37	V	V	Val	M
58	H	H	His	P
93	Y	Y	Tyr	C/H/N

Legend table S3. Table S3 lists the known DAA resistance associated mutations in NS5a which was assessed in this study.

Table S4. Mutations known to be associated with resistance in NS5b

Amino acid position (per gene)	Sensitive aa	Amino acid in ref. H77 1a AF011751	3-letter code in H77	Resistant aa
282	S	S	Ser	T
316	C	C	Cys	Y/N
365	S	S	Ser	?
368	S	S	Ser	T
395	A	A	Ala	G
411	N	N	Asn	S
414	M	M	Met	T
419	L	L	Leu	S
421	A	A	Ala	V
422	R	R	Arg	K
423	M	M	Met	V/T
444	N	N	Asn	K
445	C	C	Cys	F/N/G
448	Y	Y	Tyr	H/C
451	C	C	Cys	S/R
482	I	I	Ile	L
486	A	A	Ala	V
494	V	V	Val	?
495	P	P	Pro	S/L

496	P	P	Pro	?
499	V	A	Ala	?
553	A	A	Ala	V
556	S	S	Ser	F/N/G
559	D	D	Asp	G
565	S	S	Ser	F
585	I	I	Ile	V

Legend table S4. Table S4 lists the known DAA resistance associated mutations in NS5b which was assessed in this study.

Table S5. Overview of sequence results

Patient ID	Diagnostic value IU/ml	Note	Total no. of reads	On-target reads (%)	Coverage of ref. (%)	Mean read depth	Consensus Sequence (bp) length	Out-come
Eb02	9.55E+04	baseline	642,798	69.24	89.02	2,060.07	8546	partial
Ec03	5.24E+05	baseline	1,480,434	95.46	97.93	8,426.85	9401	complete
Ed04	1.67E+06	baseline	1,237,426	95.33	97.75	10,976.69	9384	complete
Ee05	4.87E+07	baseline	1,227,084	97.34	98.43	14,376.27	9449	complete
Ef06	1.91E+07	baseline	1,078,260	96.63	96.85	19,075.34	9298	complete
Eh08	1.00E+07	baseline	1,305,908	96.89	97.63	12,147.69	9372	complete
Ei09	8.13E+05	baseline	1,249,624	89.5	97.72	6,812.59	9381	complete
ma01	5.95E+05	baseline	892,498	72.53	96.82	5,331.41	9295	complete
Na01	6.85E+05	baseline	1,035,510	96.06	97.09	20,855.58	9321	complete
Oa01	4.27E+04	baseline	634,808	92.45	92.97	2,987.47	8587	complete
Ra01	1.20E+04	baseline	285,792	47.59	92.69	484.47	8898	complete
Oc03	1.05E+04	baseline	1,157,600	90.71	98.20	9,595	9427	complete
Of06	1.43E+04	baseline	821,494	42.25	98.10	2,485	9418	complete
Od04	3.38E+05	baseline	1,239,850	92.93	98.30	10,985	9437	complete
Ob02	8.06E+06	baseline	1,343,942	70.57	98.33	14,385	9440	complete
Og07	8.08E+04	baseline	928,300	39.4	91.39	2,863	8773	complete
Oe05	1.27E+07	baseline	1,033,120	94.9	86.10	14,460	8266	partial
Aa01	2.81E+05	baseline	1,459,442	93.39	98.17	15,485	9424	complete
Ab02	6.03E+06	baseline	918,110	93.33	98.28	10,985	9435	complete
Ek11	1.07E+05	baseline	904,798	67.98	98.02	6,056	9410	complete
Em13	1.61E+05	baseline	1,104,180	83.41	99.15	9,612	9518	complete
Eo15	2.31E+07	baseline	1,036,060	92.87	97.60	13,013	9370	complete
Ep16	2.80E+07	baseline	1,084,814	94.28	98.00	12,755	9408	complete

Eq17	2.65E+05	baseline	1,305,270	81.85	98.02	9,250	9410	complete
Es19	5.23E+06	baseline	896,744	87.2	97.73	10,045	9382	complete
Ea01	7.49E+05	baseline	1,577,244	94.03	95.11	15,671	9131	complete
Ac03	9.81E+04	baseline	947,992	85.39	97.74	6,508	9383	complete
Ad04	1.69E+07	baseline	870,658	89.66	96.86	12,099	9299	complete
Ae05	3.10E+05	baseline	829,314	87.15	98.53	5,587	9459	complete
Af06	6.71E+04	baseline	1,079,900	91.65	96.00	9,890	677, 4366, 4174	complete
Ag07	5.55E+04	baseline	1,037,550	81.82	97.79	4,345	9388	complete
Mc03	9.40E+07	baseline	934,552	95.53	98.19	11,596	9426	complete
Md04	1.21E+05	baseline	911,260	89.46	98.04	5,864	9412	complete
Me05	4.52E+06	baseline	1,002,154	91.98	97.79	7,637	9388	complete
Mf06	2.51E+04	baseline	1,019,344	94.08	98.15	8,227	9422	complete
On14	1.64E+07	baseline	898,712	96.92	96.42	11,054	7643, 1613	complete
Op16	1.00E+07	baseline	789,490	94.68	90.65	10,643	599, 2702, 2309, 3093,	complete
Oq17	1.07E+04	baseline	594,852	20.25	88.75	572	7595, 925	partial
Os19	2.36E+06	baseline	1,014,772	92.6	98.59	11,406	9465	complete
Ot20	2.66E+06	baseline	875,084	95.74	97.99	10,521	9407	complete
Ec03	5.04E+05	week 24	926,832	91.63	94.00	7,208	9024	complete
Oe05	3.90E+06	week 24	1,250,586	97.19	97.97	12,173	9405	complete
Sc03	8.59E+04	baseline	716,228	96.14	97.68	3,716	9377	complete
Sd04	2.40E+05	baseline	1,022,048	92.17	97.92	6,415	9400	complete
Ua01	1.32E+07	baseline	1,006,068	97.52	98.02	12,330	9410	complete
Ub02	1.01E+07	baseline	2,075,844	92.81	98.13	18,395	9420	complete
Uc03	7.10E+05	baseline	1,530,604	94.73	97.85	11,909	9394	complete
Ad04	2.30E+07	week 4	979,162	96.94	96.86	19,789.72	9299	complete
Md04	2.14E+05	week 24	956,070	94.4	93.70	4,688	9383	complete
On14	5.79E+03	week 4	638,858	15.77	93.70	364	9205	complete

Ol12	4.58E+03	baseline	2,410,436	30.97	95.97	16,081	9214	complete
En14	3.73E+03	baseline	3,607,426	6.58	95.86	1,448.10	9203	complete
Ou21	6.00E+03	baseline	2,135,530	28.53	96.36	4,308.14	9251	complete
Eg07	8.59E+03	baseline	1,942,860	10	91.20	1,520.65	8755	complete
El12	3.07E+03	baseline	1,799,658	0.78	31.00	0.92	-	fail

Legend table S5 .The column on-target reads (%) refers to the proportion of the total number of reads from each sample mapping to HCV. Coverage of ref. (%) is the fraction of the HCV genome recovered from the WGS and is estimated from the consensus sequence (bp) length divided by 9600 (bp), which is an approximated size of the HCV genome. The mean read depth is the average sequence coverage across the genome. Sequence outcome is defined as complete when >90-100% of the HCV genome is recovered (coverage of ref.), partial when >80-90% of the HCV genome is recovered and fail when <50% of the HCV genome is recovered.