



Abdelrahman, Tamer. (2015) *Quasispecies dynamics and treatment outcome during early hepatitis C infection in a cohort of HIV-infected men*. PhD thesis.

<http://theses.gla.ac.uk/7402/>

Copyright and moral rights for this work are retained by the author

A copy can be downloaded for personal non-commercial research or study, without prior permission or charge

This work cannot be reproduced or quoted extensively from without first obtaining permission in writing from the author

The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the author

When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given

Enlighten:Theses
<http://theses.gla.ac.uk/>
theses@gla.ac.uk

Quasispecies Dynamics and Treatment Outcome During Early Hepatitis C Infection in a Cohort of HIV-infected Men

Tamer Abdelrahman

MBBch, MSc

Submitted for the fulfilment of the requirements of degree
of Doctor of Philosophy

School of Medicine
College of Medical, Veterinary and Life Sciences
University of Glasgow

November 2015

Abstract

Hepatitis C virus (HCV) is emerging as one of the leading causes of morbidity and mortality in individuals infected with HIV and has overtaken AIDS-defining illnesses as a cause of death in HIV patient populations who have access to highly active antiretroviral therapy.

For many years, the clonal analysis was the reference method for investigating viral diversity. In this thesis, a next generation sequencing (NGS) approach was developed using 454 pyrosequencing and Illumina-based technology. A sequencing pipeline was developed using two different NGS approaches, nested PCR, and metagenomics. The pipeline was used to study the viral populations in the sera of HCV-infected patients from a unique cohort of 160 HIV-positive patients with early HCV infection. These pipelines resulted in an improved understanding of HCV quasispecies dynamics, especially regarding studying response to treatment.

Low viral diversity at baseline correlated with sustained virological response (SVR) while high viral diversity at baseline was associated with treatment failure. The emergence of new viral strains following treatment failure was most commonly associated with emerging dominance of pre-existing minority variants rather than re-infection.

In the new era of direct-acting antivirals, next generation sequencing technologies are the most promising tool for identifying minority variants present in the HCV quasispecies populations at baseline. In this cohort, several mutations conferring resistance were detected in genotype 1a treatment-naïve patients. Further research into the impact of baseline HCV variants on SVR rates should be carried out in this population.

A clearer understanding of the properties of viral quasispecies would enable clinicians to make improved treatment choices for their patients.

Table of contents

ABSTRACT	II
TABLE OF CONTENTS	III
LIST OF TABLES	VIII
LIST OF FIGURES.....	X
ACKNOWLEDGEMENTS	XIV
AUTHOR'S DECLARATION.....	XV
DEFINITIONS/ ABBREVIATIONS	XVI
PUBLICATIONS	XIX
CHAPTER 1: INTRODUCTION	20
1.1 HEPATITIS C VIRUS (HCV)	20
1.2 HCV DISCOVERY AND CLASSIFICATION.....	20
1.3 HCV GENOME.....	20
1.3.1 5' Untranslated region (5'UTR).....	21
1.3.2 3' Untranslated region (3'UTR).....	21
1.3.3 Frame-shift protein	21
1.3.4 Core protein.....	21
1.3.5 Envelope glycoproteins	22
1.3.6 Protein p7	22
1.3.7 Non-structural proteins.....	22
1.4 HCV REPLICATION	25
1.5 HCV REPLICONS	26
1.5.1 Adaptive mutations	26
1.5.2 Genotype 1a replicons.....	27
1.5.3 Reporter genes.....	27
1.5.4 Permissive cell lines for HCV replication	28
1.5.5 Applications of the HCV replicon system	28
1.6 EPIDEMIOLOGY OF HEPATITIS C VIRUS INFECTION.....	29
1.7 STUDYING THE VIRAL POPULATION WITHIN INFECTED PATIENTS.....	30
1.7.1 Replicative homeostasis hypothesis	33
1.7.2 Quasispecies theory	33
1.7.3 Methods of quasispecies analysis	34
1.8 NEXT GENERATION SEQUENCING	37
1.8.1 Digital data.....	37
1.8.2 Sources of noise and error models.....	38
1.8.3 454 Pyrosequencing	38
1.8.4 Illumina® technology.....	39

1.8.5 Ion Torrent™ technology.....	40
1.8.6 Single-molecule DNA sequencing.....	41
1.9 ANALYSIS OF DEEP SEQUENCE READS	42
1.9.1 Mapping	43
1.9.2 Haplotype reconstruction	43
1.10 HIV AND HEPATITIS C VIRUS CO-INFECTION.....	44
1.10.1 The influence of HCV on HIV infection.	46
1.10.2 The influence of HIV on HCV infection.	47
1.11 TREATMENT OF ACUTE HCV IN HIV-POSITIVE PATIENTS.....	48
1.11.1 Predictors of treatment response in HIV-positive individuals with acute HCV infection.....	49
1.12 A NEW ERA OF ANTIVIRALS FOR HCV INFECTION	52
1.12.1 Simeprevir.....	52
1.12.2 Daclatasvir	53
1.12.3 Sofosbuvir	53
1.12.4 PrOD combination (Paritaprevir, Ombitasvir, and Dasabuvir)	54
1.13 THE HCV AVAILABLE TREATMENT REGIMENS	55
1.13.1 Interferon-containing regimens.....	55
1.13.2 Interferon-free options.....	57
1.14 HCV DRUG RESISTANCE	66
1.14.1 Genetic barrier for resistance	67
1.14.2 Natural polymorphisms.....	68
1.14.3 Persistence of resistance-associated variants.....	73
1.14.4 Factors influencing resistance before treatment	74
1.15 PREVENTION OF TREATMENT FAILURE WITH DAAS	78
1.15.1 HCV drug monitoring resistance tools	78
1.16 RETREATMENT OF HCV	79
1.17 AIMS AND HYPOTHESES.....	82
CHAPTER 2: MATERIALS AND METHODS	83
2.1 PATIENT COHORT	83
2.1.1 Clinical groups and definitions	83
2.1.2 Treatment outcome subgroups.....	84
2.1.3 Antiviral resistance study cohort	85
2.2 EXTRACTION OF VIRAL RNA FROM PLASMA SAMPLES	85
2.2.1 Manual RNA extraction	85
2.2.2 Automated RNA extraction.....	86
2.3 cDNA SYNTHESIS	86
2.3.1 Superscript III® reverse transcriptase	86

2.3.2 Maxima H Minus Reverse Transcriptase®	86
2.4 DNA SECOND STRAND SYNTHESIS	87
2.5 PCR AMPLIFICATION	87
2.5.1 PCR error rate estimation using different enzymes.....	87
2.5.2 Amplification of the E2 HVR1 region	89
2.5.3 Quantitative real-time PCR assays.....	90
2.6 GEL ELECTROPHORESIS	92
2.7 PURIFICATION OF DNA	92
2.7.1 Isolation and Purification of DNA from Agarose Gels	92
2.7.2 DNA purification using Agencourt AMPure XP® beads.....	93
2.8 FLUOROMETRIC MEASUREMENT OF NUCLEIC ACID CONCENTRATION	93
2.9 NUCLEOTIDE SEQUENCING AND ANALYSIS	94
2.9.1 Sanger sequencing and analysis	94
2.9.2 454 Pyrosequencing analysis.....	94
2.9.3 Illumina® sequencing and analysis	95
2.10 FULL GENOME SEQUENCING	104
2.10.1 Nested PCR-based full genome sequencing of HCV	104
2.10.2 Metagenomic sequencing	111
2.11 BACTERIAL CLONING	111
2.11.1 TOPO-TA® Cloning.....	111
2.11.2 CloneJET® PCR cloning	112
2.11.3 One shot® Top 10 cells	112
2.11.4 NEB 5-alpha competent <i>E. coli</i> ®	112
2.11.5 NEB 10-beta competent <i>E. coli</i> ®	113
2.12 PREPARATION OF DNA FOLLOWING BACTERIAL CLONING	113
2.12.1 Small scale plasmid preparation from transformed bacteria	113
2.12.2 Large scale plasmid preparation from transformed bacteria	114
2.13 MOCK COMMUNITY.....	116
2.13.1 Preparation of the mock community.....	116
2.13.2 MiSeq sequencing	116
2.13.3 Haplotype reconstruction.....	116
2.14 STATISTICAL ANALYSIS	117
2.15 BIOINFORMATICS METHODS FOR VIRAL SEQUENCE ANALYSIS	117
2.15.1 Sequence alignment	117
2.15.2 Measures of diversity	117
2.15.3 Mapping.....	119
2.15.4 Construction of phylogenetic trees	119
2.16 RESISTANCE MUTATION DATABASE	120

2.17 HUMAN HEPATOMA CELLS (HUH7.5)	123
2.18 <i>IN VITRO</i> REPLICATION ASSAYS.....	123
2.18.1 Site-Directed Mutagenesis	123
2.18.2 Introduction of mutations into the HCV-1a subgenomic replicon ...	124
2.18.3 Restriction enzyme digestion of DNA	125
2.18.4 DNA ligation.....	125
2.18.5 <i>In vitro</i> transcription	125
2.18.6 Electroporation of RNA.....	129
2.18.7 Firefly luciferase activity assay.....	129
2.18.8 Prediction model for detecting resistance mutations for SOF	129
2.18.9 Electroporation of RNA.....	130
2.18.10 Firefly luciferase activity assay	131
CHAPTER 3: INTRA-HOST VIRAL POPULATION STRUCTURES	133
3.1 BACKGROUND	133
3.2 RESULTS	135
3.2.1 PCR error rate estimation	135
3.2.2 454 pyrosequencing error rate estimation	135
3.2.3 Measuring viral diversity using HVR1 region.....	135
3.2.4 Viral complexity	136
3.2.5 Comparison of sequencing techniques	136
3.2.6 Full genome sequencing of HCV.....	143
3.3 PCR AMPLIFICATION STRATEGY.....	149
3.3.1 Optimization of nested PCR	149
3.3.2 The PCR amplification process	154
3.3.3 Metagenomic sequencing.....	161
3.3.4 Illumina® sequencing of HCV full genome	166
3.3.5 Mock community experiment.....	178
3.4 DISCUSSION	182
3.4.1 The role of NGS in understanding HCV viral population structure ...	182
3.4.2 HCV full genome sequencing	185
3.4.3 Bioinformatics framework.....	188
3.5 CONCLUSIONS	191
CHAPTER 4: HEPATITIS C VIRUS DIVERSITY.....	192
4.1 BACKGROUND	192
4.2 RESULTS	195
4.2.1 Viral diversity as a predictor of treatment outcome	195
4.2.2 Superinfection versus relapse	205
4.3 DISCUSSION	225

4.3.1 Predictors of outcome in patients treated with PegIFN α /RBV.	225
4.3.2 Transmission diversity	227
4.3.3 Definition of relapse and re-infection in HCV	228
4.3.4 HCV compartmentalisation	231
4.4 CONCLUSION	233
CHAPTER 5: HCV ANTIVIRAL DRUG RESISTANCE	234
5.1 BACKGROUND	234
5.1.2 Replicative fitness.....	239
5.1.3 Importance of resistance testing.....	239
5.2 RESULTS	244
5.2.1 Natural polymorphisms in the NS3 protease region	244
5.2.2 Natural polymorphism in the NS5A gene.....	245
5.2.3 Natural polymorphisms in the NS5B gene.....	251
5.2.4 Design of a genotype 1a replicon-based luciferase assay	264
5.3 DISCUSSION	281
5.3.1 Prevalence of natural polymorphisms at resistance-associated residues.	282
5.3.2 Prediction of <i>in vitro</i> resistance mutations within NS5B	286
5.3.3 Resistance profile in HIV/HCV co-infected patients	288
5.3.4 The potential role of NGS in antiviral resistance testing	289
5.3.5 Clinical utility of resistance testing	290
5.4 CONCLUSIONS	291
CHAPTER 6: CONCLUSIONS AND FURTHER WORK.....	293
6.1 CONCLUSIONS	293
6.1.1 Sequencing of the HCV genotype 1a open reading frame.....	293
6.1.2 The role of NGS in studying HCV quasispecies	293
6.1.3 The effect of viral dynamics on treatment outcome	294
6.1.4 Antiviral resistance in HCV.....	295
6.1.5 Replicon based assay for RAVs replication capacity	295
6.2 FUTURE WORK	296
6.2.1 Sequencing of the HCV genotype 1a open reading frame.....	296
6.2.2 The effect of viral dynamics on treatment outcome	296
6.2.3 Mock communities.....	297
6.2.4 HCV antiviral resistance	299
6.2.5 Cost - effectiveness of NGS in the diagnostic setting.....	299
APPENDICES	302
BIBLIOGRAPHY	335

List of Tables

Table 1-1: Directly acting antivirals efficacy studies (Part 1)	58
Table 1-2: Directly acting antivirals efficacy studies (Part 2)	59
Table 1-3: Directly acting antivirals efficacy studies (Part 3)	60
Table 1-4: Directly acting antivirals efficacy studies (Part 4)	61
Table 1-5: EASL Recommendations on Treatment of Hepatitis C 2015.	62
Table 2-1: Genotype-specific primers used for nested PCR.	91
Table 2-2: Phred quality scores.	96
Table 2-3: Primer sets and optimised conditions for amplicon 1.	107
Table 2-4: Primer sets for amplicon 2.	108
Table 2-5: Primer set for amplicon 3.	109
Table 2-6: Primer sets for amplicon 4.	110
Table 2-7: Mutagenesis primers.	127
Table 2-8: Sequencing primers.	127
Table 2-9: Ligation of insert DNA into plasmid vector DNA.	132
Table 2-10: Recircularization of linear DNA using Rapid DNA Ligation Kit.	132
Table 3-1: Overview of experiments discussed in chapter 3.	137
Table 3-2: Viral complexity and viral diversity using different sequencing techniques.	140
Table 3-3: The distribution of reads generated following Superscript III® cDNA synthesis.	168
Table 3-4: The distribution of reads generated following Maxima®RT.	169
Table 3-5: Comparison between random hexamers and HCV specific primer based cDNA synthesis.	170
Table 3-6: Analysis of reads generated by metagenomic and nested PCR-based sequencing.	176
Table 3-7: Genetic distance between consensus sequences.	176
Table 3-8: Mock community; pairwise distance among 13 clones.	179
Table 4-1: Characteristics of the study groups of subjects.	197
Table 4-2: Comparison of quasispecies diversity in TF group versus SVR group.	197
Table 4-3: Measures of diversity in the treatment failure group (TF).	198
Table 4-4: Measures of diversity in responder group (SVR).	198
Table 4-5: Clinical characteristics of the treatment failure cohort.	207
Table 4-6: Characteristics of viral population dynamics and treatment response in patients with treatment failure.	209

Table 5-1: Prevalence of resistance-associated variants within the NS3 gene in HIV/HCV co-infected patients and distribution at the intra-host level. ---- 250

Table 5-2: Prevalence of RAVs within NS5A in HIV/HCV co-infected patients and distribution at the intra-host level. ----- 258

Table 5-3: Prevalence of resistance-associated variants within NS5B protein in HIV/HCV co-infected patients and distribution at intra-host level. ----- 263

Table 5-4: List of mutations detected in predicted residues. ----- 265

List of Figures

Figure 1-1: HCV genome.	24
Figure 1-2: prevalence of hepatitis C virus genotypes in different regions.	31
Figure 1-3: Global distribution and prevalence of hepatitis C virus genotypes.	32
Figure 1-4: Schematic diagram representing the process of reconstructing viral haplotypes from next-generation sequencing data.	45
Figure 2-1: Design of fusion primers.	91
Figure 2-2: Workflow for PCR Purification using Agencourt AMPure XP® beads.	96
Figure 2-3: Nextera XT workflow.	100
Figure 2-4: Workflow of Illumina sequencing.	103
Figure 2-5: PCR (Amplicon) amplification strategy.	106
Figure 2-6: Examples of measures of diversity.	121
Figure 2-7: Relationship between diversity measures.	122
Figure 2-8: Genotype 1a HCV subgenomic replicon APP238 pH/SG-Neo (L+I).....	126
Figure 2-9: Schematic representation of mutations introduced into the replicon.	128
Figure 3-1: Estimation of the error rate of different polymerase enzymes.	138
Figure 3-2: A chromatogram showing an example of miscalling.	139
Figure 3-3: Illustrative example of multiple variants detected by CS and NGS.	141
Figure 3-4: Illustrative example of the superiority of NGS compared to Sanger sequencing in detecting viral variants.	142
Figure 3-5: Illustrative example of clonal analysis detecting a minority variant (0.06%).	144
Figure 3-6: Clonal analysis may not detect variants detected by deep sequencing. ---	145
Figure 3-7: Distribution of the number of variants detected by clonal analysis (CS) versus 454 pyrosequencing (NGS).	146
Figure 3-8: Comparison of variants detected using different NGS platforms.	147
Figure 3-9: Comparison of the performance of different RNA extraction methods.	151
Figure 3-10: Optimisation of incubation temperature for cDNA synthesis.	151
Figure 3-11: Comparison of the yield of cDNA synthesized using different HCV-specific primers.	152
Figure 3-12: Comparison of cDNA synthesis using Maxima® versus SuperScript III®.	152
Figure 3-13: Optimisation of modified primer sets for whole genome amplification of HCV genotype 1a.	155
Figure 3-14: Optimisation of PCR amplification of Amp1.	156
Figure 3-15: Optimisation of PCR amplification of Amp 2.	157
Figure 3-16: Optimisation of PCR amplification of Amp 3.	158

Figure 3-17: Optimisation of PCR amplification of Amp 4.	159
Figure 3-18 : PCR Amplification process.	160
Figure 3-19: Example of chromatograms derived from Sanger sequencing of HCV amplicons.	162
Figure 3-20: Effect of input DNA concentration in Nextera XT® on final library size. --	165
Figure 3-21: RNA yield following extraction by different nucleic acid extraction methods.	167
Figure 3-22: Number of mapped HCV reads generated using different cDNA synthesis systems.	169
Figure 3-23: Comparison of primer choice for reverse transcription.	170
Figure 3-24: Metagenomic approach process.	171
Figure 3-25: Data analysis pipeline.	173
Figure 3-26: Quality of reads generated using the amplicon sequencing approach.	174
Figure 3-27: An example of read coverage using the amplicon sequencing approach. -	175
Figure 3-28: An example of read coverage using the metagenomic approach.	175
Figure 3-29: Comparison between consensus sequences generated by metagenomic and amplicon approaches.	177
Figure 3-30: Performance of different quasispecies assembler using the mock community	180
Figure 3-31: Reconstructed haplotypes using PredictHaplo in one of the triplicate (3A).	181
Figure 3-32: Estimated frequency cut-off for detectability by different techniques. --	186
Figure 4-1: Comparison of diversity measures between non-responders (TF) and responders (SVR) groups.	199
Figure 4-2: Illustrative example of the difference in Shannon entropy between SVR group (P6) and TF group (P81).	200
Figure 4-3: HVR1 amino acid diversity.	201
Figure 4-4: Illustrative example of the difference in viral complexity in SVR (P6) and TF (P81) groups.	203
Figure 4-5: Illustrative example of Poisson Fitter from both groups.	204
Figure 4-6: Phylogenetic tree of direct Sanger sequences.	208
Figure 4-7: Comparison of viral complexity in paired serum samples in P38.	210
Figure 4-8: Comparison of viral complexity in paired serum samples in P63.	211
Figure 4-9: Comparison of viral complexity in paired serum in P67.	212
Figure 4-10: Comparison of viral complexity in paired serum samples in P81	213
Figure 4-11: Comparison of viral complexity in paired serum samples in P112.	214
Figure 4-12: Comparison of viral complexity in paired serum samples in P118.	215

Figure 4-13: Comparison of viral complexity in paired serum samples in P31.	216
Figure 4-14: Comparison of viral complexity in paired serum samples in P21.	217
Figure 4-15: Comparison of viral complexity in paired serum samples in P105.	218
Figure 4-16: Comparison of viral complexity in paired serum samples in P57.	219
Figure 4-17: Comparison of viral complexity in paired serum samples in P141.	220
Figure 4-18: Comparison of viral complexity in paired serum samples in P76.	221
Figure 4-19: Comparison of viral complexity in paired serum samples in P75.	222
Figure 4-20: Comparison of viral complexity in paired serum samples in P101.	223
Figure 4-21: Comparison of viral complexity in paired serum samples in P131.	224
Figure 4-22: Viral dynamics during treatment failure.	232
Figure 5-1: Schematic representation of Protease inhibitors resistance associated mutations detected <i>in vivo</i>	237
Figure 5-2: Schematic representation of NS5A inhibitors resistance associated mutations detected <i>in vivo</i>	240
Figure 5-3: Schematic representation of NS5B inhibitors resistance associated mutations detected <i>in vivo</i>	241
Figure 5-4: Model of HCV resistance in case of pre-existing minority RAV.	243
Figure 5-5: Prevalence of natural polymorphisms in NS3 of HIV/HCV co-infected patients and HCV mono-infected patients.	246
Figure 5-6: Frequency of RAVs within NS3 in HIV/HCV co-infected patients.	247
Figure 5-7: Distribution of baseline RAVs within NS3 of HIV/HCV co-infected patients.	248
Figure 5-8: Distribution of baseline RAVs in NS3 region of HCV mono-infected patients.	249
Figure 5-9: Prevalence of natural polymorphisms within NS5A in HIV/HCV coinfecting and HCV mono-infected patients.	253
Figure 5-10: Frequencies of resistant and non-resistant variants within NS5A in HIV/HCV Co-infected Patients.	254
Figure 5-11: Frequencies of resistant and non-resistant variants within NS5A region in HCV mono-infected patients.	255
Figure 5-12: Distribution of baseline RAVs in NS5A of HIV/HCV co-infected patients. -	256
Figure 5-13: Distribution of baseline RAVs in NS5A of HCV mono-infected patients. ---	257
Figure 5-14: Prevalence of Natural Polymorphisms in NS5B Region of HIV/HCV Co-infected Patients (HIV/HCV) and HCV mono-infected patients (HCV). ----	259
Figure 5-15: Frequencies of resistant and non-resistant variants within NS5B in HIV/HCV Co-infected Patients.	260

Figure 5-16: Frequencies of Resistant and Non-resistant Variants Within NS5A Region in HCV mono-infected Patients	261
Figure 5-17: Distribution of baseline RAVs in NS3 region of HCV mono-infected patients.	262
Figure 5-18: Predicted NS5B resistance associated mutations based on HCV structure and interaction sites.	265
Figure 5-19 Cloning strategy.	267
Figure 5-20: Restriction digest of replicon APP238 pH/SG-Neo (L+I).	268
Figure 5-21: Schematic illustration of a restriction digest of APP238 pH/SG-Neo (L+I) using <i>AgeI</i> and <i>KpnI</i>	269
Figure 5-22: pGFP-C1 as intermediate vector to introduce the Luciferase firefly cassette	270
Figure 5-23: Schematic illustration of TA/SG/Interim/HCV-1a with different mutations to be introduced.	271
Figure 5-24: E1726R mutation (GAG)-(GGG)	272
Figure 5-25: K1691R mutation (AAG) - (AGG)	273
Figure 5-26: TA/SG/Luc/HCV-1a culture adapted replicon.	274
Figure 5-27: Optimisation of a transient replication assay for the subgenomic replicon TA/SG/Luc/HCV-1a.	276
Figure 5-28: Replication capacity of mutated replicon (S96T, S282T).	277
Figure 5-29: Replication capacity of the mutated replicon (G493A, P495A, P496A). ---	278
Figure 5-30: Replication capacity of mutated replicon (R32A, R32H).	279
Figure 5-31: Replicative fitness of replicons containing predicted RAVs.	280
Figure 6-1: The structure of mock communities	298
Figure 6-2: Workflow of cost-effectiveness study of NGS in diagnostic setting	301

Acknowledgements

All praises to Allah for his guidance and blessings in every step of my life.

I would like to express my sincere gratitude to my supervisors Dr John McLauchlan and Dr Emma Thomson for the continuous support, patience, motivation, vast knowledge, and their contribution in developing me as a researcher. I would particularly like to thank Emma for making the PhD journey as smooth, boosting my confidence in my abilities and shaping my career and John for his open door policy and introducing me to the world of molecular biology. It was a pleasure exploring HCV secrets with him.

I would like to extend my gratitude to all my friends and colleagues in the MRC-University of Glasgow Centre for Virus Research. It has been a pleasure working in such a friendly environment, and I believe I was lucky to be surrounded by excellent researchers who taught me so much. Special thanks to Sultan Al-Otaibi for the sleepless nights we worked together and for all the fun we have had in the last four years.

This thesis is a product of long years of professional development where I have been blessed with great mentors, Prof Mohamed Shoeir, Prof Laila Makhoulouf, Prof Ermanno Candolfi, and Dr Elizabeth McCrudden. Today I want to thank them all for making me reach this stage.

Last but not least, I would like to thank my angels; Souade, Myriam, and Camelia who had suffered a lot to make this work possible. They were very supportive and understanding in the last few years and kept me going. I would like to thank my family and friends in Egypt and France; Nader, Wafaa, Yousry for their endless support.

A final word to my mother *“Mum I did it”*.

Tamer Abdelrahman

Author's declaration

This work was completed at the University of Glasgow between October 2011-2015 and has not been submitted for another degree. All work presented in this thesis was obtained by the author's own efforts, unless otherwise stated.

All experimental work presented in this thesis was generated by Dr Tamer Abdelrahman under the supervision of Dr Emma Thomson and Dr John McLauchlan at the MRC-University of Glasgow Centre for Virus Research.

The cloning strategy for construction of a genotype-1a replicon was designed in collaboration with Dr Vera Schregel and Dr Ana Filipe; postdoctoral research associate at the MRC- University of Glasgow Centre for Virus Research.

Sanger sequencing and 454 pyrosequencing were carried out by Beckman Coulter Genomics, UK. Illumina sequencing was carried out in High Throughput Sequencing (HTS) facility in MRC- University of Glasgow Centre for Virus Research in collaboration with Dr Gavin Wilkie.

The bioinformatics pipeline used in data analysis was developed by Dr Sreenu Vattipally and Dr Joseph Hughes, research bioinformaticians in the Bioinformatics unit, MRC- University of Glasgow Centre for Virus Research.

The HCV mono-infection dataset was sequenced using Illumina sequencing by Dr Chris Hinds and Dr Walt Adamson.

The Mock community design and data analysis were performed in collaboration with Dr Melanie Schirmer, postdoctoral research associate at The Broad Institute of MIT and Harvard, USA.

The systematic review of antiviral resistance mutations was carried out jointly with two master students under my supervision; Weronika Witkowska and Yangie Pinanga.

Definitions/ Abbreviations

ATP	Adenosine triphosphate
aa	Amino acids
bp	Base pair(s)
cDNA	complementary DNA
DCV	Daclatasvir
DSV	Dasabuvir
°C	Degrees celsius
DAAs	Directly Acting Antivirals
DMEM	Dulbecco's modified Eagle's medium
EMCV	Encephalomyocarditis virus
<i>E. coli</i>	<i>Escherichia coli</i>
FCS	Foetal calf serum
HD	Hamming distance
HCVcc	Cell-culture-derived infectious HCV
HCV	Hepatitis C Virus
HCC	Hepatocellular carcinoma
HAART	Highly active antiretroviral therapy
Huh-7	Human hepatoma cell line
hVAP-33	human vesicle-associated protein-33
H77	Hutchinson strain
HVR	Hypervariable region
IFN	Interferon
IRF-3	Interferon regulatory factor 3
ISDR	Interferon sensitivity-determining region
IRES	Internal ribosome entry site
LDV	Ledipasvir
ML tree	Maximum likelihood tree
MSM	Men who have sex with men
μ	Micro (10 ⁻⁶)
Min	Minutes
Neo	Neomycin phosphotransferase
NGS	Next Generation Sequencing
NS	Non-Structural
NF-κB	Nuclear factor kappa-light-chain-enhancer of activated B cells
nt	Nucleotides
OBV	Ombitasvir
ORF	Open reading frame

Pi	Pairwise diversity
PTV	Paritaprevir
PrOD	Paritaprevir-ritonavir, Ombitasvir, Dasabuvir
PegIFNα	Pegylated Interferon-alpha
PWID	People who inject drugs
PBMC	Peripheral blood mononuclear cells
PCR	polymerase chain reaction
PI	Protease inhibitor
QD	Quasispecies diversity
RLU	Relative light units
RAMs	Resistance-associated mutations
RAVs	Resistance-associated variants
RIG-I	Retinoic acid inducible gene-I
RT	Reverse transcription
RBV	Ribavirin
RdRp	RNA-dependent RNA polymerase
Sec	Seconds
SBS	Sequencing by synthesis
SMV	Simeprevir
SNP	Single nucleotide polymorphism
SOF	Sofosbuvir
SGR	subgenomic replicons
SVR	Sustained virological response
TVR	Telaprevir
EC50	The half maximal effective concentration
IC50	The half maximal inhibitory concentration
3D	Three-dimensional
TLR	Toll-like receptor
TRIF	Total reflection fluorescence
UTR	Untranslated region
WT	Wild type
WHO	World Health Organisation
ZMWs	Zero-mode waveguides

Amino Acid Names Abbreviations

Amino acid	Three letter code	One letter code
Alanine	Ala	A
Arginine	Arg	R
Asparagine	Asn	N
Aspartic acid	Asp	D
Cysteine	Cys	C
Glutamine	Gln	Q
Glutamic acid	Glu	E
Glycine	Gly	G
Histidine	His	H
Isoleucine	Ile	I
Leucine	Leu	L
Lysine	Lys	K
Methionine	Met	M
Phenylalanine	Phe	F
Proline	Pro	P
Serine	Ser	S
Threonine	Thr	T
Tryptophan	Trp	W
Tyrosine	Tyr	Y
Valine	Val	V

Publications

➤ Publications arising from this thesis

Abdelrahman T, Hughes J, Main J, McLauchlan J, Thursz M, Thomson EC. Correspondence: Reply to Martin *et al.* *Hepatology*. 2015 Apr;61(4):1438.

Abdelrahman T, Hughes J, Main J, McLauchlan J, Thursz M, Thomson EC. Deep sequencing sheds light on the natural history of hepatitis C infection in patients that fail treatment. *Hepatology*. 2015 Jan;61(1):88-97.

➤ Publications obtained from personal contribution to other studies

Shepherd S, Abdelrahman T, MacLean AR, Thomson EC, Aitken C, Gunson RN. Prevalence of HCV NS3 pre-treatment resistance associated amino acid variants within a Scottish cohort. *Journal of clinical virology*. 2015 Apr;65:50-3.

Leggewie M, Vattipally SB, Abdelrahman T, Leitch EC, Wilkie GS, Klymenko T, Muir D, Thursz M, Main J, Thomson EC. Natural NS3 resistance polymorphisms occur frequently prior to treatment in HIV-positive patients with acute hepatitis C. *AIDS*. 2013 Sep 24;27(15):2485-8.

➤ Submitted manuscript

Vattipally SB, Alotaibi S, Abdelrahman T, Bruncker K, Orton R, Klymenko T, Wilkie G, Davison A, Thomson EC, TANOTI: a BLAST guided divergent read mapper for small genomes.

Chapter 1: Introduction

1.1 Hepatitis C virus (HCV)

Hepatitis C virus (HCV) is a single-stranded RNA virus that belongs to the *Hepacivirus* genus within the *Flaviviridae* family (Simmonds et al., 1994). The positive sense genome has a length of 9.6 kb.

1.2 HCV discovery and classification

Hepatitis C virus was identified as the causative agent of non-A, non-B hepatitis in 1989 (Choo et al., 1989). The virus is highly heterogeneous, due to the lack of proofreading ability of the RNA-dependent RNA polymerase (RdRp) that it encodes.

HCV is classified into seven genotypes and more than 90 different subtypes (Nakano et al., 2012). HCV genotypes differ at 30-35 % of nucleotide positions on average over the complete genome. Each of these seven major genetic groups contains a cluster of more closely related subtypes that differ in their nucleotide sequences by 20-25 % (Simmonds, 2004). This variation is likely the result of the error-prone replication of HCV and the high replication rate of 10^{12} viral particles/day (Neumann et al., 1998). The seventh genotype sequence has been deposited in the NCBI databases, but no clinical details associated with this clinical isolate have been reported (Gottwein et al., 2009).

Based on the time of divergence of the different genotypes, it is estimated that HCV originated between 500 and 2000 years ago (Simmonds, 2001). The inclusion of subgenomic sequences from infected patients in 1953 in evolutionary analyses provides new estimates of the common ancestor of HCV (Gray et al., 2013).

1.3 HCV genome

The ORF encodes ten proteins, including Core (C), Envelope (E1 and E2), p7

protein and six non-structural (NS) proteins (NS2, NS3, NS4A, NS4B, NS5A, and NS5B). Another protein termed F protein is translated from a frame-shift in the core coding region (Figure 1-1) (Thurner et al., 2004).

1.3.1 5' Untranslated region (5'UTR)

The 5'UTR contains 341 nucleotides (nt). It contains four domains, numbered I to IV (Brown et al., 1992). The domains II, III, IV and the first 12-30 nt of the core coding region form the internal ribosome entry site (IRES) that mediates the translation of genomic RNA by binding the 40S ribosomal subunit to form a stable pre-initiation complex to initiate the HCV polyprotein translation (Honda et al., 1996).

1.3.2 3' Untranslated region (3'UTR)

The 3'UTR contains around 225 nucleotides; it is organised into three regions; a variable site of around 30-40 nt, a long poly(U)-poly(U/UC) tract and a highly conserved 3'X region of 98 nt (Kolykhalov et al., 1996). It interacts with the NS5B RdRp and is essential for RNA replication (Lee et al., 2004).

1.3.3 Frame-shift protein

The F (frame-shift) protein is encoded from a -2/+1 ribosomal frame shift in the N-terminal core-encoding region. The protein is produced during infection, and anti-F protein antibodies have been detected in clinical samples (Walewski et al., 2001). The role of the F protein remains unclear although it may play a role in viral persistence (Baril and Brakier-Gingras, 2005).

1.3.4 Core protein

HCV core is a highly conserved protein that constitutes the viral nucleocapsid. The immature form of the core protein is composed of three domains; the last 20 aa works as a signal peptide for the E1 protein (Grakoui et al., 1993b). Domain D1 is mainly involved in RNA binding. Domain D2 is responsible for targeting core to the surface of lipid droplets (Suzuki et al., 2005).

It also interacts with some cellular proteins and pathways during the virus life cycle (McLauchlan, 2000).

1.3.5 Envelope glycoproteins

E1 and E2 are type I transmembrane glycoproteins that are essential for viral entry and fusion (Nielsen et al., 2004, Bartosch et al., 2003a). They have several properties including membrane anchoring, ER localization and heterodimer assembly (Cocquerel et al., 2000). E2 contains hypervariable regions ((HVR1, HVR2, and HVR3).

HVR1 consists of 27 aa, is a major neutralizing epitope, and its amino acid sequences differ by up to 80% between HCV genotypes (Farci et al., 1996). However, its overall conformation is highly conserved across all genotypes, indicating a vital role in the virus life cycle (Penin et al., 2001). The HVR1 also plays a major role in viral attachment through interaction with negatively charged molecules at the cell surface via positively charged residues (Flint and McKeating, 2000, Bartosch et al., 2003b).

1.3.6 Protein p7

Protein p7 is a small, 63 aa polypeptide that has ion channel activity (Kalita et al., 2015). Cleavage of p7 from the polyprotein is mediated by host signal peptidase (Lin et al., 1994). The ion channel activity of p7 is critical in the HCV life cycle as mutations or deletions in the p7 protein suppress infectious virus production (Jones et al., 2007).

1.3.7 Non-structural proteins

1.3.7.1 NS2

NS2 is a transmembrane protein, which has two internal hydrophobic segments that are responsible for ER membrane association (Yamaga and Ou, 2002). NS2 forms an autoprotease with the N-terminal domain of the NS3 protein that cleaves at the junction between NS2 and NS3 (Grakoui et al., 1993a).

Various studies have identified adaptive mutations in NS2 that enhance HCVcc virus production, but NS2 is not required for RNA replication (Jones et al., 2007).

1.3.7.2 NS3-NS4A

NS3 has several functions; it contains an N-terminal serine protease domain and a C-terminal helicase/NTPase domain. NS4A is a cofactor of NS3 protease activity. The NS3/4A complex plays an indispensable role in HCV life cycle and pathogenesis (Pawlotsky, 2006).

➤ NS3/NS4A Protease

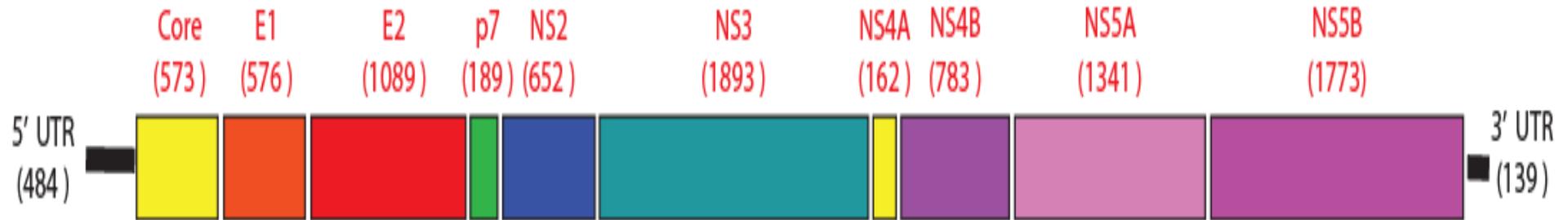
The NS3/NS4A protease cleaves the HCV polyprotein at the NS3/NS4A, NS4A/NS4B, NS4B/NS5A and NS5A/ NS5B junctions. NS4A serves as a cofactor for NS3 serine protease activity, enabling its localisation at the ER membrane as well as cleavage-dependent activation (Bartenschlager et al., 1995).

The NS3/4A serine protease contributes to HCV immune evasion through blocking the phosphorylation of the interferon regulatory factor 3 (IRF-3) pathway, which affects interferon induction in response to a viral infection (Foy et al., 2003). Moreover, NS3/4A-mediated cleavage of TRIF inhibits signalling through the Toll-like receptor 3 (TLR3) pathway (Li et al., 2005).

➤ NS3 Helicase-NTPase

The NS3 helicase-NTPase domain consists of 442 aa at the C-terminus of the NS3 protein. It is a multifunctional protein, as it is involved in RNA binding, and unwinding of secondary structure, it also has RNA-stimulated NTPase activity (Cho et al., 1998, Gwack et al., 1997). The introduction of adaptive mutations in both protease and helicase regions of NS3 improve the RNA replication of replicons in cell culture (Lohmann et al., 2003).

Figure 1-1: HCV genome.



Open reading frame produces structural and non-structural proteins flanked by two untranslated regions. The length of each region is expressed in brackets as the number of nucleotides.

1.3.7.3 NS4B

NS4B is a 261 aa integral membrane protein, it acts as a membrane anchor for the replication complex (Elazar et al., 2004), and inhibits cellular protein synthesis (Florese et al., 2002, Piccininni et al., 2002). Different substitutions in NS4B were reported to enhance or abolish RNA replication indicating its role in the regulation of HCV NS5B RdRp activity (Lohmann et al., 2003). NS4B is reported to play a role in virus assembly (Jones et al., 2009).

1.3.7.4 NS5A

NS5A is a hydrophilic phosphoprotein, its role in HCV replication is not entirely understood. RNA replication is inhibited by mutations within the NS5A sequence (Elazar et al., 2003). NS5A associates with lipid rafts derived from intracellular membranes, and this is essential for the HCV replication (Gao et al., 2004). Moreover, NS5A phosphorylation may play a fundamental role in the HCV life (Appel et al., 2005).

1.3.7.5 NS5B RNA-dependent RNA polymerase

NS5B acts as an RNA-dependent RNA polymerase (RdRp). It was shown that the RdRp has a classical "fingers, palm and thumb" structure. It has a typical right-hand structure with a central palm domain harbouring the catalytic GDD motif and the fingers and thumb domains on either side; the resulting conformation allows the binding of a single-stranded RNA template (Lesburg et al., 1999, Bartenschlager et al., 2004).

1.4 HCV replication

The combination of viral proteins, cellular components, and nascent RNA strands forms a replication complex that involves lipid rafts through protein-protein interactions between human vesicle-associated protein-33 (hVAP-33) and both NS5A and NS5B (Gao et al., 2004, Shi et al., 2003).

In HCV replication, the positive-sense genome RNA serves as a template for the synthesis of a negative-sense replication intermediate that acts as a template to produce further positive-sense RNA strands. The new positive-sense RNA strands may be directly translated, used as a template for the synthesis of replication intermediate, or packaged into new virions (Bartenschlager et al., 2004).

The characterization of the NS5B RdRp revealed that NS5B-mediated RNA polymerization is initiated by priming on the template via a 'copy-back' mechanism. RdRp is also capable of initiating *de novo* RNA synthesis (Zhong et al., 2000).

1.5 HCV replicons

HCV propagation in cell culture using infection with virus-containing inoculum is hindered by low and transient replication levels that prevent studies on HCV replication with natural isolates (Bartenschlager and Lohmann, 2001). However, it is possible to recapitulate replication through transfection of HCV RNAs transcribed *in vitro* from cDNA clones containing viral sequences that encode the viral NS3-NS5B replicase unit (Bartenschlager and Lohmann, 2001).

The first functional HCV subgenomic replicons were derived from consensus genotype 1b (Con1) sequences that were detected in the liver of an HCV-infected patient. It contained: i) the HCV 5' UTR and the first 12 codons of the core protein linked in-frame with the neomycin phosphotransferase cassette (Neo) to permit selection in the presence of the cytotoxic agent G418; ii) the IRES region from encephalomyocarditis virus (EMCV) that initiates translation of the downstream HCV NS3-5B polyprotein; iii) the HCV 3' UTR (Lohmann et al., 1999).

1.5.1 Adaptive mutations

The original Con1 subgenomic replicons produced a low frequency of G418-resistant cells (~1 colony per 10^6 transfected cells) (Lohmann et al., 1999, Blight et al., 2000). The low frequency was attributed to the low number of cells

supporting efficient HCV replication and the necessity of acquiring adaptive mutations to replicate efficiently in the permissive cell line. Sequence analysis of Con1-derived HCV RNAs replicating in the selected cells identified highly adaptive mutations in the nonstructural region, mainly clustering in NS5A (Blight et al., 2000, Krieger et al., 2001, Lohmann et al., 2003, Lanford et al., 2003, Enomoto et al., 1995, Enomoto et al., 1996).

1.5.2 Genotype 1a replicons

Genotype 1a replicons derived from the Hutchinson strain (H77) require at least two adaptive mutations to replicate efficiently in cell culture systems (Blight et al., 2003, Grobler et al., 2003, Gu et al., 2003, Yi and Lemon, 2004). The first H77 replicons contained a mutation (S2204I) that allowed efficient replication in Huh-7.5 cell line (Blight et al., 2002, Blight et al., 2003). Analysis of these replicons revealed that the improved replication capacity of subgenomic H77 is associated with the presence of an additional substitution in NS3 gene (P1496L) (Blight et al., 2003).

Voitenleitner *et al.* described the generation of a robust H77 replicon through the introduction of extra fitness mutations, NS4A (K1691R) and NS4B (E1726G). These mutations considerably enhanced the signal to noise ratio, leading to more robust replication in transient transfections (Voitenleitner et al., 2012).

1.5.3 Reporter genes

After the development of HCV replicons, transient RNA replication assays were developed to permit rapid analysis of replication efficacy. Reporters such as luciferase and β -lactamase were used to monitor replication following transfection (Blight and Norgard, 2006).

Luciferase activity is a reliable marker of replication as it directly correlates with the levels of HCV RNA synthesis. The Firefly luciferase was introduced to bicistronic subgenomic replicons, thus enabling assessment of

replication at different time points by measuring enzyme activity relative to a polymerase-defective replicon, which contains a mutation in the GDD RNA polymerase motif (GDD to GND). The adapted Con1 replicons produce luciferase activity at 48-72 hours that is about 100-fold higher than the negative GND control (Krieger et al., 2001).

1.5.4 Permissive cell lines for HCV replication

Although hepatocytes are the main site of HCV replication, other extrahepatic cells are also reported to harbour the virus, including lymphocytes, monocytes and dendritic cells (Laskus et al., 2000, Goutagny et al., 2003).

Huh-7.5 cells were produced by treatment of HuH-7 cells harbouring a subgenomic replicon with human interferon- α (IFN α) at a high concentration for long periods. They have the capacity to support high levels of subgenomic HCV replication in more than 75% of transfected cells, rendering it the most permissive subline available to date.

The higher permissiveness is attributed to the mutational inactivation of the retinoic acid-inducible gene-I (RIG-I), a protein that is involved in the induction of type I IFN production (Blight et al., 2002, Sumpter et al., 2005).

The performance of transient assays varies between different passages of Huh-7 cells, which may cause up to 100-fold difference in replication efficiencies of subgenomic RNAs. This difference is independent of the adaptive mutations introduced or the quality of RNA (Lohmann et al., 2003). Moreover, there is an association between stage of the cell cycle and the replication of subgenomic replicons (Blight et al., 2002, Blight et al., 2000).

1.5.5 Applications of the HCV replicon system

The development of the replicon system has been instrumental in defining the functions of individual proteins, and enabling studies on virus-host interactions. One of the applications is estimating the *in vitro* replication fitness of replicons

after the introduction of resistance-associated mutations (RAMs) by comparing replication efficacy with that of the wild-type (WT) replicon based on transient replication assays.

1.6 Epidemiology of Hepatitis C virus infection

An estimated 185 million individuals have been infected with HCV worldwide (Thomas, 2013). The host immune response against HCV infection fails to prevent chronicity in 50%-80% of cases resulting in approximately 135 million people with chronic infection.

Our understanding of global HCV epidemiology is imperfect, as it has been shaped by seroprevalence studies performed in selected populations, such as blood donors. Although population-based studies representative of an entire community would be useful, they are not feasible to perform, so in many countries, the exact magnitude of the problem is not clear.

Currently available data show that the prevalence of HCV varies across the globe with the highest reported in Egypt where the seroprevalence rate is estimated at 22% of the total population (Shepard et al., 2005). This high prevalence rate is a result of the national antischistosomal treatment programme, which until the 1970s involved intravenous administration of drugs using reusable syringes (Frank et al., 2000).

The geographic distribution of HCV is available using the Los Alamos online database at <http://hcv.lanl.gov> (Figure 1-2). The prevalence of HCV is 3-4% in some Asian countries, almost 2% in North America, and above 10% in regions of Central Africa. The seroprevalence in England and Wales in the adult population aged 15-59 is estimated to be less than 0.6% (Thomson, 2009).

Genotypes 1-3 have a worldwide distribution and account for almost all infections in developed countries with 1a and 1b being the most common (Figure 1-3). Genotypes 4 and 6 remain the most diversified genotypes (Nakano et al., 2012).

HCV genotype 5a has been transmitted in South Africa and Belgium independently Africa for more than a century (Verbeeck et al., 2006), while HCV genotype 7 is reportedly originating from central Africa (Murphy et al., 2015).

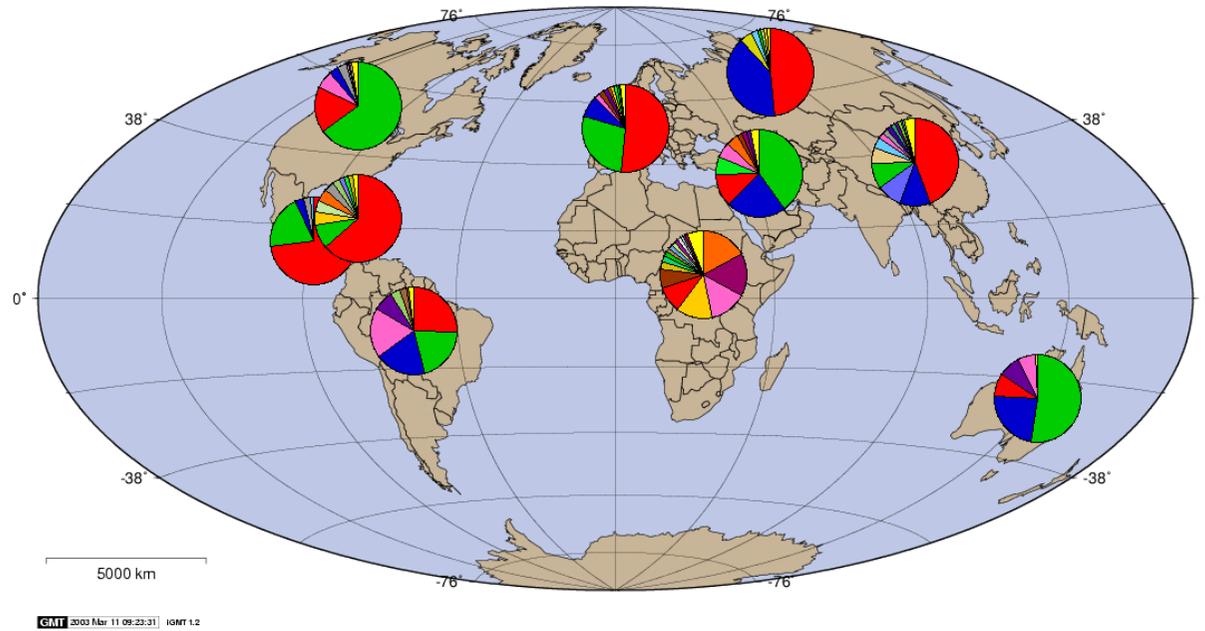
Patients who have received injected medical treatments including blood products before the introduction of screening blood for HCV are at highest risk globally for HCV infection. In developed countries, the vast majority of new cases are associated with unsafe injections in people who inject drugs (PWID). Sexual transmission of HCV infrequently occurs in HIV-negative couples but is more common in HIV-positive patients, particularly in men who have sex with men (MSM) (Danta et al., 2007).

This thesis is based on the study of a cohort of HIV-positive MSM with acute HCV infection in reported in urban centers in the UK, Germany, France, the Netherlands and the USA (Gilleece et al., 2005, Vogel et al., 2005, Dominguez et al., 2006). The majority of these patients have been asymptomatic and have been diagnosed after detection of deranged liver function tests at routine HIV follow-up clinics. Sexual transmission in these patients was associated with exposure risk factors including sexual practices with a high risk of mucosal trauma, unprotected anal intercourse, multiple partners, sex under the influence of drugs, and internet-arranged sex (Danta et al., 2007). Heterosexual spread of HCV in HIV-infected patients may be higher than in HIV-negative couples. The higher transmission could be attributed to a higher HCV viral load in genital secretions (D'Oliveira et al., 2005).

1.7 Studying the viral population within infected patients

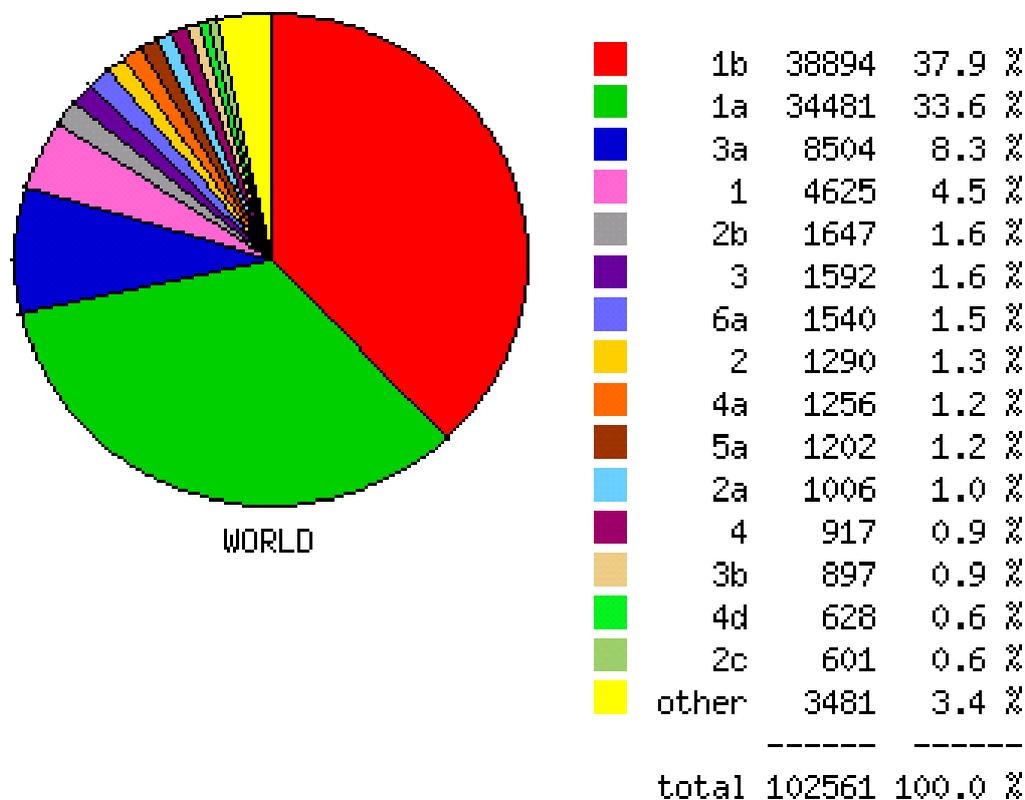
The RNA viruses do not follow the evolutionary dynamics of traditional population genetics, as RNA viruses are characterised by high replication rate, large (close to infinite) population size, and high mutation rates (Holland et al., 1982).

Figure 1-2: prevalence of hepatitis C virus genotypes in different regions.



The figure represents the frequency in the NCBI HCV Database; the color code is the same used in Figure 1-2. Geographic and subtype distribution is shown for the 102516 sequences available online at <http://hcv.lanl.gov/components/sequence/HCV/geo/geo.comp>, accessed on 05/09/2015.

Figure 1-3: Global distribution and prevalence of hepatitis C virus genotypes.



The figure represents the frequency in the NCBI HCV Database as an estimate of population prevalence. The geographic and subtype distribution is shown for the 102516 sequences <http://hcv.lanl.gov/components/sequence/HCV/geo/geo.comp>, accessed on 5/9/2015.

1.7.1 Replicative homeostasis hypothesis

The replicative homeostasis hypothesis (RH) states that dynamic RNA polymerase has a reduced replication fidelity, which results in a high intracellular concentration of mutant genomes, and consequently a spectrum of mutant proteins. The protein products interact with RdRp to regulate its processivity and fidelity. This regulation leads to the establishment of replicative homeostasis (Sallie, 2007). This regulatory mechanism links the dynamic RdRp with quasispecies diversity, creating a stable but reactive replicative equilibrium (Sallie, 2007).

The replicative equilibrium provides a sequence space that tolerates mutations; the variation in viral sequence space is controlled by factors such as viral fitness and the host's immune system (Sallie, 2005). The homeostatic regulation requires three main components: i) a pathway that initiates a response to any disturbance of the replicative equilibrium; ii) a feedback pathway that measures the system's response to this disturbance; iii) mechanism(s) of viral auto-regulation (Sallie, 2005).

1.7.2 Quasispecies theory

Quasispecies theory has been defined as 'dynamic distributions of non-identical but closely related mutant and recombinant viral genomes subjected to a continuous process of genetic variation, competition and selection, and which act as a unit of selection' (Domingo et al., 2005).

Quasispecies theory described the evolutionary dynamics of RNA viruses and was validated experimentally in model systems (Duffy et al., 2008). Quasispecies theory considers certain characteristics of RNA viruses: i) a high mutation rate; ii) a small genome size; iii) a large population; iv) an equilibrium state where the effects of mutations and selection are equal (Comas et al., 2005, Biebricher and Eigen, 2006).

The defining feature of quasispecies theory is ‘the survival of the flattest’, as the high error rates result in a mutational drive, which in turn leads to natural selection at the population level rather than at the individual level. This selection results in the outgrowth of populations with variants exhibiting similar viral replication capacity rather than a population in which variants have a higher replicative fitness. This feature distinguishes the process within the quasispecies from the traditional notion of ‘survival of the fittest’ (Wilke et al., 2001).

Data are suggesting that quasispecies theory does not apply to RNA viruses as, although RNA viruses mutate rapidly, mutation rates may not reach a sufficient threshold to have a significant impact on quasispecies dynamics (i.e. ‘the survival of the flattest’) (Holmes, 2010). However, the circulation of multiple genetically related but not identical variants within an infected host has major implications on the outcome of infections and the progression of the disease.

1.7.3 Methods of quasispecies analysis

The major issues associated with the identification of viral variants present in any clinical sample are the viral complexity and the detection of minority (low frequency) variants. The development of next-generation sequencing (NGS) platforms has provided a better understanding of intra-host viral populations, rendering NGS a potential alternative to conventional methods (Cruz-Rivera et al., 2013).

1.7.3.1 Sanger sequencing

Viral RNA is extracted from a clinical sample or from an *in vitro* system and amplified by reverse transcription-polymerase chain reaction (RT-PCR). The amplified PCR product (amplicon) is then sequenced directly using Sanger sequencing. Data analysis determines the consensus sequence of all variants within the population.

Automated sequencing platforms identify the most common base at each position in the sequence. If two or more bases are present with similar frequencies at a particular position, automated sequencing software may fail to identify the single most common base (an issue known as miscalling). In some cases, such discrepancies can be resolved by reviewing the chromatogram accompanying the sequencing results. Doing this can also provide indications as to the variation within the population.

1.7.3.2 Clonal analysis

Clonal analysis is a powerful research tool, as it can be used to measure viral diversity by sampling sequences that are circulating in the host. However, the process is labour-intensive and expensive. The PCR product is cloned into a plasmid vector that is introduced to bacteria. Transformed bacteria are grown on agarose plates at a low density to allow for the selection of individual colonies. The individual colonies are assumed to contain one plasmid only and thus only one viral variant. The sequence of a single variant can be obtained by sequencing the plasmid DNA extracted from a single bacterial colony.

The number of clones required to identify all variants circulating in the viral quasispecies is not well defined. It has previously been estimated that 20 clones are sufficient to sample 95% of the major variants (those with at least 10% frequency in the population) (Fishman and Branch, 2009). Another factor that can influence the number of clones needed to identify the circulating variants could be the variability of the studied region of the genome (Fishman and Branch, 2009). For studying HVR1, it has been reported that 99 clones are required to identify 95% of any variants present at a frequency above 3% within the viral population (McCaughan et al., 2003). However, most minor variants would remain undetected using this threshold, therefore excluding a significant proportion of the viral complexity.

PCR may introduce errors if a non-proofreading enzyme is used which creates a biased profile of the viral population. This bias can be avoided by using a limiting dilution PCR approach, but it is a time-consuming process, and not

suitable for large-scale studies. Other sources of errors include reverse transcription, bacterial cloning, and the sequencing reaction.

Bidirectional sequencing may reduce the error rate at the sequencing stage; however, this approach results in an increase in cost. Primer selection bias may also be a problem, particularly when primers are designed to anneal with highly diverse regions such as HVR1. This design will result in some variant strains being omitted and unamplified. The use of multiple primer sets or degenerate primers may help to overcome this problem.

1.7.3.3 Next generation sequencing (NGS)

The development of NGS has revolutionised the analysis of viral populations. NGS provides a detailed analysis of circulating variants with much higher coverage compared to conventional Sanger sequencing. However, it requires a short PCR step that may introduce errors, and different sequencing platforms have their instrumental errors (Poh et al., 2013). Another difference between Sanger sequencing and NGS data is the read length (the number of nucleotides obtained from each sequenced fragment). The signal-to-noise ratio limits read length on NGS platforms, all of which produce shorter reads than Sanger sequencing.

NGS technology has introduced the possibility of using metagenomics to discover viruses. The metagenomics has led to the discovery of the aetiology of Merkel cell cancer. A viral transcript corresponding to a novel polyomavirus was identified, and the virus was named as the Merkel cell polyomavirus (Feng et al., 2008). Additional contributions were the identification of a bunyavirus in patients with thrombocytopenia and leukopenia syndrome (Xu et al., 2011). Metagenomic sequencing was also used to determine the human virome in acute febrile diarrhea in children (Wylie et al., 2012). The possibility of detection of viruses in chronic infection is a major challenge that requires highly sensitive methods, as the viral burden may be diminished with disease progression. An example of this has been demonstrated in cirrhotic patients (Duvoux et al., 1999).

1.8 Next generation sequencing

The NGS process is a sequence of reactions including: i) the addition of nucleotides; ii) a detection step to identify the incorporated nucleotides on each fragment being sequenced; iii) a wash step to remove fluorescent labels or blocking groups to prepare for the next reaction (Mardis, 2011).

1.8.1 Digital data

NGS platforms use a library of fragmented DNA bound covalently to synthetic DNA sequences (adapters). These adapters are universal sequences that are subsequently used to amplify library fragments during the sequencing process. Amplification is carried out on a solid surface (either a bead or a flat glass microfluidic channel depending on platform) that is covalently bound to adapter sequences complementary to those on the library fragments. Each amplified fragment produces a single cluster of data (Mardis, 2011).

Paired-end sequencing has improved the NGS sequencing results: it produces sequence data from both ends of each library fragment, improving the alignment step in data analysis by offering a duplicate of each nucleotide site, thus confirming the placement of each read relative to the reference genome. Most alignment programs take into account the average length of fragments in the sequencing library to achieve the most accurate placement (Korbel et al., 2007).

The digital nature of NGS data means that each read originates from a single cluster, which in turn is created by the amplification of a single library fragment. This process enables the quantification of abundance, hence in population-based studies, where NGS is used to study the individual species present in an isolate, the presence of each species may be quantified as a proportion of the total population (Human Microbiome Project, 2012).

1.8.2 Sources of noise and error models

As observed in Sanger sequencing, enzymatic amplification may introduce systematic errors during library preparation. These errors are independent of the instrument used and could be minimised by using a high-fidelity polymerase and/or by reducing the number of amplification cycles. The signal to noise ratio is the main limitation to read length in NGS platforms. Noise accumulates as the signal from the nucleotide incorporation results from a false signal from a prior reaction or incorrect incorporation episodes (Mardis, 2013).

The 'error model' is defined by the interaction between different sources of noise and the resulting sequencing errors. It is instrument- and chemistry-specific. An important variable in determining the error model is the depth (number of sequence reads covering the analysed region) (Mardis, 2008). A control sample of known sequence is usually used in each sequencing run to determine read-length limitations and error types (Hillier et al., 2008). Using the control enables: i) identification of the error model; ii) the detection of different types of error (substitutions, insertions, or deletion errors); iii) coverage bias (the complete or partial lack of representation of certain regions) (Mardis, 2013).

1.8.3 454 Pyrosequencing

The first NGS platform was launched in 2005 by 454 Life Sciences. It is based on the detection of DNA polymerase activity using a chemoluminescent enzyme. It has an average read length up to 800bp.

The library preparation involves adapter ligation, the short adapters providing priming sequences for both the amplification and the sequencing of library fragments. One adaptor contains a biotin tag to permit the immobilisation of the DNA library on streptavidin-coated beads. The bead-bound library undergoes an emulsion PCR for library preparation (Zheng et al., 2010). The beads are added into a picotiter plate. The wells contain a mixture of enzymes; DNA polymerase, ATP sulfurylase, and luciferase.

The library is sequenced by using the single-stranded DNA fragments as a template for the synthesis of complementary strands. Nucleotides are added to the plate, one nucleotide at a time (Voelkerding et al., 2009).

Nucleotide incorporation releases pyrophosphate, which is converted to adenosine triphosphate (ATP) by ATP sulfurylase. ATP acts as a substrate for the luciferase-mediated conversion of luciferin to oxyluciferin. This process causes the emission of light that is detected by a camera and analysed in a pyrogram. (Ronaghi et al., 1998). The detected signal is proportional to the amount of ATP. Hence, the light intensity corresponds to the number of incorporated bases. If several similar nucleotides may be incorporated in a single cycle, the accurate number of incorporated bases cannot be estimated. Thus, in regions containing long homopolymers, a high percentage of insertion and deletion errors may be introduced. The insertion and deletion errors were estimated to represent 65%-75% and 20%-30% of all sequencing errors respectively (Astrovskaya et al., 2011).

1.8.4 Illumina[®] technology

This Illumina technology was launched in 2007 by Solexa and was afterwards acquired by Illumina[®] (Bentley et al., 2008). The library workflow starts by fragmentation of DNA and ligation of specific adapters. Illumina[®] uses a flow cell composed of flat glass with eight microfluidic channels as a solid surface. The library fragments are amplified on the flow cell surface by bridge amplification, resulting in clusters for sequencing.

The Illumina[®] platform uses reversible dye terminator sequencing. In this process, all four fluorescent-labelled nucleotides are present in each sequencing cycle, and the fragmented DNA (which all carries the same adapter) is primed with a complementary synthetic DNA primer to provide a free 3-OH group, which can be extended in subsequent stepwise sequencing reactions. Each single nucleotide has a blocking group at the 3-OH position of the ribose sugar to prevent chain elongation.

The sequencing step includes a series of events: i) the addition of a fluorescent-labelled nucleotide; ii) a wash step for the removal of unincorporated nucleotides; iii) recording the fluorescent signal; iv) the chemical cleaving of fluorescent groups; v) chemical de-blocking at the 3-OH position.

Sequencing errors are mainly substitution errors, in which an incorrect nucleotide identity is assigned due to two main sources of noise: i) incomplete de-blocking or a lack of de-blocking in prior cycles causes some fragments within the same cluster to fall out of phase; ii) interference noise resulting from incomplete fluorescent label cleavage in previous cycles (Mardis, 2013).

1.8.5 Ion Torrent™ technology

Ion Torrent platform was commercialised in 2010 (Rothberg et al., 2011). The Ion Torrent™ platform is based on the detection of hydrogen ions during the sequencing reaction. The change in pH due to the release of hydrogen ions as by-products of nucleotide incorporation is quantified by coupled silicon detector.

The library preparation includes DNA fragmentation and adapter ligation. After quantification, the library fragments are mixed in equimolar concentration with beads and PCR reagents to undergo emulsion PCR (Dressman et al., 2003).

The emulsion PCR uses beads as a solid surface for the amplification reaction; the beads have covalently-linked adapter complementary sequences on their surfaces to facilitate amplification of the fragments. The emulsion is formed by a mixture of beads and library fragments in oil micelles that contain PCR reagents.

After amplification, the retrieved beads are deposited into the wells of a specialised silicon chip (Ion Chip), designed to detect pH changes within individual wells as the sequencing reaction progresses. The sequencing reaction starts by introducing native nucleotides in a systematic order.

The calibration of pH measurement is performed by measuring the signal strength due to the incorporation of four single bases downstream of the primer's 3-OH in the adapter sequence in a sequence that matches the first four individual nucleotide added to the chip.

The average read length is 200 bp, produced as single-end reads. The error model of Ion Torrent™ sequencing is defined by insertion or deletion errors that are most prevalent at fragments with multiple bases of the same identity (homopolymers) due to issues surrounding accurate quantitation and ultimately the saturation of the pH detector. Substitution errors occur at a very low frequency due to contamination effects from the prior incorporation cycle. Overall, an average error rate of the Ion Torrent is 1 in 100 bases (Mardis, 2013).

1.8.6 Single-molecule DNA sequencing

Several issues must be considered when working with sequencing platforms that use polymerase-amplified libraries, such as primer selection bias, the introduction of polymerase errors, and the preferential amplification of particular fragments in the library. Currently, the signal to noise ratio limits the read length to 200-800 bp. Therefore, there is a need for a superior platform that obtains sequence data from individual molecules of a DNA isolate and has longer reads.

The single-molecule real-time (SMRT) method was developed in 2010 by Pacific Bioscience (Menlo Park, CA, USA). It makes use of modified enzymes and sequencing is obtained through the direct observation of the enzymatic reaction in real time (Eid et al., 2009).

SMRT Sequencing has some features that enabled the technology to overcome the major challenges facing single-molecule DNA sequencing. The method has been used to detect accurately the low signal produced from an individual molecule during the sequencing process using zero-mode waveguides (ZMWs), phospholinked nucleotides, and engineered DNA polymerase.

The library fragments are bound with DNA polymerase molecules immobilised at the bottom of a ZMWs chip (SMRT Cell) using a diffusion-mediated process. DNA polymerase produces an entirely natural DNA strand by incorporating the nucleotide into the complementary strand and cleaves off the fluorescent dye previously linked with the nucleotide. ZMWs restrict illumination to the bottom of the well where the polymerase/template is fixed. The optics system scans the active site of each ZMW-bound polymerase during the addition of fluorescent nucleotides to the SMRT cell surface, thus detecting the nucleotide incorporation into synthesised DNA strand, by sensing the signal from its laser light-stimulated emission (Liu et al., 2012).

The read lengths obtained can be quite long (up to 10,000 nucleotides). The error model is mainly considering insertion and deletion errors. There are several sources of errors that contribute to an overall high error rate in single-molecule sequencing, mostly related to the detection method used. An error in nucleotide preparations can lead to the lack of the fluorescent label, and hence, nucleotides are not detected when incorporated during sequencing.

The interaction between the nucleotides and the DNA polymerase active may cause different types of errors: i) insertion error occurs when nucleotides remain in the active site long enough to be detected but are not incorporated or falsely indicate that multiple incorporations have taken place; ii) deletion error happens when nucleotides that do not remain in the active site long enough to be detected yet are incorporated into the synthesized strand (Mardis, 2013).

The main advantages of SMRT sequencing are the long read lengths and the random nature of the error process. The error rate is approximately 15 bases per 100 (Mardis, 2013). As the errors are randomly distributed across the target, it is possible to obtain an accurate consensus sequence if sufficient depth is achieved (Roberts et al., 2013).

1.9 Analysis of deep sequence reads

Following the emergence of NGS technologies, many tools have been created for

the large-scale mapping of short fragment reads to a reference sequence. The identification of single nucleotide polymorphisms (SNPs) and the detection of variants can be carried out using a variety of methods. Such tools are extremely powerful but are limited by biases inherent in NGS technology (Flicek and Birney, 2009, Metzker, 2010, Schwartz et al., 2011).

1.9.1 Mapping

Deep sequencing generates short reads. Various mapping tools can analyse these short reads. These tools use two different types of algorithm: i) Burrows-Wheeler transform (BWT); ii) Hash-seed extension (Miller et al., 2010, Wajid and Serpedin, 2012). The BWT based mapping programs (e.g. BWA) are fast and require low computational power, but they are less sensitive than hash-based programs (e.g. Novoalign).

There are two common biases that interfere with mapping NGS data: i) nucleotide per cycle bias where the distribution of sequenced nucleotides varies across the position of each read; ii) mappability bias as regions vary in complexity, resulting in varying coverage across the target genome (Hillier et al., 2008, Dohm et al., 2008).

1.9.2 Haplotype reconstruction

A haplotype is a group of genes that are inherited from one parent. In this study, the term is used to refer to the viral genome. New algorithms are required to achieve the complete reconstruction of individual viral haplotypes to explore the dynamics of viral quasispecies. NGS enables the detection of SNPs with an ability to detect minority variants in the viral population even in low abundance haplotypes. However, to reconstruct haplotypes accurately, the real diversity needs to be distinguished from noise signals in the NGS data. In addition, the limited read length produced by current NGS platforms makes it difficult to determine which SNPs are present in the same haplotype. Therefore, viral diversity is only assessed by SNPs calling or the detection of several mutations within the read length produced by the platform.

Numerous bioinformatics tools have been developed to apply error correction models to the error-prone NGS data and to reconstruct, from an alignment of short reads, the haplotypes that constitute the viral population (Giallonardo et al., 2014). HCV represents a major challenge in dealing with this issue due to its high mutation rate and quasispecies diversity (Astrovskaya et al., 2011).

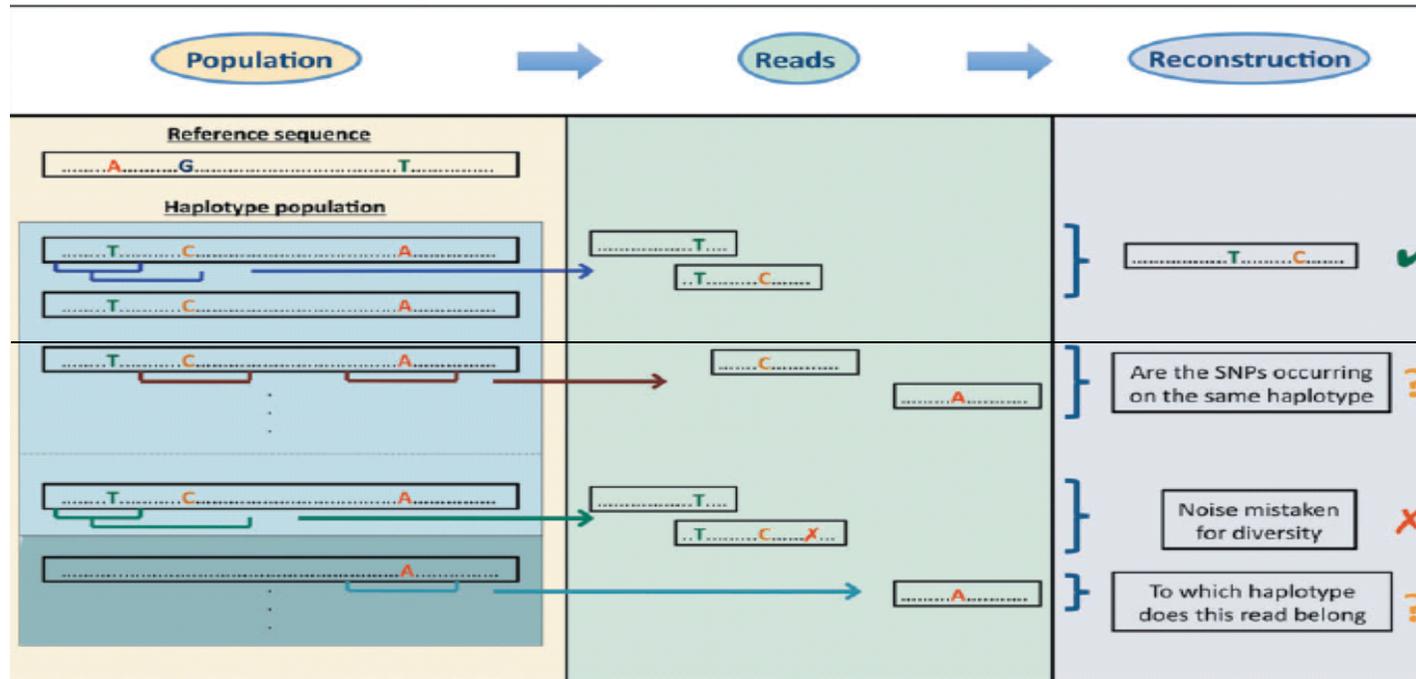
De novo assembly methods are designed to reconstruct a single genome sequence but are not suitable for reconstructing a viral quasispecies due to the closely related genomes of the circulating variants. Mapping-based approaches to HCV haplotype reconstruction are preferred to *de novo* assembly since reference genomes are available and the ability to reconstruct closely related haplotypes. However, the mapping-based approach requires high depth across the genome, which is feasible with the current NGS platforms due to the short length of viral genomes.

Eriksson *et al.* proposed an algorithm that involves three steps: i) error correction via clustering; ii) haplotype reconstruction via chain decomposition; iii) haplotype frequency estimation (Eriksson et al., 2008). An example of tools developed for quasispecies reconstruction is ShoRAH, which has been applied to HIV data (Eriksson et al., 2008, Zagordi et al., 2010a). The phases of the process of reconstructing a population from NGS data and the importance of overlaps in the data are described in Figure 1-4. Prediction programs perform better when they analyse paired-end data with overlaps (Schirmer et al., 2014).

1.10 HIV and hepatitis C virus co-infection

It is estimated that 3-4 million people are infected with HCV each year (WHO, 2011). Around 35 million people are living with HIV (WHO, 2009). Liver disease has recently overtaken AIDS-defining illnesses as a cause of death in many patient populations following the successful introduction of highly active antiretroviral therapy (HAART) for the treatment of HIV (Bica et al., 2001).

Figure 1-4: Schematic diagram representing the process of reconstructing viral haplotypes from next-generation sequencing data.



The first column represents two haplotypes occurring at different abundances with one SNP in common. The next column displays a set of observed reads obtained from NGS technologies including sequencing noise. The third presents different scenarios that can occur during reconstruction. In the first scenario, the reconstruction is successful. Two reads that contain SNPs and have sufficient overlap are assembled correctly into a contig of the first haplotype. In the second, the distance between SNPs exceeds the read length that means that the reads do not map to a haplotype based on read overlap. In the third, noise is mistaken for diversity. In the fourth, we cannot infer the origin of the read as the SNP occurs on both haplotypes. Reproduced with permission from Dr Melanie Schirmer (Schirmer et al., 2014).

Co-infection with HCV is observed in up to 30% of HIV-infected individuals with higher prevalence rates in different populations (Vallet-Pichard and Pol, 2006). At least 4-5 million people have HCV/HIV co-infection, this is attributed to shared routes of transmission (Alter, 2006). The prevalence of HCV infection in HIV-infected haemophiliacs is 60-90% and 50-70% in HIV-infected PWID (Rockstroh and Spengler, 2004).

An epidemic of sexually transmitted acute HCV infection in HIV-positive MSM has been reported in the Netherlands, Germany, France, the USA, and the UK in the last decade. More than 1000 cases of acute HCV/HIV-positive MSM have been reported worldwide (Vogel et al., 2011). Phylogenetic analysis of HCV variants obtained from infected patients provided evidence of an extensive international network of HCV transmission in this population (van de Laar et al., 2009).

1.10.1 The influence of HCV on HIV infection.

The impact of HCV on the progression of HIV-disease remains controversial. The Swiss HIV Cohort study revealed that HCV infection was significantly associated with an increased risk of progression to AIDS, although a similar use of HAART in both HIV mono-infected group and HCV/HIV co-infected group (Greub et al., 2000). HIV mono-infected patients are more likely to achieve a CD4 count rise of at least 50 cells/ml within one year of starting HAART than HCV/HIV co-infection patients with a hazard ratio of 0.79 (0.72-0.87) for HCV-seropositive patients (Greub et al., 2000).

The HCV/HIV co-infected patients have an increased likelihood of having a CD4+ cell count of <200 cells/mm³, compared with infection with HIV-1 alone (hazard ratio, 1.52; 95% confidence interval, 1.07-2.17) (Stebbing et al., 2005, Sullivan et al., 2006). Several studies have reported no increased risk of clinical progression to AIDS in HCV/HIV co-infected individuals. However, the mortality rate due to liver disease is increased in HCV/HIV co-infected patients (Rockstroh et al., 2005, Sullivan et al., 2006).

1.10.2 The influence of HIV on HCV infection.

There is a reported delay in the diagnosis of acute HCV in HIV-positive individuals due to delayed seroconversion. There was no difference in seroconversion time between spontaneous clearers and progressors, but the delay in diagnosis and treatment of HCV may affect the outcome of the infection. Seroconversion time in spontaneous clearers and progressors was not significantly different (Thomson et al., 2009). The introduction of HCV core antigen testing as a screening assay for HIV-infected patients may shorten this diagnostic window period.

There is a potentially higher rate of HCV transmission in HCV/HIV co-infected patients as they have higher HCV viral loads by a magnitude of 0.3-1.08 log RNA copies/ml in blood and other body fluids compared with HIV-1 mono-infected subjects (Shire and Sherman, 2005, Dionne-Odom et al., 2009, Mohsen et al., 2002). Persistence of HCV infection is more common in HCV/HIV co-infected patients, which is probably related to a failure of host immune responses to HCV infection (Danta et al., 2008).

The level of CD4 immunosuppression has emerged as one of the most important determinants of progression to liver fibrosis. Patients with a low CD4 count or who have had an AIDS diagnosis are at increased risk of severe liver disease (Mohsen et al., 2002). There is growing evidence that progression to liver cirrhosis occurs at a faster rate in HCV/HIV co-infected patients. The median time to cirrhosis was estimated at 32 years and 23 years from the time of acquisition in HCV-infected, and HCV/HIV co-infected individuals, respectively (Mohsen et al., 2003).

The HIV infection accelerates the development of hepatocellular carcinoma (HCC) in patients with HCV-related cirrhosis, as it occurs at a younger age and within a shorter time scale, compared to the estimated rate of 1-4% per annum in HCV mono-infected patients (Garcia-Samaniego et al., 2001, Singal and Anand, 2009).

Antiretroviral therapy does not induce a decline in HCV viral load during the first six months of treatment. However a decrease in concentration of HCV RNA of about 1 log₁₀ was seen in HIV/HCV co-infected individuals after receiving 12 months of HAART, the decline being attributed to the improvement of the host immune response (Rockstroh and Spengler, 2004).

1.11 Treatment of acute HCV in HIV-positive patients

Interferon-based HCV antiviral therapy has established but limited efficacy and a decreased likelihood of success in HIV/HCV co-infection (Gilleece et al., 2005, Dominguez et al., 2006, Serpaggi et al., 2006, Braitstein et al., 2004, Torriani et al., 2004, Hadziyannis et al., 2004). Lack of early virological response (EVR) at Week 12 of therapy is highly predictive of treatment failure among HIV/HCV co-infected patients.

Standard therapy in most of the world is still a combination of PegIFN α /RBV; the duration of treatment varies from 16-72 weeks, based on HCV genotype, baseline viral load and early viral response to therapy (Soriano et al., 2011).

The World Health Organisation (WHO) will issue updated HCV treatment guidelines that will not recommend IFN-based therapy as the first line of treatment due to: i) low efficacy, less than 50% of individuals infected with HCV genotypes 1 or 4 achieved an SVR with IFN-based therapy as compared with 85% of patients infected with genotypes 2 or 3; ii) high toxicity, as PegIFN α /RBV therapy is frequently associated with serious undesirable side effects, including depression, anaemia, and decompensation with advanced disease (Soriano et al., 2011).

In numerous clinical trials, rates of treatment discontinuation have ranged between 15% and 30%, and substantial proportions of patients required dose reductions of PegIFN α or RBV (Rockstroh and Spengler, 2004).

In the new era of direct-acting antivirals (DAAs), differences in treatment response between HCV mono-infected and HCV/HIV co-infected patients have not been detected (Sulkowski et al., 2014b, Sulkowski et al., 2013). Thus, treatment recommendations for both patient groups are the same although extra consideration needs to be given to drug-drug interactions.

1.11.1 Predictors of treatment response in HIV-positive individuals with acute HCV infection.

Due to the current limited number of studies, the identification of predictors of treatment response in HIV-infected patients with acute HCV is difficult. Acute HCV has a favourable outcome in the majority of HIV-negative patients but is less easy to treat in HIV-positive patients (Gilleece et al., 2005). Early treatment of HCV/HIV co-infected patients with PegIFN α /RBV results in improved SVR rates (59%) but does not match the 98% treatment success rate reported in HIV-negative patients (Torriani et al., 2004, Gilleece et al., 2005). The reason for this discrepancy is not clear, but impaired immune control and high rate of viral evolution in HIV-infected patients are possible explanations. SVR rates have improved dramatically since the introduction of DAAs (Cooper and Klein, 2014).

1.11.1.1 HCV genotype

HCV genotype has been established as the strongest predictor of successful treatment in chronic HCV mono-infection. However, the small number of acute HCV/HIV cases has limited the studies that can demonstrate an effect of genotype on treatment outcome in acute HCV/HIV co-infected patients.

The majority of participants in studies reporting the outcome of therapy in the HCV/HIV population have been infected with HCV genotype 1a or 4d with an overall SVR rate of 57%. This result compares to an overall SVR rate of 87% in genotype 2/3 participants and suggests that genotype may have an effect on treatment response in a similar way to that observed in chronic HCV mono-infection (van de Laar et al., 2010).

These data are less concerning following reports of the high efficacy of DAAs in HIV/HCV co-infection with SVR rates of 76-98% across genotypes (Sulkowski, 2014, Wyles et al., 2015).

1.11.1.2 Viral load dynamics during early HCV infection

Following acute infection, three different patterns of HCV RNA evolution have been reported: i) persistent high viral load; ii) rapid decline in viral load with subsequent clearance; iii) fluctuating viral load. Spontaneous clearance usually occurs within three months of diagnosis in both HIV-infected and uninfected cohorts (Santantonio et al., 2006, Page et al., 2009). The pattern of viraemia may help to identify those patients who are likely to clear HCV spontaneously (Danta et al., 2008). In a UK-based cohort of HCV/HIV co-infected patients, the drop in HCV viral load within 100 days of the first positive test was strongly predictive of spontaneous clearance with an odds ratio (OR) per \log_{10} drop=1.78 (Thomson et al., 2011).

1.11.1.3 Viral diversification within the host

Clonal analysis was used to study viral diversity during HCV infection. Studies of the HVR1 region in a small group of acute HCV infection have shown that high quasispecies diversity is associated with progressive disease while spontaneous clearance could be predicted in the presence of a narrowed quasispecies repertoire (Thomson et al., 2011). The increased prevalence of escape mutations and resistant variants will lead to progression to chronicity and treatment failure.

1.11.1.4 Transmission bottleneck

Spontaneous clearance of HCV is more likely in HIV-infected patients who acquired HCV sexually than those infected via intravenous drug use (21.9% and 11.6 % respectively). This could be explained by smaller, less diverse inocula, and possibly due to local effects at the mucosal level.

However, individuals who reported homosexual contact have a clearance rate similar to PWID, which could be interpreted by increased mucosal trauma during sex (Shores et al., 2008).

1.11.1.5 Compartmentalization

Compartmentalisation of peripheral blood mononuclear cells (PBMCs)-specific variants has been reported in HIV-infected individuals. These variants were genetically distinct from those detected in plasma. The virus detected in these patients may originate from a cellular reservoir including NK cells, monocytes, and B-cells. Detection of minority variants is further compounded by the fact that viral strains may also be compartmentalised and replicate in PBMCs and extrahepatic compartments including the central nervous system (Forton et al., 2004a, Forton et al., 2004b).

1.11.1.6 Host genetic factors

A single nucleotide polymorphism (rs12979860) has been identified in the IL28B gene that encodes the type III interferon IFN- λ 3. The CC genotype is associated with the spontaneous clearance of HCV infection; it is the strongest genetic predictor of spontaneous clearance of HCV infection which highlight the possible role of IL28B in clearance of HCV infection (Thomas et al., 2009).

The favourable impact of IL-28B genotype CC was genotype dependent as it was stronger in patients infected with genotype 1 or 4 than those who acquired genotype 3 infection. The same conclusion was reported with treatment outcome as patients bearing IL-28B genotype CC had a favourable treatment outcome (Neukam et al., 2011).

Polymorphisms in both IL6 and tumor growth factor (TGF) genes have been reported to be associated with treatment outcome in acute HCV/HIV co-infection (van de Laar et al., 2010). The development of a decision-making model integrating clinical variables and IL28B genotype would be of value in predicting the treatment outcome (O'Brien et al., 2011).

1.12 A new era of antivirals for HCV infection

The non-structural NS3/4A, NS5A and NS5B proteins of HCV are the primary targets of DAAs. NS3/4A protease inhibitors (PIs) are competitive inhibitors by blocking the protease active site or allosteric inhibitors that affect the substrate cleavage by interfering with conformational changes (Steinkuhler et al., 1998, Delang et al., 2013, Romano et al., 2012, Rupp and Bartenschlager, 2014). NS5A inhibitors interact with domain 1 of NS5A, although the exact mechanism of action is still unclear. Nucleos(t)ide analogue NS5B polymerase inhibitors are incorporated into the elongating RNA strand and cause chain termination by inhibiting the incorporation of the next nucleotide; non-nucleoside NS5B polymerase inhibitors inhibit polymerase activity by allosteric mechanisms of action through interaction with either the thumb 1, thumb 2, palm 1, or palm 2 domain of NS5B (Pawlotsky, 2013b, Gerber et al., 2013).

DAAs currently licensed for use in the European Union are Simeprevir (SMV), Daclatasvir (DCV), Sofosbuvir (SOF), Ledipasvir (LDV), Ritonavir-boosted Paritaprevir (PTV), Ombitasvir (OBV), and Dasabuvir (DSV). This section will describe treatment options for genotype 1a in both HCV mono-infection and HCV/HIV co-infection according to the recent recommendation on the treatment of HCV published by European Society for Study of the Liver (EASL).

1.12.1 Simeprevir

SMV is an NS3/4A protease inhibitor. It is metabolised by the hepatic CYP3A system and eliminated mostly via biliary excretion. SMV has an excellent tolerability profile with no reported differences in the type and incidence of adverse events between SMV and placebo groups.

In clinical trials, the most common side effects (>10%) reported during a 12 weeks full course in combination with SOF were fatigue (25%), headache (21%), nausea (21%), insomnia (14%), pruritus (11%) and rash (11%). When treatment was extended for 24 weeks, dizziness and diarrhea were reported in 16% of patients (EASL, 2015).

There is a lack of data on the efficacy and safety of SMV in end-stage renal disease (Lawitz et al., 2014a). Few patients have reported mild reversible transient increases in bilirubin levels while receiving SMV treatment (Hayashi et al., 2014b).

Drug-drug interaction is an important issue with some antiretroviral drugs in HCV/HIV co-infected patients. As the primary enzyme involved in the metabolism of SMV is CYP3A4, co-administration of SMV with substances that are moderate or strong inducers or inhibitors of CYP3A4 is not recommended to avoid any significant effect on the exposure of SMV. A number of drugs are contra-indicated in patients receiving SMV, including any HIV protease inhibitor, boosted or not by ritonavir (EASL, 2015).

1.12.2 Daclatasvir

Daclatasvir is an NS5A inhibitor. DCV is mainly excreted in the faeces. DCV does not require any dose adjustment for patients with any degree of renal or hepatic impairment (EASL, 2015). The most common adverse events are fatigue, nausea, and headache. DCV exposure is reduced if co-administered with strong CYP3A4 or P-gp inducers while drugs that strongly inhibit the CYP3A4 system increase the plasma levels of DCV. Therefore, dose adjustments of DCV are required (Bunchorntavakul and Reddy, 2015). DCV can be administered safely with antiretroviral drugs including raltegravir, dolutegravir and maraviroc (Wyles et al., 2015).

1.12.3 Sofosbuvir

Sofosbuvir is an oral nucleotide inhibitor of the HCV NS5B RdRp enzyme. SOF is the main component in available combination therapy in all HCV genotypes due to its pan-genotypic activity (Abraham and Spooner, 2014).

SOF is mainly excreted via the kidneys. End stage renal disease leads to high drug exposure while liver impairment increases SOF plasma level slightly. Therefore, there is no dose recommendation for patients with severe renal

impairment. The most common adverse events observed during SOF treatment are mainly due to the combination with PegIFN α /RBV such as fatigue, headache, nausea, insomnia, and anaemia. There is no reported contraindication to administration of SOF with antiretroviral agents (EASL, 2015).

SOF is available in a two-drug fixed dose combination with LDV in a single tablet. LDV is eliminated via biliary excretion. Cirrhosis (including decompensated cirrhosis) has no clinically relevant effect on exposure to LDV. The most frequent side effects reported with SOF/LDV combination are fatigue and headache. Since the combination contains LDV and SOF, any interactions identified with the individual drugs will apply to the combination. As both LDV and SOF are transported by P-gp, co-administration of any potent P-gp inducers will decrease drug exposure of both drugs and consequently a reduced efficacy of the regimen. However, LDV/SOF may be given with all antiretrovirals (EASL, 2015).

1.12.4 PrOD combination (Paritaprevir, Ombitasvir, and Dasabuvir)

The PrOD combination consists of ritonavir-boosted paritaprevir (PTV), ombitasvir (OBV), and dasabuvir (DSV). PTV is an NS3-4A protease inhibitor. It is eliminated mainly in the faeces. As PTV is metabolised by CYP3A4, it is co-administered with ritonavir, which is a CYP3A inhibitor, resulting in increased PTV drug exposure, allowing an administration of lower dose and a once daily regimen. OBV is an NS5A inhibitor that is predominantly eliminated in the faeces. DSV is a non-nucleoside NS5B polymerase inhibitor that is mostly cleared by biliary excretion and in the faeces.

The most common adverse events encountered with the PrOD combination were pruritus, fatigue, nausea, asthenia and insomnia. The more frequent side effects are considered to be related to RBV, but pruritus is considered to be related to PrOD. Severe adverse events occurred in <2.5% of cases and treatment discontinuation due to adverse events occurred in 1-2%. Only 1-2% discontinued treatment in clinical trials, and severe adverse events occurred in less than 2.5% of patients.

The PrOD combination is contraindicated in patients with severe hepatic impairment (Child-Pugh C) (EASL, 2015). The administration of ritonavir increases the likelihood of occurrence of many drug-drug interactions. As ritonavir is a strong inhibitor of CYP3A4, a number of drugs that are metabolised by this enzyme are contraindicated to avoid toxicity and serious adverse events including HIV PIs (EASL, 2015).

1.13 The HCV available treatment regimens

The treatment options for HCV have evolved rapidly in recent years, and DAA regimens have reached SVR rates of up to 97-100% in genotype 1a patients. The main clinical trials studying the efficacy of various DAAs in achieving viral clearance in HCV genotype 1 patients are discussed in this section including interferon-containing regimens and interferon-free options. Published results are listed below (Table 1-1 to Table 1-4).

Due to the rapid evolution of available treatments, options are regularly updated with the latest EASL recommendations published in May 2015. A brief overview of currently recommended regimens is shown in Table 1-5.

1.13.1 Interferon-containing regimens

1.13.1.1 Simeprevir/ Pegylated Interferon/Ribavirin

Treatment-naïve patients, who received SMV for 12 or 24 weeks with PEG-IFN α /RBV for 24 or 48 weeks (according to response-guided therapy), achieved SVR rates of 77-92% (Hayashi et al., 2014a). A similar response rate was achieved in treatment-experienced patients, who received SMV for 12, 24, or 24 weeks with PegIFN α /RBV for 48 weeks, with SVR rates of 61-80% (Akuta et al., 2014).

SMV in a combination with PegIFN α /RBV is well tolerated and associated with improved SVR rates in most treatment-naïve, HCV genotype 1 infected patients (Hayashi et al., 2014b). Two Phase III clinical trials in treatment-naïve patients; Quest-1 and Quest-2, demonstrated overall SVR rates in genotype 1

infected patients of 80% and 81%, respectively, with a similar outcome for patients infected with subtype 1b (85%) and subtype 1a (84%). However, there was a reduction in SVR rate (58%) when a Q80K substitution was detectable as a dominant variant at baseline. SMV was associated with a higher SVR rate (84%) in non-cirrhotic patients compared to 60% in cirrhotic patients (Jacobson et al., 2014, Manns et al., 2014).

SMV was used to retreat patients who relapsed after PegIFN α /RBV therapy; genotype 1a patients achieved an SVR24 of 70% (Forns et al., 2014). In HCV/HIV co-infected patients, SVR was achieved in 79% of treatment-naïve patients receiving SMV/PegIFN α /RBV. A direct comparison between TVR and SMV in retreatment of a group of patients who failed to respond to PegIFN α /RBV showed that SVR12 was achieved in 70% of prior partial responders and 44% of null responders who received a triple combination including SMV. In the TVR arm, SVR12 occurred in 68% of prior partial responders and 46% of prior null responders (Reddy et al., 2015).

1.13.1.2 Sofosbuvir/ Pegylated Interferon/ Ribavirin

A combination of PegIFN α /RBV and SOF for 12 weeks was administered to genotype 1 treatment-naïve patients in the NEUTRINO trial. The overall SVR rate was 92% for subtype 1a and 82% for subtype 1b. Cirrhotic patients had a lower response rate (80%) than the non-cirrhotic patient (92%) (Lawitz et al., 2013).

Real life observational studies showed variability in treatment outcome. SVR4 was achieved in 90% of non-cirrhotic patients compared to 70% reported in cirrhotic patients (Jensen et al., 2014). Similar outcomes occurred in a similar mixed cohort that included 58% treatment-naïve and 42% treatment-experienced patients, as 77% of the treatment-experienced non-cirrhotic patients achieved SVR12 while 62% of cirrhotic patients achieved SVR12 (Dieterich et al., 2014a).

1.13.2 Interferon-free options

1.13.2.1 Sofosbuvir/ Ledipasvir

This combination achieved excellent efficacy in the ION-1 study; an SVR12 of 99% was achieved in treatment-naïve patients after a course of 12 weeks of the fixed-dose combination of SOF/LDV/RBV. SVR12 rates were 99% and 98% after 24 weeks of therapy with and without RBV respectively. The study group included 16% of patients with compensated cirrhosis (Afdhal et al., 2014b).

The duration of treatment may be further reduced to 8 weeks without a high drop in SVR rate; SVR12 rates were 94% without RBV for 8 weeks, 93% with RBV for 8 weeks and 95% without RBV for 12 weeks. In this study, baseline viral load was used to guide the duration of treatment, and further analysis suggested that a shorter duration of treatment (8 weeks) was successful in patients with an HCV RNA level <6 million (6.8 Log_{10}) IU/ml at baseline (Kowdley et al., 2014). The lack of the accuracy of viral load measurement and inter-laboratory variation with this range of value may restrict the use of such a threshold for deciding the treatment strategy. Further studies are needed to confirm the efficacy of the 8-week regimen.

Table 1-1: Directly acting antivirals efficacy studies (Part 1)

Trial Name	Patients	Study groups	Study outcome (SVR)	References
Simeprevir (SMV)				
CONCERTO-1	Treatment-naïve Genotype 1	SMV 12 Wks + PR 24/48 Placebo 12Wks + PR48	SVR12: SMV 12 Wks + PR 24/48 (88.6%) Placebo 12Wks + PR48 (61.7%)	(Hayashi et al., 2014a)
PILLAR	Treatment-naïve Genotype 1	SMV 75/150 mg 12/24 Wks + PR 24/48 Placebo 24 Wks + PR48	SVR12: SMV (74.7%-86.1%) Placebo (64.9%)	(Fried et al., 2013)
ATTAIN	Treatment failure (PegIFN α /RBV)	SMV 150mg 12 Wks+ PR 48 Wks TVR150mg 12 Wks+ PR 48 Wks	SVR12: SMV (54%) vs TVR (55%)	(Reddy et al., 2015)
QUEST-1	Treatment naïve Genotype 1	SMV 150 mg 12 Wks + PR 24/48 Placebo 24 Wks + PR48	SVR12: SMV+PR (80%) Placebo + PR (50%)	(Jacobson et al., 2014)
QUEST-2	Treatment naïve Genotype 1	SMV 150 mg 12 Wks + PR 24/48 Placebo 12 Wks + PR48	SVR12: SMV+PR: SVR12 [81%] Placebo + PR: SVR12 [50%]	(Manns et al., 2014)
NCT01479868	HIV +ve with Genotype 1 Treatment naïve or experienced	SMV 150 mg 12 Wks + PR 24/48	SVR12: Treatment-naïve (79.2%), Null responders (57%), Relapsers (87%)	(Dieterich et al., 2014b)
The Dragon	Treatment-naïve Genotype 1	SMV 50/100 mg 12/24 Wks + PR 24 Control: PR48	SVR24 SMV (77%-92%), PR48 (45.5%)	(Hayashi et al., 2014b)

The table is limited to clinical trials addressing the clinical efficacy of DAAs against HCV genotype 1 with published results. Source: Clinical trials.org, last accessed 07/11/2015. (PR): PegIFN α /RBV, (Wks): weeks.

Table 1-2: Directly acting antivirals efficacy studies (Part 2)

Trial Name	Patients	Study groups	Study outcome (SVR)	References
Paritaprevir (PTV), Ombitasvir (OBV), and Dasabuvir (DSV)				
PEARL IV	Treatment-naïve - Genotype 1	PTV+OBV+DBV ± RBV 12Wks	PTV+OBV+DSV (90%) PTV+OBV+DSV+RBV (97%)	(Ferenci et al., 2014)
TURQUOISE-II	Cirrhotic treatment-naïve and experienced - Genotype 1	PTV+OBV+DSV + RBV 12/24 Wks	PTV+OBV+DSV+RBV 12 Wks (92%) PTV+OBV+DSV+RBV 24 Wks (96.5%)	(Poordad et al., 2014)
SAPPHIRE-II	Treatment failure	PTV+OBV+DSV + RBV 12 Wks	SVR 12: PTV+OBV+DSV+RBV (96.3%)	(Zeuzem et al., 2014b)
Sofosbuvir (SOF)				
Fission	Treatment-naïve Genotype 1, 4, 5, or 6	SOF+RBV 12 Wks PR 24 Wks	SVR12 SOF+RBV (67%), PR 24 (67%)	(Lawitz et al., 2013)
Neutrino	Treatment-naïve Genotype 1, 4, 5, or 6	SOF+ PR12 Wks	SVR12 (91%)	(Lawitz et al., 2013)
ATOMIC	Treatment-naïve Genotype 1, 4, 5, or 6	SOF+PR 12 Wks SOF+PR 24 Wks	SVR 12: SOF+PR 12 Wks (90.4%) SOF+PR 24 Wks (92.7%)	(Kowdley et al., 2013)
PHOTON-1	Treatment naïve (TN) (or) Treatment experienced (TE) (Genotypes 1,2,3) HIV positive patients	SOF+RBV 12/24 weeks	SVR12: TN Gen1 SOF + RBV 24 Wks (76%) TN Gen2 SOF + RBV 12 Wks (88%) TN Gen3 SOF + RBV 12 Wks (67%) TE Gen2 SOF + RBV 24 Wks (92%) TE Gen 3 SOF + RBV 24 Wks (88%)	(Sulkowski et al., 2014b)

The table is limited to clinical trials addressing the clinical efficacy of DAAs against HCV genotype 1 with published results. Source: Clinical trials.org, last accessed 07/11/2015. (PR): PegIFN α /RBV, (Wks): weeks.

Table 1-3: Directly acting antivirals efficacy studies (Part 3)

Trial Name	Patients	Study groups	Study outcome (SVR)	References
Sofosbuvir (SOF) + Ledipasvir (LDV)				
ION-1	Treatment-naïve	SOF+LDV 12/24 Wks ± RBV	SVR24: LDV+SOF 12 Wks (98.6%) LDV+SOF + RBV 12 Wks (97.2%) LDV+SOF 24 Wks (98.2%) LDV+SOF+RBV 24 Wks (99%)	(Afdhal et al., 2014b)
ION-2	Treatment failures on PR ± PIs	SOF+LDV 12/24 Wks ± RBV	SVR24: LDV+SOF 12 Wks (93.6%) LDV+SOF + RBV 12 Wks (96.4%) LDV+SOF 24 Wks (99%) LDV+SOF+RBV 24 Wks(99%)	(Afdhal et al., 2014a)
ION-3	Treatment-naïve	SOF+LDV 8 weeks ± RBV SOF+LDV 12 weeks	SVR24: LDV + SOF 8 weeks (94%) LDV + SOF + RBV 8 weeks (93%) LDV-SOF 12 weeks (96%).	(Kowdley et al., 2014)
LONESTAR	Cohort A non-cirrhotic, treatment naïve, Cohort B PI failures (55% compensated cirrhosis)	SOF+LDV± RBV 8/12 weeks	SVR24: In cohort A, SOF + LDV 8 Wks (95%) SOF+ LDV+RBV 8 Wks (100%) SOF+LDV 12 Wks (95%) In cohort B, OF + LDV 12 Wks (95%) SOF + LDV + RBV 12 Wks (100%).	(Lawitz et al., 2014b)
SIRIUS	Cirrhotic treatment experienced (Gen 1)	SOF+LDV+ RBV 12 weeks SOF+LDV 24 weeks	SVR 24: LDV+SOF+ RBV 12 weeks (96%) LDV+SOF 24 weeks (97.4%)	(Bourliere et al., 2015)
Sofosbuvir (SOF) + Simeprevir (SMV)				
The COSMOS	Non-responders: F0-F2, Treatment naïve: F3-F4	SMV 150mg + SOF 400 mg 12/24 weeks +/- RBV	SVR12: Non-responders (92%) SVR12: Treatment naïve (94%)	(Lawitz et al., 2014a)

The table is limited to clinical trials addressing the clinical efficacy of DAAs against HCV genotype 1 with published results. Source: Clinical trials.org, last accessed 07/11/2015. (PR): PegIFN α /RBV, (Wks): weeks.

Table 1-4: Directly acting antivirals efficacy studies (Part 4)

Trial Name	Patients	Study groups	Study outcome (SVR)	References
Daclatasvir (DCV)				
HEPCAT	Treatment-naïve Genotype 1 and 4	DCV 20/60 mg 12 Wks + PR 24/48 Placebo 12 Wks + PR 48	SVR24 DCV (60%), Placebo (37.5%)	(Hezode et al., 2015)
Sofosbuvir (SOF) + Daclatasvir (DCV)				
A1444040	Treatment-naïve or experienced (genotypes 1,2,3)	Different dose regimens SOF+DCV	SVR12: 98% in Genotype 1 SVR12: 92% in Genotype 2 SVR12: 89% in Genotype 3	(Sulkowski et al., 2014a)
ALLY 1	Post-liver transplant Cirrhotic patients	DCV+SOF+RBV for 12 Wks	SVR12: 95.1% in post-transplant SVR12: 82.2% in cirrhotic patients	(Poordad et al., 2015)
ALLY-2	HIV co-infected Treatment-naïve or experienced	DCV 60mg +SOF 400mg 8/12 Wks	SVR12: 96.4% in treatment-naïve SVR12: 97.7% in treatment experienced	(Wyles et al., 2015)

The table is limited to clinical trials addressing the clinical efficacy of DAAs against HCV genotype 1 with published results. Source: Clinical trials.org, last accessed 07/11/2015. (PR): PegIFN α /RBV, (Wks): weeks.

Table 1-5: EASL Recommendations on Treatment of Hepatitis C 2015.

IFN-free regimens	
Options	Genotype
SOF+RBV	2,3
SOF/LDV (\pm RBV)	1,4,5,6
OBV/Paritaprevir/Ritonavir+ DSV (\pm RBV)	1
SOF + SMV (\pm RBV)	1,4
SOF + DCV (\pm RBV)	All
OBV/ Paritaprevir/ Ritonavir (\pm RBV)	4
IFN-containing regimens	
PegIFN α + RBV + SOF	All
PegIFN α + RBV + SMV	1,4

Treatment recommendation for HCV/HIV co-infected patients with chronic hepatitis C with compensated (Child-Pugh A) cirrhosis.

The combination achieved high SVR rates in treatment-experienced patients, including 20% with cirrhosis; after 12 weeks of treatment with and without RBV, the SVR12 rates were 96% and 94%, respectively. After prolonging treatment for another 12 weeks, the SVR rates were the same in both groups at 99% (Afdhal et al., 2014a). The combination of SOF/LDV was given for 12 weeks to HCV/HIV co-infected patients, and the SVR12 rate was 98% (Townsend et al., 2014).

The administration of SOF/LDV combination with or without RBV in different trials showed an overall SVR rate of 95% after 12 weeks of treatment. The prolonged 24-week course increased the overall SVR rate to 98%. No safety issues were reported in this patient population and the safety profile was similar to that reported in non-cirrhotic patients (Bourliere et al., 2014b). The addition of RBV to the combination did not have an impact on SVR12 in treatment-naïve patients (SVR12 rates between 96% and 100%). However, the duration of treatment and the addition of RBV improved the SVR12 in treatment-experienced patients. The overall SVR was 90% after 12 weeks of treatment without RBV and increased to 100% after prolonging the treatment to 24 weeks with RBV included in the combination (Bourliere et al., 2014b).

In the SIRIUS study, four different regimens of SOF/LDV were evaluated in patients with compensated cirrhosis who failed to achieve an SVR after triple therapy containing either TVR or BOC. The patients were assigned to two groups; SOF/LDV for 12 weeks with RBV or 24 weeks without RBV. The SVR12 rates were 96% and 97%, respectively (Bourliere et al., 2014a). Thus, the addition of RBV can shorten the treatment duration in cirrhotic patients.

Clinical trials have demonstrated the safety and efficacy of LDV/SOF in HCV/HIV co-infected patients receiving antiretroviral therapy including tenofovir and emtricitabine with rilpivirine, raltegravir or efavirenz. These patients were infected with HCV genotypes 1 and 4 (Osinusi et al., 2015).

1.13.2.2 Paritaprevir, ritonavir, Ombitasvir, Dasabuvir (PrOD)

The PrOD combination was evaluated in the SAPPHIRE-I study. In this study, a cohort of non-cirrhotic, treatment-naïve patients was treated with PrOD and RBV for 12 weeks. SVR12 rates were 95% in subtype 1a and 98% in subtype 1b patients (Feld et al., 2014). The efficacy of PrOD combination was evaluated in both subtype 1a and 1b separately.

In PEARL-IV, treatment-naïve non-cirrhotic subtype 1a patients had SVR12 rates of 97% and 90%, with and without RBV, respectively. In PEARL-III, treatment-naïve non-cirrhotic subtype 1b patients achieved SVR12 rates of 99% both with and without RBV (Ferenci et al., 2014).

The TURQUOISE-I study included HCV/HIV co-infected, treatment naïve patients. SVR12 rates were 93% and 91% after 12 or 24 weeks of treatment, respectively; SVR12 was achieved in 91% of subtype 1a and 100% of subtype 1b patients (Wyles et al., 2014b). The patients received antiretroviral regimens containing atazanavir or raltegravir.

Non-cirrhotic, treatment-experienced (PegIFN α /RBV failures) patients were included in the SAPPHIRE-II trial in which patients were treated with PrOD and RBV for 12 weeks. There was no difference in response between subtype 1a and subtype 1b; the SVR12 rates were 96% and 97%, respectively. (Zeuzem et al., 2014b). The addition of RBV in this cohort of patients is considered unnecessary as SVR12 was achieved in 100% of cases without RBV and 97% with RBV in patients infected with subtype 1b receiving this combination in the PEARL-II trial (Andreone et al., 2014).

RBV was added to the PrOD combination to treat patients with compensated cirrhosis. The study cohort included treatment-naïve and treatment-experienced patients, and overall SVR rates were 92% after 12 weeks and 96% after 24 weeks of the combination of PrOD and RBV (Poordad et al., 2014).

1.13.2.3 Sofosbuvir and Simeprevir

This combination has been evaluated in several patient populations. In one study, non-cirrhotic, prior null responders to PegIFN α /RBV therapy were offered treatment for 12 or 24 weeks, with or without RBV. SVR12 rates were 93% and 96% for 12 weeks of therapy with and without RBV, respectively, and 93% and 79% for 24 weeks of therapy with and without RBV, respectively.

In the second cohort, treatment-naïve patients and treatment -experienced (null responders) with a METAVIR score of F3-F4 were treated with the same regimens as above. The SVR12 rates were 93% and 93% for 12 weeks of therapy without and with RBV, respectively, and 100% and 93% for 24 weeks of therapy without and with RBV, respectively (Lawitz et al., 2014c).

Two large-scale real-life studies with SOF and SMV indicated that the SOF/SMV regimen is well tolerated, but SVR rates were lower than those reported in the COSMOS trial, especially in individuals with advanced stages of liver disease (EASL, 2015). In both studies, the overall SVR4 rate was comparable; 81% and 89 %. However, a higher SVR was reported in non-cirrhotic patients (92%) compared to cirrhotic cases (87%) in the HCV TARGET study. SVR4 was more frequent in subtype 1b than 1a patients (Jensen et al., 2014). Similar results were shown in the preliminary analysis of the TRIO real-life study, as in treatment-naïve patients SVR was higher in non-cirrhotic patients (88%) compared to cirrhotic cases (75%) (Dieterich et al., 2014a).

1.13.2.4 Sofosbuvir and Daclatasvir

An interferon-free treatment regimen containing SOF and DCV with or without RBV was investigated for previously untreated genotype 1-3 patients and genotype 1 patients with prior TVR or BOC triple therapy treatment failure. Genotype 1-infected patients achieved a higher SVR compared to genotype 3-infected patients, SVR rates were 98% and 89% respectively (Sulkowski et al., 2014a). SVR rates were 100% with or without RBV in treatment-naïve patients with 24 weeks of therapy, compared with 98% in treatment-naïve patients who

received 12 weeks of therapy without RBV. In the treatment-experienced group who did not respond to triple therapy with either TVR or BOC, SVR rates were 100% and 95% with and without RBV, respectively (Sulkowski et al., 2014a).

Data were derived from studies that included 335 HCV treatment-naïve and treatment-experienced HIV/HCV-co-infected persons as well as those with or without cirrhosis. The efficacy and safety of SOF/LDV for 12 weeks have been confirmed in clinical trial data in HIV-infected individuals with HCV genotype 1 or 4 infection, similar efficacy and safety were demonstrated by PrOD combination [paritaprevir (PTV), ombitasvir (OMV) and dasabuvir (DSV)] with or without RBV for 12 weeks (Osinusi et al., 2015, Sulkowski et al., 2015).

1.14 HCV drug resistance

As discussed above, DAAs target different viral functions, specifically the NS3/4A protease, the NS5A protein and the RNA-dependent NS5B polymerase. Drug resistance may play a fundamental role in patients with failure to DAA-containing regimens (Schneider and Sarrazin, 2014). Current combinations of DAAs are highly effective; however, antiviral potency, safety issues, posology, drug interactions, and resistance are major determinants of treatment success.

Resistance-associated mutations (RAMs) occur in patients treated with DAAs due to the error-prone nature of the HCV RdRp either before treatment as naturally occurring polymorphisms or as *de novo* mutations under drug selection pressure. The risk of outgrowth of resistance-associated variants (RAVs) is related to multiple factors including high replication rates, error-prone RdRp, selective pressure of the drug, role of replication space, replication capacity of RAVs, the genetic barrier to drug resistance (Halfon and Locarnini, 2011). Poor adherence is another indirect factor as it lowers drug exposure thereby affecting antiviral efficacy (Sarrazin and Zeuzem, 2010).

Resistance-associated mutations (RAMs) have been identified by *in vitro* replicon systems and, less frequently, in patients who fail DAA treatment in clinical trials.

The discrepancies between the reported *in vitro* and *in vivo* RAMs are mainly attributed to two factors: i) the replication capacity of the RAVs and their ability to persist in the quasispecies; ii) the level of resistance conferred by RAMs as low level of resistance can still respond to suppressive effect of the drug (Soriano et al., 2011).

1.14.1 Genetic barrier for resistance

The genetic barrier is ‘the threshold probability that the virus will mutate and escape from the selective action of the drug’. A simplified definition is the number of mutations required to acquire antiviral resistance. The genetic barrier for resistance is correlated with the number of RAMs needed to confer resistance against a certain class of drugs (Halfon and Locarnini, 2011).

A low genetic barrier resistance is one that involves a primary mutation that comes at a low fitness cost to the virus and can emerge quickly. In contrast, a higher genetic barrier is one that involves mutations with a high fitness cost to generate “total” resistance. Moreover, other mutations may be required to restore fitness of these low-fitness/high resistance variants. Drugs with a high genetic barrier to resistance are less likely to be associated with clinically meaningful resistance (Halfon and Locarnini, 2011).

In the case of the PIs, the genetic barrier is influenced by the HCV subtype (either 1a or 1b). Two nucleotide changes are required to generate an amino acid change in position 155 in subtype 1b isolates: R155K [CGG---AAG] while only one (R155K [AGG---AAG]) is needed for subtype 1a (Sarrazin and Zeuzem, 2010). As of today, the circulating HCV variants did not demonstrate any emergence of drug resistance against the current nucleoside analogue polymerase inhibitors; this may be due to significantly reduced replication capacity of the resistant variants (S282T of NS5B) or different binding sites of nucleoside and non-nucleoside inhibitors (Hang et al., 2009).

1.14.2 Natural polymorphisms

The most common form of genetic variation is a SNP. A central issue in HCV is whether the presence of minority subpopulations containing RAVs would help to determine the treatment regimen. Both Sanger sequencing and NGS have identified mutations that confer resistance in patients who have not previously been exposed to DAAs (Gregori et al., 2013). Dominant, naturally occurring DAAs resistance mutations are common in HCV genotype 1 infected treatment-naïve patients.

The likelihood that a DAA will select for and allow outgrowth of RAVs depends on the DAA's genetic barrier to resistance, the level of drug exposure, the patient population and the viral fitness of the resistant variant (Fridell et al., 2010, Gao, 2013). The prevalence of baseline polymorphisms, which are natural substitutions that may differ in their prevalence by genotype, subtype, and subtype clade, can also greatly affect the efficacy of specific DAAs (Paolucci et al., 2012).

The detection of RAVs depends primarily on the method used for sequencing as different platforms have different sensitivities. For Sanger sequencing, clonal analysis, and NGS sequencing, variants below frequencies of approximately 25%, 5% and 0.5% respectively cannot be detected (Halfon and Sarrazin, 2012).

Viral loads are high in the majority of patients carrying natural polymorphism mutations, suggesting that such viruses might achieve replicative capacities comparable to non-resistant strains *in vivo* (Kuntzen et al., 2008).

1.14.2.1 NS3/4A Protease Inhibitors

The HCV protease is characterised by a solvent-exposed substrate binding site, requiring small molecule inhibitors to rely on few specific interactions to achieve tight binding with the enzyme (Romano et al., 2012). The 3D structure has shown that RAVs bearing substitutions located in close vicinity to the NS3

protease catalytic triad reduce the drug inhibitory effect by lowering the drug's affinity for the enzyme's catalytic site (Pawlotsky, 2011).

Protease inhibitor-RAVs have recently been shown to have reduced *in vitro* fitness due to inability to produce infectious virions (Shimakami et al., 2011). However, there is growing evidence that acquisition of RAMs may not necessarily be associated with reduced *in vivo* viral fitness (Kuntzen et al., 2008). This is demonstrated in NS5A RAVs, where RAVs persist for a long time after stopping the treatment.

Naturally occurring amino acid changes in NS3, associated with reduced drug susceptibility, have been observed at baseline in treatment-naïve patients (Bartels et al., 2013, Kuntzen et al., 2008). In fact, it is foreseen that all single drug-resistant mutations and double/triple combinations pre-exist before treatment in most patients (Rong and Perelson, 2010). Because of the intrinsic fitness cost of resistant mutations, RAVs regularly emerge, but are not fixed in the absence of selective pressure, and remain detectable at low frequency in untreated patients (Rong and Perelson, 2010).

Bartonili *et al.* observed natural amino acid changes in NS3 associated with reduced protease inhibitor susceptibility using Pyrosequencing in treatment-naïve chronically infected genotype 1 patients with or without HIV coinfection. RAVs were observed in 85.7% of patients (Bartolini et al., 2013).

Recently developed PIs, such as SMV and PTV, exhibit improved genetic barriers to resistance and enhanced antiviral activity against different HCV genotypes. However, the efficacy of these PIs is still substantially reduced by substitutions at amino acids R155 and D168 (Romano et al., 2012, Schinazi et al., 2014). For PTV and SMV, the rate of naturally occurring NS3 resistance mutations is low (0.1-3%). However, for SMV one additional variant to confer medium level resistance (Q80K) is observed in 9-48% of treatment naïve genotype 1a infected patients with the highest incidence in North America. Patients with the Q80K variant achieve consistently lower SVR rates with SMV triple therapy (Schneider and Sarrazin, 2014).

The most commonly observed substitutions in genotype 1a-infected patients who do not achieve SVR with NS3 inhibitors are R155K and D168E/V. NS3 amino acid substitutions frequently detected in genotype 1b-infected patients who do not achieve SVR are Q80R and D168E/V. For both subtypes, the substitutions Q80K/R, S122A/G/I/T, R155Q, and D168F were not observed at the time of virological failure alone but rather in combination with other substitutions at positions 80, 122, 155, and/or 168 (Lenz et al., 2010, Manns et al., 2014, Jacobson et al., 2014, Forns et al., 2014).

The impact of the presence of resistance mutations to HCV NS3 protease inhibitors in <1% of the viral quasispecies is minimal as >1000-fold viral load reduction upon treatment is achieved; this low impact is attributed to the reduced replicative fitness of the variants carrying these mutations *in vitro*.

Meanwhile, an R155K protease mutation was detected as the dominant variant in a treatment-naïve patient, raising concerns about pre-existing PI drug resistance (Kuntzen et al., 2008). The PROMISE study in relapsers as well as the QUEST trials in treatment-naïve subjects illustrated the impact of pre-existent Q80K in genotype 1 patients on treatment outcome (Lenz et al., 2015, Jacobson et al., 2014, Forns et al., 2014).

In the majority of patients (83.7%) in phase IIb and III studies with genotype 1a and baseline Q80K, an emerging single R155K-variant at the time of treatment failure was detected, suggesting that the presence of Q80K alone is not sufficient to explain treatment failure (Schneider and Sarrazin, 2014).

For PTV, the NS3 amino acid substitutions most commonly observed in genotype 1a-infected patients, who did not achieve SVR were D168A/V/Y. NS3 amino acid substitutions frequently detected in genotype 1b-infected patients, who did not achieve SVR were Y56H and D168V (Feld et al., 2014, Zeuzem et al., 2014b, Poordad et al., 2014).

1.14.2.2 NS5A Inhibitors

The mechanism of action of NS5A inhibitors is not entirely understood. Current NS5A inhibitors are characterised by broad genotypic coverage and a relatively low barrier to resistance. Certain key RAMs like Y93H are detected as natural baseline polymorphisms in some patients (Gao, 2013, Fridell et al., 2011, Dore et al., 2015, Lawitz et al., 2015b).

NS5A amino acid substitutions most commonly observed in genotype 1a-infected patients who do not achieve SVR after treatment with DCV are M28T, Q30E/H/R, L31M, H58D, and Y93H/N. The NS5A amino acid substitutions most commonly observed in genotype 1b-infected patients who do not achieve SVR are L31M/V and Y93H (Nelson et al., 2015, Dore et al., 2015).

Genotype 1a-infected patients with LDV treatment failure most commonly have Q30E/R, L31M, and Y93C/H/N RAVs. The NS5A amino acid substitution most commonly observed in genotype 1b-infected patients who do not achieve SVR is Y93H (Lawitz et al., 2012, Wong et al., 2013, Afdhal et al., 2014b).

Amino acid substitutions most commonly observed in genotype 1a- infected patients who do not achieve SVR despite treatment with OBV are M28T and Q30R. Y93H is the most commonly observed NS5A mutation in genotype 1b-infected patients who do not achieve SVR (Poordad et al., 2014, Zeuzem et al., 2014b, Feld et al., 2014).

1.14.2.3 NS5B Polymerase Inhibitors

Nucleos(t)ide inhibitors have a high resistance barrier (Gentile et al., 2014b, Schinazi et al., 2014). NS5B non-nucleoside analogue inhibitors bind outside of the polymerase active site and can be further sub-classified based on their allosteric binding sites (Palm 1, Palm 2, Thumb 1, and Thumb 2) (Gerber et al., 2013). *In vitro* studies have shown that the substitution of serine 282 with threonine (S282T) is the main resistance mutation that results in reduced susceptibility of HCV to SOF (Wohnsland et al., 2007). This mutation confers reduced replicative capacity due to decreased functional ability of the NS5B

polymerase, a phenomenon also demonstrated by the absence of the S282T mutation in treatment-naïve patients (Kuntzen et al., 2008). In the ELECTRON study, one patient with genotype 2b infection who relapsed following treatment with SOF monotherapy developed the S282T mutation, but the mutation was undetectable 12 weeks after completion of treatment (Gane et al., 2013). Additionally, SOF maintains activity against HCV with mutations conferring resistance to other classes of agents (Abraham and Spooner, 2014).

L159F and V321A substitutions were selected in several subjects during a pooled analysis of SOF phase III trials including HCV genotype 3 infected patients who failed treatment. These variants alone conferred 1.2- to 1.6-fold reduced phenotypic susceptibility to SOF *in vitro*. However, these variants require further investigation to elucidate their association with viral resistance (Zeuzem et al., 2014a, Donaldson et al., 2015, Svarovskaia et al., 2014).

While pre-existence of naturally occurring resistance S282T mutations have not been described in treatment naïve or PegIFN α /RBV failure patients even by deep sequencing analysis (Schneider and Sarrazin, 2014), other treatment-emergent NS5B substitutions (L159F, E341D, L320F) have frequently been observed in patients relapsing with SOF-based regimens. None of these substitutions have been associated with a phenotypic change in SOF or RBV susceptibility, however (Degasperi and Aghemo, 2014).

The S282T mutation is difficult to detect *in vivo*; Pyrosequencing has been used to detect NS5B RAVs in 16 treatment naïve, genotype 1 patients. None of the patients harboured the S282T variant, which is in agreement with the reported high genetic barrier of nucleoside analogue NS5B polymerase inhibitors (Franco et al., 2013). Similarly, the S282T mutation was not present at baseline in 71 treatment-naïve HCV-infected patients tested using 454 pyrosequencing (Margeridon et al., 2011).

C316N/H/F substitutions were found to be present at baseline in six subjects infected with HCV genotype 1b who failed treatment. These substitutions were also observed during treatment in one HCV genotype 1a-

infected patient who experienced a relapse. However, more studies are needed to assess the role of this substitution in resistance to SOF (Donaldson et al., 2015).

DSV is a non-nucleoside analogue inhibitor of polymerase I. NS5B amino acid substitutions most commonly observed in genotype 1a-infected patients who do not achieve SVR are M414T and S556G. The NS5B amino acid substitution, most commonly observed in genotype 1b-infected patients who failed to achieve SVR is S556G (Poordad et al., 2014, Zeuzem et al., 2014b, Feld et al., 2014).

1.14.3 Persistence of resistance-associated variants

In contrast to HIV infection, HCV RAVs are not archived as integrated proviral DNA, so the possible long-term effect of selected variants needs further investigation.

Protease inhibitor-RAVs progressively clear and become undetectable by Sanger sequencing within a few months to 2 years after drug discontinuation. The EXTEND study has shown that nearly 90% of drug resistance mutations selected on TVR therapy disappeared within the first two years following the end of treatment. However, minority variants might have been missed in this study (Soriano et al., 2011).

In the long-term follow-up, patients who failed TVR treatment had undetectable RAVs using population sequencing, after 22 months (85%) and 89% after 29 months (Schneider and Sarrazin, 2014, Sherman et al., 2011). A faster disappearance of RAVs in subtype 1b versus 1a variants has been observed (Sullivan et al., 2013). Similar results have been reported with BOC (Pawlotsky, 2011), but faster clearance of RAVs and reversion to the wild-type virus were reported in the majority of patients 6 to 14 months after drug discontinuation (Barnard et al., 2013). RAVs emerging during SMV treatment at NS3 positions 80, 122, 155, and/or 168 were no longer detectable by population sequencing in 50% of patients after a median follow-up of 28.4 weeks (Lenz et al., 2015).

The median time until loss of emerging R155K mutation was 36 and 24 months for genotype 1a and 1b, respectively. Interestingly, the median time to lose the mutation for the R155K variant without baseline Q80K was 64 months compared to 32 months for patients who had emerging R155K and baseline Q80K substitutions (Schneider and Sarrazin, 2014).

In contrast, viruses resistant to NS5A inhibitors are fit and can remain dominant, after they have been selected by a regimen including an NS5A inhibitor (McPhee et al., 2013, Sullivan et al., 2013, Wang et al., 2013, Lenz et al., 2015, Krishnan et al., 2014, Sarrazin et al., 2014).

Pre-existence of NS5A inhibitor-resistant substitutions compromise the response to DCV, and treatment failure is associated with the emergence of both NS5A-L31/Y93 and NS3-D168 variants. NS5A-L31/Y93 variants remain at high frequency 103 to 170 weeks post-treatment (Yoshimi et al., 2015).

1.14.4 Factors influencing resistance before treatment

1.14.4.1 The role of natural occurring RAVs

Naturally occurring RAVs in treatment-naïve patients are known to influence treatment outcome in some patients (Kuntzen et al., 2008). The incidence of resistant variants is variable and depends on the binding domain and HCV genotypes and subtypes present. Nearly all described RAVs within the NS3/ 4A gene can be detected using NGS analysis (Verbinnen et al., 2010).

To date, RAVs at very low frequencies have not been found to have an impact on treatment response. Treatment-naïve patients with TVR-resistant variants before treatment achieved similar SVR-rates compared to patients without RAVs (Bartels et al., 2013).

Further analysis of TVR and BOC phase III studies emphasised that treatment response is independent of the presence of pre-existing RAVs if there is responsiveness to the PegIFN α /RBV backbone.

On the other hand, patients with baseline RAVs, who are also poor PegIFN α /RBV responders (<1 log₁₀ decrease in HCV-RNA during lead-in-phase) showed lower SVR rates compared to poor PegIFN α /RBV responders without baseline RAVs (22% vs. 37%) (Schneider and Sarrazin, 2014). Prior null responders with the pre-existing variants T54S or R155K treated with TVR in the REALIZE-study always failed therapy whereas prior relapsers achieved SVR in most cases (De Meyer et al., 2012).

1.14.4.2 The influence of the HCV genotype/subtype

The phase 3 studies of TVR, BOC, and SMV showed lower SVR rates for HCV-genotype 1a compared to genotype 1b (TVR 71% vs. 79%, BOC 59-62% vs. 66-73%, FDV 69-76% vs. 83-84%, SMV 71% vs. 90% for treatment-naïve patients) (Jacobson et al., 2011, Schneider and Sarrazin, 2014). Pre-existing dominant resistance mutations are more common in treatment-naïve patients infected with genotype 1a (cumulative incidence 8.6% vs. 1.4%) compared to genotype 1b- infected patients (Kuntzen et al., 2008).

1.14.4.3 Cross-resistance

Cross-resistance occurs when resistance mutations are selected that are common to more than one drug within each class. This is typical for inhibitors that bind the same pocket but not necessarily for inhibitors with the same mechanism of action (Halfon and Locarnini, 2011).

The genotyping analysis shows an overlapping cross-resistance profile for PIs. The A156T RAV confers high levels of resistance with reduced viral fitness. It is mainly selected *in vitro*, but also occasionally occurs *in vivo*. The R155K mutation confers low levels of resistance to linear PI compounds but results in high-level resistance to the macrocyclic group and is frequently found *in vivo*. This change is costly for the virus resulting in loss of fitness.

Prevention of DAA-resistant virus outgrowth is based on the use of combinations of potent antiviral drugs with no cross-resistance (Halfon and Sarrazin, 2012).

There is a growing database identifying the profiles of resistance and cross-resistance between antiviral agents; this will be necessary for decisions on treatment choices to achieve the highest cure rate.

1.14.4.4 *In vivo* fitness of viral populations

Replication fitness has been defined as the ability to produce offspring in the setting of natural selection (Richman, 2000). Viral fitness is defined as the replication efficiency of resistant variants, in proportion to the replication efficiency of WT-HCV. The viral fitness of resistant variants influences the emergence of resistance.

The replication capacity of HCV variants is typically assessed *in vitro* using a transient replicon system, or can be examined by comparing colony formation efficiency of the mutant replicon RNA with that of WT variants in co-culture growth competition assays (Bartenschlager and Lohmann, 2000). It is important to consider that replicon-based assays may underestimate the loss of fitness caused by PI-resistance mutations because some mutations in the NS3 protease domain specifically impair late steps in the viral life cycle that involve intracellular assembly of infectious virus (Shimakami et al., 2011).

RAVs must have the capacity to propagate to fill the replication space left vacant by susceptible wild-type virus during drug exposure. The poor viral fitness decreases the clinical significance of highly resistant variants, hence a less resistant but fitter virus affect the outcome of treatment if it can replicate efficiently in the presence of the drug. However, the viral fitness could be restored after the introduction of compensatory mutations that permit the resistant variant to replicate efficiently in the presence of the drug. (Pawlotsky, 2011).

1.14.4.5 Selective pressure of HCV treatments

EC₅₀ (50% effective concentration) and IC₅₀ (50% inhibitory concentration) are used early in the discovery process to evaluate the suitability and performance of drugs.

Drug exposure is the *in vivo* concentration achieved by the administered drug relative to the EC_{50} / IC_{50} values of resistant variants. RAVs will be inhibited if a drug exposure is sufficiently higher than its EC_{50} / IC_{50} values, even if they confer resistance *in vitro*.

Antiviral efficacy *in vivo* may not be affected if a resistant variant naturally replicates at low levels and/or if the drug retains partial efficacy (particularly if drug exposure is high) (Pawlotsky, 2011).

The incidence of resistance is inversely correlated with SVR; the compound with the highest antiviral activity results in the strongest suppression of viral replication, and complete suppression of viral replication prevents resistance emergence because mutagenesis is replication dependent (Halfon and Locarnini, 2011).

Pre-existing RAVs can be selected to outgrow other circulating variants under selective drug pressure, and can become the dominant variant in the viral quasispecies. The identification of the resistance pattern for an antiviral agent is performed by studying the viral quasispecies before, during and after treatment; RAVs are identified by detection of new variants that emerged under drug-selective pressure. This can enable tailoring treatment to avoid treatment failure (Sarrazin and Zeuzem, 2010).

Since monotherapy exerts varying degrees of antiviral activity directed at a single target site, it results in the highest probability of selecting for drug resistance (Sarrazin et al., 2007). The ideal combination therapy exerts antiviral activity at variable steps in the viral life cycle to decrease the possibility of the emergence of RAVs.

Replication space for a virus can be described as the potential of the liver (hepatocytes) to accommodate new transcriptional templates for that virus. This means that the outgrowth of any possible RAV requires the clearance of WT variant and the proliferation of hepatocytes to produce new space for infective RAVs.

Hepatocyte turnover in the normal liver is slow, displaying an average half-life of over 100 days, but this can be reduced to less than ten days in the setting of increased necroinflammatory activity or associated toxicity (Halfon and Locarnini, 2011).

1.15 Prevention of treatment failure with DAAs

In vivo, viral resistance is influenced by three major factors: i) the genetic barrier to resistance; ii) viral fitness of the variant population, and 3) drug exposure (Pawlotsky, 2011). Prevention of DAA-resistant virus outgrowth is based on the use of combinations of potent antiviral drugs with no cross-resistance (Pawlotsky, 2011). Prevention of resistance can be achieved by adopting a strategy with several rules: i) drug choice with the lowest incidence and prevalence of drug resistant mutants in clinical trials; ii) high patient adherence; iii) use combinations of drugs as in the HIV treatment model (Halfon and Locarnini, 2011).

The appearance of RAMs after a few days of treatment limits the usage of DAAs monotherapy because RAMs result in virological rebound and treatment failure. Thus, clinicians should use DAA combinations targeting different viral functions for which there is no cross resistance, an approach that has been successfully applied to prevent resistance in HIV treatment.

1.15.1 HCV drug monitoring resistance tools

Antiviral resistance testing is performed using two methods; genotypic and phenotypic assays. Data from clinical trials carried out so far have indicated that sensitive methods should be adopted to assess HCV drug resistance (Halfon and Locarnini, 2011). NGS is a potential tool for detection of RAVs due to its higher sensitivity in detecting minority variants (Verbinnen et al., 2010).

The FDA has recommended that resistance testing begins in preclinical development to provide a tool for monitoring as drugs are introduced to clinical practice. It is possible that pre-treatment screening for baseline RAMs might be

warranted to enable individual tailoring of treatment. This is more controversial, however than the use of screening for RAVs in those who have already failed DAA treatment (Thompson et al., 2009).

Except testing for baseline Q80K polymorphisms in genotype 1a HCV-infected patients before treatment with SMV, routine testing for RAVs before initial treatment is not currently recommended. Emerging data suggest that assessment for RAVs in patients whose treatment with NS5A-containing regimens failed is warranted for those who require retreatment (AASLD/IDSA, 2015).

Previous reports have described the advantages of NGS, including faster processing and large-scale sequencing, in addition to providing a better understanding of the dynamics of variants in HCV quasispecies (Hiraga et al., 2011, Nasu et al., 2011, Ninomiya et al., 2012). A recent study based on TVR-based therapy showed that it was difficult to predict the emergence of TVR-resistant variants during triple therapy at baseline testing, even with the use of NGS sequencing (Akuta et al., 2013).

1.16 Retreatment of HCV

Several studies demonstrated that the previous exposure to treatment with PegIFN α /RBV does not influence the treatment outcome with IFN-free regimens when compared to treatment-naïve patients (EASL, 2015). Thus, such patients should be retreated with an IFN-free regimen (Zeuzem et al., 2014b, Bourliere et al., 2015).

The experience of retreatment of patients infected with HCV genotype 1 who did not achieve an SVR after treatment with the triple combination of PegIFN α , RBV, and either TVR or BOC with the combination of SOF and SMV, with or without RBV, for 12 weeks is limited to on-going observational real-life cohorts (EASL, 2015). In the TARGET study, the impact of the presence of RAVS before starting retreatment is unknown, but the previous failure of PI-containing regimen was a significant negative predictor of SVR4 (Jensen et al., 2014).

Meanwhile, in the TRIO Network study, retreatment with a combination of SOF and SMV achieved an SVR rate of 82% in patients who failed on triple combination therapy, which was close to SVR rate (80%) achieved in patients who failed on PegIFN α and RBV alone (80%). Retreatment with the combination of PegIFN α /RBV and SOF of such patients yielded SVR rates of 73% and 67%, respectively (Dieterich et al., 2014a).

In non-cirrhotic patients who failed triple combination therapy, 24 weeks of combined SOF and DCV yielded SVR rates of 95% and 100% with and without RBV, respectively (Sulkowski et al., 2014a). In the ION-2 trial, the SVR rates in patients without cirrhosis re-treated with SOF and LDV for 12 weeks, without or with RBV, were 96% and 100%, respectively. Finally, the SVR rate was 97% and 100% after 24 weeks of therapy with and without RBV, respectively (Afdhal et al., 2014a).

Interestingly, in the ION-2 trial, SVR rates in cirrhotic patients retreated with SOF and LDV for 12 weeks, with or without RBV, were 86% and 85%, respectively. SVR rates increased to 100% after prolonged treatment of 24 weeks regardless of using RBV (Afdhal et al., 2014a). In the SIRIUS study, the SVR rates with SOF plus LDV, for either 12 weeks with RBV or 24 weeks without RBV, were 96% and 97%, respectively (Bourliere et al., 2014a).

Clinically resistant HCV variants have been reported with SOF, which rapidly disappear after treatment cessation. Re-treatment options should, therefore, include SOF as re-treatment with 12 weeks of SOF plus LDV with RBV yielded SVR in 98% of genotype 1 patients who failed prior treatment with SOF plus placebo, or SOF plus RBV, or SOF plus PegIFN α -a and RBV (Wyles et al., 2014a).

Patients who fail on the PrOD combination should be re-treated with a SOF-based regimen. The value and safety of re-treatment strategies combining three drugs, including SOF, a protease inhibitor, and an NS5A inhibitor, is unknown.

Current retreatment recommendations are based on indirect evidence (HCV genotype, known resistance profiles of the administered drugs, the number of drugs used, use of RBV, treatment duration, and disease status). The role of resistance testing before retreatment remains unknown, although it is recommended in NS5A treatment failure in the USA (AASLD/IDSA, 2015).

Patients who fail on a DAA-containing regimen should be retreated with an IFN-free combination including an agent with a high barrier to resistance, plus one or two other drugs, ideally with no cross-resistance with drugs previously administered. Based on results in difficult-to-cure patient populations, retreatment should be for 12 weeks with RBV or extended to 24 weeks with or without RBV (EASL, 2015).

1.17 Aims and hypotheses

A cohort of 160 HIV-positive MSM with acute HCV was available to address the following research hypotheses:

- HCV viral diversity predicts the outcome of treatment of early HCV infection in a cohort of HIV-infected individuals.
- Next generation sequencing is superior to conventional Sanger sequencing for detection of low-frequency variants in serum samples and is a better diagnostic tool.
- Natural direct-acting antiviral resistance-associated variants are frequently seen in HIV/HCV co-infected individuals

The aims were to:

Develop a next generation sequencing pipeline for full HCV genome sequencing using the Illumina[®] platform (Chapter 3).

Understand viral dynamics in HCV/HIV co-infected patients who failed treatment including the role of viral diversity in the prediction of treatment outcome (Chapter 4).

Develop a new tool to study natural polymorphisms associated with resistance to new antivirals against HCV (Chapter 5).

Develop an *in vitro* HCV genotype 1a replicon model for studying the replicative fitness of known resistance-associated mutations and to predict *in silico* new mutations conferring resistance to DAAs (Chapter 5).

Chapter 2: Materials and Methods

2.1 Patient cohort

A cohort of 160 patients with HIV and acute HCV was prospectively recruited between 2005 and 2014 in a single centre (St Mary's Acute HCV Cohort). Plasma and peripheral blood mononuclear cell (PBMC) samples were obtained prospectively at 1-3-monthly intervals. Patients were offered 24-48 weeks of treatment with PegIFN α and weight-based RBV (800-1200mg) 12-24 weeks following HCV diagnosis administered according to British HIV Association (BHIVA) guidelines available at the time of treatment (Brook et al., 2010) unless this was contraindicated or spontaneous clearance occurred.

Informed consent in writing was obtained from each patient, and the study protocol conformed to the ethical guidelines of the 1975 Declaration of Helsinki. This work was conducted as part of a study approved by the St Mary's Hospital Research Committee and the Riverside Research Ethics Committee (Reference number 05/Q0401/17) granted ethical approval. Each patient was assigned a study number starting with letter (P) followed by a unique number (e.g. P101). Blood samples were obtained in EDTA-containing vacutainer tubes. Plasma was separated by centrifugation and stored in 2ml tubes at -80°C.

2.1.1 Clinical groups and definitions

Acute HCV diagnosis was defined as a positive HCV RNA result by RT-PCR within six months of a preceding negative RT-PCR or an antibody test. The date of infection was estimated as the midpoint between the last negative and first positive test.

2.1.1.1 Patient groups based on viral load monitoring

Samples for measuring HCV viral load and liver function (ALT, bilirubin, and albumin) were taken at monthly intervals for three months following diagnosis and at three monthly intervals thereafter. CD4 count and HIV viral load were assessed at three monthly intervals.

Spontaneous clearance of HCV was defined as two successive negative RT-PCR tests (<12 IU/ml) 3 months apart. **Sustained virological response (SVR)** was defined as a negative HCV RNA by RT-PCR (<12 IU/ml) 24 weeks after the end of treatment with PegIFN α and RBV. Patients who did not respond to treatment were divided by mode of failure. **The null response** was defined as a decrease of less than 2 log₁₀ IU/ml in HCV viral load at week 12 of therapy. **Partial response** was defined as a reduction of at least 2 log₁₀ IU/ml in HCV RNA by week 12 of therapy but detectable HCV RNA at week 24. **Viral breakthrough** was defined as the recrudescence of HCV RNA at any time during treatment after a negative result. **Relapse** was defined as detectable HCV RNA within 24 weeks after the end of treatment following negative results on treatment (EASL, 2014).

2.1.1.2 Treatment outcome groups based on phylogenetic analysis

Sequences obtained from paired samples pre- and post-treatment were considered similar or different based on two criteria: i) **Phylogenetic signal** defined as a monophyletic or non-monophyletic lineage; ii) **Genetic distance** of greater than 10% between sequences; this was calculated as the pairwise distance between aligned sequences. Based on these criteria, treatment outcomes were defined as i) *Persistent infection*: the presence of at least one variant present in the pre-treatment sample persisting after treatment. This could be associated with new dominance; the outgrowth of a minority strain found in the pre-treatment sample or the presence of a new variant representing superinfection; ii) *Reinfection*: the presence of a new variant(s) in the post-treatment sample with no evidence of similar pre-existing strains (Abdelrahman et al., 2015).

2.1.2 Treatment outcome subgroups

The first study subgroup included ten patients with an SVR matched with ten patients who did not respond to treatment. This subgroup was used to investigate the role of viral diversity in the prediction of treatment outcome as detailed in Section 4.2.1.

A second subset (carried out later in the study) analysis of 15 treatment non-responders was performed to look for evidence of mixed infection before and after treatment. Of these, six were null responders, three partial responders and six were relapsers. Paired samples from each patient pre- and post-treatment were analysed as described in Section 4.2.2.

2.1.3 Antiviral resistance study cohort

Serum samples were obtained from 16 HCV genotype 1a co-infected treatment-naïve HIV-positive patients in the St Mary's Acute HCV cohort (Thomson et al., 2011) and 18 HCV genotype 1a treatment naïve patients in the STOP HCV project. The samples obtained from STOP HCV project were sequenced by Dr Chris Hinds and Dr Walt Adamson and the sequencing data were provided as raw data. Local Research Ethics Committees granted ethical approval.

2.2 Extraction of viral RNA from plasma samples

Plasma stored at -80°C was thawed on ice and nucleic acid extraction performed using two different methods; manual extraction using a QIAamp[®] Viral RNA kit (Qiagen) and automated extraction using magnetic bead technology (EasyMAG[®]). Automated nucleic acid extraction was selected once available to improve workflow and decrease variability.

2.2.1 Manual RNA extraction

Manual RNA extraction was carried out with a QIAamp[®] Viral RNA kit (Qiagen). An aliquot (140 µL) of plasma was added to 560 µL cell lysis solution (buffer AVL) containing 5.6 µL carrier RNA and incubated for 10 min at room temperature. Ethanol (560 µL) was added, and the sample applied to columns and centrifuged at 8,000g. Two further centrifugation steps were carried out using 500 µL of wash buffers AW1 and AW2. Finally, extracted viral RNA was eluted in 60 µL of buffer AVE (RNase-free water containing 0.04% sodium azide).

2.2.2 Automated RNA extraction

Extraction with the easyMAG[®] NucliSENS extractor platform was carried out according to manufacturer's recommendations; 400 μ L of each sample was placed in a disposable sample vessel that was loaded onto the extractor. After initial lysis incubation, 100 μ L of magnetic silica was added to each sample, and the extractor was restarted. Samples were eluted in 60 μ L. All samples were transferred to a 1.5 ml microcentrifuge tube and stored at -80°C .

2.3 cDNA Synthesis

cDNA synthesis was carried out using Superscript III or Maxima H kits.

2.3.1 Superscript III[®] reverse transcriptase

A Superscript III RT-PCR kit (Invitrogen, Life Technologies, Carlsbad, CA) was used to reverse transcribe extracted RNA into complementary DNA (cDNA) using random hexamers. 15 μ L of RNA, 1 μ L of random hexamers and 9 μ L of 10mM dNTPs were mixed and heated at 70°C for 5 min. The mixture was then placed on ice for 2 min. A reverse transcription master mix was made by adding 9 μ L of 5X First-strand Buffer, 2 μ L of 0.1M DTT, 4 μ L of RNase OUT (20u/ μ L) and 1 μ L of Superscript III enzyme. This mixture was added to the previous reaction mix and incubated at 50°C for 1.30 hrs. Following this, 1 μ L of Superscript III was added and the temperature increased to 55°C for a further 1.30 hrs. The reverse transcriptase enzyme was inactivated at 70°C for 15 min, then 1 μ L of RNase H was added and incubated for 20 min at 37°C . Finally, cDNA was stored at -20°C .

2.3.2 Maxima H Minus Reverse Transcriptase[®]

Complementary DNA was also prepared using Maxima H minus Reverse Transcriptase[®] (Thermo Scientific). A master mix was prepared using random hexamers and nucleoside triphosphates (NTPs); 1 μ L of random hexamers, 1 μ L of NTPs were added to 13 μ L of total RNA, then incubated at 65°C for 5 min.

The reaction was chilled on ice, briefly centrifuged and placed back on the ice. Four μL of 5x Reverse Transcriptase buffer, 1 μL Reverse Transcriptase enzyme were added and then incubated for 10 min at 25°C followed by 60 min at 65°C. Finally, the reaction was terminated by heating at 85°C for 5 min. The cDNA was stored at -20°C.

2.4 DNA second strand synthesis

Double stranded cDNA was prepared with a NEBNext[®] mRNA Second Strand Synthesis Module kit (New England Biolabs). 15 μL from the inactivated cDNA reaction was placed on ice and 48 μL nuclease-free water, 8 μL of 10X Second Strand Synthesis Reaction Buffer and 4 μL Second Strand Synthesis Enzyme Mix were added. The reaction was then incubated in a thermal cycler for 2.5 hours at 16°C.

2.5 PCR amplification

Universal precautions to avoid PCR contamination were followed at all times and PCR mixes were prepared in an allocated room separate from clinical samples.

2.5.1 PCR error rate estimation using different enzymes

Error rates of several PCR enzymes were estimated using clonal sequence analysis of a single plasmid containing a single HVR1 region insert. Bacterial cloning was carried out as described in Section 2.11.2. The enzyme that demonstrated the highest fidelity was used to perform all further PCR reactions. The error rate was calculated by counting all nucleotide variants of plasmid reads in the alignment that did not correspond to the sequence of the clone determined by Sanger sequencing. The three different enzymes tested in this study were; AccuStart[®] Taq DNA Polymerase HiFi and the proof-reading enzymes Phusion[®] High-Fidelity PCR enzyme and KOD Hot Start DNA Polymerase[®]. These were compared with an error rate of MegaMix[®] Blue PCR Taq polymerase estimated by Dr Emma Thomson using the same method.

2.5.1.1 MegaMix[®] Blue PCR Taq polymerase

MegaMix[®] Blue (Microzone) contains recombinant Taq polymerase in reaction buffer (2.75 mM MgCl₂ with 220 μM dNTPs, blue agarose loading dye & stabiliser). Cycling conditions were as follows: initial denaturation step: 95°C for 3 minutes (min) followed by 30 cycles of 1. Denaturation: 95°C for 30 seconds (sec), 2. Annealing: Optimal primer annealing temperature for 30 sec, 3. Extension: 72°C for 60 sec and a final extension at 72°C for 10 min then hold at 4°C.

2.5.1.2 Phusion[®] High-Fidelity PCR enzyme

Phusion[®] High-Fidelity PCR Master Mix with HF Buffer (New England Biolabs) contains a high fidelity enzyme. 25 μL of 2X Phusion Master Mix, 2.5 μL of 10 μM Forward Primer, 2.5 μL of 10 μM Reverse Primer, DNA template, and nuclease-free water were added to produce a final volume of 50 μL. Cycling conditions were as follows: initial denaturation step: 98°C for 2 min followed by 30 cycles of 1. Denaturation: 98°C for 10 sec, 2. Annealing: Optimal primer annealing temperature for 30 sec, 3. Extension: 72°C for 15-30 sec per kb, and a final extension at 72°C for 10 min, then hold at 4°C.

2.5.1.3 KOD Hot Start DNA Polymerase[®]

KOD Hot Start DNA Polymerase[®] (Novagen) is a premixed complex of the high fidelity KOD DNA Polymerase and two monoclonal antibodies that inhibit the DNA polymerase and 3'→5' exonuclease activities at ambient temperatures.

The PCR reaction contains 5 μL of 10X Buffer for KOD Hot Start DNA polymerase, 3 μL of 25 mM MgSO₄, 5 μL of dNTPs (2 mM each), 1.5 μL of 10 μM Forward Primer, 1.5 μL of 10 μM Reverse Primer, DNA template, 1 μL KOD Hot Start DNA Polymerase (1U/μL) and nuclease-free water, resulting in a final volume of 50 μL. Cycling conditions were as follows: polymerase activation step: 95°C for 2 min followed by 30 cycles of 1. Denaturation: 95°C for 20 sec, 2. Annealing: Optimal primer annealing temperature for 10 sec, 3. Extension: 70°C for 10 sec per kb, and a final extension at 72°C for 10 min, then hold at 4°C.

2.5.1.4 AccuStart® Taq DNA Polymerase HiFi

AccuStart® Taq DNA Polymerase HiFi is an enzymatic mixture of recombinant Taq DNA polymerase, a thermostable DNA polymerase with 3'→5' exonuclease activity, and monoclonal antibodies that bind to the polymerase and keep it inactive before PCR thermal cycling.

The PCR reaction contains 5 µL of 10x OptiBuffer, 2 µL of MgCl₂ (50 mM Solution), 2.5 µL of 20 mM dNTP Mix, 1.5 µL of 10 µM Forward Primer, 1.5 µL of 10 µM Reverse Primer, DNA template, 2 µL of DAp GoldStar® DNA polymerase (4 U/ µL) and nuclease-free water resulting in a final volume of 50 µL.

Cycling conditions were as follows: polymerase activation step: 95°C for 2 min followed by 30 cycles of 1. Denaturation: 95°C for 20 sec, 2. Annealing: Optimal primer annealing temperature for 10 sec, 3. Extension: 70°C for 60 sec per kb, and a final extension at 72°C for 10 min, then hold at 4°C.

2.5.2 Amplification of the E2 HVR1 region

Amplification of a 220 bp region including the E2 hypervariable region-1 (HVR1) was carried out by nested PCR using a combination of genotype-specific primers as shown in Table 2-1 (Thomson et al., 2011). The first-round PCR products were further amplified with “fusion primers” composed of the 454 primer keys, with different multiple identifiers for each sample, and the HCV HVR1-specific primers. Primer binding sites, adapters and multiplex identifiers (MIDs) for sample barcoding were included in the inner primer design to create fusion primers compatible with pyrosequencing (Figure 2-1). The outer primer set contained genotype-specific versions of primers E4 and E5 and the inner primer set contained genotype-specific 214 and E3 primers.

The resulting PCR products were purified using 2% agarose gel electrophoresis containing SYBR safe® DNA gel stain (Life Technologies) with lane markers and a 100bp small fragment ladder® (Thermo Scientific). DNA bands were visualised under ultraviolet light and bands of appropriate size (225 bp)

were extracted and purified using a GeneJet[®] extraction kit (Thermo Scientific). Purified amplicons were quantified using Qubit[®] fluorometric quantification. Amplicons were diluted to create a multiplexed library in equimolar concentrations.

The library was sent for Roche 454 FLX second-generation pyrosequencing using Titanium chemistry (Roche Applied Science, Indianapolis, IN) by Beckman Coulter Genomics, USA.

2.5.3 Quantitative real-time PCR assays

Real-time PCR (RT-PCR) reactions were amplified under universal conditions on a 7500 Fast Real-Time[®] PCR machine and data analysed using 7500 Fast System Software (SDS v1.3.1, ABI). The RT-PCR reaction mix total volume was 18 μ L, prepared using: 1 μ L of 18 μ M forward primer; TCTGCGGAACCGGTGAGTAC (final concentration 900nM), 1 μ L of 18 μ M reverse primer; GCACTCGCAAGCACCTATC (final concentration 900 nM), 1 μ L of 5 μ M FAM Probe; FAM-AAAGGCCTTGTGGTACTG-MGB (250nM), 10 μ L of TaqMan Fast Universal Mix[®] (2x), 5 μ L of nuclease-free water and 2 μ L cDNA in a 96 well plate. The plate was then sealed with an adhesive cover and centrifuged for one minute at 1000 RPM.

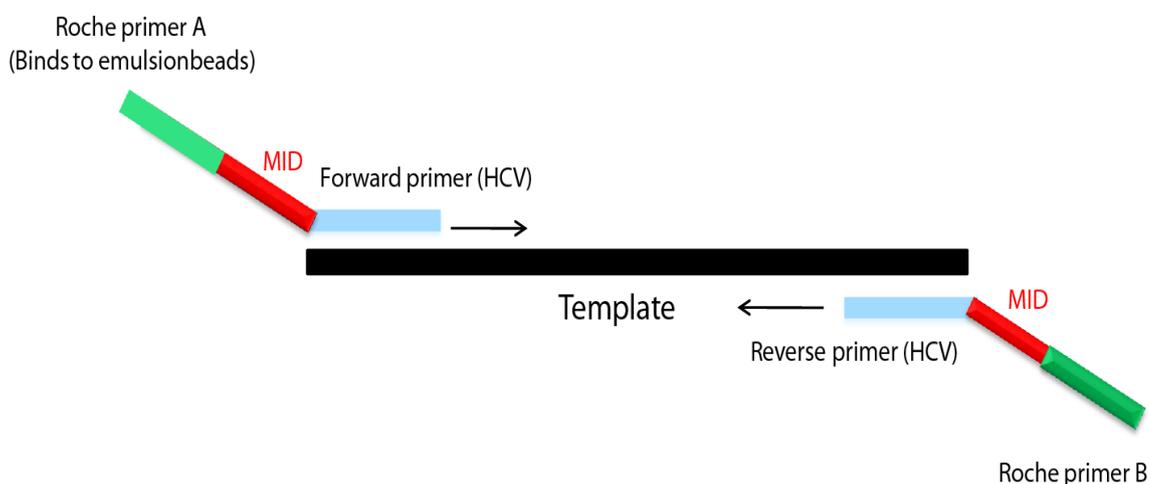
The following thermal cycler protocol cycling conditions were used: Stage 1: hold at 95°C for 20 sec; stage 2 (40 cycles): step 1: hold at 95°C for 3 sec, step 2: hold at 60°C for 30 sec.

Table 2-1: Genotype-specific primers used for nested PCR.

Primer	Primer sequence	Nucleotide position, relative to H77
214	5'-CACTGGGGAGTCCTGGCGGGC-3'	1395-1415
214(2)	5'-CAYTGGGGYGTSRTGTTYGGC-3'	1395-1415
E3	5'-GGGCAGTGCTGTTGATGT-3'	1603-1620
E3(1a)	5'-AGGCCGTGCTATTGATGT-3'	1603-1620
E3(2)	5'-GGCMGTSCGGTTKATGTGCC-3'	1604-1623
E3(4)	5'-GGGCAGTCCTATTTATATGCC-3'	1603-1623
E4	5'-GGTGTGGAGGGAGTCATTGCAGTT-3'	1623-1646
E5	5'-GCTTGGGATATGATGATGAACTGGTC-3'	1296-1321

The outer primer set (E4, E5), the inner primer set (214, E3).

Figure 2-1: Design of fusion primers.



454 primer keys, containing different multiple identifiers for each sample and HCV HVR1-specific primers. Primer binding sites, adapters and multiplex identifiers (MIDs) for sample barcoding were included in the inner primer design to create fusion primers compatible with pyrosequencing.

2.6 Gel electrophoresis

This method was employed to purify DNA fragments produced by PCR or following restriction enzyme digestion. PCR products were mixed with 0.1 volumes of 10X BlueJuice[®] Gel Loading Buffer (Invitrogen) before being loaded into the wells of a 1% or 2% agarose gel containing Syber Safe[®] (Invitrogen) or ethidium bromide (1µg/ml) with lane markers and a 100bp or 1Kb ladder (Promega) depending on the size of fragments that needed to be excised.

Gels were prepared using 1 x TAE buffer (a buffer solution containing a mixture of Tris base, acetic acid, and EDTA). Gels were typically run at 110 V; DNA bands were visualised under ultraviolet light and bands of appropriate size were cut out using a clean scalpel blade. These bands were next purified to remove extraneous products.

2.7 Purification of DNA

2.7.1 Isolation and Purification of DNA from Agarose Gels

DNA was purified from excised gel slices using a GeneJet[®] gel extraction kit (Thermo Scientific). The gel slice was placed in a pre-weighed 1.5 ml tube and weighed, then a 1:1 volume of binding buffer added to the gel slice (volume: weight) and the gel mixture incubated at 60°C for 10 min or until the gel slice was completely dissolved. The chaotropic agent in the binding buffer dissolves agarose denatures proteins and promotes DNA binding to the silica membrane in the column. 800 µL of solubilised gel solution at a time was transferred into a purification column and centrifuged for 1 min at 13,000 rpm in a microcentrifuge. The flow-through was discarded, and the column placed back into the same collection tube. 100 µL of binding buffer was next added to the purification column and centrifuged for 1 min at 13,000 rpm.

The flow-through was discarded, and the column placed back into the same collection tube. 700 µL of wash buffer was added to the purification column and centrifuged for 1 min. The flow-through was discarded, and the column placed

back into the same collection tube. The empty purification column was centrifuged for an additional 1 min to remove residual wash buffer completely. For elution, 50 μL of elution buffer was added to the centre of the purification column membrane, the purification column transferred into a clean 1.5 ml microcentrifuge tube and centrifuged for 1 min. The purified DNA was then stored at -20°C .

2.7.2 DNA purification using Agencourt AMPure XP[®] beads

The volume of Agencourt AMPure XP[®] for a given reaction was calculated using the following equation: *(Volume of Agencourt AMPure XP per reaction) = 1.8 x (Reaction Volume)*, the constant (1.8) was changed according to the size selection protocol. The sample was mixed by pipetting up and down ten times; then it was placed on a magnetic plate for 5-10 min to separate the beads from the solution. The cleared supernatant was aspirated from the reaction plate and discarded. 200 μL of fresh 80% ethanol was used to wash the beads twice without disturbing the beads. Finally, purified DNA was eluted in 40 μL nuclease-free water (Figure 2-2).

2.8 Fluorometric measurement of nucleic acid concentration

A fluorescent dye working solution was prepared in a plastic tube, using a mix of 200 μL of buffer and 1 μL of dye for every sample. After vortexing, an aliquot of 190 μL of working solution was added to two 0.5 ml Eppendorf tubes and are labelled standards, then 10 μL of each Standard 1 (0 ng/ μL) and standard 2 (10 ng/ μL) was added to the corresponding tube and mixed by vortexing. Meanwhile, a 198 μL aliquot of working solution in 0.5 ml Eppendorf tubes was mixed with 2 μL of each sample by vortexing.

After incubation at room temperature for 2 min, the nucleic acid concentration was measured using a Qubit[®] 2.0 Fluorometer; readings were always carried out following calibration with two standards. To measure DNA quantity, a Qubit[®] ds DNA HS Assay Kit was used, while for RNA quantification, a Qubit[®] RNA Assay Kit was used.

2.9 Nucleotide sequencing and analysis

2.9.1 Sanger sequencing and analysis

Sanger sequencing was performed commercially by Beckman Genomics, UK using BigDye[®] version 3.1 chemistry (Applied Biosystems, Carlsbad, CA) using a corresponding forward or reverse primer. Chromatograms were checked for miscalled nucleotides by visual inspection of chromatograms using BioEdit v7.1.3 software. Sequences were aligned using MUSCLE (Edgar, 2004), and maximum likelihood phylogenetic trees constructed using MEGA 6 (Tamura K et al., 2011). Trees were generated following gap exclusion and corrections for multiple substitutions using the best-fit substitution model for the data set detected by MEGA 6 (Kimura, 1980).

The statistical robustness and reliability of the branching order within each phylogenetic tree were confirmed by bootstrap analysis using 1000 replicates. Bootstrap values >70% were considered to be reliable.

2.9.2 454 Pyrosequencing analysis

Amplicon sequencing was performed on a 454/Roche GS FLX platform following manufacturer's instructions. Each sample was amplified independently with fusion primers, including 454 primer keys (A and B for forward and reverse primers, respectively), a different multiple identifier (MID) for each sample and HCV-specific primers.

454 pyrosequencing reads were de-multiplexed using a custom Perl script that identified the forward and reverse barcodes allowing a single mismatch in the reverse barcode. Each read was compared to a reference set of sequences from the Los Alamos HCV database and quality checked by comparing it in a pairwise alignment to the best reference match (Kuiken et al., 2008). The quality of the reads was assessed by a Phred quality score which is a measure of the quality of the identification of the bases generated by automated DNA sequencing (Table 2-2). Phred quality scores are defined as a property which is logarithmically related to the base-calling error probability (Ewing and Green,

1998, Ewing et al., 1998). A read was excluded from the dataset if it had mutations relative to the reference below a Phred score of 25, and if there was only a single copy of the read, thus keeping a conservative set of sequences with high-quality scores.

All scripts were created by Dr Joseph Hughes and are available on Github (<https://github.com/josephhughes/HCVtoolbox>). The sequences obtained from each patient sample were then aligned against a complete reference set of HCV sequences using MAFFT (Katoh and Standley).

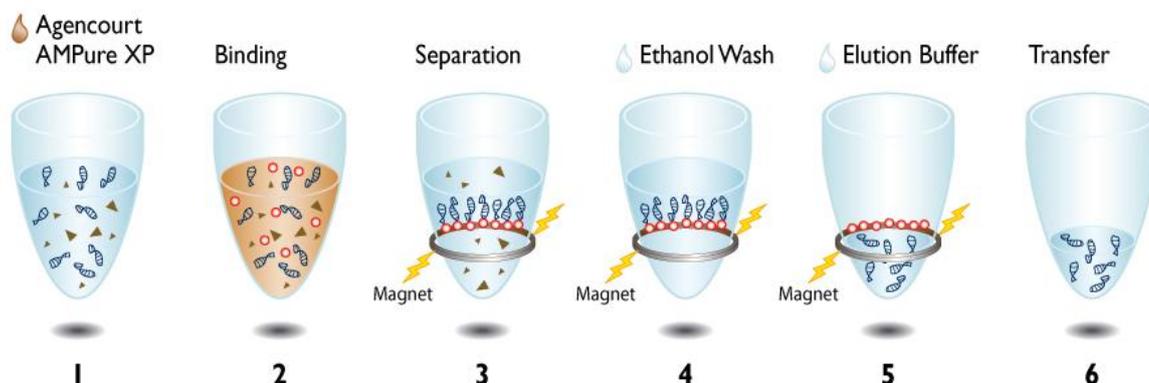
All valid reads were clustered using CD-HIT with a parameter of similarity of 90% to assign different variants detected in each sample (Fu et al., 2010). These variants were aligned with post-treatment variants that had been detected using clonal analysis as well as reference sequences of different genotypes from Los Alamos HCV database using MUSCLE (Edgar, 2004).

Maximum likelihood phylogenetic trees were constructed using MEGA 6 (Tamura K et al., 2011). All sequences generated were submitted to the European Nucleotide Archive (ENA; study accession PRJEB4613).

2.9.3 Illumina[®] sequencing and analysis

For the preparation of libraries for Illumina[®] deep sequencing, pooled amplicons in equimolar concentration were generated using either a Nextera XT[®] DNA sample preparation kit (Illumina[®]) or a KAPA HiFi Real-Time PCR Library Amplification[®] and Index kit (Illumina[®]) according to modified versions of the manufacturer's protocol. The optimisation of the sequencing platform was carried out in High Throughput Sequencing (HTS) facility in MRC- University of Glasgow Centre for Virus Research in collaboration with Dr Gavin Wilkie.

Figure 2-2: Workflow for PCR Purification using Agencourt AMPure XP® beads.



1. AMPure XP. 2. Bind DNA fragments to paramagnetic beads. 3. Separation of beads + DNA fragments from contaminants. 4. Wash beads + DNA fragments twice with 70% Ethanol to remove contaminants. 5. Elute purified DNA fragments from beads. 6. Transfer to a new tube (Source: Beckman Coulter user guide 2013).

Table 2-2: Phred quality scores.

Phred Quality Score	Probability of incorrect base call	Base call accuracy
10	1 in 10	90%
20	1 in 100	99%
30	1 in 1000	99.90%
40	1 in 10,000	99.99%
50	1 in 100,000	100.00%
60	1 in 1,000,000	100.00%

The scores are logarithmically linked to error probabilities.

2.9.3.1 KAPA[®] HiFi Real-Time PCR Library Amplification kit

PCR amplicons were washed, purified and size-selected using Agencourt Ampure XP[®] beads (Beckman Coulter) and subsequently quantified on a Qubit[®] 2.0 Fluorometer using a Qubit[®] dsDNA High-Sensitivity kit (Invitrogen, Life Technologies). Library preparation including end repair, A-tailing, and adapter ligation was performed using the KAPA[®] HiFi Real-Time PCR Library Amplification kit for Illumina[®] libraries (KAPA Biosystems), with wash and purification steps, carried out using Agencourt Ampure XP[®] beads.

The A-tailed PCR amplicons were ligated to NEBNext[®] Multiplex Oligos for Illumina[®] (New England Biolabs) using a strict 10:1 molar ratio of adapter to PCR amplicon. End-repair of the forked adapters was carried out using either Index Primer Set 1 or 2 (New England Biolabs) and the qPCR reaction was stopped between fluorescent standards 2 and 3 as recommended in the KAPA HiFi Real-Time PCR Library Amplification kit.

Adapter ligated products were size-selected using Ampure XP[®] beads and visualised on an Agilent 2200 TapeStation system using a D1K ScreenTape[®] (Agilent Technologies). The adapter-ligated libraries were then quantified using the KAPA SYBR[®] FAST ABI Prism qPCR library quantification kit (KAPA Biosystems). A library with equimolar concentrations of different samples was then processed on the MiSeq platform.

2.9.3.2 Nextera XT[®] DNA sample preparation

The Nextera XT[®] DNA Sample Preparation kit contains a synthetically engineered transposome, which is used to cut DNA strands into an average of 300 base pair (bp) fragments and simultaneously tag the DNA with adapter sequences. The adapter sequences contain binding sites for dual index sequences unique to each sample library as well as two flow cell attachment sites (Figure 2-3).

➤ Input DNA Quantitation and tagmentation

The Nextera XT[®] DNA Sample Preparation library preparation procedure utilizes

a biological DNA fragmentation step and is a more sensitive way to fragment DNA than mechanical fragmentation methods. In addition, only 1 ng input DNA is needed. A fluorometric based method specific for duplex DNA (Qubit[®] dsDNA BR Assay) was used for accurate quantification of the DNA library as detailed in Section 2.8. During the tagmentation step, DNA is “tagmented” (tagged and fragmented) using the Nextera XT[®] transposome. The transposomes fragments input DNA and add adapter sequences to the ends of each DNA fragment. The DNA fragments are amplified by non-specific PCR in subsequent steps.

First, 10 μL of Tagment DNA Buffer (TD) Buffer was added to a 0.2 ml tube, then 5 μL of input DNA at 0.2 ng/ μL (1 ng total) added to each tube, followed by 5 μL of Amplicon Tagment Mix (ATM). Next, the reaction was mixed and centrifuged at 280g for 1 minute at 20°C.

Finally, the reaction was placed in a thermocycler as follows: 55°C for 5 min, then held at 10°C. Once at 10°C, neutralization was started by adding 5 μL of Neutralisation Tagment Buffer (NT) to each reaction, mixed gently and centrifuged at 280g for 1 minute. The samples were left at room temperature for 5 minutes.

➤ PCR Amplification:

During this step, tagmented DNA is amplified via a limited-cycle PCR program. In addition to the PCR step, index 1 (i7) and index 2 (i5) and sequences required for cluster formation are added. After thawing the Nextera PCR Master Mix (NPM) and the index primers at room temperature for 30 min, 15 μL of NPM was added to each reaction tube containing 5 μL of each index and mixed gently.

Next, the reaction was centrifuged at 280g at 20°C for 1 minute, PCR was performed in a thermocycler using the following protocol: 72°C for 3 min, 95°C for 30 sec, then 12 cycles of: denaturation at 95°C for 10 sec, annealing at 55°C for 30 sec, extension at 72°C for 30 sec, followed by 72°C for 5 min and a final hold at 10°C.

➤ **PCR Clean-Up**

In this step, the amplified DNA was purified using AMPure XP magnetic beads as described in Section 2.7.2.

2.9.3.3 Sequence library quality control

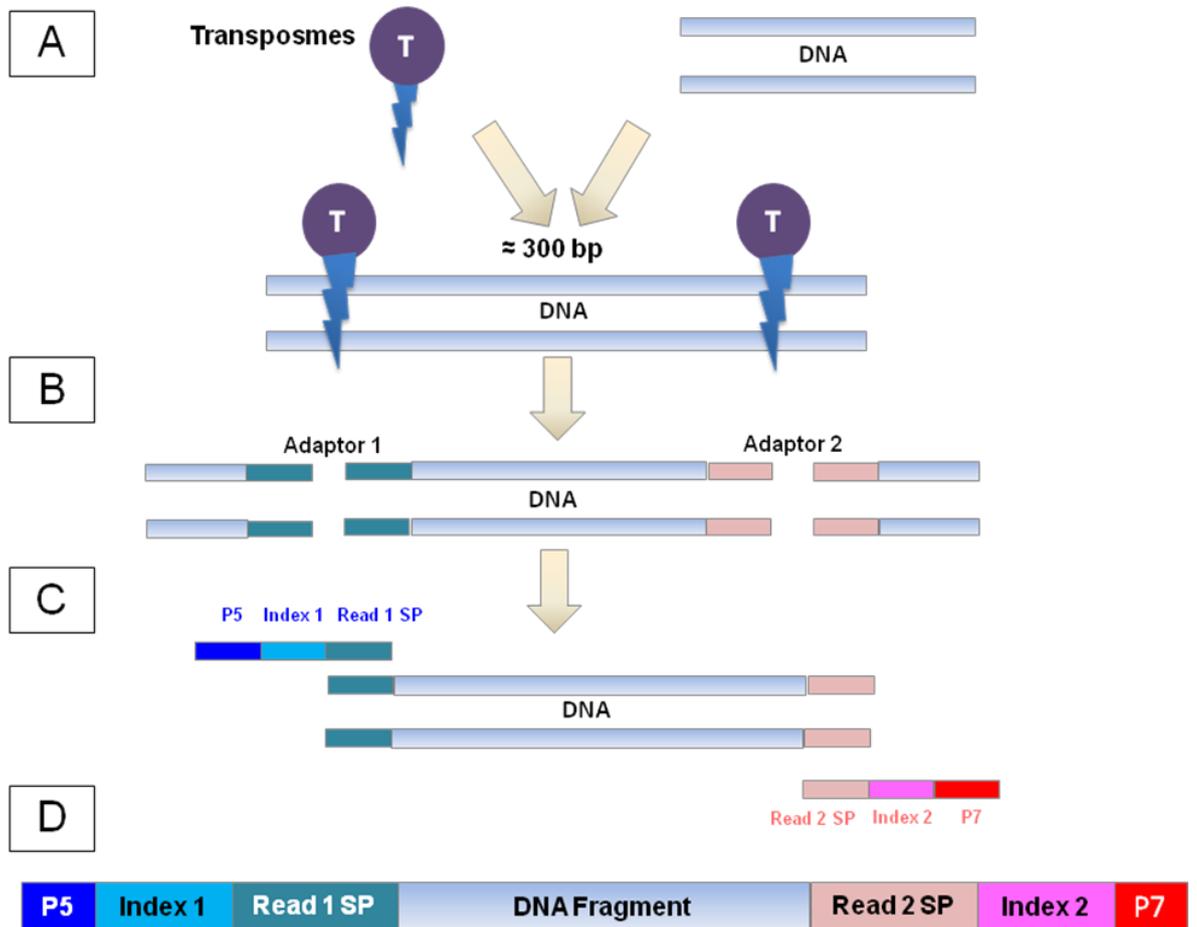
The Agilent 2200 TapeStation[®] system (Agilent Technologies) automates RNA and DNA sample quality control (QC), including sample loading, separation, and imaging. Different screen tapes are available for RNA and DNA analysis performed on an automated gel electrophoresis system.

For RNA analysis (post-extraction), 1 μ L of sample buffer R6K was added to 2 μ L of RNA, incubated at 72°C for 3 min, placed on ice for 2 min and loaded into TapeStation[®] microtubes. DNA quality and size were tested using a similar technique; 2 μ L of DNA was added to 2 μ L of D1K sample buffer and loaded into the TapeStation.

2.9.3.4 Illumina[®] sequencing using the MiSeq platform

The sequence of the DNA strands was determined using a four colour cyclic reversible termination technique, based on the use of dye-labelled modified nucleotides. The colour of the dye, and hence the type of nucleotide, is detected via total reflection fluorescence (TRIF) imaging using two lasers. The flow cell attachment sites randomly bind to complementary oligonucleotides on the flow cell. This random ligation with either P5 or P7 attachment sites allows for paired-end sequencing (i.e. both in forward and reverse direction). The following process of bridge amplification causes the formation of amplified clusters of single-stranded DNA (Figure 2-4).

Figure 2-3: Nextera XT workflow.



A) Nextera XT transposome with adapters is combined with template DNA, B) Tagmentation to fragment and addition of adapters, C) Limited cycle PCR to add sequencing primer sequences and indices; SP: sequencing primer. D) Sequence-Ready Fragment.

➤ Denaturation of library

To denature the samples, 1 ml of 0.2 N NaOH was prepared in a 1.5 ml eppendorf tube by mixing 800 μL of laboratory-grade water (Qiagen) and 200 μL from 1.0 N Stock NaOH (Sigma, UK). The tube was inverted several times to mix. Following that, 10 μL of 4nM sample DNA and 10 μL of 0.2 N NaOH were combined in a microcentrifuge tube.

The sample was mixed by vortexing followed by centrifugation for 1 minute and incubated for 5 min at room temperature to allow the denaturation of DNA into single strands. Next, 980 μL of pre-chilled hybridization buffer (HT1) was added to 20 μL denatured DNA to result in a 20 pM denatured library in 1ml of 0.2M NaOH and placed on ice. The final DNA concentration was 12 pM, using 360 μL of 20 pM of denatured DNA and 240 μL of HT1. The final reaction was mixed by inverting the tube several times followed by centrifugation. The denatured and diluted DNA was placed on ice until it was loaded into the MiSeq reagent cartridge.

➤ Phi X control

A Phi X 174 (phiX) bacteriophage genome library provided a quality control for cluster generation, sequencing, and alignment and a calibration control for cross-talk matrix generation, phasing, and prephasing. It can be rapidly aligned to estimate relevant sequencing by synthesis (SBS) metrics such as phasing and error rate. Phi X control DNA (Illumina[®], UK) was diluted to 4 nM by adding 2 μL of a 10nM stock solution to 3 μL H₂O.

The Phi X library was denatured by adding 5 μL of 0.2 N NaOH to 5 μL Phi X, then vortexed and incubated at room temperature for 5 min. This was then diluted to a concentration of 20 pM by adding 980 μL HT1 buffer and diluted again to the same loading concentration as the DNA library. 100 μL of Phi X was added to 90 μL DNA library to provide a final library with a 10% concentration of Phi X.

➤ **Loading the flow cell and buffer.**

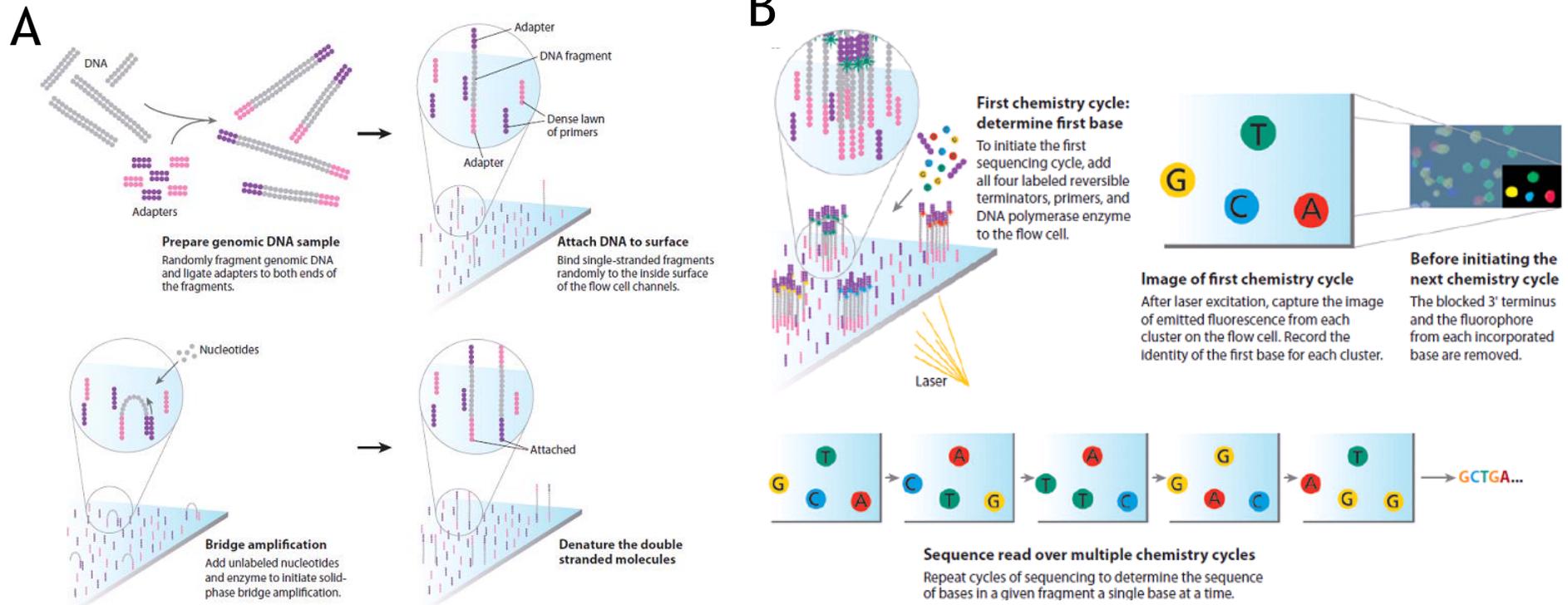
The flow cell (provided with the MiSeq reagent kit) was removed from its buffer and washed with distilled H₂O. It was dried with a lint-free wipe, ensuring that the inlet and outlet ports were clear, and the surface was clean and dust-free. The flow cell was loaded onto the MiSeq instrument. The chiller compartment was opened and wash bottle removed and replaced with incorporation buffer (PR2). The reagent cartridge was loaded into the machine, and the run started.

2.9.3.5 Bioinformatic analysis of Illumina® sequence data

Sequence reads were demultiplexed using the unique dual indexes attached to either end of each read and checked for quality using FastQC® (Babraham Bioinformatics). Reads below a Phred quality score of Q20 (99% accuracy) were discarded and the remaining reads then trimmed to remove low-quality areas towards the end of the sequence, leading to a final quality score of Q30 (99.9%) or above using the Phred score algorithm. Remaining reads were then trimmed to remove low-quality areas towards the end of the sequence.

Sequence data were uploaded in fastq format and mapped to consensus reference sequences on a dedicated computer server. Full-length genomes were assembled using an in-house alignment program based on the BLAST algorithm (Tanoti) developed by Dr Sreenu Vattipally. Alignments were visualised using Ugene® and regions of interest including resistance mutation sites were annotated. Variant frequencies were calculated using an in-house script (Mutation report) designed by Dr Sreenu Vattipally.

Figure 2-4: Workflow of Illumina sequencing.



Cluster strands created by bridge amplification are primed and all four fluorescently labeled, 3'-OH blocked nucleotides are added to the flow cell with DNA polymerase. B) Optics system scans each lane of the flow cell by imaging units called tiles. Reproduced with permission from Dr Elaine Mardis (Mardis, 2008).

2.10 Full genome sequencing

Two protocols for full genome sequencing of HCV were developed using the Illumina[®] platform; a PCR-based (Amplicon) and a metagenomic approach.

2.10.1 Nested PCR-based full genome sequencing of HCV

To maximize PCR sensitivity, the genome was divided into four amplicons that were numbered sequentially 1 to 4 starting from the 5' end of the genome. Each amplicon was less than 3 kb and overlapped with adjacent amplicons (Figure 2-5).

2.10.1.1 Primer nomenclature and design

Primer nomenclature included the amplicon number A (1-4), polarity (F-R) and relative position on the H77 reference genome. Genotype 1a specific primers were designed that were effective for amplifying all amplicons in a reproducible manner; the design of nine primers was based on those published previously either completely matched or modified (Yao and Tavis, 2005).

We optimized more than one combination for each amplicon. The three anti-sense primers must reside 3' to all three sense primers for the downstream amplicon to prevent gaps between the amplicons.

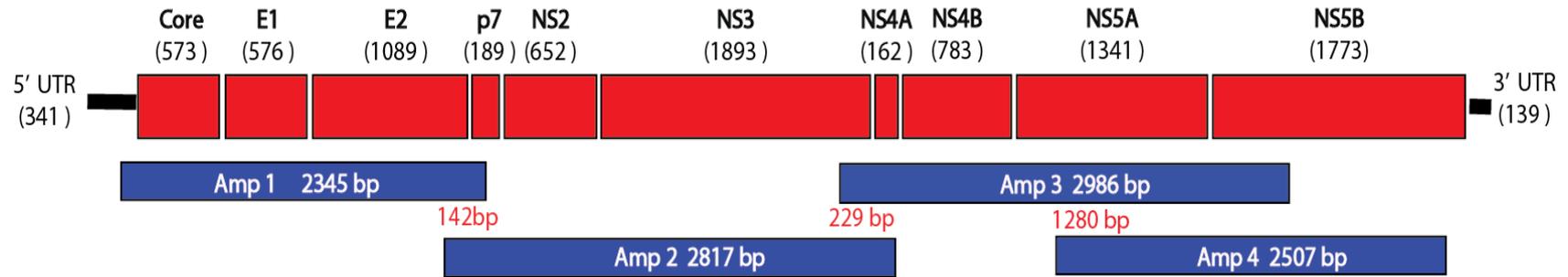
All primer combinations were selected based on melting temperature and length of primers while avoiding sequences prone to dimer or hairpin formation or self-complementary primers using web-based software; <http://www.thermoscientificbio.com/webtools/multipleprimer>. Candidate primers of 16-20 nucleotides were designed and compared to the 1a alignment from which the reference sequence was generated. For positions with unavoidable variability within the primer, degenerate positions were introduced in primers with no more than 4 degenerate sites included for each primer.

2.10.1.2 Optimisation of PCR conditions

Conditions with the largest impact on assay performance included the use of genome-specific primers in reverse transcription and the addition of a second nested PCR reaction to increase the sensitivity of the assay to ~1000 copies/ μL of HCV RNA. Nested PCR reactions were assembled in two volumes; 25 μL and 50 μL , including 5 μL of cDNA from the RT reaction as template for the first round PCR and different volumes (1 μL , 2 μL and 5 μL) of first round PCR product as template for the second PCR, for every pair of primers.

The PCR conditions were tested using touchdown PCR (ramping annealing temperatures) for increased specificity. Varying annealing temperatures were retested once a primer pair was optimised (42°C, 48°C, 50°C, 52°C, 56°C, 60°C), and different extension times and number of cycles were tested. A PCR hood and aerosol barrier tips were used for assembly of all reactions to avoid cross-contamination. Negative controls lacking template were included for each pair of primers. If any negative control was positive, all PCR reactions in that set were deemed to be contaminated and were discarded. Phusion[®] enzyme (Finnzymes) was used as it had the lowest error rate as described in Section 3.2.1 and showed high sensitivity for this large-scale sequencing project.

Figure 2-5: PCR (Amplicon) amplification strategy.



The HCV genome was covered by four overlapping amplicons (Amp 1, Amp 2, Amp 3, Amp 4).

Table 2-3: Primer sets and optimised conditions for amplicon 1.

NAME	SEQUENCE	(T _m) Finnzymes	(T _m) Allawi	GC%	Nt
T1F20	CGACACTCCACCATGAATCAC	65.9	56.1	52.4	21
T1F30	CCATGAATCACTYCCCTGT	61.6	53	50	19
T1F81	AGCCATGGCGTTAGTATGAG	62.1	54	50	20
T1F158	GGTGAGTACACCGGAATTG	61	53.4	52.6	19
T1F198	TTGGATAAACCCGCTCAAT	62.3	51.9	42.1	19
T1F210	GCTCAATGCCTGGAGATTT	62.3	53.3	47.4	19
T1F463	GCCCTAGATTGGGTGTG	59.6	53	58.8	17
T1R2354	CTGYGTGGTGGACARCAG	63.6	56.4	61.1	18
T1R2376	GARCACGGRAGRACCTG	61	54.1	61.8	17
T1R2555	CATCATCCACAARCAGGAG	61.1	52.4	50	19
T1R2576	CTCYRCYTGGGATATGAG	57.2	50.4	52.8	18
T1R2573	CGCYTGGGATATGAGTARC	60.2	52.8	52.6	19
T1R2670	CCAYGCAAAGCAGAAGAAC	62.7	53.9	50	19

The annealing temperature was calculated using two equations (Allawi, Finnzymes), primers shaded in dark grey were adapted or modified from those published by Yao et Tavis 2005, GC% is the GC content in each primer, Nt represent the number of nucleotides in each primer.

Table 2-4: Primer sets for amplicon 2.

NAME	SEQUENCE	(T _m) Finnzymes	(T _m) Allawi	GC%	Nt
T2F2084	CCCCACTGAYTGYTTC	56.5	50.3	56.3	16
T2F2413	CCTCCACCAGAACATYGTG	63.4	54.8	55.3	19
T2F2471	GTCCTGGRCCATYAAGTG	60.3	53.2	55.6	18
T2F2580	GCTTTGGARAACCTYG	55.7	28.3	50	16
T2F2657	CTGCTTTGCRTGGTAYCTG	62.4	54.3	52.6	19
T2R4548	GYTCGTCRCACTTYTTCT	57.2	51.5	47.2	18
T2R4651	GTCGACACGACGACAACAT	62.9	55	52.6	19
T2R5020	CCCTCCCAAAAYTCAAGRTG	64.7	54.2	50	20
T2R5039	GTGAGRCCYGTRAAGACG	60.6	54.2	58.3	18
T2R5230	GCRCCCAGTCTGTAYAGC	60.8	55.8	61.1	18
T2R5252	CAGGGTGAYYTCATTCTG	57.7	50.4	50	18
T2R5469	CTTCCATCTCATCGAACTC	57.8	50.2	47.4	19
T2R5536	GCCTTCTGCTTGAAGTCTG	57.6	51.5	52.9	17

The annealing temperature was calculated using two equation (Allawi, Finnzymes), primers shaded in dark grey are adapted or modified from those published by Yao et Tavis in 2005, GC% is the GC content in each primer, Nt represent the number of nucleotides in each primer.

Table 2-5: Primer set for amplicon 3.

NAME	SEQUENCE	(T _m) Finnzymes	(T _m) Allawi	GC%	Nt
T3F4725	GTCGAYTTYAGCCTTGACC	61	53.7	52.6	19
T3F5067	ACAAAGCAGAGTGGGGARA	62.8	55.1	50	19
T3F4911	GGCTGYGCTTGGTATGAG	62.5	55.2	58.3	18
T3F4742	CCCTACCTTYACCATTGAG	58.3	51.4	50	19
T3F5520	CAGTTCAAGCAGAAGGC	57.6	51.5	52.9	17
T3R7109	GAARGAGTCCAGAAACAC	54.8	50	50	18
T3R7547	GTCGCTRAGATCCGGATC	62.1	53.7	58.3	18
T3R7402	CTGCCAAARCTYTTGGTG	61.2	52.5	50	18
T3R6987	GCAAGTTGCYTTGAGRG	58.8	51.7	52.9	17
T3R7079	GA CTCRACCCTGGTGATG	60.9	54	58.3	18

The annealing temperature was calculated using two equations (Allawi, Finnzymes, GC% is the GC content in each primer, Nt represents the number of nucleotides in each primer. Primers shaded in dark grey are adapted or modified from those published by Yao et Tavis 2005.

Table 2-6: Primer sets for amplicon 4.

NAME	SEQUENCE	(T _m) Finnzymes	(T _m) Allawi	GC%	Nt
T4F6290	GATATGCGAGGTGYTGAG	58.6	51.8	52.8	18
T4F6707	CACAGAAYTGGACGGGGT	64.2	56.1	58.3	18
T4F7085	CAAAGTGGTGRTTCTGG	56.6	49.6	50	17
T4F7238	GACGTGGAAAAAGCCKGAC	65	55.9	55.3	19
T4R9214	GGAGTGAGTTTRAGCTTTGT	57.3	51.8	42.5	20
T4R9575	CGTGACTAGGGCTAAGATGG	61.7	54.8	55	20
T4R9638	TGCAGAGAGGCCAGTATCA	62.8	55.3	52.6	19

The annealing temperature was calculated using two equations (Allawi, Finnzymes), primers shaded in dark grey are adapted or modified from those published by Yao et Tavis 2005, GC% is the GC content in each primer, Nt represents the number of nucleotides in each primer.

2.10.2 Metagenomic sequencing

RNA was extracted from 400 μ L of plasma using the automated easyMAG[®] system according to manufacturer's instructions and eluted in a final volume of 60 μ L of elution buffer. For RNA library preparation, first strand cDNA was synthesized from half of the RNA volume obtained during purification (concentration usually undetectable using Qubit[®]) using reverse transcriptase (SuperScript III, Invitrogen). Double-stranded cDNA was synthesised using a NEBNext[®] mRNA Second Strand Synthesis Module (New England Biolabs).

DNA samples were purified using Agencourt AMPure magnetic beads[®] (Beckmann Coulter). The dsDNA was then prepared for Illumina[®] sequencing using a Nextera XT kit[®]. The resulting libraries were size selected prior to processing on the MiSeq[®] platform.

2.11 Bacterial cloning

Bacterial cloning was used in a number of experiments including clonal analysis of purified PCR products and replicon assays. For the clonal analysis of PCR amplicons, two different kits were used; the TOPO-TA[®] Cloning kit (Invitrogen) and CloneJet[®] (Thermo Scientific). A variety of *Escherichia coli* (*E. coli*) cells were used; One Shot[®] TOP10 Chemically Competent *E. coli* (Invitrogen), NEB 5-alpha Competent *E. coli*[®] and NEB 10-beta Competent *E. coli*[®].

2.11.1 TOPO-TA[®] Cloning

An aliquot (2-4 μ L) of PCR product was added to 0.5-1 μ L of salt solution (1.2M NaCl plus 0.06M MgCl₂) and 1 μ L of TOPO[®] vector and incubated at room temperature for 5-30 min. Ligation of the PCR product disrupts expression of the lethal *E. coli* suicide gene *ccdB* permitting growth of only positive colonies upon transformation in One Shot[®] TOP10 cells (Invitrogen).

The cloning reaction (2 μ L) was added to a vial of chemically competent One Shot[®] TOP10 *E. coli* cells and incubated on ice for 30 min. The cells were

then heat-shocked for 30 sec at 42°C in a water bath and transferred back to ice. 250 µL of room temperature SOC medium (2% tryptone, 0.5% yeast extract 10 mM NaCl, 2.5 mM KCl, 10 mM MgCl₂, 10 mM MgSO₄, 20 mM glucose) was added to each vial, capped and shaken horizontally (300 rpm) at 37°C for 1 hour.

An aliquot of 100 µL from each transformation was spread on a pre-warmed agar plate containing 100 µg/ml ampicillin or 30 µg/ml tetracycline depending on the plasmid antibiotic selection gene and incubated overnight at 37°C. Incorporation of the insert into the plasmid disrupts transcription of the suicide gene, enabling selection of all clones containing the product.

2.11.2 CloneJET[®] PCR cloning

For cloning blunt-end PCR products generated by proofreading DNA polymerases, the purified DNA fragment was ligated in a 3:1 molar ratio into the vector pJET1.2/blunt (Appendix 7.33). The ligation reaction was prepared on ice by adding: 10 µL of 2X Reaction Buffer, pJET1.2/blunt Cloning Vector (50 ng/µL), T4 DNA Ligase, purified PCR product and nuclease-free water to make the reaction volume up to 20 µL. The reaction mix was vortexed briefly, centrifuged for 5 sec and incubated at room temperature (22°C) for 5 min, then used directly for transformation.

2.11.3 One shot[®] Top 10 cells

Two µL of the TOPO[®] cloning reaction was added to a vial of One Shot[®] Chemically Competent *E. coli* and mixed gently, followed by incubation on ice for 30 min. The cells were then heated for 30 sec at 42°C without shaking (heat-shock). Immediately, the tubes were transferred to ice, and 250 µL SOC medium at room temperature added. After capping the tubes tightly, they were shaken horizontally (300 rpm) at 37°C for 1 hour. 100 µL of cells in the medium was then spread on a pre-warmed Agar plate and incubated overnight at 30°C.

2.11.4 NEB 5-alpha competent *E. coli*[®]

A tube of NEB 5-alpha competent *E. coli* cells was thawed on ice for 10 min,

followed by the addition of 1-5 μL of H_2O containing an average of 50 ng of plasmid DNA to the cell mixture. The tube was flicked 4-5 times to mix the cells and DNA, then placed on ice for 30 min without mixing. The mixture was exposed to heat shock at 42°C for 30 sec; then cells were placed on ice for 5 min. 950 μL of room temperature SOC was next added to the mixture and incubated at 37°C for 60 min while shaking vigorously (300 rpm). 100 μL from each transformation was spread on a pre-warmed selective plate and incubated at 30°C for 24-36 hours.

2.11.5 NEB 10-beta competent *E. coli*[®]

A tube of NEB 10-beta Competent *E. coli* cells was thawed on ice for 10 min, followed by the addition of 1-5 μL of H_2O containing 1 pg-100 ng of plasmid DNA to the cell mixture. The tube was flicked 4-5 times to mix cells and DNA, then placed on ice for 30min, then exposed to a heat shock at 42°C for 30 sec and left on ice for 5 min. 950 μL of room temperature SOC was added to the mixture at 37°C for 60min while shaking vigorously at 300 rpm. 100 μL from each transformation was spread on a pre-warmed selective plate and incubated at 30°C for 24-36 hours.

2.12 Preparation of DNA following bacterial cloning

Purified DNA was quantified using a Qubit[®] machine, and spectrophotometry (Nanodrop[®]) was carried out to assess the purity of DNA produced using the ratio of values obtained at A260/A280. Successful transformation was confirmed by restriction digest, which cuts the plasmid at two sites, generating two linear fragments of DNA. Digested products were run on a 1-2% agarose gel stained with ethidium bromide and visualised under ultraviolet light.

2.12.1 Small scale plasmid preparation from transformed bacteria

A single colony from a freshly streaked selective agar plate was selected and used to inoculate a 5 ml culture of LB with selective antibiotic (tetracycline 30 $\mu\text{g}/\text{ml}$). Following culture for 24-48 h at 30°C with vigorous shaking (180 rpm),

the bacteria were centrifuged in an Eppendorf microcentrifuge at 13,000 rpm and the DNA extracted from the bacterial pellet using the GeneJET[®] Plasmid Miniprep Kit (Thermo Scientific) according to the manufacturer's instructions.

All purification steps were carried out at room temperature. All centrifugation steps were carried out in a microcentrifuge at 13,300 rpm and pelleted cells were resuspended in 250 μ L of Resuspension Solution. To ensure complete suspension of bacteria, vortexing and pipetting up and down were carried out until no cell clumps were apparent. 250 μ L of Lysis Solution was next added and mixed thoroughly by inverting the tube 4-6 times until the solution became viscous and slightly clear. This was followed by the addition of 350 μ L of Neutralization Solution and mixed immediately and thoroughly by inverting the tube 4-6 times. The mixture was then centrifuged for 5 min to pellet cell debris and chromosomal DNA.

The supernatant was transferred to a supplied GeneJET[®] spin column by decanting or pipetting and centrifuged for 1 min. A wash step was carried out by adding 500 μ L of Wash Solution to the GeneJET[®] spin column, then centrifuged for 30-60 sec and the flow-through discarded. This wash step was repeated, and the column was centrifuged for an additional 1 min to remove residual Wash Solution. The elution step was carried out by adding 50 μ L of elution buffer to the centre of the GeneJET[®] spin column membrane after transferring the GeneJET[®] spin column into a fresh 1.5 ml microcentrifuge tube. The tube was incubated for 2 min at room temperature and then centrifuged for 2 min. The purified plasmid DNA was stored at -20 °C.

2.12.2 Large scale plasmid preparation from transformed bacteria

A single colony from a freshly streaked selective agar plate was selected and used to inoculate a 5 ml starter culture of LB-selective antibiotic (tetracycline 30 μ g/ml). Following a 24 hour incubation, the starter culture was diluted 1:1000 into 100 ml of LB with antibiotics and cultured for 24-36 h (to an OD 600 of 2-3) at 30 °C with vigorous shaking (180 rpm). The bacteria were then harvested by centrifugation at 5,000 g for 10 min at 4 °C.

A large-scale DNA preparation was then made from the bacteria using a GeneJET[®] Plasmid Midi Prep Kit.

Pelleted cells were resuspended in 2 ml of resuspension solution as above and incubated for 3 min at room temperature. 2 ml of Neutralization Solution was added and mixed immediately and thoroughly by inverting the tube 5-8 times, followed by addition of 0.5 ml of endotoxin binding reagent. The reaction was mixed immediately as above then incubated for 5 min at room temperature. After adding 3 ml of 96% ethanol, the tube was mixed by inverting the tube 5-6 times and then centrifuged for 40 min at 4,000-5,000Xg to pellet cell debris and chromosomal DNA.

The supernatant was transferred to a 15 ml tube by decanting and pipetting, and 3 ml of 96% ethanol added and mixed by inverting the tube 5-6 times. Part of the sample (~ 5.5 ml) was transferred to the supplied column pre-assembled with a collection tube (15 ml), then centrifuged for 3 min at 2,000g in a swinging bucket rotor. The step was repeated after discarding the flow-through and placing the column back into the same collection tube until all remaining lysate was processed through the purification column. 4ml of wash solution I was added to the purification column, followed by centrifugation for 2 min at 3,000Xg in a swinging bucket rotor. A second wash step was carried out by adding 4 ml of wash solution II to the purification column, followed by centrifugation for 2 min at 3,000Xg in a swinging bucket rotor; this step was repeated to remove residual wash solution.

For elution, the column was transferred to a fresh 15 ml collection tube, 0.35 ml of Elution Buffer added to the centre of the purification column membrane and incubated for 2 min at room temperature followed by centrifugation for 5 min at 3,000Xg to elute plasmid DNA. The purified plasmid DNA was stored at -20°C.

2.13 Mock community

The focus of this project was to test the suitability of Illumina[®] sequencing for haplotype reconstruction by assessing coverage, error profiles, trimming/filtering strategies, barcode switching and two haplotype reconstruction programmes; QuRe and PredictHaplo.

2.13.1 Preparation of the mock community

One sample from an HCV-infected patient (genotype 1a) was amplified using the Amp 1 primer set described earlier to produce a PCR product of an average size of 2.5kb. These PCR products were cloned using CloneJet[®] (Thermo Scientific). A mock community was created using 13 clones containing around 2.5 kb fragments of HCV viral variants (sequence divergence 0-0.9%). The 13 clones were mixed with log-normal abundance with a range of frequency of [3.2-30.3%].

2.13.2 MiSeq sequencing

All clones were prepared for Illumina[®] sequencing using Nextera XT[®], and one dual index for the whole community. The sample was sequenced in triplicate.

2.13.3 Haplotype reconstruction

Quasispecies assembly was carried out with QuRe software v0.9994 (<http://sourceforge.net/projects/quire/>), and PredictHaplo v0.4 software (<http://bmda.cs.unibas.ch/HivHaploTyper>). The quasispecies reconstruction method of QuRe employs a method based on an overlap graph constructed over sliding windows, selecting candidate variants using an algorithm based on overlap consistency and similarity of frequency distributions of variants in each window (Prosperi and Salemi, 2012), whilst PredictHaplo extends the sliding window Bayesian clustering approach to a global quasispecies inference based on a hidden Markov model (Schirmer et al., 2014).

2.14 Statistical analysis

Statistical analysis was performed using the statistical computer package GraphPad prism[®] V6. P-values of less than or equal to 0.05 were considered to be statistically significant.

2.15 Bioinformatics methods for viral sequence analysis

2.15.1 Sequence alignment

Sequences were aligned in fasta format in an open reading frame using MUSCLE (www.ebi.ac.uk/muscle) as an integrated tool in Mega 6 software. Chromatograms were visualised and sequences edited manually using BioEdit[®] or CLC Genomics 6[®].

2.15.2 Measures of diversity

Viral diversity can be measured using a variety of phylogenetic and statistical techniques based on sequence differences (Figure 2-6). Kimura introduced the theory of "neutral evolution" based on the observation that many genetic mutations are random events and may not confer phenotypic change (Kimura, 1980). This model may be used in the analysis of variation in HCV viral quasispecies, that occur largely as stochastic (random) events due to the error prone RNA-dependent RNA polymerase.

Provided the population size is large enough, some mutations may occur which are beneficial and increase in frequency in a population because of natural selection. Others may be neutral and persist due to random genetic drift; these changes, however, take longer to arise than adaptive mutations. This phenomenon can be measured by comparing genetic sequences and taking into account whether or not RNA sequence changes encode changes in the translated amino acid sequence.

Different measures of diversity were estimated using a script (FastaDiversity.sh) designed by Dr Joseph Hughes. It uses a fasta file of

sequences to measure i) *Richness*: how many different sequences are present within the dataset, ii) *Shannon Index*: When all sequences are equally abundant, the Shannon index takes the value of $\log(R)$. The more unequal the abundance e.g. with almost all abundance concentrated in one type and other very rare sequences, the Shannon index approaches zero when the Shannon entropy is zero iii) *Simpson Index*: the probability that two sequences taken at random from the dataset are the same (Figure 2-7).

Genetic diversity was calculated using the Hamming distance, which is defined as the number of amino acid differences between two sequences using the formula $(1 - s) \times 100$, where s is the fraction of shared sites in two aligned nucleotide sequences.

Genetic distance was calculated as the mean percentage difference between all samples at any time point. Mean intra-strain corrected genetic distance, and Hamming distances were also calculated (separate strains were defined as those originating from distinct branches following phylogenetic analysis with >10% difference from other strains derived from the same patient) (Ota and Nei, 1994).

Intra-host sequence diversity was measured using mean pairwise diversity (P_i) and compared between groups using t-tests and two-sample tests controlling for dependency in the data structure as described previously (Gilbert et al., 2005). The sequences from each patient were also placed on a reference phylogeny using the maximum-likelihood placement methods available in pplacer (Matsen et al., 2010). The quadratic diversity that combines abundance and relatedness data and phylogenetic entropy that only characterize relatedness were estimated from the phylogenies and compared between groups as detailed previously (Rao, 1982, Allen et al., 2009).

2.15.3 Mapping

Recent advancements in deep sequencing technologies have resulted in the generation of longer sequence reads (>150 bp) providing an opportunity to map them using a BLAST-based approach, optimised by simultaneous searching. Using this approach, Dr Sreenu Vattipally has developed an in-house assembler named “Tanoti” (<http://bioinformatics.cvr.ac.uk/Tanoti>), the first fast mapping program to successfully map long reads using BLASTN. Batch processing and efficient memory management make Tanoti extremely robust, requiring minimal computer resource usage, without compromising speed and sensitivity (Vattipally B. Sreenu).

It was evident from viral genome mapping comparisons that Tanoti is faster and a more sensitive mapper than its counterparts and can reliably deal with data containing insertions and deletions. It achieves fast search speed by adopting batch processing and well-regulated simultaneous searching of reads. Local search algorithms, better gap handling, low consumption of computational resources and high speed and accuracy make Tanoti an efficient program for mapping divergent reads from small genomes.

2.15.4 Construction of phylogenetic trees

Phylogenetic trees may be produced using exhaustive search methodology or stepwise clustering. Exhaustive search methods evaluate all possible tree topologies and apply criteria to select the most likely candidate. However, with increasing numbers of sequences, increasingly large numbers of possible trees may be produced, resulting in a lengthy analysis. Stepwise clustering methods overcome this problem by grouping sequence data together to form clusters with each new sequence compared to a cluster, thereby reducing computing time (e.g. Neighbour-Joining (NJ) tree).

Examples of exhaustive search methods are maximum parsimony (selects a tree with the smallest number of mutations required to generate internal nodes for each sequence) and maximum likelihood (ML) trees.

As Maximum likelihood outperforms neighbour joining under comparable conditions, ML trees were used in the construction of phylogenetic tree in this study (Huelsenbeck, 1995).

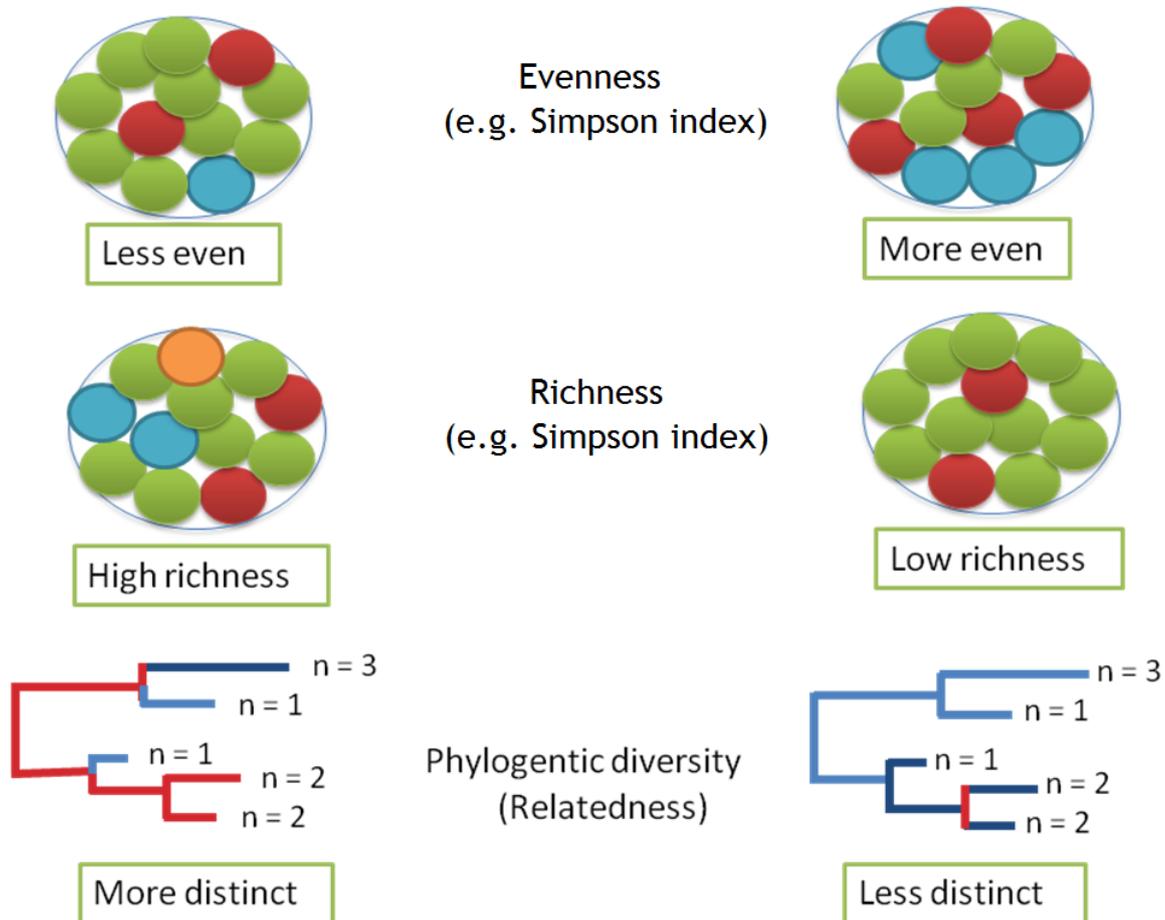
Sequences were aligned using MUSCLE. Phylogenetic trees were constructed with MEGA 6.0 using the maximum likelihood method, and each data set was tested for best-fit substitution model, Kimura two-parameter distance for all substitutions was the best model in all of our data sets, and the inferred phylogenies were tested with 1,000 bootstrap replications.

2.16 Resistance mutation database

A systematic review was performed jointly with Weronika Witkowska and Yangie Pinanga. Published articles were retrieved from Pubmed, OvidSP, Embase, and The Web of Knowledge website without date restriction. The following keywords were used: Simeprevir, Paritaprevir, Ledipasvir, Ombitasvir, Daclatasvir, Dasabuvir, or Sofosbuvir and 'resistance'. The former names of these drugs were also used as keywords: TMC 435, ABT-450, GS-5885, ABT-267, BMS-790052, ABT-333, or GS-7977. Oral and poster presentations from recent scientific conferences/ symposiums related to HCV such as AASLD, EASL, CROI, the International Workshop on Clinical Pharmacology of HIV and Hepatitis Therapy, International Symposium on Viral Hepatitis and Liver Diseases (ISVHLD) were also reviewed.

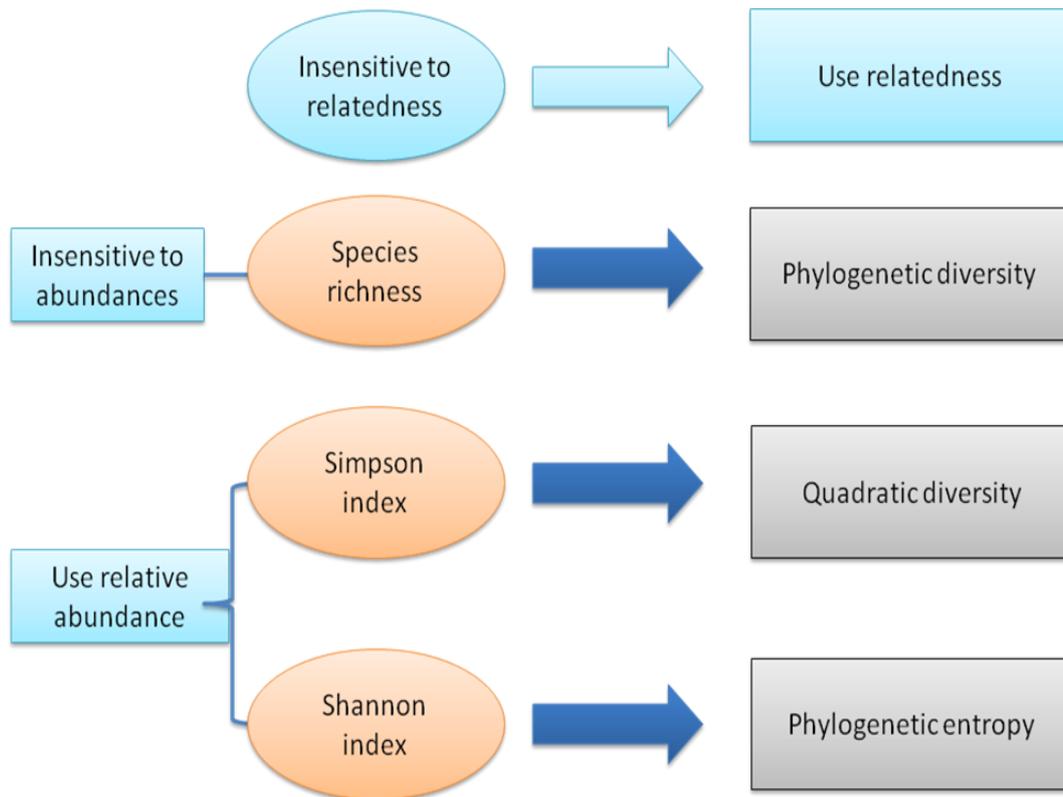
Resistance-associated mutations (RAMs) obtained from the literature were listed in mutation tables. The tables contain RAMs with reference to the relevant DAAs, along with the change in fold-resistance compared to wild type variant and the methods used to detect the RAMs. Tables were later utilized to create mutation reports for HCV-infected patients in this study.

Figure 2-6: Examples of measures of diversity.



Measures of diversity are based on different parameters including richness, evenness, and relatedness.

Figure 2-7: Relationship between diversity measures.



Measures of diversity according to their sensitivity to abundance and relatedness, Modified from (Allen et al., 2009).

2.17 Human Hepatoma Cells (Huh7.5)

Human hepatoma Huh7.5 cells (Huh7 cells cured of the HCV replicon with interferon) (Blight et al., 2002), were propagated at 37 °C in an atmosphere of 5% CO₂ in Dulbecco's modified Eagle's medium (DMEM[®]) supplemented with 10% fetal calf serum (FCS) and 1% penicillin/streptomycin.

Cell lines were grown in 80 cm² or 175 cm² tissue culture flasks (Nunc[®]). Passage of cells (twice weekly) was carried out when cells reached 80% confluency by first gently washing cells with 5 ml of ice-cold PBS to the flask and pass over the cell surface by gently rotating the flask. This is a wash step to remove dead cells and any residual fetal bovine serum. Removal of cells was achieved by the addition of 3ml of 0.05% Trypsin /EDTA to the T150 flask (Sigma[®]). Cells were then resuspended in 10 ml of complete DMEM before re-seeding or use in experiments.

2.18 *In vitro* replication assays

A bicistronic, subgenomic replicon system has been developed to examine the RNA synthesis of hepatitis C virus (HCV) genotype 1a (Blight et al., 2000). This replicon contains an HCV IRES that directs expression of the neomycin resistance gene in the first cistron and an encephalomyocarditis virus (EMCV) IRES to direct production of the HCV non-structural proteins NS3-NS5B in the second cistron as shown in (Figure 2-8). This replicon was used as a template for the construction of a genotype 1a transient replication system as described in Section 2.18.2.

2.18.1 Site-Directed Mutagenesis

Mutagenesis reactions were performed using the QuickChange XL[®] Site-Directed Mutagenesis kit (Stratagene). Forward and reverse primers for mutagenesis (Table 2-7) were designed to incorporate the desired mutation in the middle of the primer sequence and were between 25 and 45 bases in length, according to manufacturer's instructions (Figure 2-9).

A Quick-Change PCR reaction was prepared by mixing 5 μL of 10X reaction buffer, DNA Plasmid (10 ng), 1.25 μL Forward primer (10 μM), 1.25 μL Reverse primer (10 μM), 1 μL of dNTP mix, 3 μL of QuickSolution reagent, 1 μL of PfuUltra HF DNA polymerase and nuclease-free water added to complete a final volume of 50 μL . Cycling conditions in the thermocycler were as follows: 1 min at 95°C, 18X (50 sec at 95°C, 50 sec at 60°C, 1min/kb at 68°C), 7min at 68°C.

Following the PCR reaction using QuickChange II XL[®], each reaction was chilled on ice for 2 min followed by the addition of *DpnI* (10 U) to digest the non-mutated dam-methylated parental DNA. Reactions were mixed by pipetting, then centrifuged in an Eppendorf microcentrifuge at 13,000 rpm for 1min, followed by incubation at 37 °C for 1 h. *DpnI*-treated DNA was transformed into 50 μL NEB 10-beta cells as described in Section 2.11.

2.18.2 Introduction of mutations into the HCV-1a subgenomic replicon

All mutations were individually introduced into the TA/SG/Luc/HCV-1a replicon by site-directed mutagenesis. Six colonies from each agar plate were selected and used to prepare plasmid mini-preps as described in Section 2.12.2. Plasmids were sequenced using the appropriate sequencing primers (Table 2-8). Positive clones were then restriction digested with *Clal* and *Bsu36I* to remove the fragment containing the mutated NS5B sequence, which was subsequently ligated back into the parental TA/SG/Luc/HCV-1a vector backbone.

The ligation mixture was then transformed into bacteria and used to prepare a plasmid maxi-prep. A diagnostic digest was performed on the plasmid using *Clal* and *Bsu36I* to confirm the ligation of the desired sequence. In addition, the same sequencing primers; for adaptation mutations (Adaptation-F645), for NS5B mutations (Robin-F3047, Robin-F4239, Robin-F3502), and Lucy-Forward were used to verify that no mutation occurred accidentally during the subcloning process.

2.18.3 Restriction enzyme digestion of DNA

All restriction enzyme digests of plasmid DNA were carried out at 37°C for at least 1h unless otherwise specified by the manufacturer. Typically, 10U of each enzyme per µg DNA were used in a total volume of 50µL. All reactions were performed using the appropriate enzyme buffers and BSA if necessary.

2.18.4 DNA ligation

A Rapid DNA Ligation Kit[®] (Thermo Scientific) was used for DNA ligation reactions. It was used to introduce an insert into a plasmid vector in the subcloning process to insert mutant fragments into the TA/SG/Luc/HCV-1a backbone (Table 2-9 and Table 2-10).

2.18.5 *In vitro* transcription

In vitro transcription was carried out using RiboMAX[®] Large Scale RNA Production Systems (Promega) following manufacturer's instructions.

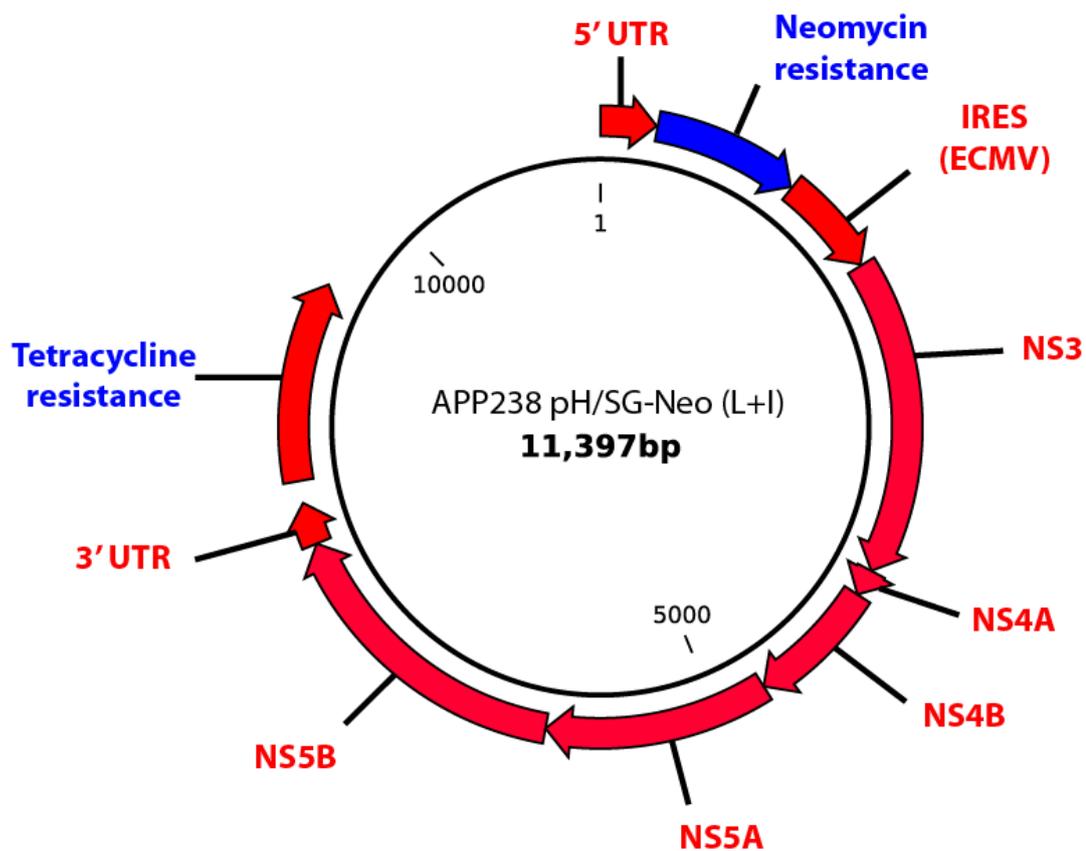
2.18.5.1 DNA linearization

Clal enzyme digestion linearised plasmids in a 100 µL reaction in 1.5 ml RNase-free tube (Ambion) followed by treatment with Mung Bean nuclease[®] (Promega) treatment to digest the sticky ends (30°C for 30 min). The concentration of linear DNA template was then determined using Qubit[®].

2.18.5.2 Transcription Procedure

The reaction components were assembled at room temperature in a 1.5 ml microcentrifuge tube by adding 1 µg of linear DNA template, 20µL of T7 Transcription 5X Buffer, 30µL of rNTPs (25mM ATP, CTP, GTP, UTP, 10µL of Enzyme Mix (T7) plus nuclease-free water up to 100µL final volume. After all the components were added and mixed by pipetting gently, the reaction was incubated at 37°C for 2 hours. RNA was visualized by gel electrophoresis on a 1% agarose gel; finally, the amount of RNA was measured by Qubit[®].

Figure 2-8: Genotype 1a HCV subgenomic replicon APP238 pH/SG-Neo (L+I)



The plasmid encoding the genotype 1a HCV subgenomic was supplied by Apath; USA.

Table 2-7: Mutagenesis primers.

Primer	Sequence
qc H77GNDfor	caccatgctcgtgtgtggcAacgacttagtcgttatctg
qc H77GNDrev	cagataacgactaagtcgtTgccacacacgagcatggtg
qc H77 K1691R fore	caggattgtcttgccgggaGgccggcaattatacctg
qc H77 K1691R rev	caggtataattgccggcCtcccggacaagacaatcctg
qc H77 E1726G fore	gagcaagggatgatgctcgtggtggcagttcaagcagaaggcc
qc H77 E1726G rev	ggccttctgctgaactgcccagcgagcatcatcccttgctc
qc S96T Foreward	gcctgacgccccacatacagccaaatc
qc S96T Reverse	gatttggctgtatgtggggcgctcaggc
qc S282T Foreward	caggtgccgcgcgAccggcgctac
qc S282T Reverse	gtacgccggctcgcgggcacctg
qc R32A Foreward	actgagcaactcgttgctagcccatcacaatctgggtg
qc R32A Reverse	tgactcgttgagcaacgatcgggtagtgtagaccac
qc G493A Foreward	gcctcagaaaacttgccgggtcccgcctt
qc G493A Reverse	aagggcgggaccgcaagtttctgaggc
qc P495A Reverse	gctcgcaagggcgcgaccccaagttt
qc P495A Foreward	aaaacttggggctcgcccttgcgagc
qc P496A Foreward	cttgggggtcccGccttgcgagctt
qc P496A Reverse	aagctcgcaaggccgggaccccaag
qc R32H Forward	gagcaactcgttgctacaccatcacaatctgggtg
qc R32H Reverse	acaccagattgtgatggtgtagcaacgagttgctc
qc R32A Forward	actgagcaactcgttgctagcccatcacaatctgggtg
qc R32A Reverse	caccagattgtgatgggtagcaacgagttgctcagt

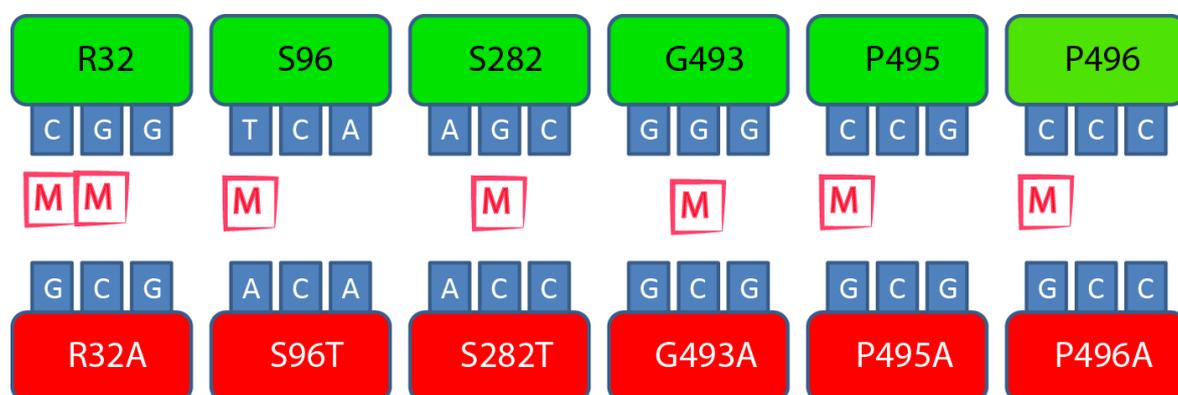
Eleven residues were targeted, representing reported SOF-resistance associated mutations and predicted resistance mutations.

Table 2-8: Sequencing primers.

Robin -F3047	CAATGTCTTATTCTGGAC
Robin-F4239	GACAGCAAGACACACTCC
Robin-F3502	GGTCGTAAGCCAGCTCGTC
Lucy- Forward	CATG GAA GAC GCC AAA AAC
Adaptation- F645	CCAACACCCCTGCTATAC

These primers were used for verification of the introduction of desired mutations in the replicon. Sanger sequencing was carried out by Beckman Coulter Genomics.

Figure 2-9: Schematic representation of mutations introduced into the replicon.



WT residues are shown in green and mutated residues in red. (M) represent mutated sites introduced with the Quickchange XL protocol.

2.18.6 Electroporation of RNA

Huh7.5 cells (80% confluent) in a T75 flask were trypsinised and transferred into a 50ml Falcon tube for counting. 3×10^6 cells were transferred into a 15 ml falcon tube and centrifuge at 1000 rpm for 5 min. After discarding the supernatant, the cells were washed twice by adding 10 ml of PBS to resuspend the pellet and centrifuge at 1000 rpm for 5 min, then resuspend the cell pellet in 1 ml of fresh PBS. For electroporation, cells were transferred to 4 mm electroporation cuvette and ~5 μ g of RNA was added per electroporation.

The electroporation apparatus was set at 270v and 950 capacitance setting the resistance to ∞ . Treated cells were then added to 20 ml of medium and distributed into 12 wells of a 24-well plate (1ml per well). Extracts from cells were made at 24 hourly time interval up to 120h to measure luciferase activity in triplicates.

2.18.7 Firefly luciferase activity assay

D-luciferin is the substrate for firefly luciferase (*Photinus pyralis*). Luciferase catalyzes the oxidation of luciferin to oxyluciferin in the presence of ATP and magnesium, resulting in bioluminescence. After removal of culture medium and washing the cells twice with 1 ml of PBS buffer, cells were lysed by adding 80 μ L of cell culture lysis reagent[®] (Promega) for 2 min. After scraping the bottom of the plate to ensure complete lysis, 40 μ L cell lysate was transferred per well into an Eppendorf tube containing 80 μ L luciferin (Promega). The sample was mixed by vortexing, and the resulting luminescence analysed using a GloMax[®] Single Tube Luminometer (Promega). All luciferase assays were carried out in triplicate, and representative data are given for each experiment. All calculations and graphical conversions were performed using GraphPad Prism 6[®].

2.18.8 Prediction model for detecting resistance mutations for SOF

The HCV RNA polymerase crystal structure in complex with UTP crystal structure (PDB ID: 1GX6) from the protein databank (<http://www.rcsb.org>). Residues interacting with UTP were obtained using an in-house C program.

It uses spatial coordinates of each atom and measures van der Waals forces with all other atoms that are within 5 angstroms radius from the atomic centre. With this information, it is possible to predict key residues in the protein with their high contact interactions.

Since SOF acts as a nucleotide analogue, it is likely that residues interacting with UTP will also interact with SOF. From the RNA polymerase and UTP complex crystal structure, residues that interact with UTP and have low contact interactions with other residues for the detection of possible resistance mutations were identified. The model was designed by Dr Sreenu Vattipally.

2.18.8.1 Transcription Procedure

The reaction components were assembled at room temperature in a 1.5 ml microcentrifuge tube by adding 1 µg of linear DNA template, 20µL of T7 Transcription 5X Buffer, 30µL of rNTPs (25mM ATP, CTP, GTP, UTP, 10µL of Enzyme Mix (T7) plus nuclease-free water up to 100µL final volume. After all the components were added and mixed by pipetting gently, the reaction was incubated at 37°C for 2 hours. RNA was visualized by gel electrophoresis on a 1% agarose gel; finally, the amount of RNA was measured by Qubit[®].

2.18.9 Electroporation of RNA

Huh7.5 cells (80% confluent) in a T75 flask were trypsinised and transferred into a 50ml Falcon tube for counting. 3×10^6 cells were transferred into a 15 ml falcon tube and centrifuge at 1000 rpm for 5 min. After discarding the supernatant, the cells were washed twice by adding 10 ml of PBS to resuspend the pellet and centrifuge at 1000 rpm for 5 min, then resuspend the cell pellet in 1 ml of fresh PBS. For electroporation, cells were transferred to 4 mm electroporation cuvette and ~5 µg of RNA was added per electroporation.

The electroporation apparatus was set at 270v and 950 capacitance setting the resistance to ∞ . Treated cells were then added to 20 ml of medium and distributed into 12 wells of a 24-well plate (1ml per well).

Extracts from cells were made at 24 hourly time interval up to 120h to measure luciferase activity in triplicates.

2.18.10 Firefly luciferase activity assay

D-luciferin is the substrate for firefly luciferase (*Photinus pyralis*). Luciferase catalyzes the oxidation of luciferin to oxyluciferin in the presence of ATP and magnesium, resulting in bioluminescence. After removal of culture medium and washing the cells twice with 1 ml of PBS buffer, cells were lysed by adding 80 μ L of cell culture lysis reagent[®] (Promega) for 2 min. After scraping the bottom of the plate to ensure complete lysis, 40 μ L cell lysate was transferred per well into an Eppendorf tube containing 80 μ L luciferin (Promega).

The sample was mixed by vortexing, and the resulting luminescence analysed using a GloMax[®] Single Tube Luminometer (Promega). All luciferase assays were carried out in triplicate, and representative data are given for each experiment. All calculations and graphical conversions were performed using GraphPad Prism 6[®].

Table 2-9: Ligation of insert DNA into plasmid vector DNA.

Linearized vector DNA	10-100 ng
Insert DNA (<i>at 3:1 molar excess over vector</i>)	variable
5X Rapid Ligation Buffer	4 μ L
T4 DNA Ligase, 5 U/ μ L	1 μ L
Water, nuclease-free	to 20 μ L
Total volume	20 μ L

The reaction was carried out in a 1.5 ml microcentrifuge tube and incubated at 22 °C for 5 min.

Table 2-10: Recircularization of linear DNA using Rapid DNA Ligation Kit.

Linearized vector DNA	10-50 ng
5X Rapid Ligation Buffer	10 μ L
T4 DNA Ligase, 5 U/ μ L	1 μ L
Water, nuclease-free	to 50 μ L
Total volume	50 μ L

The reaction was carried out in a 1.5 ml microcentrifuge tube and incubated at 22 °C for 5 min.

Chapter 3: Intra-host viral population structures

3.1 Background

The viral load of HCV may reach 10^7 copies/ml in immunocompetent patients, with the production and clearance of 10^{10} - 10^{12} genomes per day (Gregori et al., 2014). The HCV error-prone RdRP leads to a high viral mutation rate that is estimated to be 2.5×10^{-5} mutations per nucleotide per genome replication cycle (range 1.6 - 6.2×10^{-5}) (Ribeiro et al., 2012). This high mutation rate is consistent with the high degree of intra-host genetic diversity observed in HCV-infected patients (Cruz-Rivera et al., 2013).

Studies of intra-host viral diversity may explain the evolutionary dynamics of HCV infection including, replication rate, natural selection and random genetic events (Gray et al., 2012). Heterogeneity at the nucleotide sequence level is not distributed evenly across the HCV genome. The most variable part of the HCV genome is the hypervariable region 1 (HVR1) (Moreau et al., 2008). HVR1 has been investigated in many HCV quasispecies studies, given that a high degree of diversity increases the likelihood of distinguishing one HCV variant from another.

The targeted region was amplified using multiplexed genotype-specific primers to compare different sequencing techniques (direct Sanger sequencing, bacterial clonal analysis, and deep sequencing) for the detection of quasispecies heterogeneity as described in Section 2.5.2. In addition, for more detailed and accurate information about population structure, two methods of sequencing the whole genome were developed using two approaches; PCR-based amplicon sequencing and metagenomic sequencing. The optimisation of these approaches is discussed in detail in this chapter.

Viral complexity and diversity indices may be measured in sequential samples in one patient (intra-host diversity) or groups of patients. It is important that a consistent and non-biased framework be used to estimate the measure of diversity thus a number of techniques for measuring viral diversity and

reconstructing haplotypes were assessed. In the new era of direct-acting antivirals (DAAs), studying the HCV viral population before and after treatment is of great value, as viral complexity within infected individuals may result in treatment failure due to either mixed infection or the *de novo* evolution of resistance during therapy.

The prevalence of mixed genotype infection in HCV-infected subjects ranges from 5% in a cohort of patients co-infected with HCV/HIV to 39% in a cohort of PWID (Pham et al., 2010, McNaughton et al., 2013). There may be an increased need for improved screening, by applying sequencing techniques at baseline to ensure the detection of multiple variants infection of low frequency in the viral population which are missed by current diagnostic methods (Abdelrahman et al., 2015, McNaughton et al., 2013).

A potential candidate is the use of next generation sequencing (NGS) (Barzon et al., 2011). Further work is required to assess whether the detection of low-frequency variants is associated with treatment failure - at present, it is not clear how relevant the detection of such variants is, but the evolution of new sequencing technologies will allow a clearer view of viral population structures and their evolution during treatment.

The development of NGS technologies opens up significant challenges; a major challenge is to reconstruct circulating viral haplotypes from NGS data. Two of the currently available haplotype reconstruction programmes (PredictHaplo and QuRe) are assessed using a variety of artificially constructed mock communities as described in Section 2.13.

In this chapter, different methods of assessing HCV population structures are compared using direct sequencing, clonal sequencing, and deep sequencing approaches. The development of whole genome sequencing approaches using PCR and metagenomic deep sequencing are detailed. Finally, a variety of bioinformatic approaches for the reconstruction of viral populations is assessed (Table 3-1).

3.2 Results

3.2.1 PCR error rate estimation

Three different enzymes were tested using clonal sequence analysis of a single DNA insert in the pJET1.2/blunt vector as a template for sequencing. The produced clones were directly sequenced, and error rate was calculated according to the formula; Error rate = number of errors/ (target size × number of clones).

Error rate estimates were determined to be as follows; Phusion® High-Fidelity DNA Polymerase ($<7.1 \times 10^{-6}$ error/bp), KOD Hot Start DNA Polymerase® (5.2×10^{-6} error/bp), AccuStar® DNA polymerase (1.32×10^{-5} error/bp). These error rates were lower than that of MegaMix® Blue PCR Taq polymerase (1.9×10^{-5} error/bp). The proofreading enzyme Phusion® High-Fidelity DNA Polymerase ($<7.1 \times 10^{-6}$ error/bp) was selected for use in all further PCR reactions (Figure 3-1).

3.2.2 454 pyrosequencing error rate estimation

An alignment of control HVR1 plasmid sequences resulted in a mean depth of 16,918 at each nucleotide site. The most frequently found sequence was the original sequence in the plasmid. Based on the polymorphisms detected, the calculated error rate was 0.002 per base pair. Either these errors could arise during sequencing, during reverse transcription step or during PCR amplification that may affect the interpretation of the presence of polymorphism; hence, an abundance cut-off was instated above the error rate value to avoid any false results.

3.2.3 Measuring viral diversity using HVR1 region

A 183 base pair region of the E1/E2 gene including the HVR1 region was amplified from samples obtained from 16 patients with early HCV and chronic HIV infection using indexed fusion primers as described in Section 2.5.2.

We analysed viral diversity in paired samples using three different techniques; direct Sanger sequencing, bacterial cloning followed by Sanger sequencing, and 454 pyrosequencing. An additional three samples were used to compare two deep sequencing platforms; 454 pyrosequencing and Illumina® sequencing.

3.2.4 Viral complexity

We used two criteria for differentiating viral variants, a pairwise distance of >10% and phylogenetic signal (Abdelrahman et al., 2015).

3.2.5 Comparison of sequencing techniques

After sequencing 16 samples using Sanger sequencing, visual inspection of chromatograms generated using Bio-Edit v.7.1.3 software revealed that 8/16 patients had evidence of miscalling due to double peaks suggestive of mixed strain infection (Figure 3-2).

Using deep sequencing, multiple variants were detected in all patients, with two or more variants found in every sample. These were undetected by direct sequencing and partially detected using clonal analysis (Table 3-2). Mixed genotype or subgenotype infection was detected in 37.5% of samples (6/16) using deep sequencing, two of which were undetected by clonal analysis. An example of multiple variant infections is illustrated in Figure 3-3.

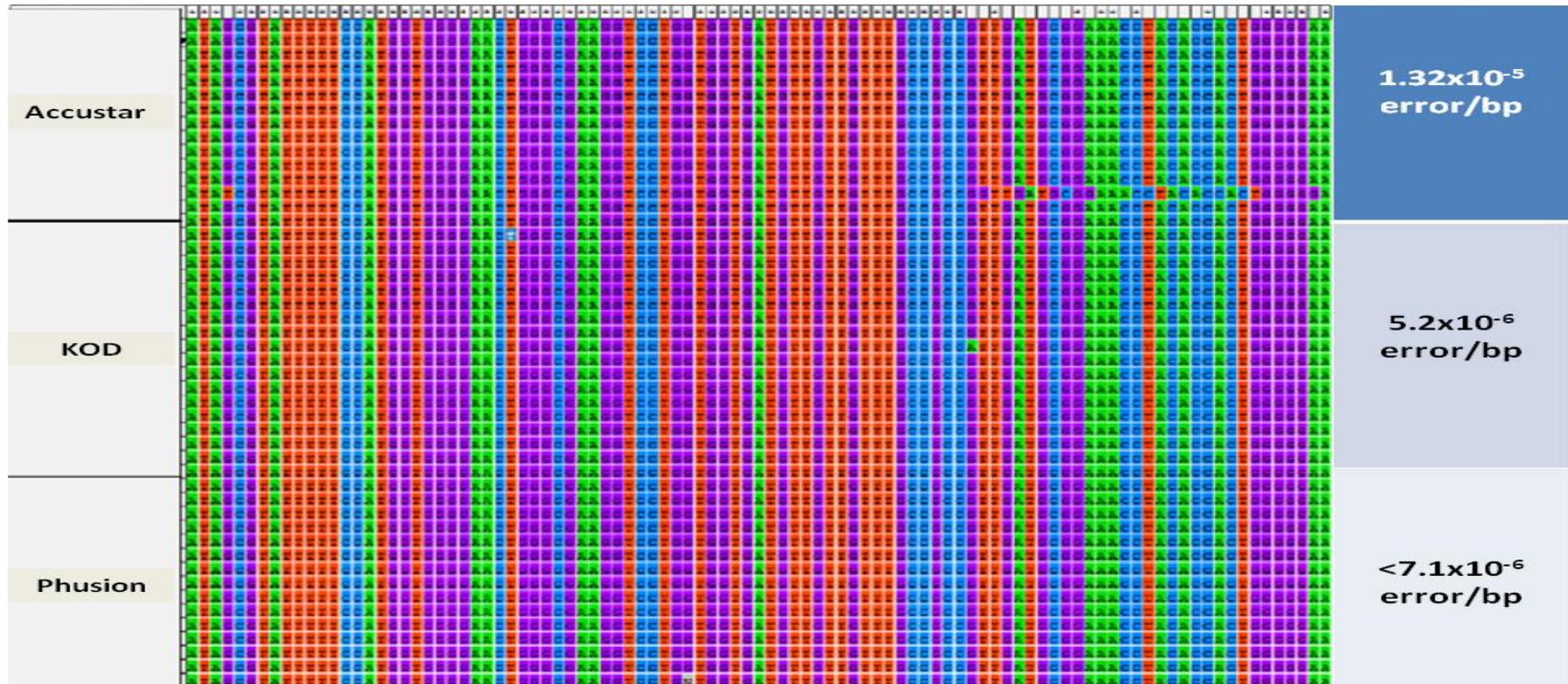
Clonal analysis and deep sequencing only partially correlated with LiPA genotyping, as the latter did not result in detection of mixed-genotype infections. In all patients, the variant identified by either LiPA or direct sequencing was detected by both clonal analysis and deep sequencing (Figure 3-4).

Table 3-1: Overview of experiments discussed in chapter 3.

Experiment	Samples	Method	Results
PCR error rate estimation	One sample from (P13)	E1/E2 amplified using four different polymerase enzymes as details in Section 2.5.1	Section 3.2.1
Comparison of different methods to study viral diversity	16 samples from patients (P21,P31,P38,P55,P57,P63, P67, P112,P118, P131,P141)	Amplified region of E1/E2 (183 bp) is sequenced using different techniques; Direct sequencing, Clonal analysis, Illumina sequencing and 454 Pyrosequencing	Section 3.2.5
Full genome sequencing (Amplicon approach)	14 samples from patients (P2,P3,P6,P8,P9,P11,P12,P13,P14, P22,P55,P63,P101,P105)	The optimisation of the full genome sequencing using amplicon approach as described in Section 2.10.1	Section 3.2.6
Full genome sequencing (Metagenomic approach)	11 samples from patients (P10,P12,P13,P22,P17,P22,P34,P45,P52,P55,P75,P113)	Using Metagenomic approach as described in Section 2.10.2	Section 3.3.3
Mock community	One sample from (P13)	2.5kb region was amplified using Amp 1 primer then clonal analysis followed by NGS sequencing as described in section 2.13.1	Section 3.3.5

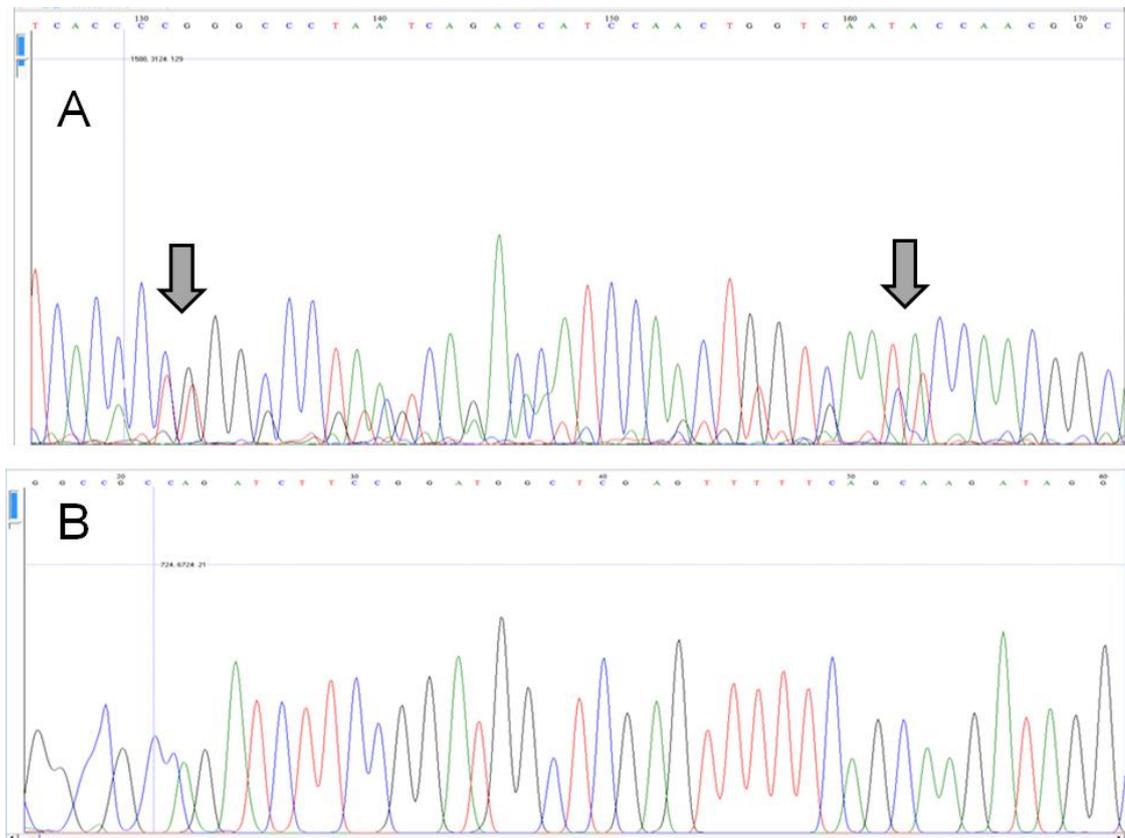
Patients were anonymised and attributed a (P) number. Samples used were either serum or plasma.

Figure 3-1: Estimation of the error rate of different polymerase enzymes.



Alignment of sequences following direct sequencing of the HVR1 region amplified by different polymerase enzymes. Three PCR enzymes were assessed using a plasmid template for amplification. The alignment shown in the second column was carried out using Clustal W embedded within MEGA 6.0 software; this represents an average of 15 clones studied for each enzyme. The third column is showing the estimated error rate of each enzyme.

Figure 3-2: A chromatogram showing an example of miscalling.



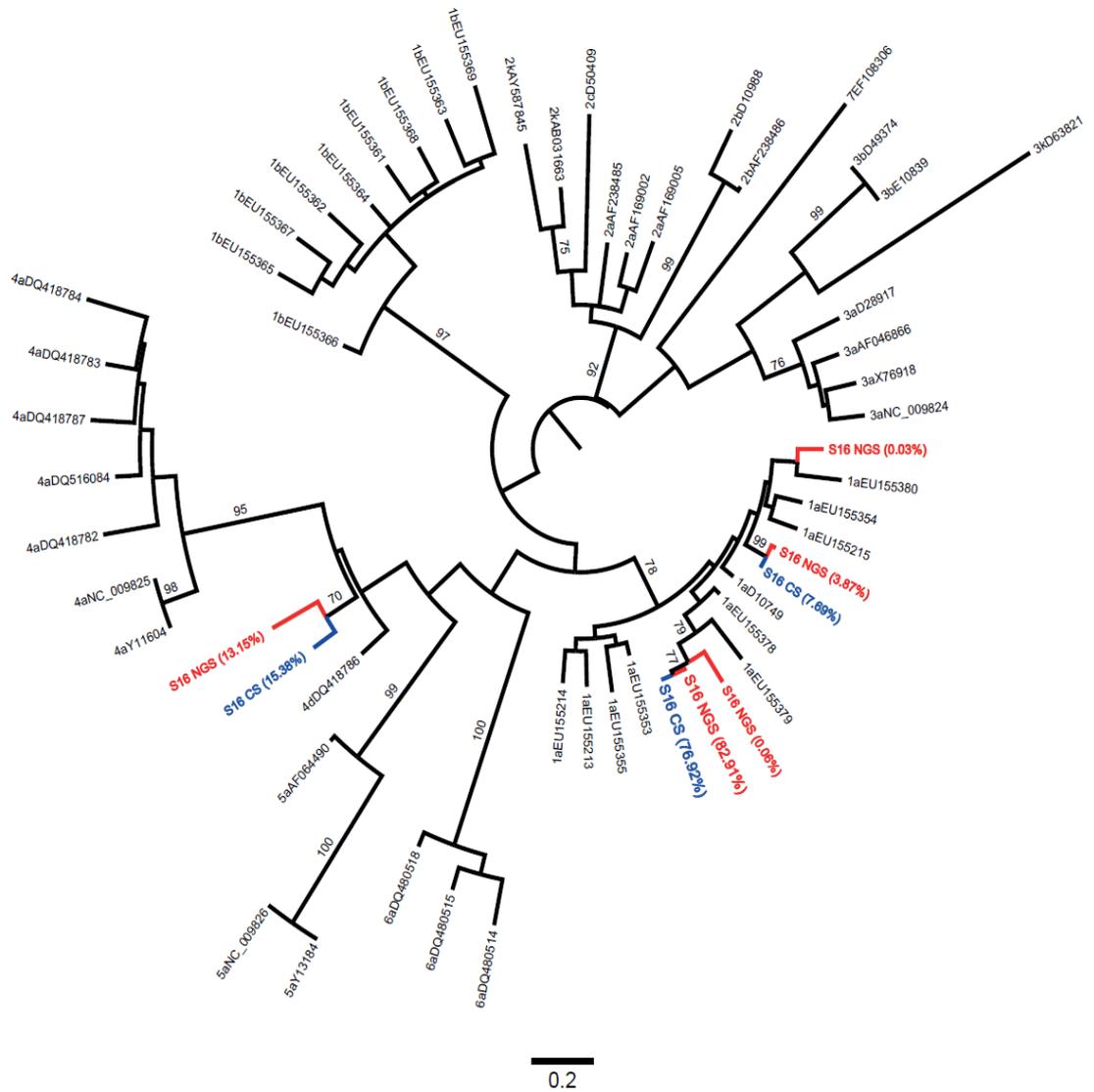
A region of a chromatogram generated by Sanger sequences A) Direct Sanger sequencing of one sample with grey arrows point to double peaks, B) Sanger sequencing after bacterial cloning demonstrating the purity of the sequence. Sequencing was carried out by Beckman Coulter Genomics, UK.

Table 3-2: Viral complexity and viral diversity using different sequencing techniques.

ID	Genotype	Clonal analysis-Sanger (CS)			454 analysis (NGS)		
	Direct Sanger (DS)	No of Clones	Variants No (Genotype)	Diversity	No of Reads	Variants No (Genotype)	Diversity
1	1*	20	1 (1a)	0	31574	4(1a)	0.067
2	1a	22	2 (1a)	0.149	36422	3(1a)	0.124
3	1a	14	3(1a)	0.082	46155	3(1a)	0.055
4	1*	20	1(1a)	0.004	23735	3(1a)	0.050
5	1a	18	1 (1a)	0.002	23042	2(1a)	0.089
6	1*	27	1(1a)	0.002	23660	3(1a),1(1b),1(3a)	0.034
7	1*	20	1(1a)	0.005	10742	2(1a), 1(1b)	0.041
8	1*	12	1(1a)	0	23639	2(1a)	0.043
9	1a	16	1(1a)	0	587	5 (1a)	0.057
10	1a	19	2(1a), 1(1b)	0.255	19165	3(1a),1(1b)	0.257
11	1a	15	1(1a)	0	21265	3(1a)	0.057
12	1*	22	1(1a)	0.001	44269	3(1a)	0.050
13	1*	18	1(1a),1(1b)	0.252	21246	2(1a), 1(1b)	0.226
14	1a	20	1(1a),1(1b)	0.135	28131	4(1a) 1(1b)	0.138
15	1a	18	1(1a)	0.017	19739	3(1a)	0.071
16	1*	13	2(1a),1(4d)	0.162	23788	3(1a), 1(4d)	0.185

Viral complexity (number of variants detected), viral diversity (overall mean pairwise distance), Direct sequencing (DS), Clonal analysis (CS), and 454 pyrosequencing (NGS). (*) Genotyping carried out by (INNO-LiPA II[®]; Innogenetics) and was not able to identify a subtype. The genotypes are highlighted in red

Figure 3-3: Illustrative example of multiple variants detected by CS and NGS.



Maximum likelihood tree showing all variants from different sequencing platforms; Clonal analysis (CS) in blue, 454 pyrosequencing (NGS) in red, with the frequency of each variant in the total population. The analysis included; 13 clonal sequences, 23,788 reads derived from 454 pyrosequencing and selected HCV references from the Los Alamos HCV database. The values on the branches represent bootstrap support.

We noted a disparity however in the representation of minority variants using clonal sequencing, with some variants over-represented and others under-represented. In one patient, for example, a variant present at a prevalence of 0.06% using pyrosequencing was detected by clonal analysis (Figure 3-5), while another variant present at a level of 25% using pyrosequencing was undetected (Figure 3-6). Deep sequencing resulted in the detection of at least one additional quasispecies variant in 93.7% of samples when compared with clonal analysis (Figure 3-7).

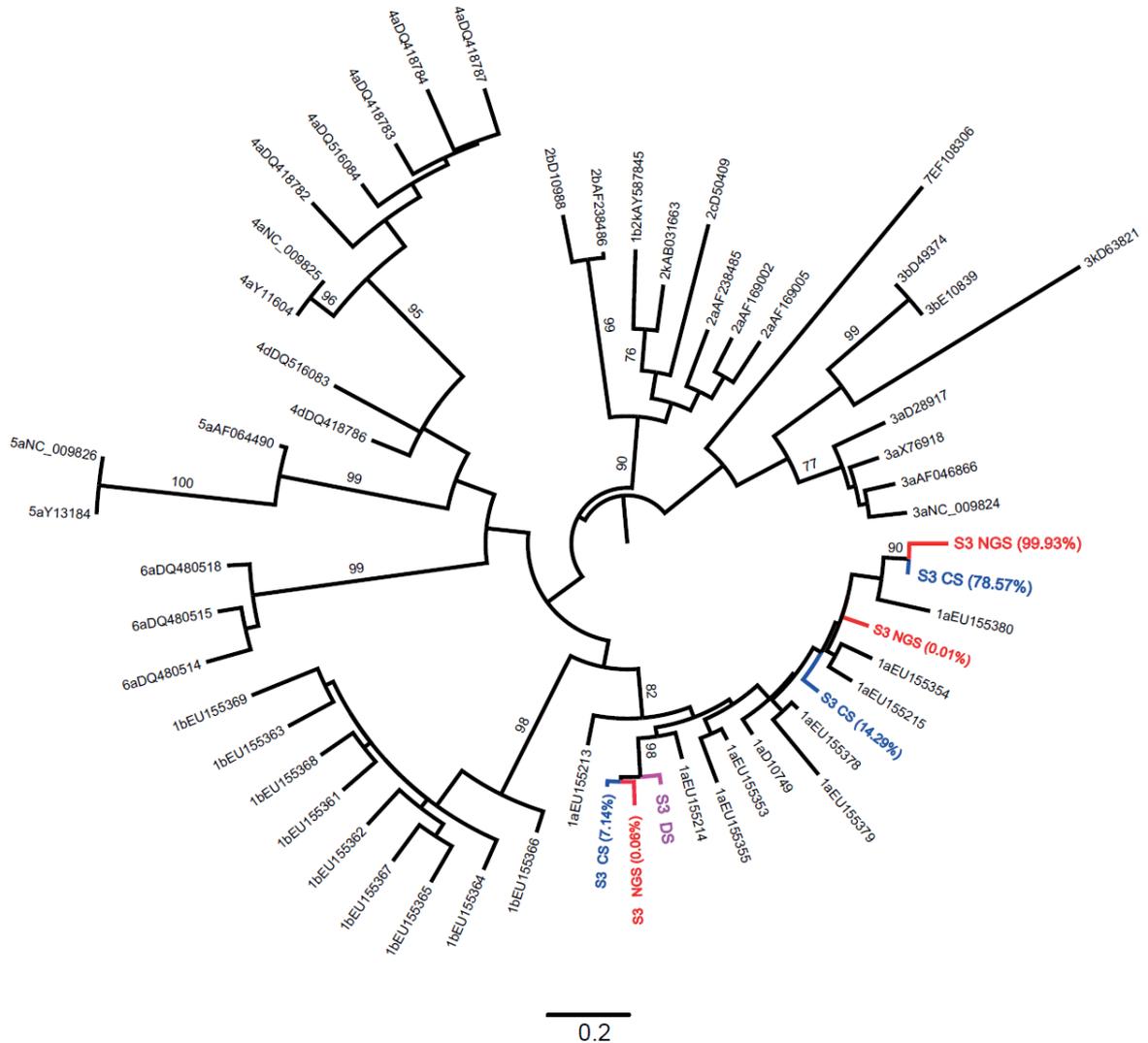
We next used three samples to compare two different deep sequencing technologies; 454 pyrosequencing as described in Section 2.9.2, and Illumina® sequencing as described in Section 2.9.3. Both technologies detected almost identical variants in these samples (Figure 3-8).

3.2.6 Full genome sequencing of HCV

We next aimed to develop and refine methods to amplify and sequence the whole HCV ORF of genotype 1 for further analysis of population structure within infected individuals. We used two approaches, a PCR-based approach, and a metagenomic approach. The aim was to optimise several steps in the process, including extraction of HCV RNA from patient plasma or serum, reverse transcription, PCR primer sequences, PCR cycling conditions, template preparation, library preparation, and finally sequencing and assembly.

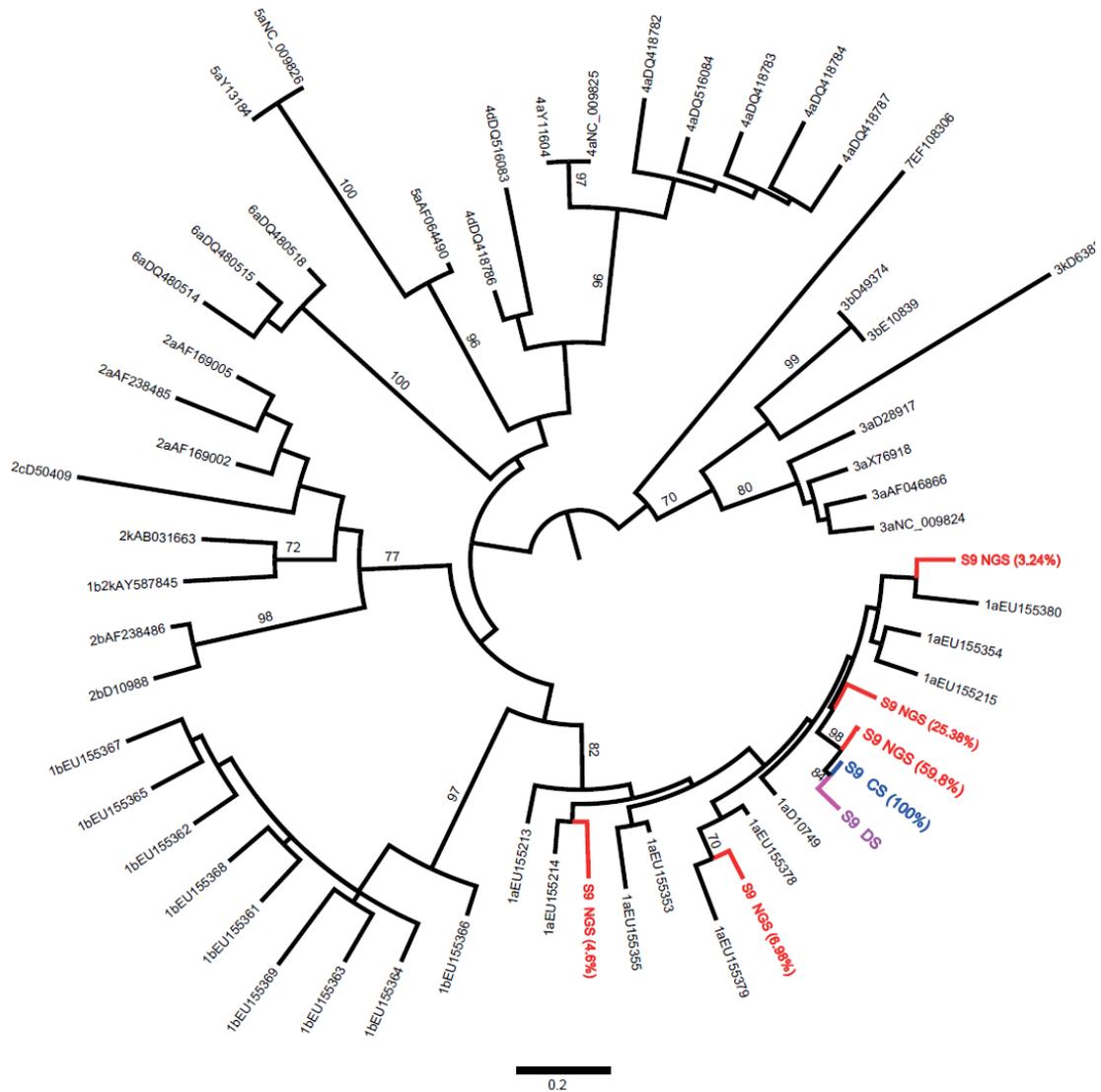
We were able to amplify the HCV ORF of genotype 1a from 93% (13/14) of samples using RT-PCR amplification to generate four amplicons (Amp 1-4) and successfully obtained viral sequences from all 13 samples that could be amplified. Samples used had a mean viral load of 25813 copies/ml, with a range of [1,512-13,163] copies/ml, the failed sample had a viral load of 13620 copies/ml; it was possible to amplify Amp 3 and Amp 4 from this sample, but not Amp1 and Amp 2.

Figure 3-5: Illustrative example of clonal analysis detecting a minority variant (0.06%).



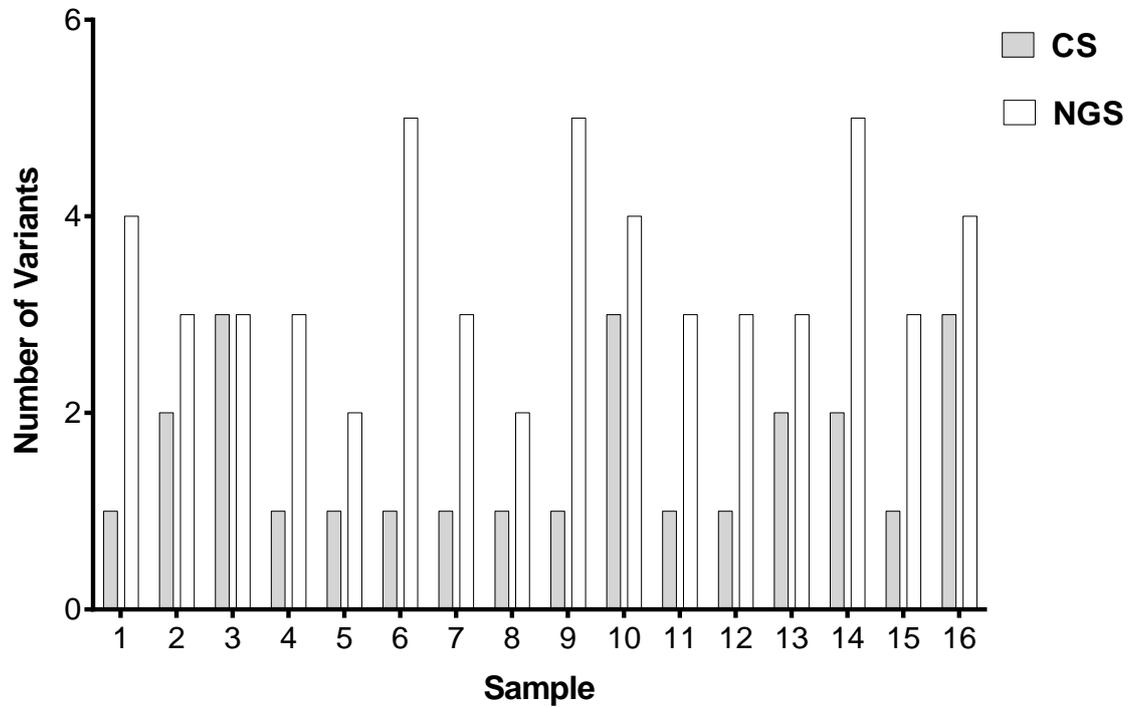
Maximum likelihood tree showing all sequences from different sequencing platforms; the direct sequence (DS) is shown in purple, clonal sequences (CS) in blue, 454 pyrosequences (NGS) in red, with the frequency of each variant in the total population shown in brackets. The analysis included 14 clonal sequences, 46155 reads derived from 454 pyrosequencing and selected HCV references from the Los Alamos HCV database. Values on the branches represent bootstrap support.

Figure 3-6: Clonal analysis may not detect variants detected by deep sequencing.



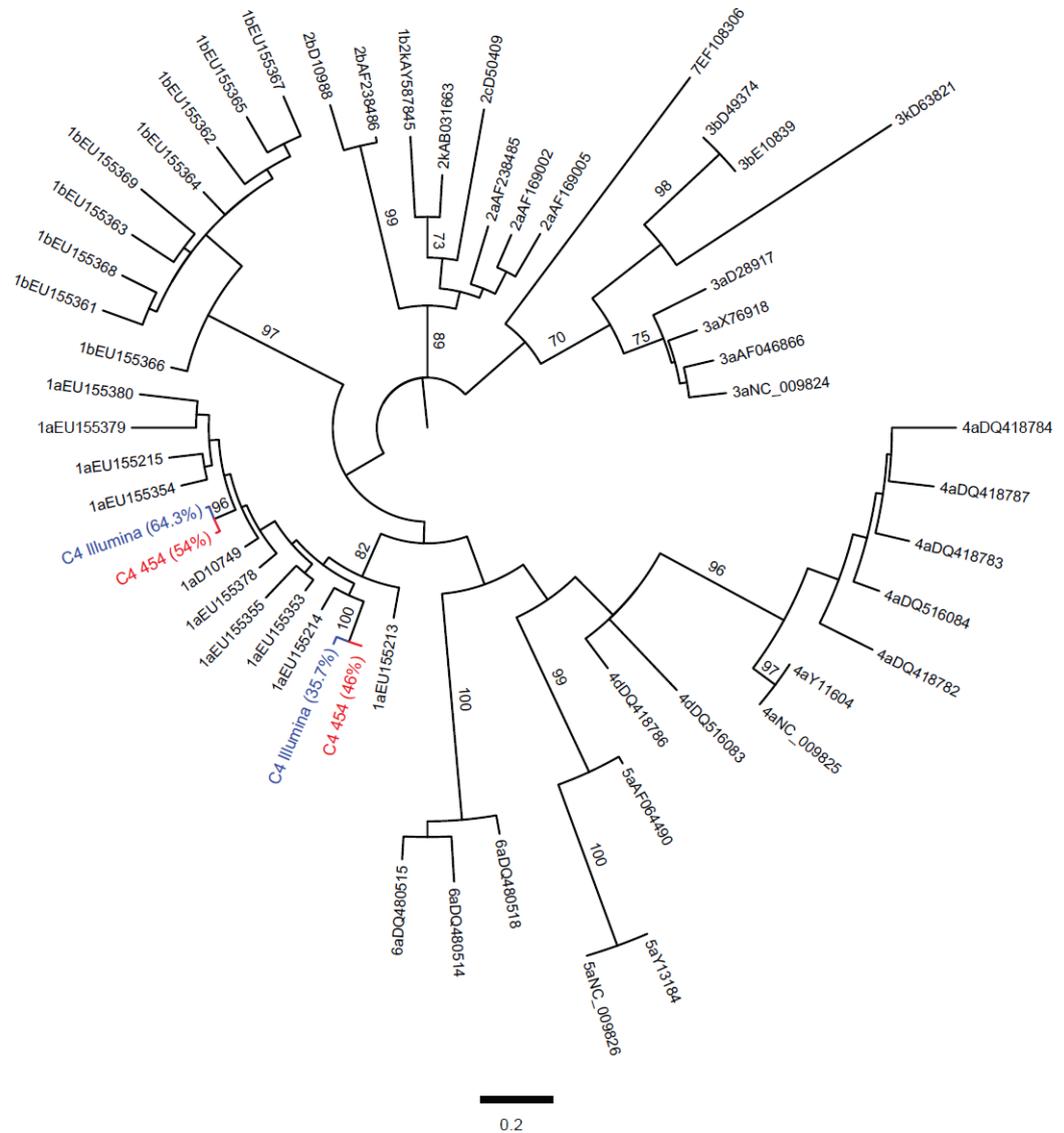
Maximum likelihood tree showing all sequences from different sequencing platforms; the direct sequence (DS) is shown in purple, clonal sequences (CS) in blue, 454 pyrosequences (NGS) in red, with the frequency of each variant in the total population shown in brackets. The analysis included; 16 clonal sequences, 587 reads derived from 454 pyrosequencing and selected HCV references from the Los Alamos HCV database. Values on the branches represent the bootstrap support.

Figure 3-7: Distribution of the number of variants detected by clonal analysis (CS) versus 454 pyrosequencing (NGS).



The comparison included 16 samples; NGS detected more variants than CS.

Figure 3-8: Comparison of variants detected using different NGS platforms.



Maximum likelihood tree showing sequences derived from different sequencing platforms; Illumina sequencing (**Illumina**) in blue, 454 pyrosequencing (**454**) in red and the frequency of each variant within the total viral population. The analysis includes; 695856 reads derived from Illumina sequencing, 16500 reads derived from 454 pyrosequencing, and selected HCV references from the Los Alamos HCV database. Values on the branches represent the bootstrap support.

In order to account for differences in circulating variants between US and UK isolates, we adapted primers and designed a selection of new primers using genotype 1a full genome sequences from the NCBI database as shown in Section 2.10.1.1.

3.2.6.1 Optimization of RNA extraction for PCR-based amplification of the HCV genome

We evaluated two RNA extraction methods prior to PCR amplification of HCV sequences; the QIAamp[®] Viral RNA Mini Kit (Qiagen) and the automated NucliSENS[®] easyMAG[®] (BioMérieux). Extractions were performed according to the manufacturer's recommendations as described in Section 2.2. Each sample was extracted, and amplification of the 5'UTR region carried out in triplicate using real-time PCR. The mean threshold cycle (C_t value) was calculated as the mean C_t value of the triplicate of each sample, and it was compared using each technique. Both methods resulted in similar results (p value=0.75) (Figure 3-9).

3.2.6.2 Optimisation of cDNA synthesis

We tested several conditions and concentrations of the reverse transcriptase enzyme SuperScript[®] III (Invitrogen) and subsequently Maxima RT with a view to optimising the synthesis of long strands of cDNA in order to facilitate reverse transcription of the entire HCV genome (9.6 kb).

➤ Choice of annealing/elongation temperature

The manufacturer's recommendation of 55°C for the elongation step was tested using a temperature gradient of 37°C, 42°C, 50°C, 52°C and 55°C as well as different elongation times. The highest yield of cDNA was detected by quantitative PCR when using a temperature of 37°C (Figure 3-10).

➤ Primer selection for cDNA synthesis

We next compared cDNA synthesis using random hexamers versus HCV-specific primers as detailed in Section 2.3.

Different oligonucleotide primers which were tested included random hexamers and primers that annealed to the NS5B or 3' untranslated region (UTR).

Random hexamers generated shorter fragment lengths of 500-800 base pairs while two specific primers T4R9214 that anneals to NS5B and T4R9575 that anneals to the 3' UTR of HCV RNA, both resulted in the production of cDNA encompassing the entire HCV ORF. Using eight samples for a direct comparison, both primers resulted in similar levels of amplified PCR product as quantified by RT-PCR (Figure 3-11).

➤ Choice of reverse transcriptase enzyme

Nine samples were tested for cDNA synthesis using SuperScript III versus Maxima[®] reverse transcriptase (RT) (Thermo Scientific). Ct values generated following real-time qPCR suggested that the enzymes were comparable with no significant difference between mean Ct values (Figure 3-12).

3.3 PCR Amplification strategy

In order to maximize PCR sensitivity, we divided the genome into four overlapping amplicons that were numbered sequentially as amplicons 1 to 4 starting from the 5' end of the genome. Each amplicon was less than 3 kb (Figure 2-5).

3.3.1 Optimization of nested PCR

Nested PCR was used for all PCR reactions, and PCR cycling conditions were optimised for each amplicon. Negative controls lacking template were included for every primer set. If any negative control was positive, all PCR reactions in that set were deemed to be contaminated and discarded.

We optimized nested PCR conditions for each amplicon independently. Initially, we tested the primer sets and cycling conditions for nested PCR reported by Yao and Tavis (Yao and Tavis, 2005); in their study, primer design was based on sequences from US isolates.

In an initial optimisation with four samples, we were only able to amplify the Amp 4 region in two samples, while failing to amplify other amplicons consistently in these samples (Figure 3-13).

We aimed to incorporate viral genetic heterogeneity into our primer design. Primers targeted conserved regions of the genome to facilitate annealing to a maximal number of isolates.

3.3.1.1 Amplicon 1

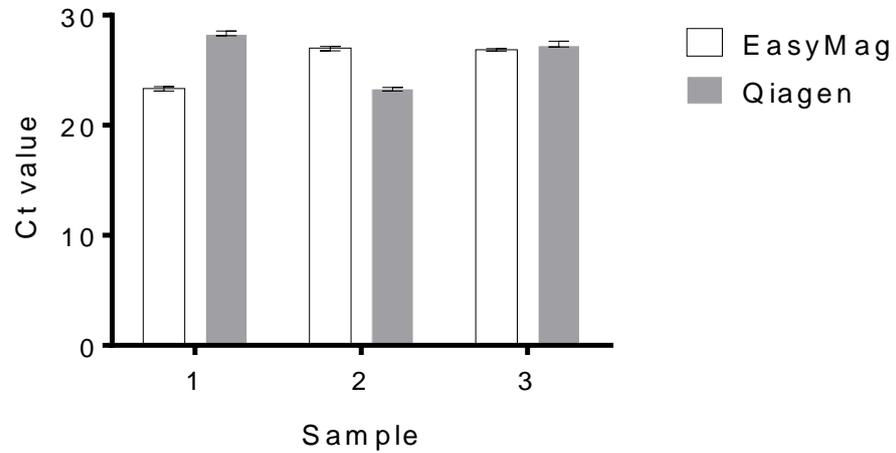
The nested PCR was optimised with the following protocol; first run, in a 50 μ L volume reaction, 25 μ L High Fidelity Phusion buffer 2X, 2.5 μ L reverse primer (T1R2576) 10 μ M, 2.5 μ L forward primer (T1F20) 10 μ M, 10000 copies minimum cDNA, adding water up to total volume of 50 μ L, cycling conditions; 98°C for 2 min, followed by 40 cycles [98°C (10 sec), 62°C (30 sec), 72°C (2:30 min)], then 72°C for 10min, finally hold at 4°C for ∞ .

For the second run, 1 μ L was added of first run product to 25 μ L of High Fidelity Phusion buffer 2X, 19 μ L water, 2.5 μ L reverse primer 10 μ M, 2.5 μ L forward primer 10 μ M, cycling conditions; 98°C 2 min, followed by 35 cycles [98°C (10 sec), 60°C (30 sec), 72°C (2 min)], then 72°C for 10min, finally hold at 4°C for ∞ . The resulting amplicon was 2345 base pairs; Figure 3-14 shows an illustrative example of the optimisation process.

3.3.1.2 Amplicon 2

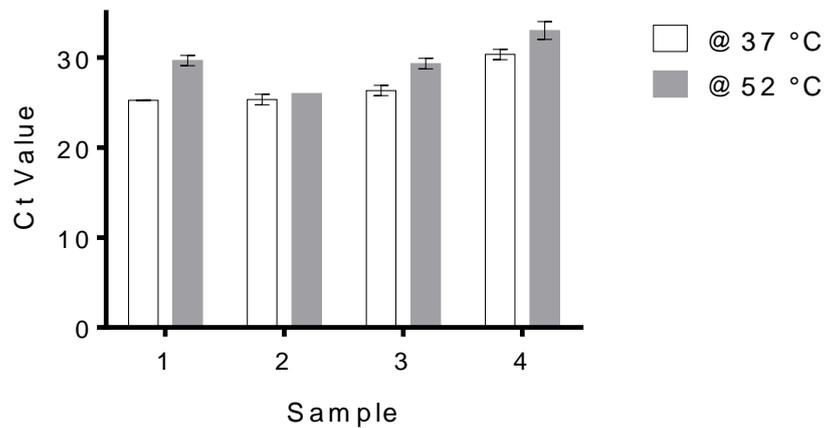
The nested PCR was optimised with the following protocol; first run, in a 50 μ L volume reaction, 25 μ L High Fidelity Phusion buffer 2X, 2.5 μ L reverse primer (T2R5536) 10 μ M, 2.5 μ L forward primer (T21275, T2F1305, T2F1397, or T2F2084) 10 μ M, 10000 copies minimum cDNA, adding water up to total volume of 50 μ L, cycling conditions; 98°C for 2 min, followed by 35 cycles [98°C (10 sec), 58°C (30 sec), 72°C (4 min)], then 72°C for 10 min, finally hold at 4°C for ∞ .

Figure 3-9: Comparison of the performance of different RNA extraction methods.



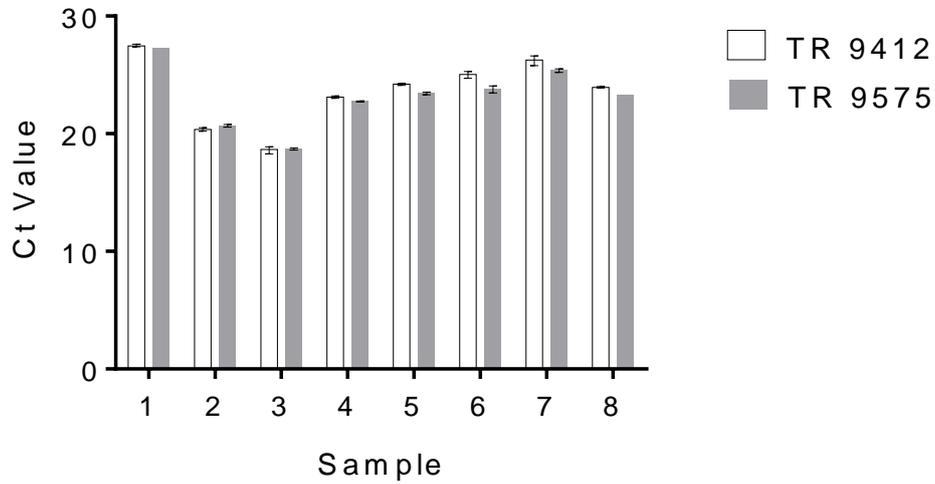
Performance was compared using the yield of RT-PCR carried out using three samples extracted by both methods; QIAamp[®] Viral RNA Mini Kit (Qiagen) and easyMAG[®] (BioMérieux) were used, with extracts processed in triplicate.

Figure 3-10: Optimisation of incubation temperature for cDNA synthesis.



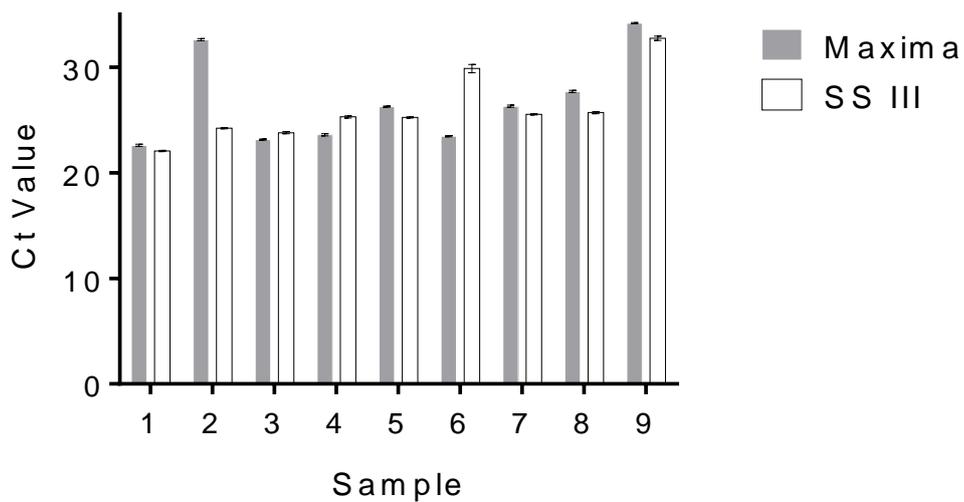
Performance was compared using the yield of RT-PCR carried out using the same sample processed for cDNA synthesis using the reverse transcriptase enzyme SuperScript[®] III (Invitrogen) at 37°C versus 52°C. Four samples were used, and each cDNA product was processed in triplicate.

Figure 3-11: Comparison of the yield of cDNA synthesized using different HCV-specific primers.



Performance was compared using the yield of RT-PCR carried out using eight samples in triplicates. The compared primers are reverse primers (TR 9412 located in NS5B and TR 9575 in 3'UTR region).

Figure 3-12: Comparison of cDNA synthesis using Maxima[®] versus SuperScript III[®].



Performance was compared using the yield of RT-PCR carried out using nine samples in triplicates.

For the second run, 1 μL of first run product was added to 25 μL of High Fidelity Phusion buffer 2X, 19 μL water, 2.5 μL reverse primer (T2R5469, T2R5020, or T2R5230) 10 μM , 2.5 μL forward primer (T2F2662) 10 μM , cycling conditions; 98°C 2 min, followed by 35 cycles [98°C (10 sec), 58°C (30 sec), 72°C (2:30 min)], then 72°C for 10 min, finally hold at 4°C for ∞ . The resulting amplicon was 2817 base pairs. Figure 3-15 shows an illustrative example of the optimisation process.

3.3.1.3 Amplicon 3

The nested PCR was optimised with the following protocol; first run, in a 50 μL volume reaction, 25 μL High Fidelity Phusion buffer 2X, 2.5 μL Reverse primer (T3R7987) 10 μM , 2.5 μL Forward primer (T3F4725) 10 μM , 10000 copies minimum cDNA, adding water up to total volume of 50 μL , cycling conditions; 98°C for 2 min, followed by 35 cycles [98°C (10 sec), 50/60°C (30 sec), 72°C (3 min)], then 72°C for 10 min, finally hold at 4°C for ∞ .

For the nested PCR, 1 μL of first run product was added to 25 μL of High Fidelity Phusion buffer 2X, 19 μL water, 2.5 μL reverse primer (T3R7109) 10 μM , 2.5 μL forward primer (T3F5001) 10 μM , cycling conditions; 98°C for 2 min, followed by 35 cycles [98°C (10 sec), 60°C (30 sec), 72°C (2 min then 72°C for 10 min, finally hold at 4°C for ∞ , or 2.5 μL Reverse primer (T3R7987) 10 μM , 2.5 μL Forward primer (T3F5001) 10 μM , cycling conditions; 98°C 2 min, followed by 35 cycles [98°C 10s, 60 °C 30s, 3:00 72°C], then 72°C for 10 min, finally hold at 4°C for ∞ . The resulting amplicon was 2986 base pairs. Figure 3-16 shows an illustrative example of the optimisation process.

3.3.1.4 Amplicon 4

The nested PCR was optimised with the following protocol; first run, in a 50 μL volume reaction, 25 μL High Fidelity Phusion buffer 2X, 2.5 μL Reverse primer (T4R9575 or T4R9214) 10 μM , 2.5 μL Forward primer (T4F6290) 10 μM , 10000 copies minimum cDNA, adding water up to total volume of 50 μL , cycling conditions; 98°C for 2 min, followed by 35 cycles [98°C 10s, 60 °C 30s, 3:00

72°C], then 72°C for 10 min, finally hold at 4°C for ∞. For the nested PCR, 1 µL of first run product was added to 25 µL of High Fidelity Phusion buffer, 19 µL water, 2.5 µL reverse primer (T3R7109) 10µM, 2.5 µL forward primer (T3F5001) 10µM, cycling conditions; 98°C 2 min, followed by 35 cycles [98°C (10 sec), 60°C (30 sec), 72°C (2 min)], then 72°C for 10 min, finally hold at 4°C for ∞, or adding 1 µL of first run product to 25 µL of High Fidelity Phusion buffer 2X, 19 µL Water, 2.5 µL reverse primer (T4R9214) 10µM, 2.5 µL forward primer (T4F6707/T4F7085) 10 µM, cycling conditions; 98°C for 2 min, [98°C (10 sec), 55°C (30 sec), 72°C (2:30 min)], then 72°C for 10 min, finally hold at 4°C for ∞. The resulting amplicon was 2507 base pairs. Figure 3-17 shows an illustrative example of the optimisation process.

3.3.2 The PCR amplification process

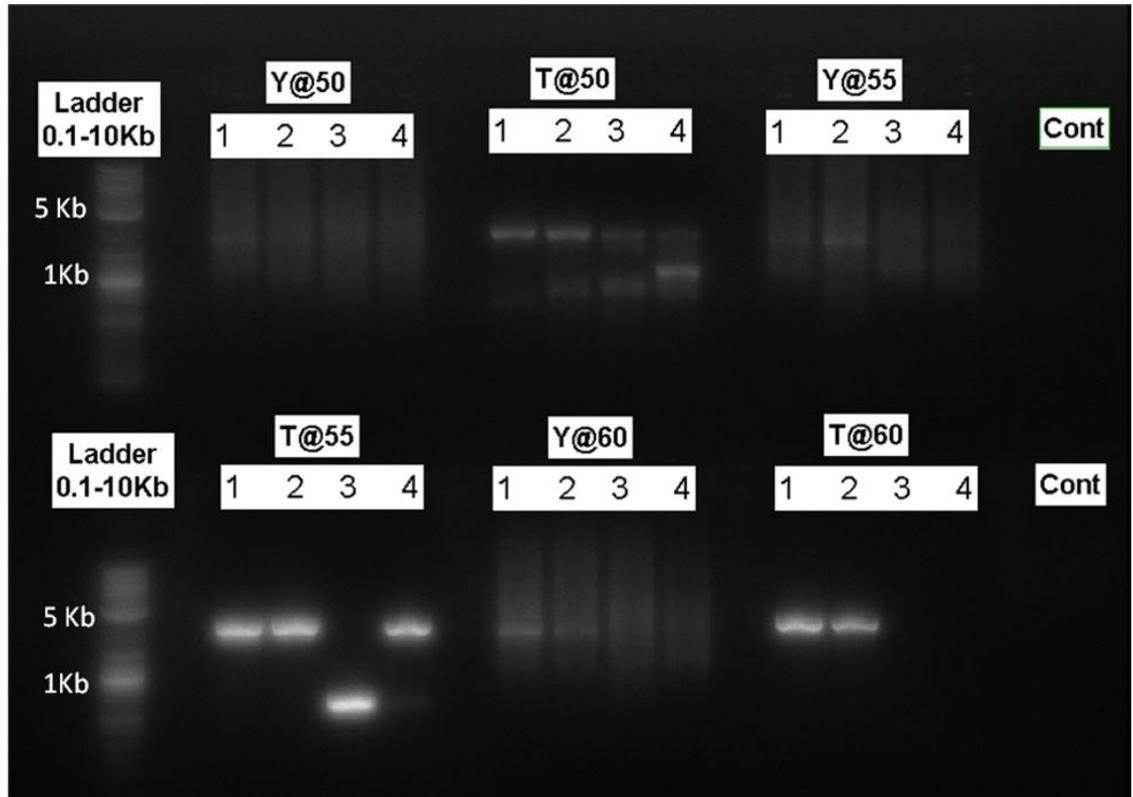
Despite the high sequence diversity of HCV, successful amplification of a near full-length (~9 kb) HCV genotype 1 genome by assembling sequences from Amp 1-4, was completed in 93% of patient plasma samples with viral RNA levels greater than or equal to 1000 copies/ml (Figure 3-18).

3.3.2.1 Sequencing

Each set of sequencing primers contained three primers; including both sense and anti-sense primers to allow complete coverage of each amplicon. Using Sanger sequencing, mean read length was approximately 1kb (Figure 3-19).

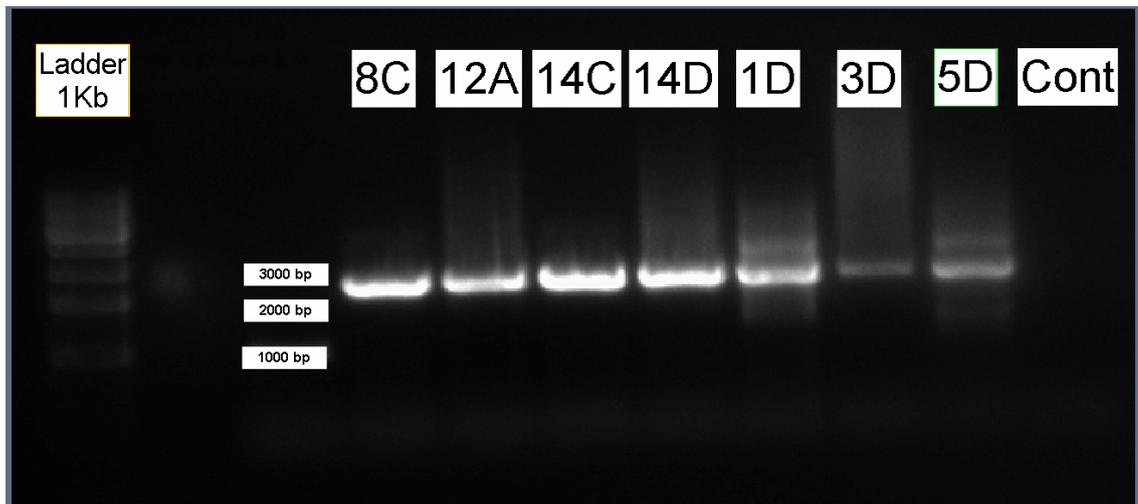
Beckman Coulter Genomics UK carried out sequencing using the ABI automated dye-terminator system. Consensus sequences were assembled and edited using CLC Genomics 6. This program automatically assembles overlapping sequence reads and identifies nucleotide positions with discrepancies between overlapping regions. Computer base-calling errors were corrected following visual inspection of the corresponding sequence chromatograms. Mixed-base positions within the HCV quasispecies were resolved by identifying the predominant base manually at each position.

Figure 3-13: Optimisation of modified primer sets for whole genome amplification of HCV genotype 1a.



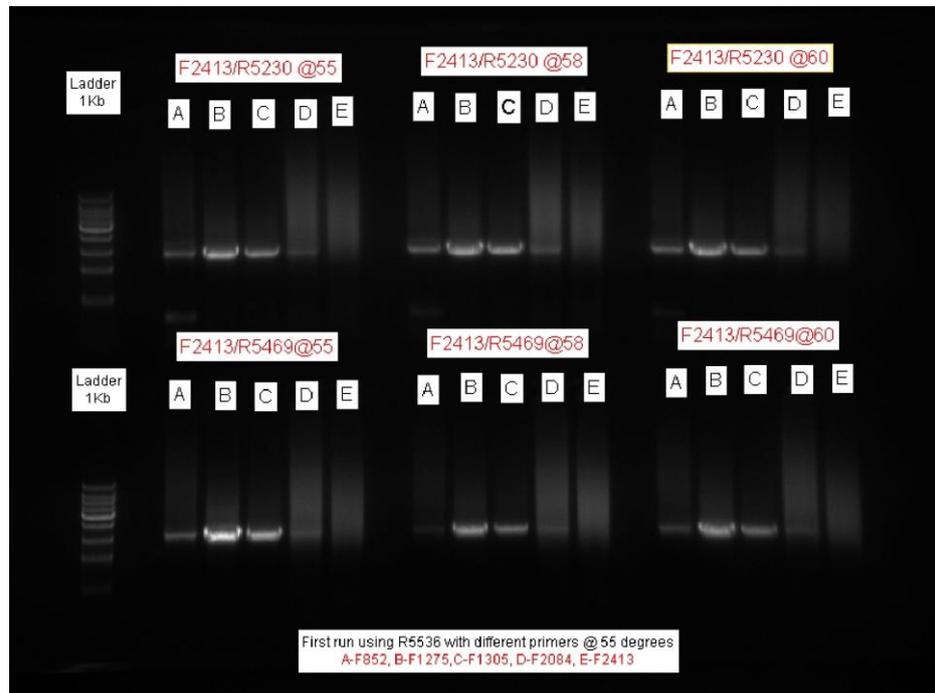
PCR amplification of Amp 4 was carried out using two sets of primers; those described by Yao and Tavis, 2005: Y and a set of modified primers: T. Primer set Y consisted of outer primers A4xF3 and Out7KbR, and inner primers A4xF3 and F7R. Primer set T (Modified primers) consisted of outer primers T4F7085 and Out7KbR, and inner primers T4F7085 and F7R, both sets were compared at different annealing temperature (50, 55, 60°C).

Figure 3-14: Optimisation of PCR amplification of Amp1.



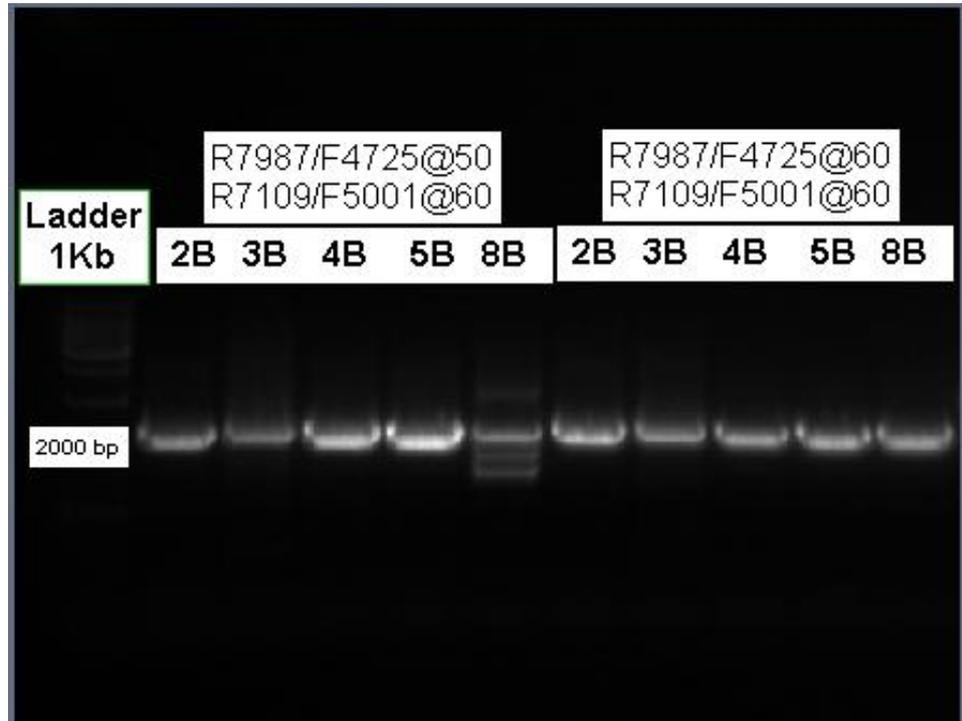
Amplicon length is 2345 bp, optimisation was carried out using seven sera samples to estimate the amplification efficiency of the primers.

Figure 3-15: Optimisation of PCR amplification of Amp 2.



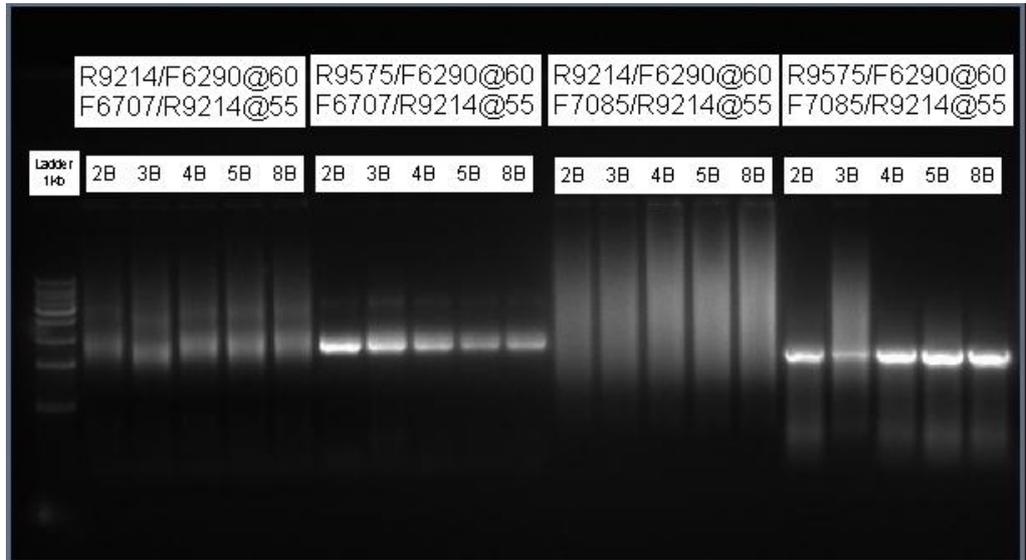
Optimisation was performed using different primer sets; Illustrated are two different outer primer sets (F2413 and R5469), and (F2413 and R5230) with the second run performed using the one reverse primer (R5536) and five different forward primers (F852, F1275, F1305, F2084, and F2413).

Figure 3-16: Optimisation of PCR amplification of Amp 3.



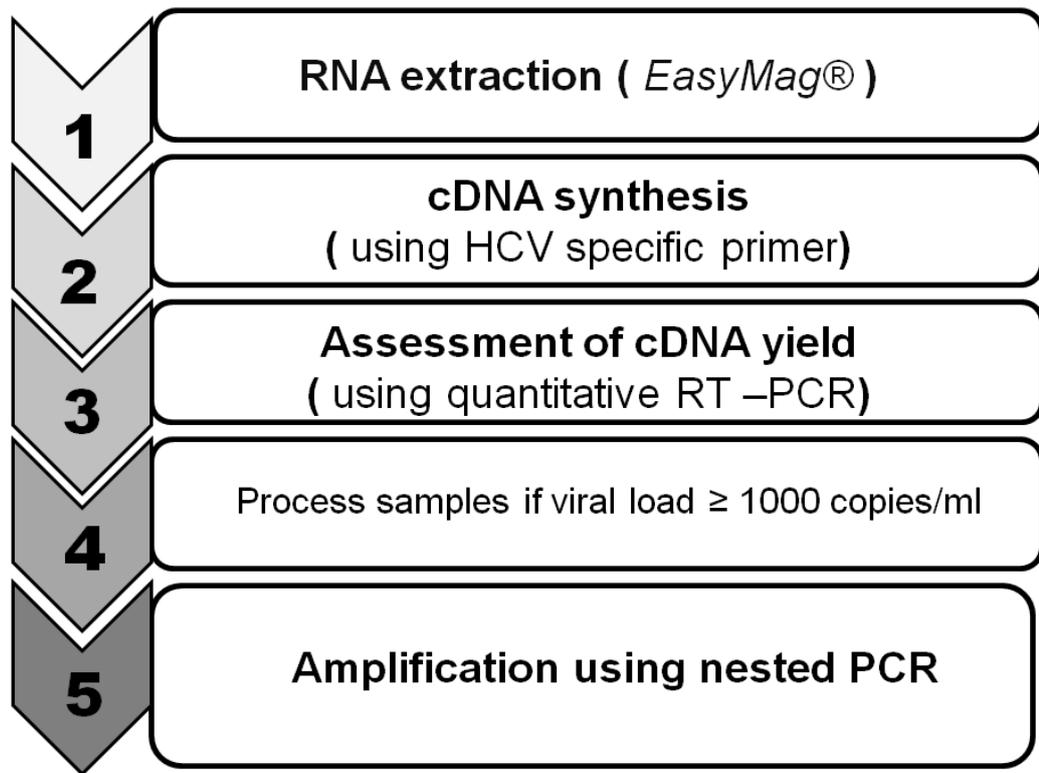
Amplicon length is 2108 bp, optimisation was carried out using five sera samples, Primer set consisted of outer primers (R7987 and F4725), and inner primers (R7109 and F5001), the set was tested using two different annealing temperature in the first run (50°C and 60°C).

Figure 3-17: Optimisation of PCR amplification of Amp 4.



Optimisation was carried out using five samples; different Primer sets were tested. Two sets were able to amplify Amp 4; outer primer set (R9575 and F6290) with two inner sets (F6707 and R9214) or (F7085 and R9214).

Figure 3-18 : PCR Amplification process.



The cut-off for successful sequencing of clinical samples using amplicon sequencing was estimated to be 1000 copies/ml.

Where necessary, additional sequencing reactions were performed to confirm the identity of a base or its predominance in the quasispecies population. For accuracy, we stipulated that each nucleotide was present in at least two unambiguous sequencing reactions.

3.3.3 Metagenomic sequencing

We next assessed the possibility of sequencing whole HCV genomes using a metagenomic sequencing approach utilising an in-house Illumina[®] MiSeq platform. The central aim was to provide a detailed method for the construction of unbiased metagenomic libraries from sera and to develop bioinformatics pipelines for the assembly of sequencing data.

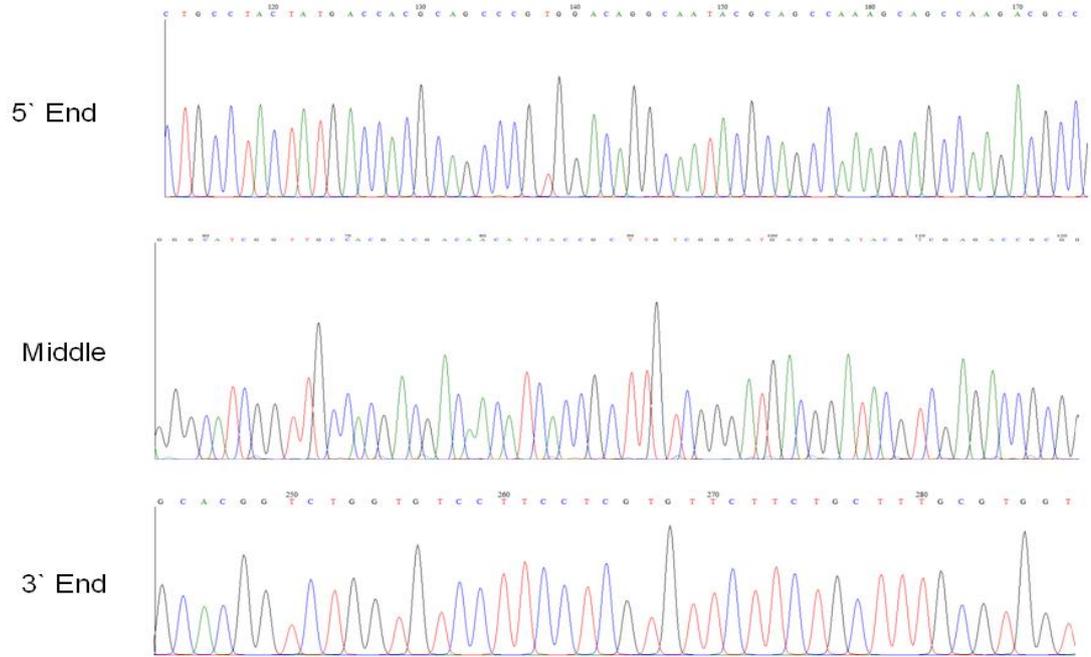
3.3.3.1 Construction of Metagenomic Sequencing Libraries

Quantitative RT-PCR on HCV RNA was carried out prior to library preparation. If the viral load was above 1000 copies/ml, double-stranded cDNA was synthesised as described in Section 2.4. This was fragmented and tagged using Nextera XT[®] (Illumina) as described in 2.9.3.2.

➤ Evaluation of library size

The quality and fragment size of each library was measured using a TapeStation platform as described in Section 2.9.3.3. Following AmpureXP[®] bead purification, purified products were visualised. Fragment lengths were dependent on the concentration of input DNA used in the Nextera XT[®] reaction. We initially used 1ng DNA as recommended by the manufacturer but noted a shift towards longer fragment length with increased DNA concentration, reaching saturation at 10 ng (Figure 3-20).

Figure 3-19: Example of chromatograms derived from Sanger sequencing of HCV amplicons.



The sequencing was performed by Beckman Coulter Genomics. Results were received in fasta files associated with their chromatograms.

The DNA input quantity was therefore varied in order to achieve the fragment length required for Illumina[®] sequencing. For 300X MiSeq cartridges, (optimal fragment length 300-400 base pairs), we used 1ng input DNA while for the 600X kit, a higher amount of DNA input (5 ng) was used (optimal fragment length 500 base pairs).

➤ **Samples and Preparation of Nucleic Acids**

❖ *Optimisation of RNA extraction for metagenomic sequencing*

As for PCR-based sequencing, we assessed two RNA extraction methods; column-based extraction using the QIAamp Viral RNA Mini Kit[®] (Qiagen) and magnetic bead extraction using the easyMag[®] (Biomerieux) platform.

Following sequencing of samples extracted using each method, metagenomic sequencing was carried out. In samples extracted using column based extraction (based on the use of carrier RNA), we found that while larger concentrations of RNA were eluted, greater quantities of carrier RNA were present in the sample as indicated by a very marked AT bias. This was particularly evident when random hexamers were used for cDNA synthesis rather than HCV-specific primers. The quality and size of RNA extracted by each method were assessed using the TapeStation system as described in Section 2.9.3.3; the TapeStation-RNA screen tape demonstrated a peak in keeping with that of carrier RNA whereas lower concentrations of RNA were found following easyMAG[®] extraction. EasyMAG[®] was subsequently adopted as the RNA extraction method of choice for metagenomic sequencing as it does not use carrier RNA that may interfere with downstream sequencing.

❖ *Optimisation of cDNA synthesis*

• **Superscript III versus Maxima Reverse Transcriptase**

Two reverse transcription systems Maxima[®]RT (Thermo Scientific) and SuperScript[®] III RT (Invitrogen) were compared for metagenomic sequencing. Ten 10 samples were processed using both systems. The results were analysed using

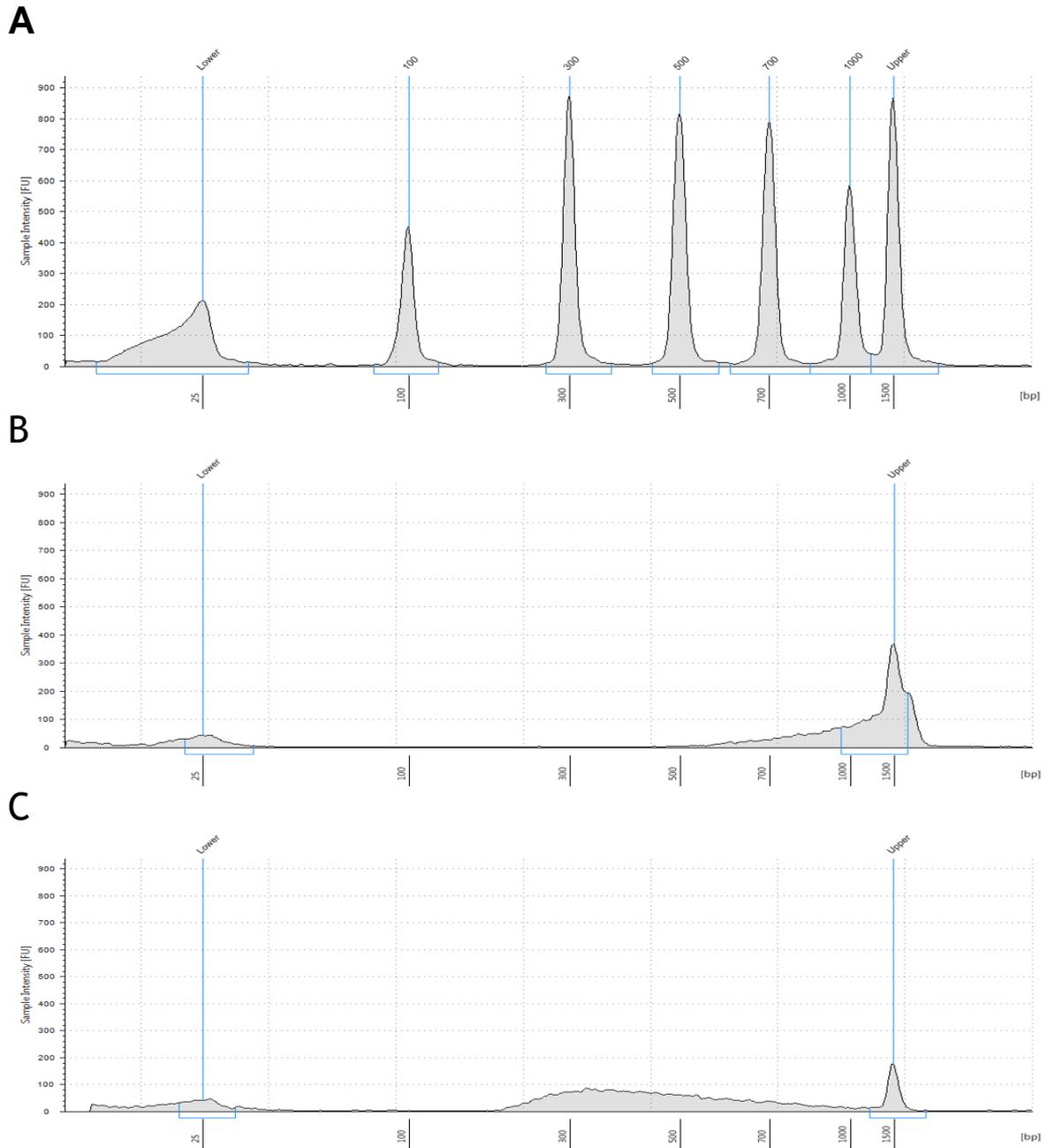
bioinformatics pipeline described in Section 3.3.4.1. The results were comparable with no evidence of inhibition of downstream sequencing. Both systems generated similar ratios of viral reads to total reads with ranges of 0.05-2% for SuperScript III[®] RT and 0.07-3.1% for Maxima[®] RT; this difference was not statistically significant with a p value of 0.08 as illustrated in Table 3-3, Table 3-4, and Figure 3-22.

- Primer choice for reverse transcription

As described above, 0.05%-3.1% of total reads generated by metagenomic sequencing was found to map to the HCV genome. We next investigated whether the use of random hexamers versus HCV-specific primers resulted in a difference in the number of HCV reads generated following metagenomic sequencing. There was a trend for the generation of more HCV reads using random hexamers. However, there was no statistical difference observed between the percentages of HCV viral reads detected when using random hexamers compared to HCV genome-specific primers (Table 3-5 and Figure 3-23).

In conclusion, a process for metagenomic sequencing of the HCV genome was developed incorporating RNA extraction using easyMAG[®], cDNA synthesis using random hexamers, quantification of input HCV using RT-PCR with a cut off of 1000 copies/ml and quality assessment using TapeStation DNA D1K[®] screen tape (Figure 3-24).

Figure 3-20: Effect of input DNA concentration in Nextera XT[®] on final library size.



Nextera library size was measured using the D1K DNA screen tape (TapeStation platform), A) DNA Ladder, B) DNA input 10ng, C) DNA input 1 ng.

3.3.4 Illumina[®] sequencing of HCV full genome

We next compared nested PCR amplification and metagenomic sequencing for sample processing using the MiSeq[®] NGS platform (Illumina[®]). We compared three samples using both approaches. In the PCR approach arm, the four amplicons were mixed in equimolar concentrations prior to library preparation using Nextera XT[®] while in the metagenomic arm the library covered the whole HCV genome.

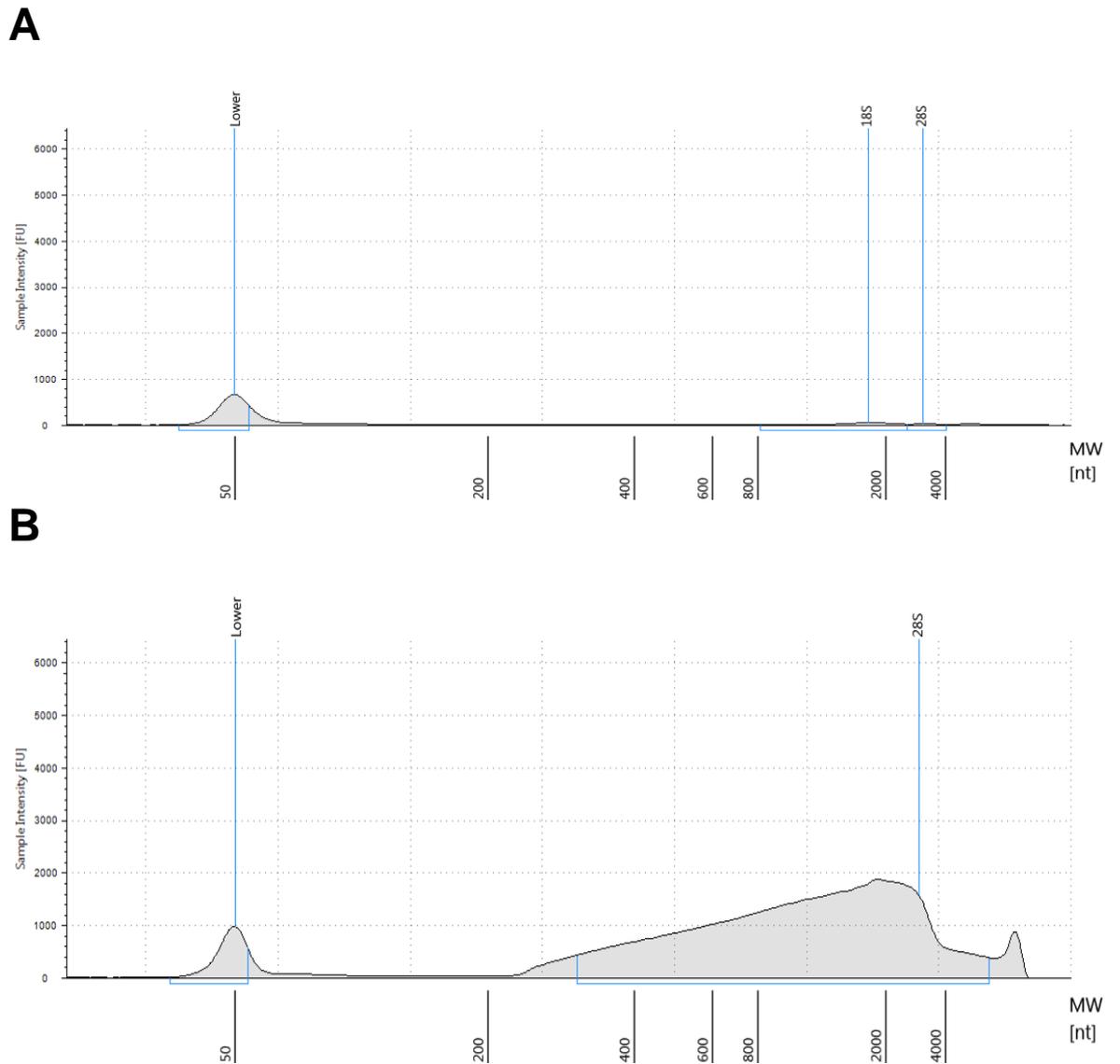
3.3.4.1 Bioinformatics analysis Pipeline

Sequences were mapped to a reference genome using an in-house bioinformatics pipeline.

➤ Quality analysis of deep sequencing raw data.

Quality screening of the raw sequence data generated in Fastq format was carried out using FastQC[®] (Babraham Bioinformatics) software. Low quality regions present at the ends of the sequences were filtered and then trimmed to reach a final overall Phred quality score of Q30 or above; during paired-end Illumina[®] sequencing, the second read was always of lower quality than the first read especially toward the end of each sequence.

Figure 3-21: RNA yield following extraction by different nucleic acid extraction methods.



RNA yield following extraction measured using TapeStation RNA screen tape. A) easyMag extract, B) QIAamp Viral RNA Mini Kit extract, the peak demonstrating large quantities of carrier RNA.

Table 3-3: The distribution of reads generated following Superscript III® cDNA synthesis.

ID	Total Reads ¹	Cleaned Reads ²	Unmapped reads ³	HCV reads (%Total) ⁴
1	1135885	941592	107654	8534 (0.75)
2	1371825	720036	116090	3158 (0.23)
3	978041	1069190	39596	6091 (0.62)
4	1092475	886308	65776	1013 (0.09)
5	987171	556186	131272	4477(0.45)
6	1267607	837250	48950	1976 (0.16)
7	711142	505972	36940	1192(0.17)
8	1189758	542722	47301	654(0.05)
9	1111914	621997	170456	22373(2)
10	923149	665188	89664	6829(0.74)

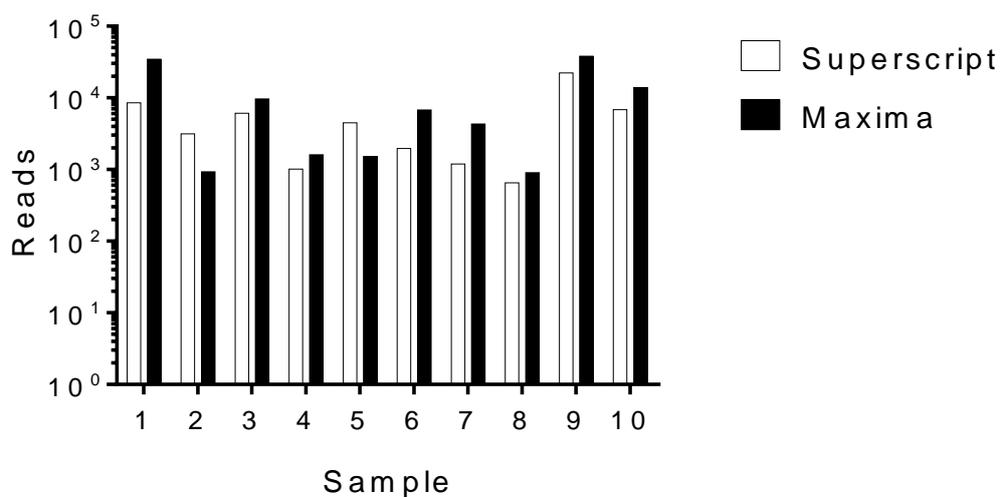
1-Number of sequence reads in the raw Fastq file following sequencing on the MiSeq platform, 2 -Number of reads retained after data clean up, 3-Number of reads not mapped to the human genome or cDNA, 4-Number of reads mapped to HCV reference from the total and its percentage.

Table 3-4: The distribution of reads generated following Maxima® RT.

ID	Total Reads ¹	Cleaned Reads ²	Unmapped reads ³	HCV (%Total) ⁴
1	1857873	941592	143939	34483 (1.86)
2	1280247	720036	73644	935 (0.07)
3	1812403	1069190	65235	9714 (0.54)
4	1519185	886308	102685	1607 (0.11)
5	1301716	556186	64086	1524 (0.12)
6	1353311	837250	115174	6750 (0.5)
7	1180880	505972	99801	4320 (0.37)
8	740678	542722	47814	904 (0.12)
9	1227695	621997	200667	38004 (3.1)
10	1168174	665188	141110	13909 (1.19)

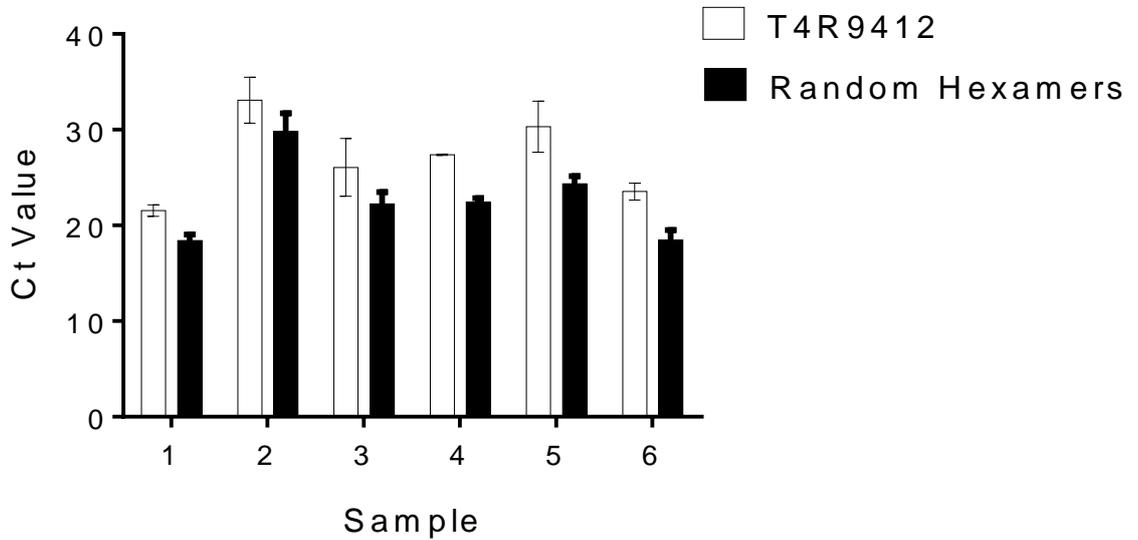
1-Number of sequence reads in the raw Fastq file following sequencing on the MiSeq platform, 2 -Number of reads retained after trimming of adapters and discarding unpaired reads using the in-house Weecleaner script, 3-Number of reads not mapped to the human genome or cDNA, 4-Number of reads mapped to HCV reference using Tanoti from the total and its percentage.

Figure 3-22: Number of mapped HCV reads generated using different cDNA synthesis systems.



The number of reads generated from each sample was not dependent on the method of cDNA synthesis.

Figure 3-23: Comparison of primer choice for reverse transcription.



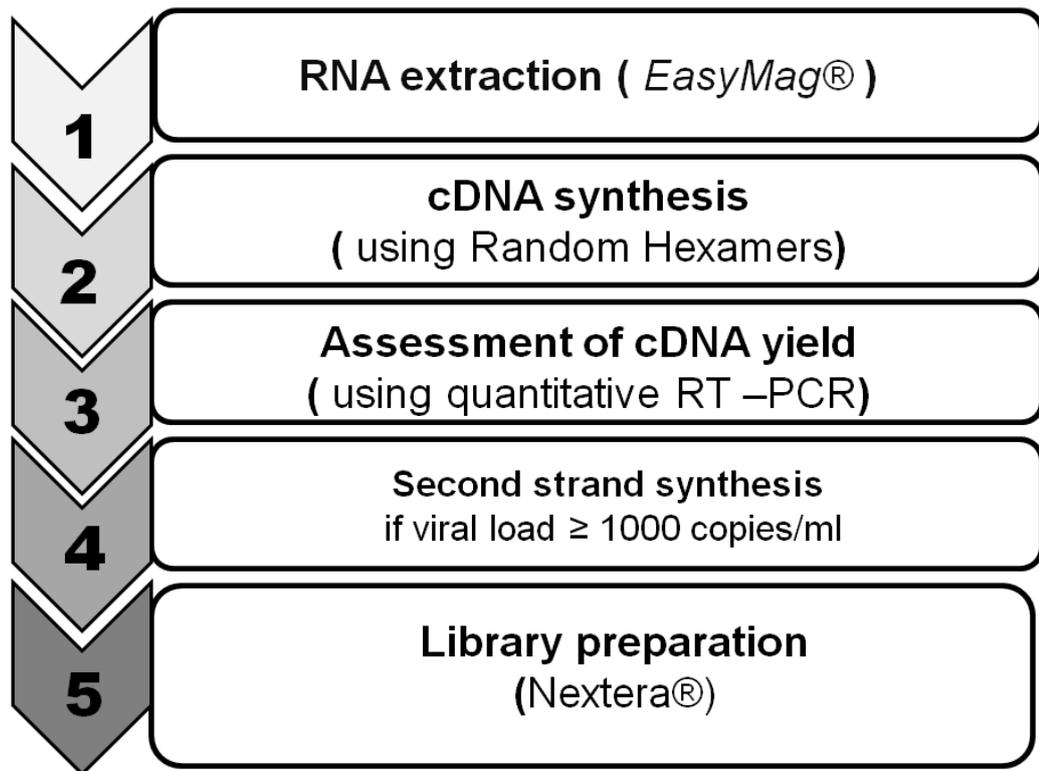
The comparison was carried out using 6 samples (1-6), then detecting the yield using qPCR targeting 5'UTR region.

Table 3-5: Comparison between random hexamers and HCV specific primer based cDNA synthesis.

	HCV specific primer (T4R9412)			Random Hexamers		
	1	2	3	1	2	3
Total reads ¹	1225095	439532	2195446	1332924	1814619	2377050
Clean reads ²	630519	152950	794551	405985	918920	732326
Unmapped reads (%) ³	158900 (13%)	71943 (16.3%)	180631 (8.2%)	127074 (9.5%)	111189 (6.1%)	144247 (6%)
HCV reads (%) ⁴	17896 (1.46%)	14347 (3.26%)	12335 (0.56%)	24172 (1.8%)	19979 (1.1%)	23024 (0.97%)

1-Number of sequence reads in the raw Fastq file following sequencing on the MiSeq platform, 2-Number of reads retained after trimming of adapters and discarding unpaired reads using the in-house Weecleaner script, 3-Number of reads not mapped to the human genome or cDNA, 4-Number of reads mapped to HCV genome using Tanoti and % of total reads.

Figure 3-24: Metagenomic approach process.



The cut-off for successful sequencing of clinical samples using metagenomic sequencing was estimated to be 1000 copies/ml.

➤ **Adapter Removal**

Different adapters were used during the library preparation prior to sequencing. We removed the standard Nextera® adapter CTGTCTCTTATACACATCT using Trim_galore (http://www.bioinformatics.babraham.ac.uk/projects/trim_galore). Cleaned reads are shown in Figure 3-26, the trimming led to shortening of the median sequence length from 150 to 131 base pairs, and improvement of the Phred quality score to Q30.

➤ **Next generation sequencing alignments**

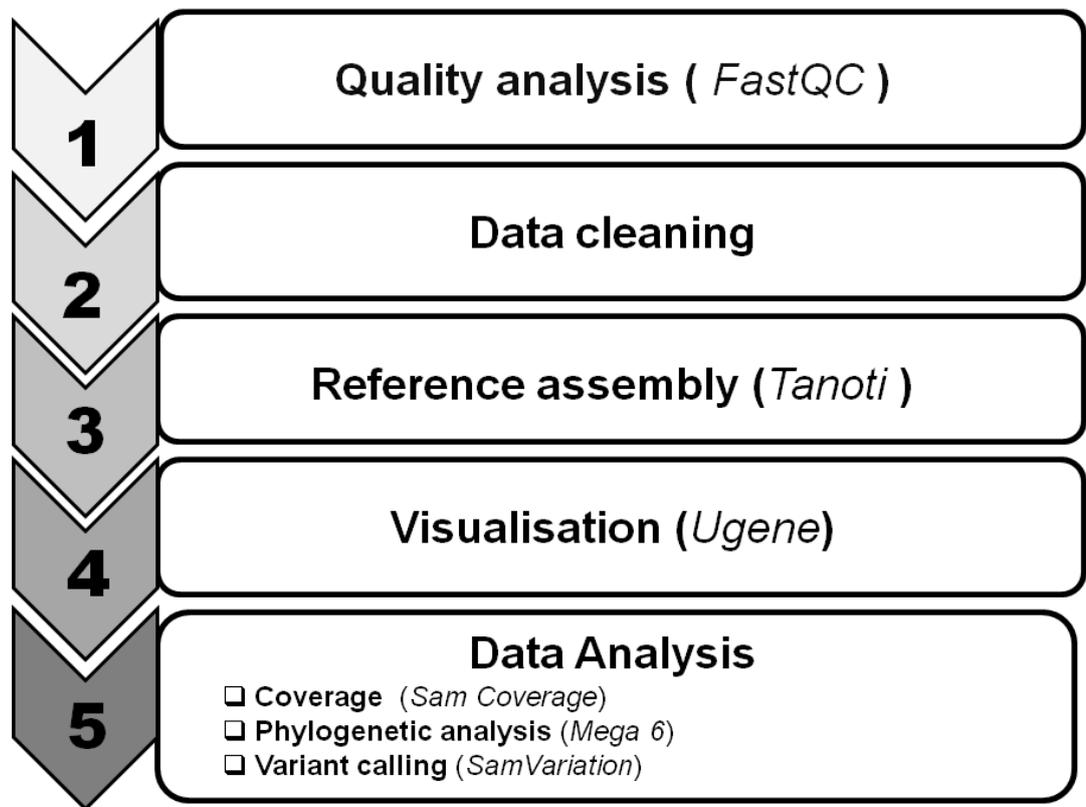
Using the reads generated by three samples P13, P22, P55, mapping with Tanoti yielded better coverage using the metagenomic approach (99.2%) than the nested PCR approach (97%). This was mainly attributable to mapping to the 5'UTR (the primers used annealed at site 210 on the reference HCV genome (H77)) (Table 3-6). However, the depth of coverage was higher using the PCR based approach (mean=17133) than the metagenomic approach (mean=172).

We next used an in-house script (SAM_Coverage) to calculate the coverage and depth of reads in each sample (Figure 3-27, Figure 3-28). The coverage in the samples processed by amplicon sequencing was bounded by the 5' and 3' locations of the primers used; both approaches covered the complete ORF, while amplicon sequencing presented a higher depth than metagenomic approach.

➤ **Quasispecies complexity**

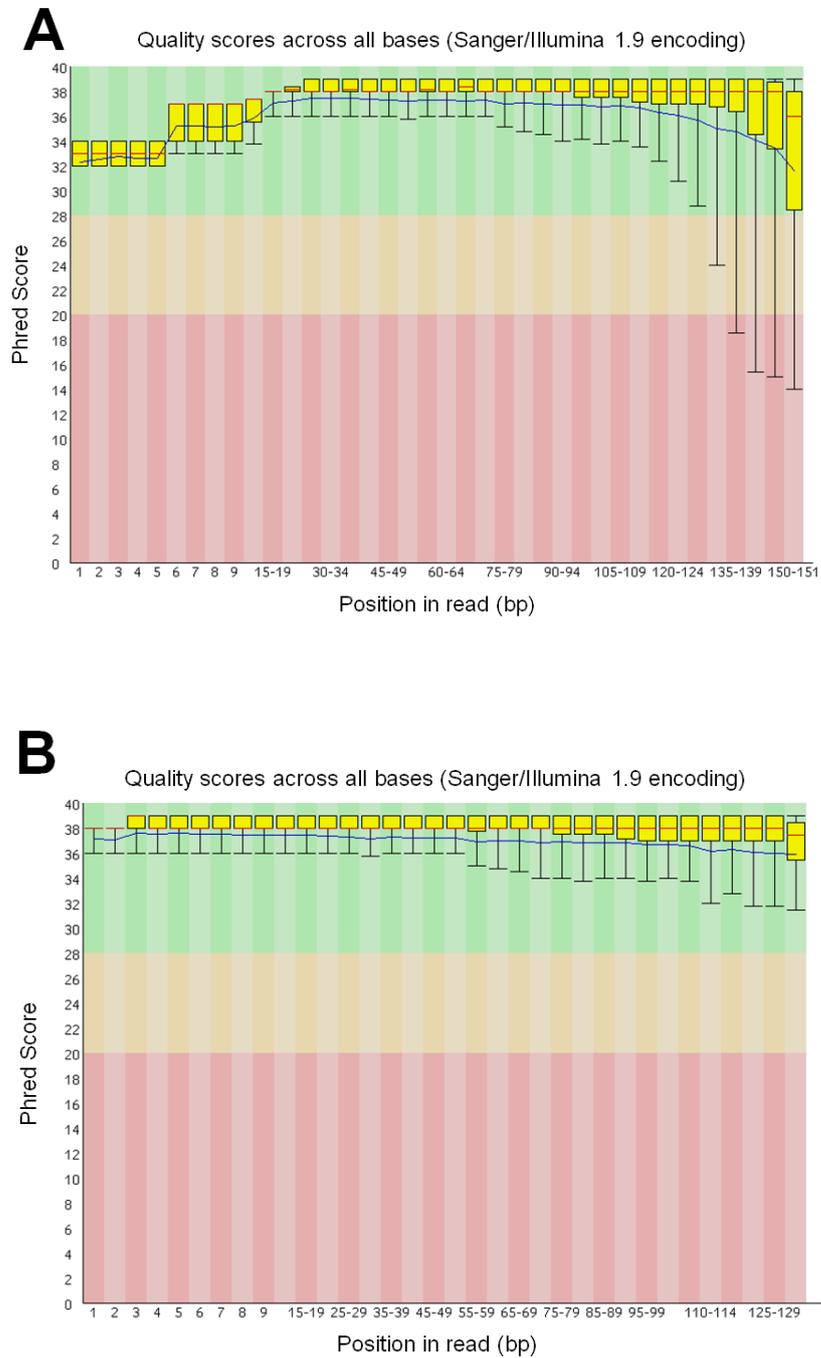
The analysis of samples processed using the metagenomic and amplicon-based approaches revealed that both generated very similar (but not identical) consensus sequences. The genetic distance between consensus sequences is illustrated in (Table 3-7). We created a maximum likelihood tree based on these consensus sequences and 10 genotype 1a reference variants from the NCBI database (Figure 3-29). It revealed that both approaches identified similar consensus sequence in the studied samples.

Figure 3-25: Data analysis pipeline.



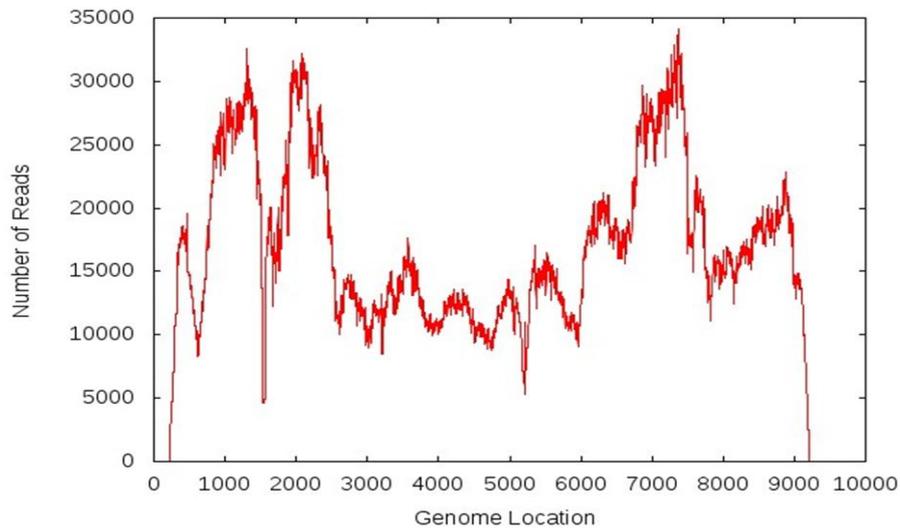
Different programmes and scripts were used in the pipeline, examples are presented in brackets.

Figure 3-26: Quality of reads generated using the amplicon sequencing approach.



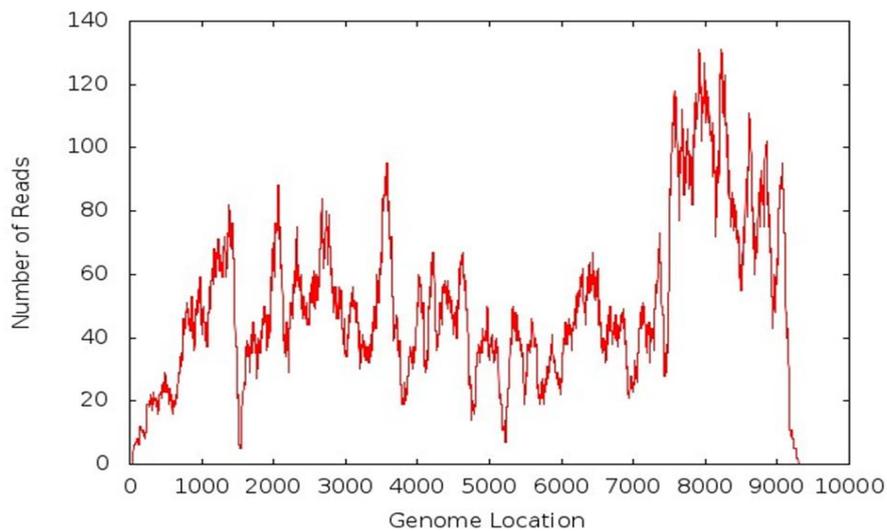
A) Raw reads (first read) generated by MiSeq run, B) cleaned reads (first read) cleaned by weecleaner.

Figure 3-27: An example of read coverage using the amplicon sequencing approach.



This was carried out using Sam_coverage script, showing coverage across HCV genome.

Figure 3-28: An example of read coverage using the metagenomic approach.



This was carried out using Sam_coverage script, showing coverage across HCV genome.

Table 3-6: Analysis of reads generated by metagenomic and nested PCR-based sequencing.

	Metagenomic approach			Nested PCR approach		
	Mapped area	Coverage	Average depth	Mapped area	Coverage	Average depth
P13	9303	98.80%	226	9120	96.86%	16980
P22	9384	99.66%	242	9129	96.95%	11615
P55	9261	98.35%	50	9185	97.55%	22805

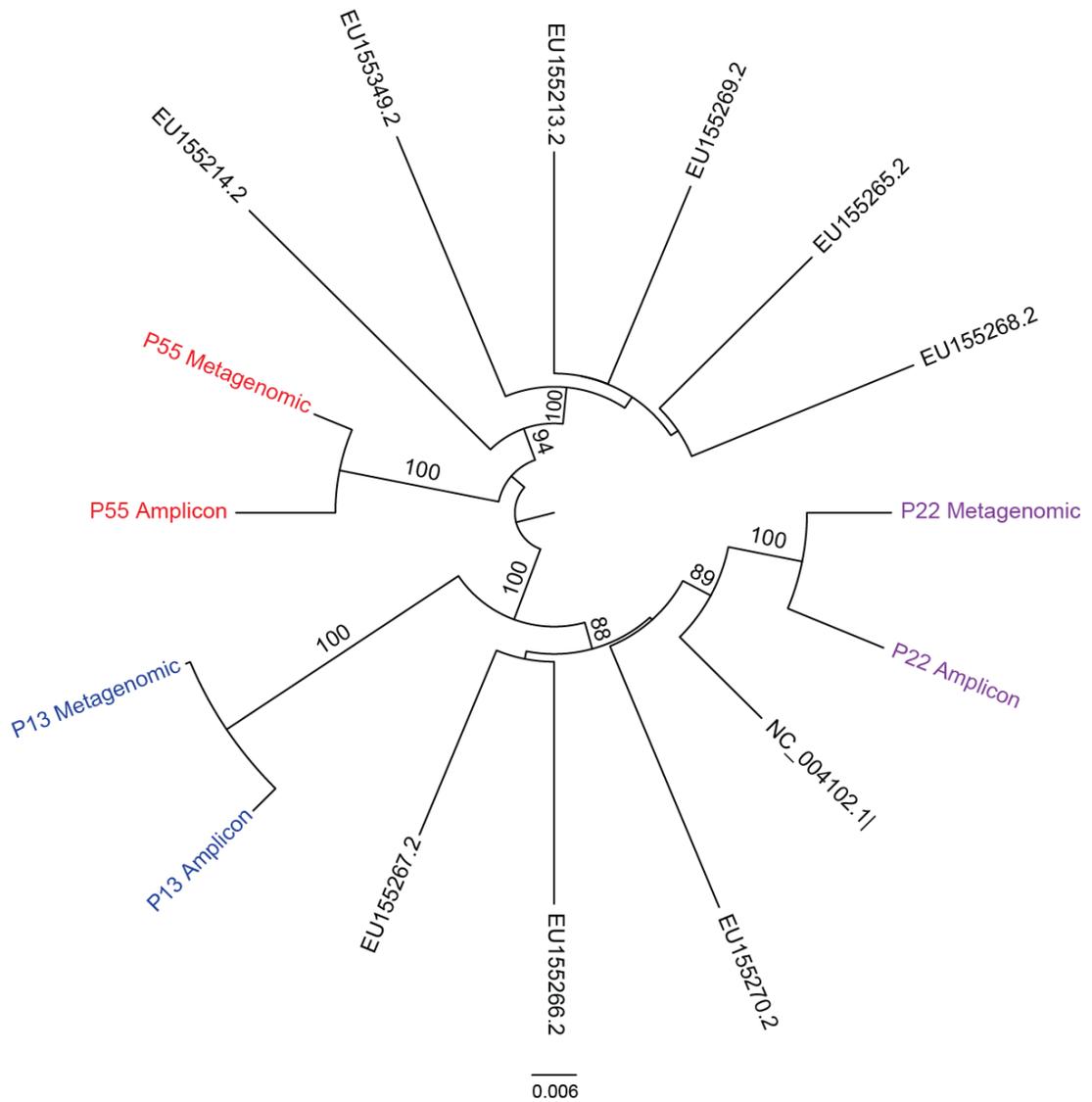
The comparison was carried out using 3 samples; The amplicon approach has limited coverage due to location of the inner primers used.

Table 3-7: Genetic distance between consensus sequences.

	P13	P22	P55
Genetic distance (Nucleotides)	0.50%	2.60%	1.90%
Genetic distance (Amino acids)	1.10%	5.90%	4.40%

All positions containing gaps and missing data were eliminated. There was a total of 9120 positions in the final dataset. The genetic distance between consensus sequences generated by metagenomic and nested PCR approaches was calculated using MEGA 6.

Figure 3-29: Comparison between consensus sequences generated by metagenomic and amplicon approaches.



Maximum likelihood tree illustrating consensus sequences derived using different approaches and selected genotype 1a HCV references from the Los Alamos.

3.3.5 Mock community experiment

We next studied two of the currently available haplotype reconstruction programmes (PredictHaplo and QuRe) and assessed their performance using an artificial mock community that we designed as described in Section 2.13.1. The mock community was created using 13 clones containing around 2.5 kb fragments of HCV viral variants which simulate the quasispecies in a clinical sample. The 13 clones were mixed with log-normal abundance with a range of frequency of [3.2-30.3%] in order to attempt to link sequences to identify whole viral genomes present within each sample.

The clones were each 2,225 - 2,377 base pairs in length. The pairwise genetic distance between the clones was 0- 0.9% as described in Table 3-8. The pJET1.2/blunt cloning vector containing the insert was fragmented using Nextera XT[®] kit, the library's average fragment length was 500 base pairs. The quality of reads was high with read quality per base over Q30, only deteriorating as expected towards the end of the second read.

3.3.5.1 Haplotype reconstruction programmes

Quasispecies assembly and evaluation of reconstruction performance of QuRe and PredictHaplo were applied to the mock community data set. QuRe calculated homopolymeric and non-homopolymeric error rates as estimated from mapping the reads to the plasmid reference (default parameters for the rest). PredictHaplo was run with default parameters.

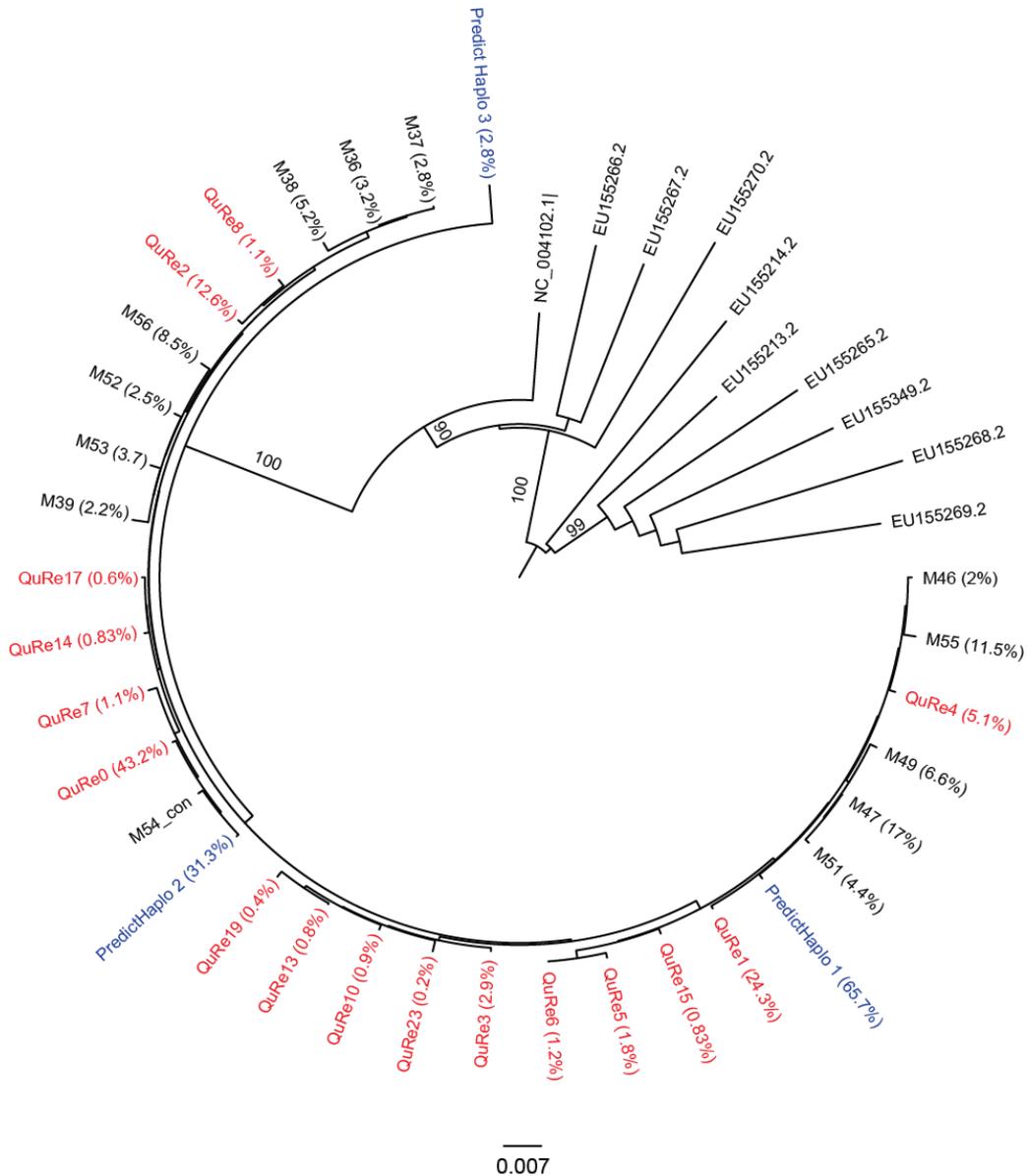
QuRe and PredictHaplo reconstructed 16 and 3 distinct variants respectively for the mock data set in two samples (3B and 3C) of the triplicate while the third sample (3A) revealed 16 and 5 distinct variants when using QuRe and PredictHaplo respectively. All variants were true variants with a precision of 100% for both algorithms. Reconstructed variants were classified as 'correct' when clustering with original Sanger clones in a phylogenetic tree at 75% bootstrap support. The analysis was done in collaboration with Dr Melanie Schrimmer.

Table 3-8: Mock community; pairwise distance among 13 clones.

Sequence		1	2	3	4	5	6	7	8	9	10	11	12
1	M36 (3.2%)												
2	M37 (2.8%)	0											
3	M38 (5.3%)	0.003	0.002										
4	M39 (2.2%)	0.005	0.005	0.006									
5	M46 (2%)	0.007	0.007	0.009	0.009								
6	M47 (17%)	0.006	0.006	0.006	0.007	0.003							
7	M49 (6.6%)	0.008	0.008	0.009	0.009	0.004	0.003						
8	M51 (4.4%)	0.008	0.007	0.008	0.009	0.003	0.002	0.004					
9	M52 (2.5%)	0.004	0.004	0.005	0.006	0.008	0.006	0.008	0.008				
10	M53 (3.7%)	0.004	0.004	0.005	0.005	0.007	0.006	0.008	0.008	0.004			
11	M54 (30.3%)	0.004	0.004	0.005	0.006	0.008	0.006	0.008	0.008	0.004	0.004		
12	M55 (11.5%)	0.007	0.007	0.009	0.009	0.004	0.004	0.006	0.004	0.009	0.008	0.009	
13	M56 (8.5%)	0.004	0.004	0.005	0.005	0.006	0.005	0.007	0.007	0.004	0.004	0.004	0.007

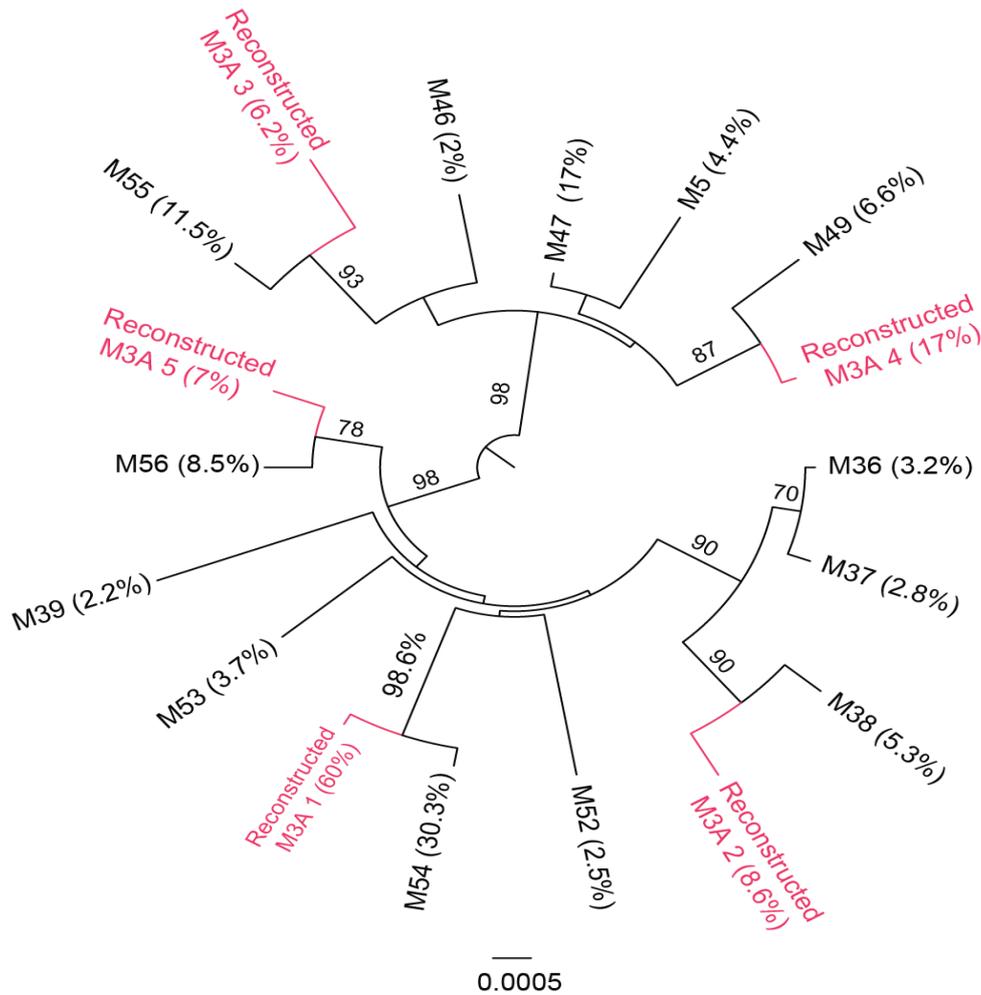
Pairwise distance calculated using Mega 6 software. The frequency of each clone in the mock community is shown in brackets.

Figure 3-30: Performance of different quasispecies assembler using the mock community



Maximum likelihood tree showing known sequences from mock community, and reconstructed haplotypes from QuRe in red, and PredictHaplo in blue. The frequency of each variant is shown in brackets. The values on the branches represent bootstrap support.

Figure 3-31: Reconstructed haplotypes using PredictHaplo in one of the triplicate (3A).



Maximum likelihood tree showing known sequences from Mock community (M3A), and reconstructed haplotypes from PredictHaplo (20119 reads); the frequency of each haplotype depicted in brackets. The values on the branches represent bootstrap support.

QuRe identified more variants and at least one variant clustered with each node on the phylogenetic tree from the original artificial set. All variant frequency estimation methods correlated significantly with the original distributions (Figure 3-30).

The haplotypes reconstructed using PredictHaplo (20119 reads), the reconstructed haplotypes matched the clones with the highest frequency in Mock community as shown in Figure 3-31.

3.4 Discussion

3.4.1 The role of NGS in understanding HCV viral population structure

Viral complexity is defined as the number of viral variants within the HCV quasispecies. A genetic distance of 15% between HCV genomes is the cut-off for the assignment of subgenotypes; a new class of subgenotype requires identification of three strains within that group (Smith et al., 2014). Genomic sequences of independent HCV isolates were reported to differ by approximately 10% (Nattermann and Tacke, 2009). However, there is no consensus on the cut-off of genetic distance to assign two isolates as different variants within a subgenotype.

In our cohort, multiple variant infections were detected in all patients, as two or more variants were detected in every sample. These were undetected by direct Sanger sequencing as it yields only a single consensus sequence. Smith *et al.* reported similar results in another cohort of 10 patients with acute HCV infection, where 80% of patients harboured more than 1 HCV subtype (Smith et al., 2010).

Clonal analysis is the standard method for studying viral quasispecies. However, this strategy lacks the sensitivity to detect minority variants. Minority variants present in viral populations may be critically relevant to the disease progression and response to treatment (Barzon et al., 2011).

Accurate genotyping of HCV is important for tailoring antiviral therapy, as well as for predicting the likelihood of response. As current diagnostic techniques fail to detect mixed infections (McNaughton et al., 2013, Abdelrahman et al., 2015), NGS methods may be increasingly adopted in the clinical setting if the emergence of minority variants is shown to occur following the larger scale use of DAAs. The first diagnostic assay likely to be performed using NGS will be those currently based on Sanger sequencing because of limited sensitivity. In the case of HCV, genotyping and antiviral resistance testing would be relevant future applications that may be developed using NGS.

Among currently used diagnostic techniques, the Abbott m2000 RealTime HCV Genotype II assay is able to resolve most HCV genotypes (~90%), but further sequencing is required to fully resolve the genotype in the remaining cases (Benedet et al., 2014). Importantly, commercial HCV genotyping assays, such as LiPA[®], underestimate the true prevalence of mixed HCV infection (Blackard and Sherman, 2007), and have error rates of HCV subtyping of 10-15.6%, with an intrinsic bias towards HCV genotype 1b (Sarrazin et al., 2007, Fan et al., 2009).

Sanger sequencing is the reference laboratory method for genotyping; while direct Sanger sequencing suffices for simple mutation patterns, it does not detect variants present at levels lower than 20% (Kwong et al., 2011). In multiple variant infections, Sanger sequencing often results in “miscalling” where the software is not able to assign a base due to the presence of double peaks. This was the case in 50% of samples in this study, where we were unable to identify the genotype in these samples using this technique.

Clonal analysis may provide a better understanding of the true viral population by avoiding miscalling and increasing numbers of clones sequenced. It is possible to sample any variant with equal chance, but the detection limit will always be more than 1/N, where N is the number of sequenced clones (Gregori et al., 2014).

The reported consensus in the literature has been between 15-20 clones sequenced, but this limits detection of drug-resistant variants to around 20-25% of the viral population in an individual sample (Barzon et al., 2011). Several studies in HIV have shown that minor RAVs are often responsible for the virological failure of a new antiretroviral treatment regimen (Barzon et al., 2011, Roquebert et al., 2006).

In our study cohort, we noticed that detection of variants by clonal analysis could be a very random process and having a cut-off is not accurate as a variant that has a frequency of 0.06% was detected by clonal analysis, while it missed another variant that had a variant frequency of 25% as reported by 454 pyrosequencing.

In this study, it was possible to detect multiple infections in all samples including, variants at low frequencies, which would go undetected by conventional methods. NGS detected mixed subtype/genotype infection in 37.5% of samples (6/16), two of which were missed by clonal analysis. The estimated frequency cut-off for detectability by different techniques is shown in Figure 3-32. Nevertheless, in order to reliably reconstruct the viral quasispecies from raw data obtained by NGS, thorough data analysis is required (Caraballo Cortes et al., 2013).

Nucleotide sequencing is subject to some notable artefacts. Firstly, sequence errors may be introduced by use of non-proofreading DNA polymerases to assess viral diversity; hence, DNA polymerases with proofreading activity should always be used when performing PCR for quasispecies analysis to minimize such errors.

Another issue is the primer bias where primers may not bind to heterologous template-binding sites leading to selective amplification of a portion of the sequences in the target population. The primer bias could be quantified and minimized through careful primer design and control experiments (Yao and Tavis, 2005, Mullan et al., 2001).

In the case of low viral load, the viral complexity may be lower than the true value due to template re-sampling during PCR (Cruz-Rivera et al., 2013).

3.4.2 HCV full genome sequencing

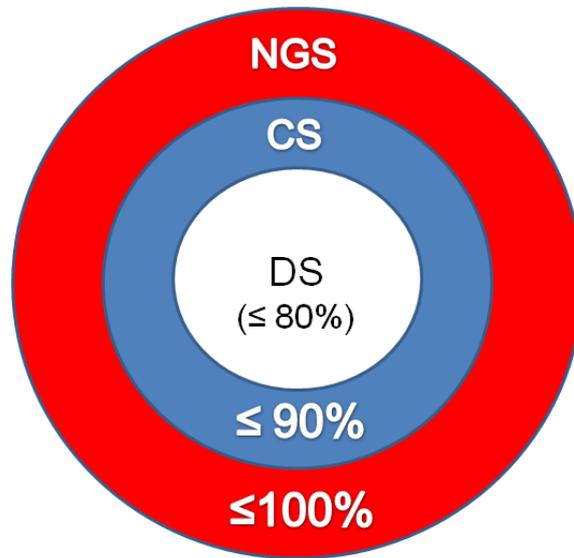
The detection of HCV sequence variation has important implications in understanding HCV biology. Significant sequence variation is present within the HCV ORF at both the nucleotide and the amino acid levels, especially in hyper-variable regions (HVR1, HVR2 and HVR3) (Kato et al., 1992).

Amplification of the HCV genome must overcome not only high quasispecies diversity, but also low viral load concentration. This is why whole genome sequencing of HCV has traditionally been very challenging, especially from heterogeneous clinical isolates. Different PCR-based strategies have been designed, usually based on multiple amplicon amplification using nested PCR to increase the sensitivity and specificity. Yao and Tavis reported an amplification strategy using five amplicons, but employed a Taq DNA polymerase that lacked the proofreading activity, which is insufficiently accurate for diversity studies (Yao and Tavis, 2005).

Following the introduction of DAAs, there is a need to develop a method to analyse the NS3, NS5A, and NS5B regions of HCV. Zhang *et al.* recently described an efficient RT-PCR method that allows viral sequencing of all regions targeted by the most common DAAs, this assay was used in clinical trials to amplify a nearly full-length HCV genotype 1 genome with a success rate of 97% including samples with low viral load (Zhang et al., 2013). However, the method used a mixture of Taq polymerase and high fidelity enzyme and is therefore likely to have been subject to both primer bias and PCR amplification error.

RNA isolation must be efficient to yield adequate amounts of pure template due to the limited amount of patient plasma or serum and the relatively low viral load. Automated extraction presented a better option to avoid contamination and decrease variability in the clinical laboratory.

Figure 3-32: Estimated frequency cut-off for detectability by different techniques.



The sensitivity of different techniques was variable in our study. The clonal analysis is a random process and only once was able to detect a variant with a frequency of $<1\%$ of the total population.

There is a consensus that the nucleic acid extracted by easyMAG[®] is of better quality for amplification than manual methods (Dundas et al., 2008). The extracted RNA using easyMAG[®] was superior in our downstream deep sequencing analysis as it lacked the carrier RNA incorporated in the RNA extraction process using Qiagen[®] that caused artefactual reads.

Two full-genome sequencing methods were developed using Illumina[®] technology, with two different approaches; the first was PCR-based amplicon sequencing using four amplicons covering the whole genome and the second employed a metagenomic approach. The consensus sequences generated by both approaches were similar but not identical suggestive of primer bias. However, the metagenomic approach was limited by a low number of reads in one region of HCV genome (only two reads), which highlights the importance of low coverage of the metagenomic approach. Depth of coverage can be improved by increasing the amount of sequencing, but this has resource implications.

In this study, we demonstrated two methods of successfully amplifying the whole HCV genome. The PCR-based approach was limited to genotype 1a while the metagenomic approach allowed the reconstruction of any HCV genotype. We assessed the complexity of HCV viral populations in the plasma of an acute HCV/HIV co-infected cohort.

The identification of viruses in NGS libraries relied on alignments to reference genomes although *de novo* assembly methods not covered here would also allow the identification of novel viruses using a metagenomic sequencing approach. NGS detected viruses that were diagnosed by traditional diagnostic methods; additionally, other genotypes were detected in several samples. Finally, we defined a threshold of abundance of viral reads to serve as a cut-off to discards false reads.

Metagenomic sequencing is a potential tool that could be extensively implemented in clinical research and diagnostics in the near future due to its high sensitivity and the ability to simultaneously detect a broad spectrum of viruses and new variants.

3.4.3 Bioinformatics framework

The raw NGS data analysis presents several challenges; efficient mapping to an appropriate reference sequence, error detection/correction, SNP identification, and genome assembly methods (Bao et al., 2011).

The reverse transcription (RT) and PCR steps create a noise level that lead to sequencing errors. A cut-off to differentiate true from erroneous mutations is required, but a major issue is where to estimate an abundance filter to exclude artefactual haplotypes and point mutations. Any data processing of amplicon sequencing requires such a threshold that is vital for studying complexity and the intrinsic diversity of the sample. In this study, we elected to exclude single sequences only.

This problem was investigated by Carballo *et al.* using three different strategies; firstly, they considered variants detected at a frequency higher than 1%; which retained only eight variants. A second strategy was to include sequences covered bi-directionally in both a forward and reverse reads; this method identified 40 variants. In the third strategy, they used ShoRAH software that had a variable outcome depending on the implemented cut-off value; “low cut-off value may result in low precision (fraction of true haplotypes among all called haplotypes) and conversely, high cut off may significantly lower recall (fraction of called haplotypes among all true haplotypes)” (Caraballo Cortes et al., 2013).

NGS has higher sensitivity than conventional methods in detecting low variants carrying SNPs. The NGS reads from any sequencing technology may contain noise from PCR amplification and platform specific noise, which must be distinguished from real diversity to be able to reconstruct the haplotypes accurately; however, the short read lengths render it difficult to determine which SNPs reside on the same haplotype.

In the near future, NGS will overcome Sanger’s limitations. A key step will be achieved when available SNP detection and assembly algorithms are

sufficiently reliable. Two approaches were implemented in this study; metagenomic and amplicon sequencing. Two possible factors can cause a difference in the viral complexity detected by both approaches, either through selection bias in the amplicon approach or underrepresentation in the metagenomic approach.

The main limitation of such analysis in the metagenomic arm is that our in-house script (Sam Variation) has a limited length (read length= 150bp), and it gathers only the reads covering the region under investigation completely. Hence, despite a depth of more than 200 reads in certain regions, the analysed reads could be only 20-30 reads and even in this group, many singletons were found. Whether these are artefacts or true reads remains a major doubt. As singletons were excluded in such analysis, only few reads (8-12) were compared with the amplicon arm in some regions.

An abundance filter cut-off (1%) was used for the amplicon reads, which is suggested by Illumina® to discard any false reads due to cross talk between indices, whether lower reads have been true ones and represent a minority variant is to be considered. Until an accurate analysis tool is available, we will not be able to address the comparison between both approaches (metagenomic and amplicon) except at SNP level, which could be of value in the case of assessing the presence of antiviral resistance mutations as discussed later in chapter 5.

Different quasispecies assembly methods have performed robustly in simulations and empirical experiments (Zagordi et al., 2010b, Prospero et al., 2011, Astrovskaya et al., 2011) This study is the first to consider HCV intra-host quasispecies, by creating an artificial viral population (mock community) by sequencing amplicon products from a clinical sample from a known HCV patient. After clonal analysis, the clones were mixed at controlled proportions and then processed using Illumina® sequencing.

Using our Mock data set, two assemblers, QuRe and PredictHaplo, were able to reconstruct the most frequent variants as demonstrated by phylogenetic analysis. PredictHaplo was found to be a conservative algorithm, with only three variants detected (23%) of the reported variants, which was similar to 25% recalled when validated using a 454 dataset (Prosperi et al., 2013).

QuRe showed the best precision/recall trade-off, although we were not able to determine a recall rate due to the low diversity of the artificial mock community data set but we would not predict it to be far from the 50% value reported by Prosperi *et al*, in a similar validation using a 454 pyrosequencing HCV dataset (Prosperi et al., 2013). In both assemblers, the estimated variant frequencies correlated with the original values stated in the mock community design.

Various factors can influence the performance of any haplotype reconstruction programme such as the instrumental sequencing error of NGS platform, read length, coverage, depth, the variant prevalence and the average diversity (Prosperi et al., 2013). Evaluation of these assemblers in different settings, using artificial data sets along with biological materials will be an important step in validating bioinformatics pipelines for studying viral populations in infected individuals.

None of the programmes that we tested was specifically designed for Illumina® reads. Hence, the higher coverage provided with Illumina® may present a challenge when using these assemblers. For example, QuRe was designed to deal with an average 20,000 reads produced by pyrosequencing. We had technical difficulties using this assembler for samples with 60,000-100,000 reads produced by Illumina® platform.

We tested the programmes on the Illumina® reads to examine their potential for viral haplotype reconstruction as Illumina® sequencing has become the most popular deep sequencing platform across all applications (Li et al., 2014). However, the lack of validated viral sequence analysis tools for the Illumina® platform remains a hurdle to the widespread adoption of Illumina®

sequencing in clinical laboratories. Viral haplotype reconstruction from a set of observed reads is one of the most challenging bioinformatics problems that requires further improvement. Further experiments are still needed to address this problem as explained in Section 6.2.3.

3.5 Conclusions

NGS is a powerful technique that has a potential role in HCV diagnostic algorithms for the detection of mixed infection currently underdiagnosed by standard methods (e.g. PCR-based hybridization assays and direct Sanger sequencing) and in screening for known antiviral resistance mutations. NGS is superior to both direct Sanger sequencing and clonal analysis in the detection of minority variants. The recent advances in NGS technology are associated with the parallel development of a large number of software and algorithms capable of handling the massive amount of data generated by NGS platforms. This will lead to faster implementation in a variety of settings (e.g. field of molecular epidemiology) (Escobar-Gutierrez et al., 2012).

Chapter 4: Hepatitis C virus diversity

4.1 Background

HCV displays high intra-host diversity, forming a pool of closely related but distinct genetic variants, the viral quasispecies (Martell et al., 1992). The immune system promotes a high degree of diversity in viral progeny and contributes to the ongoing evolution of HCV. This is confirmed by the fact that a low degree of HVR1 genomic variability has been observed in immunosuppressed patients, suggesting that the variability in HVR1 is directed by random substitutions induced by host immune pressure (de Amorim et al., 2014).

Studies of the treatment outcome in HCV-infected patients identified various viral factors as predictors of treatment response, such as genotype, viral load, and viral diversity (Wohnsland et al., 2007). Moreover, host factors were reported to affect response rates, such as age, weight, sex, race, liver function, and extent of fibrosis. IL28B genotype and clinical variables were integrated into a clinical prediction model that produced a useful individualized prediction of treatment response (O'Brien et al., 2011).

Characterization of the natural history of HCV infection has been hindered by a lack of studies due to the asymptomatic nature of early infection. Infection with the virus is often unknown until liver cirrhosis develops, usually after many years of silent infection. Acute HCV infection is conventionally defined as the initial six months of infection that induces virus-host interactions that influence the outcome of the disease.

In the last decade, an emerging epidemic of acute HCV infection among HIV-infected MSM has provided the opportunity to prospectively study the progression of early disease as patients are identified early in infection, due to the detection of abnormal liver function tests at routine follow-up visits for their HIV infection (Dominguez et al., 2006, Gilleece et al., 2005, Vogel et al., 2005, Matthews et al., 2009, Fierer et al., 2008).

In the pre-DAA era, SVR rates were lower in the HIV/HCV co-infection group (59-71%) than in HCV mono-infected patients (98%) (Gilleece et al., 2005, Serpaggi et al., 2006). Studies in this group of patients are likely to increase understanding of the determinants of spontaneous clearance versus evolving progression to chronicity.

Viral complexity in HCV/HIV co-infected patients prior to treatment has been a controversial topic, with various studies showing either greater complexity compared to those with HCV mono-infection, no difference from those with mono-infection or even less complexity (Sherman et al., 2010).

Several studies have investigated the composition and the evolution of HCV quasispecies to determine their role in predicting the outcome of antiviral therapy. Many studies have suggested a correlation between a high level of heterogeneity within HVR1 and a poor response to PegIFN α /RBV therapy (Farci et al., 2002, Chambers et al., 2005, Sandres et al., 2000).

The accurate evaluation of viral complexity may play a major role in the prediction of the outcome of antiviral treatment (Gregori et al., 2014). Current diagnostic methods vary in their sensitivity; most are likely to underestimate the complexity of viral populations due to lack of detection of low-frequency variants. The development of NGS platforms provided a new tool for detailed analysis of intra-host viral populations as shown in chapter 3 and reported previously (Cruz-Rivera et al., 2013). In this chapter, we revisited some questions about the role of viral diversity in understanding and predicting treatment outcome in a unique acute HCV/HIV cohort using an NGS-based approach. We compared this with more conventional Sanger sequencing and cloning-based approaches.

Different diversity measures were employed in this study. Shannon entropy and the Simpson's index both measure diversity based on the number of variants present (species richness) and the distribution of the number of organisms per species (species evenness). Simpson's index estimates the probability that two isolates randomly selected from a sample will be the same variant, the index

value ranges from 0 to 1, with a value of “1” indicating no diversity. However, both measures of diversity are not sensitive to variant frequency; hence, they may not reflect the presence of rare variants.

Other measures of diversity such as pairwise diversity (P_i) determines how different all variants within the population are, while considering the genetic distance between each pair of individuals in the viral population. Hamming distance, (the mean number of amino acid differences between variants) may also be used to quantify the composition of the quasispecies population (Thomson et al., 2011).

Meanwhile, phylogenetic entropy and quadratic diversity are sensitive to relatedness, using phylogeny in determining the viral diversity in the population. Phylogenetic entropy places a high value on distinctive species but also retains the rare variants (Allen et al., 2009).

We hypothesized that high viral diversity would be expected in patients who do not respond to treatment while treatment response might be predicted by low viral diversity. We further hypothesized that diversity analyzed using NGS might be a better predictor than the use of less sensitive techniques such as bacterial cloning.

Descriptions of HCV re-infection in acute HCV/HIV-infected MSM have been widely reported without detailed phylogenetic analysis (Lambers et al., Martin et al., 2013). Multiple HCV variants are commonly present in individual infected patients in this group (Thomson et al., 2011). Such variation could occur because of transmission of multiple HCV strains either around the time of initial infection or sequentially over time.

We hypothesised that re-infection rates following treatment would be over-estimated by standard Sanger sequencing due to lack of detection of varying dominance of minority variant strains present at the onset of infection and aimed to investigate this in a group of HCV/HIV co-infected patients who failed HCV treatment.

In order to test the above hypotheses, pyrosequencing was used to characterize the complexity of HVR1 as described in Section 2.9.2, to determine whether intra-patient viral diversity of HCV in HIV-positive patients during early infection determines the likelihood of SVR following treatment with PegIFN α /RBV.

The same NGS platform was used to dissect the different causes of treatment failure using pre- and post-treatment plasma samples taken from patients who failed standard of care therapy with 24-48 weeks of PegIFN α /RBV. NGS data were analysed using phylogenetic analysis of pre- and post- treatment variants obtained using clonal analysis and pyrosequencing to differentiate relapse from reinfection.

4.2 Results

The study outlined here is the first to evaluate viral complexity and diversity in an acutely infected HCV cohort using deep sequencing. To determine whether viral diversity affects treatment response, pre-treatment viral diversity was analysed using a segment of 183 nucleotides from the E1/E2 region including the HVR1 region.

4.2.1 Viral diversity as a predictor of treatment outcome

4.2.1.1 Characterization of the study group

Twenty patients were selected who had completed treatment with 48 weeks of PegIFN α /RBV. Ten patients achieved SVR following treatment, and ten failed to clear the virus (four null responders, three partial responders, and three relapsers). The groups were matched for different variables including peak ALT, pre-treatment viral load and duration of HCV infection (an estimate calculated by the difference between the date of the RNA-positive study sample and date of last negative HCV RNA PCR). All of the patients were reported by the clinical virus laboratory to be infected with genotype 1a (Table 4-1).

4.2.1.2 Diversity of viral quasispecies.

After sequencing the 20 samples using pyrosequencing, the data generated by NGS of the study group were analysed as described in 2.15.2. The genetic diversity of quasispecies sequences from patients in each group was calculated using pairwise diversity (Pi), Simpson's index, Shannon index, phylogenetic diversity and quadratic diversity. Diversity at baseline and within 150 days from infection was significantly lower in those who achieved SVR compared to the TF group as summarized in (Table 4-2).

Similar results were obtained comparing only the HVR1 region between the two groups and using a normalized set of 10,000 sequences per sample (data not shown). The measures of viral diversity in individual patients in each group are shown below (Table 4-3 and

Table 4-4); the two groups are compared in (Figure 4-1).

4.2.1.3 Entropy

The amino acid sequence variability at each position of HVR1 was evaluated using the Shannon Entropy program at the Los Alamos National Security Website (<http://hcv.lanl.gov/content/sequence/ENTROPY/entropy.html>). A detailed analysis of entropy per site of the sequences provided evidence that higher variability was observed in HVR1 in the non-responder group than in the responder group; an example comparing a patient from each group is illustrated in Figure 4-2.

Graphical representations of the patterns within a multiple sequence alignment called sequence logos provide a richer and more precise description of sequence similarity than consensus sequences and revealed the higher variability per site in the HVR1 region. The logos in Figure 4-3 were created using WebLogo software (Crooks et al., 2004).

Table 4-1: Characteristics of the study groups of subjects.

Variable	TF group Mean \pm SEM	SVR group Mean \pm SEM	P-value
Peak ALT	917.6 \pm 330.8	1039 \pm 236.6	0.7633
Days of infection	91.93 \pm 21.2	121.3 \pm 31.9	0.1339

Days of infection is an estimate of the time between last HCV RNA sample and the study sample.

Table 4-2: Comparison of quasispecies diversity in TF group versus SVR group.

	SVR group Mean \pm SEM	TF group Mean \pm SEM	P-value
Shannon's index	0.8 \pm 0.07	1.66 \pm 0.19	0.0005
Simpson's index	0.73 \pm 0.03	0.44 \pm 0.04	0.0001
Pairwise diversity	1.33 \pm 0.39	10.66 \pm 4.08	0.0356
Phylogenetic entropy	0.009 \pm 0.003	0.12 \pm 0.056	0.05
Quadratic diversity	0.003 \pm 0.001	0.08 \pm 0.04	0.05

P-value calculated using paired t-test.

Table 4-3: Measures of diversity in the treatment failure group (TF).

ID	Numbers of reads	Shannon index	Simpson index	Phylogenetic entropy	Quadratic diversity	Pairwise diversity
1	23042	1.2	0.599	0.092	0.045	0.663
2	21702	1.629	0.504	0.001	0.001	0.077
3	10742	1.368	0.43	0.005	0.002	1.54
4	23639	0.974	0.606	0.002	0.001	0.482
5	19165	1.594	0.455	0.399	0.271	1.898
6	21246	1.739	0.337	0.483	0.321	1.774
7	19739	1.972	0.439	0.037	0.023	0.89
8	20659	1.043	0.65	0.032	0.019	0.705
9	20453	2.191	0.281	0.004	0.002	0.795
10	16500	2.969	0.158	0.188	0.129	0.622

Measures of diversity were calculated using a Perl script designed by Dr Joseph Hughes.

Table 4-4: Measures of diversity in responder group (SVR).

ID	Number of reads	Shannon index	Simpson index	Phylogenetic entropy	Quadratic diversity	Pairwise diversity
1	28238	0.964	0.623	0.007	0.002	0.71
2	26719	0.97	0.616	0.009	0.002	0.903
3	38968	1.06	0.641	0.027	0.011	1.225
4	26292	0.537	0.866	0.003	0.001	0.741
5	52970	0.958	0.73	0.002	0.001	0.862
6	21424	0.801	0.715	0.004	0.001	1.042
7	41729	0.611	0.84	0.026	0.011	0.959
8	15054	0.916	0.645	0.004	0.001	0.922
9	63048	0.836	0.755	0.003	0.001	1.258
10	31036	0.432	0.897	0.005	0.002	1.014

Measures of diversity were calculated using a Perl script designed by Dr Joseph Hughes.

Figure 4-1: Comparison of diversity measures between non-responders (TF) and responders (SVR) groups.

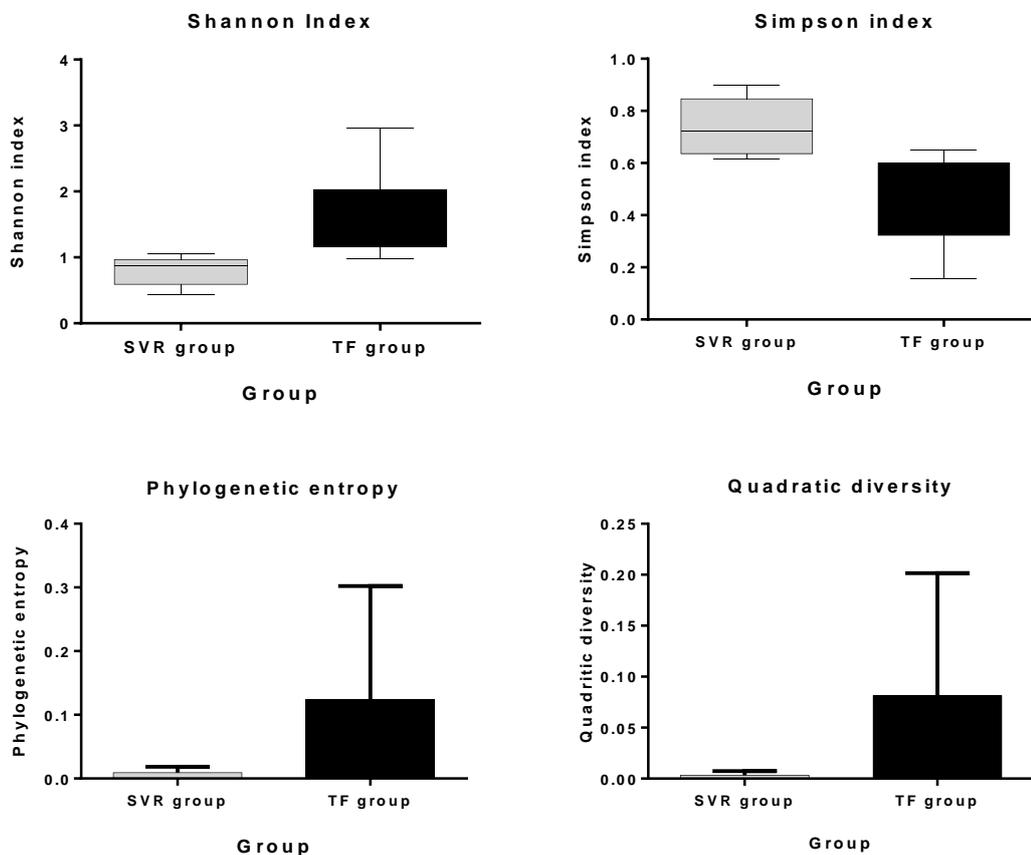
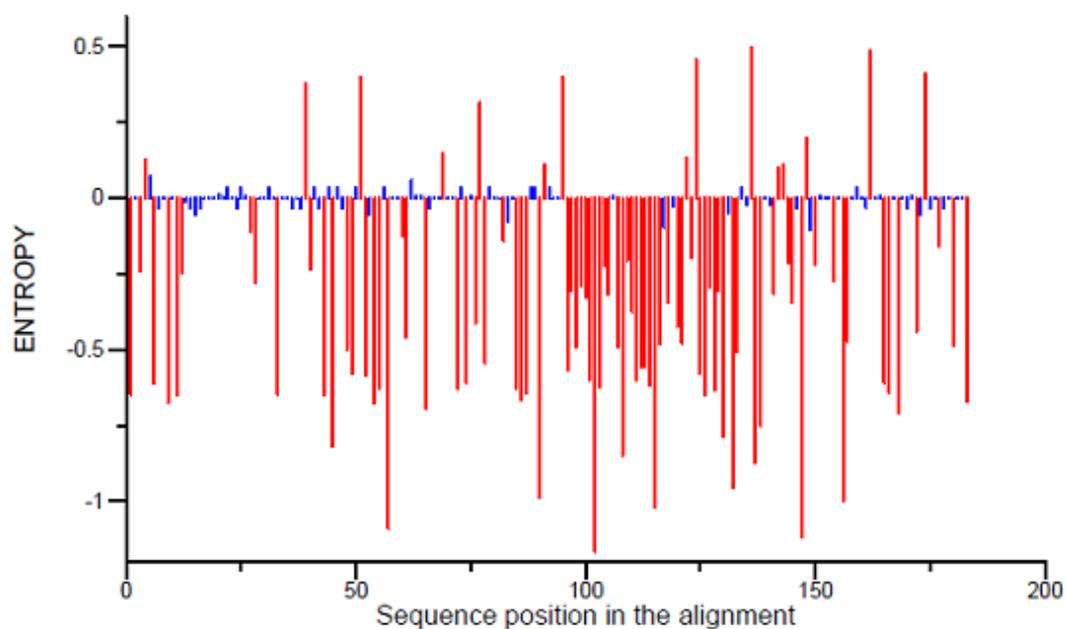


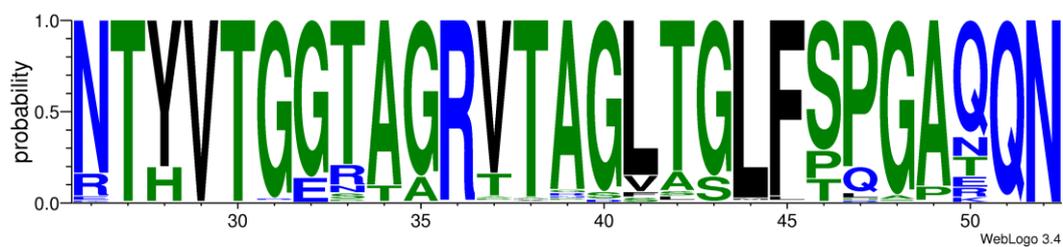
Figure 4-2: Illustrative example of the difference in Shannon entropy between SVR group (P6) and TF group (P81).



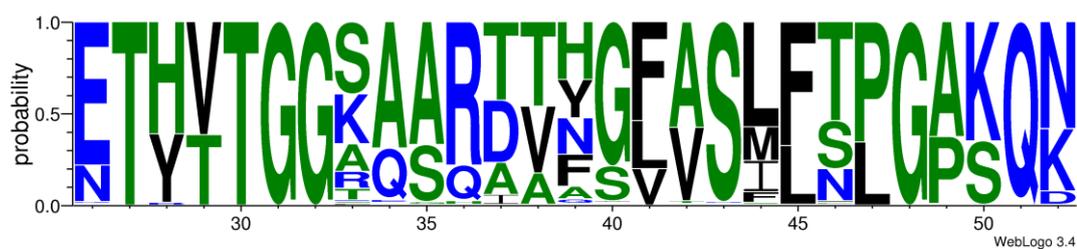
The compared region is 183 nt covering HVR1 region. Shannon entropy of P6 represented above (0) and P81 below (0), significantly different sites (p -value <0.05) are plotted in red.

Figure 4-3: HVR1 amino acid diversity.

A



B



Viral diversity A) in a patient (P6) who responded to treatment, B) in a patient (P81) who failed treatment.

4.2.1.4 Viral complexity

Phylogenetic analysis was performed using HVR1 nucleotide sequences from all patients. The phylogenetic trees suggested a pattern of higher viral complexity in the TF group compared to the SVR group, with multiple variants detected in the TF group; an illustrative example of the difference is shown in (Figure 4-4). The viral variants detected in each group are detailed in a maximum likelihood phylogenetic tree in Appendix 7.1 and Appendix 7.2.

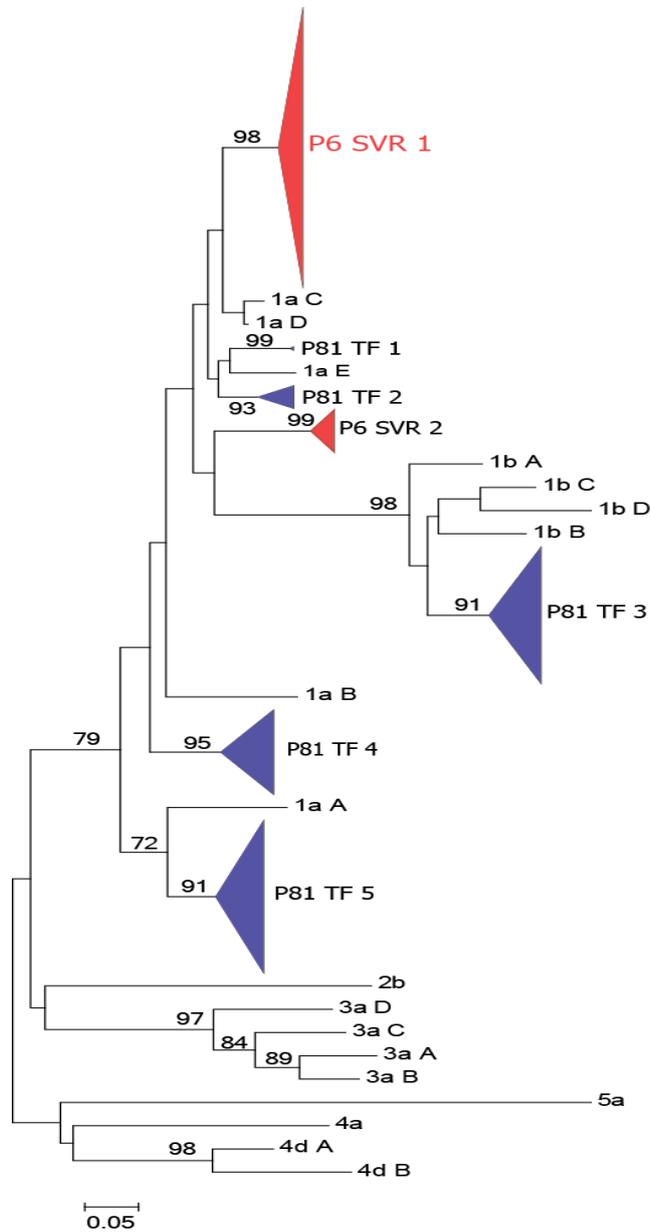
4.2.1.5 Star Phylogeny

Star phylogeny in a phylogenetic tree represents the occurrence of multiple short branches originating from an internal node. It is indicative that the viral population evolved from a common ancestor. In order to distinguish infections initiated by a single variant (homogenous infection) from those where multiple variants entered the host, we employed a tool used extensively in the HIV field for this purpose (Poisson Fitter).

Poisson Fitter analyses Hamming distance (HD) frequencies by computing the best fitting Poisson distribution and evaluating results of the Goodness of Fit test (GOF). P-values of less than 0.05 indicate divergence from a Poisson distribution and can be interpreted as transmission of multiple viral variants or the presence of positive selection. The analysis was performed using the online tool http://www.hiv.lanl.gov/content/sequence/POISSON_FITTER/ (Giorgi et al., 2010).

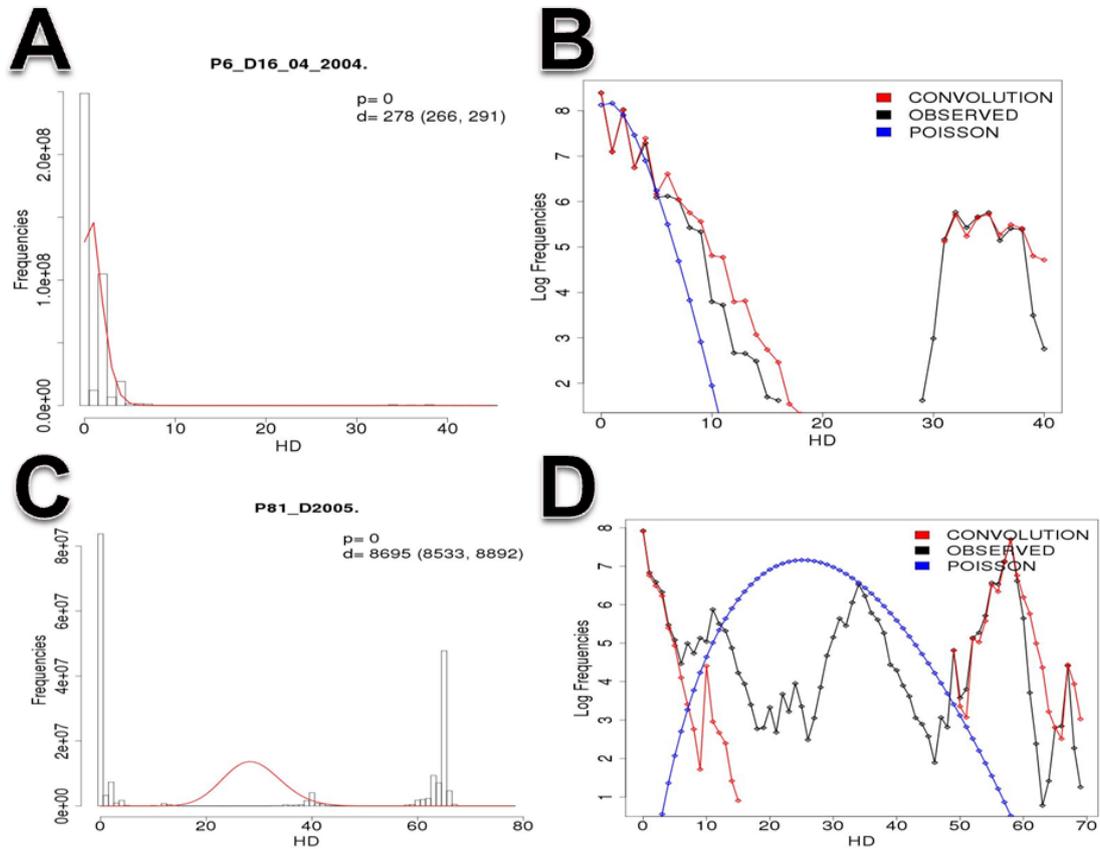
The analysis revealed that patients in the TF group were more likely to be infected with multiple founder strains than those in the SVR group. For example, the SVR group P6 sample followed a star-like phylogeny suggesting the infection originated from a single founder virion, while P81, a sample from the TF group, did not follow a star phylogeny and was suggestive of multiple variant infection as shown in Figure 4-5. The results of individual samples are detailed in Appendix 7.3 to Appendix 7.18.

Figure 4-4: Illustrative example of the difference in viral complexity in SVR (P6) and TF (P81) groups.



Maximum likelihood tree showing the variants in P6 (SVR group) and P81 (TF group). The bootstrap values are shown next to the branch points (1,000 replicates); only values higher than 70% are shown.

Figure 4-5: Illustrative example of Poisson Fitter from both groups.



P6 follows star phylogeny with one founder virus while P81 does not follow star phylogeny with possibly 5 clusters of sequences representing multiple infection at the outset. The convolution estimates are used as an internal check for star-phylogeny, the observed values represent the hamming distance calculated among circulating variants. Figures (A and C) show the histogram of observed hamming distance, Plots (B and D) demonstrate the observed pairwise HD (in black) and the theoretical (convolution) pairwise HD frequency counts (in red) if the sample were to follow a star-phylogeny, and finally the best fitting Poisson by the blue line.

4.2.2 Superinfection versus relapse

4.2.2.1 Cohort characteristics

A group of 15 patients failed to respond to treatment (which includes the patients included in the TF versus SVR study group); six null responders, three partial responders, and six relapsers. Paired samples from each patient pre- and post-treatment were analysed. Demographic and clinical parameters are shown in Table 4-5.

4.2.2.2 Viral dynamics in pre- and post- treatment paired serum samples.

The pairwise distance between variants (for deep sequencing and clonal analysis, this was calculated between the most similar pre-treatment and post-treatment strains) was significantly higher using direct rather than pyrosequencing (mean 0.221 versus 0.026 respectively; $p=0.0002$).

Using direct sequencing, evidence of a new variant was detected in 10/15 (66%) of patients post-treatment as illustrated in Figure 4-6. In contrast, comparison of pre-treatment sequences determined by pyrosequencing with clonal sequences from post-treatment samples revealed that 100% of patients had evidence of a similar variant present in pre- and post-treatment samples. A new variant (in addition to a pre-existing variant) was detected in post-treatment samples in 6/15 (40%) patients as shown in Table 4-6. The new variant was a minority variant in three patients and a majority variant in three patients.

4.2.2.3 Mixed strain infections

Multiple strains infection was detected in all 15 patients who failed treatment, the number of detected variants was 2-6 variants of genotype 1a. Seven patients had evidence of mixed subtype or genotype infection at baseline; six patients had two subtypes (1a and 1b), and one patient had a mixed genotype infection (1a and 4d). The frequency of minority strains that emerged following therapy ranged from 3% to 13% of the viral population in pre-treatment samples and reached up to 75-100% of the total viral population in post-treatment samples.

4.2.2.4 Patient groups

In null responders [six patients - P38 (Figure 4-7), P63 (Figure 4-8), P67 (Figure 4-9), P81 (Figure 4-10), P112 (Figure 4-11) and P118 (Figure 4-12)], mixed subtype infection (1a/1b) at the outset was detected in 5/6 patients and the sixth patient had multiple variants of genotype 1a. All six patients had evidence of a similar strain present pre- and post-infection. Three patients (P63, P67, P112) had persistent infection with the same pre-treatment dominant strain, one patient (P118) had new dominance of a pre-existing minority variant, and two patients (P81, P38) had evidence of a new variant in addition to pre-existing strains (minority and majority post-treatment variant respectively).

In all partial responders [P31 (Figure 4-13), P21 (Figure 4-14), P105 (Figure 4-15)], multiple variants were present (3, 4 and 5 variants respectively). One patient had persistent infection with the same pre-treatment variant (P31), and two patients (P21, P105) had persistent infection with evidence of a new previously unidentified strain in the post-treatment sample (minority and majority variants respectively).

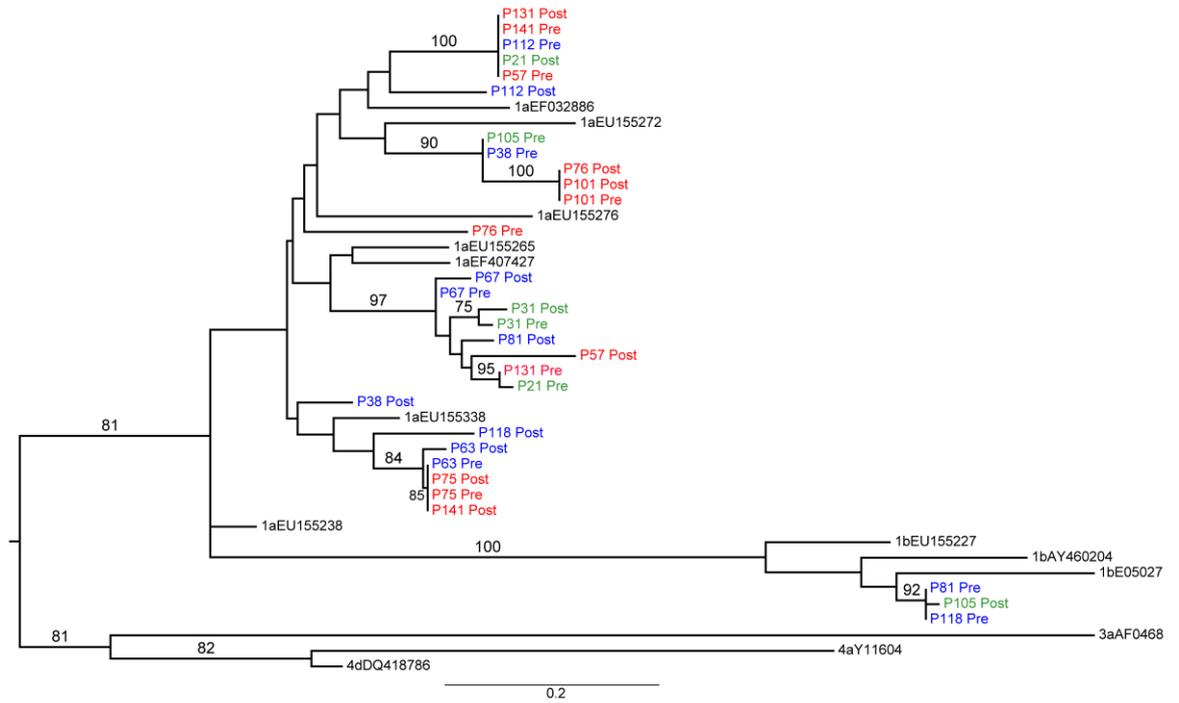
In relapsers [6 patients- P57 (Figure 4-16), P141 (Figure 4-17), P76 (Figure 4-18), P75 (Figure 4-19), P101 (Figure 4-20), P131 (Figure 4-21), all patients had evidence of persisting variants and four of them (P101, P57, P76, P141) showed new dominance of pre-existing minority strains. Two patients had evidence of new previously undetected strains; in one case the previously undetected variant became dominant in the post-treatment sample (P131) while in another patient (P75), the undetected variant was a minority strain (29%) on top of a pre-existing variant (71%).

Table 4-5: Clinical characteristics of the treatment failure cohort.

ID	Age	CD4	Nadir CD4	Route of transmission	IL28B	RVR	HAART	SVR
P21	36	770	60	UAI	-	-	Yes	No
P31	35	580	510	UAI	-	-	No	No
P38	31	830	320	UAI, INDU, PWID	CT	No	Yes	No
P57	46	710	20	UAI, INDU, PWID	CC	No	Yes	No
P63	47	400	150	UAI, INDU, PWID	CT	No	Yes	No
P67	40	340	80	UAI, INDU, PWID	CC	No	Yes	No
P75	32	220	220	UAI, INDU	TT	No	Yes	No
P76	48	390	210	UAI	-	Yes	Yes	No
P81	53	510	260	UAI	-	-	Yes	No
P101	35	640	380	UAI, INDU	TT	No	Yes	No
P105	39	790	300	UAI	CC	No	Yes	No
P112	46	220	130	UAI	-	No	No	No
P118	28	500	290	UAI, INDU	CT	No	Yes	No
P131	50	680	340	UAI	CT	Yes	Yes	No
P141	21	560	420	UAI	-	No	No	No

ID: study ID number, UAI: Unprotected Anal Intercourse, PWID - People Who Inject Drugs, INDU - Intranasal Drug Use, RVR - Rapid Virological Response, (-) unknown data.

Figure 4-6: Phylogenetic tree of direct Sanger sequences.



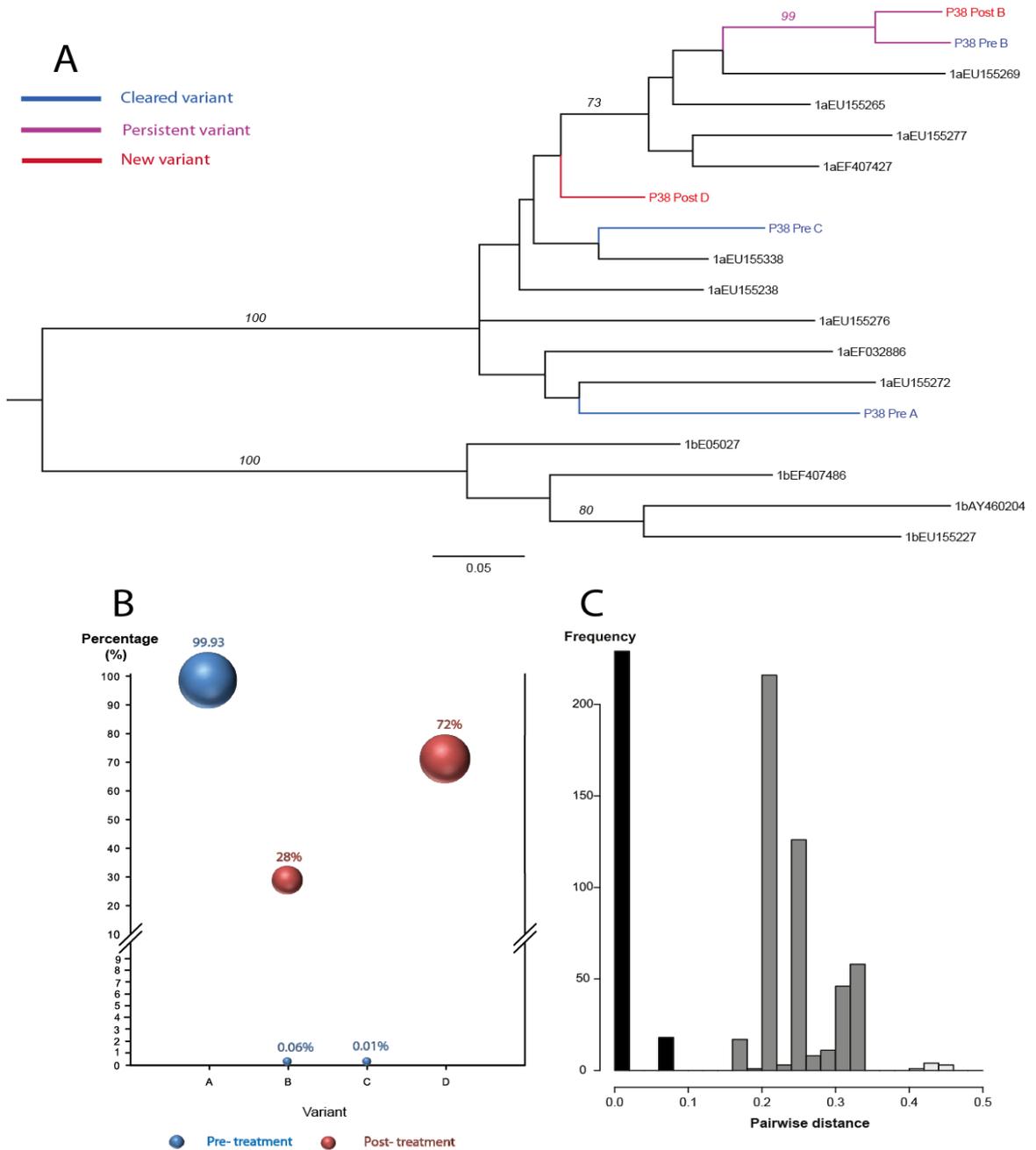
A maximum likelihood tree was constructed using nucleotide sequences determined using Sanger sequencing from paired samples and selected HCV reference sequences downloaded from the Los Alamos HCV database (Red- Relapsers, Blue-Null responders, and Green-Partial responders).

Table 4-6: Characteristics of viral population dynamics and treatment response in patients with treatment failure.

ID	Clinical outcome	Pairwise distance (Sanger) ¹	Pairwise distance (NGS) ²	New Dominance ³	New Variants ⁴	Cleared Variants ⁵	Final conclusion
P38	Null response	0.19	0.08	-	1	2	Persistent infection (New variant detected)
P63	Null response	0.03	0	-	0	1	Persistent infection
P67	Null response	0.04	0.04	-	0	2	Persistent infection
P81	Null response	0.48	0.06	13%	1	3	Persistent infection (New dominance and new variant detected)
P112	Null response	0.17	0.01	-	0	3	Persistent infection
P118	Null response	0.47	0.01	3%	0	5	Persistent infection (New dominance)
P21	Partial response	0.27	0	NA	1	2	Persistent infection (New dominance and new variant detected)
P31	Partial response	0.08	0.08	-	0	3	Persistent infection
P105	Partial response	0.46	0	-	1	2	Persistent infection (New variant detected)
P75	Relapse	0	0	-	1	1	Persistent infection (New variant detected)
P76	Relapse	0.24	0.05	3.2%	0	4	Persistent infection (New dominance)
P101	Relapse	0	0	-	0	3	Persistent infection
P57	Relapse	0.33	0.03	9%	0	2	Persistent infection (New dominance)
P131	Relapse	0.27	0.05	-	2	3	Persistent infection (New variant detected)
P141	Relapse	0.24	0.01	3.9%	0	3	Persistent infection (New dominance)

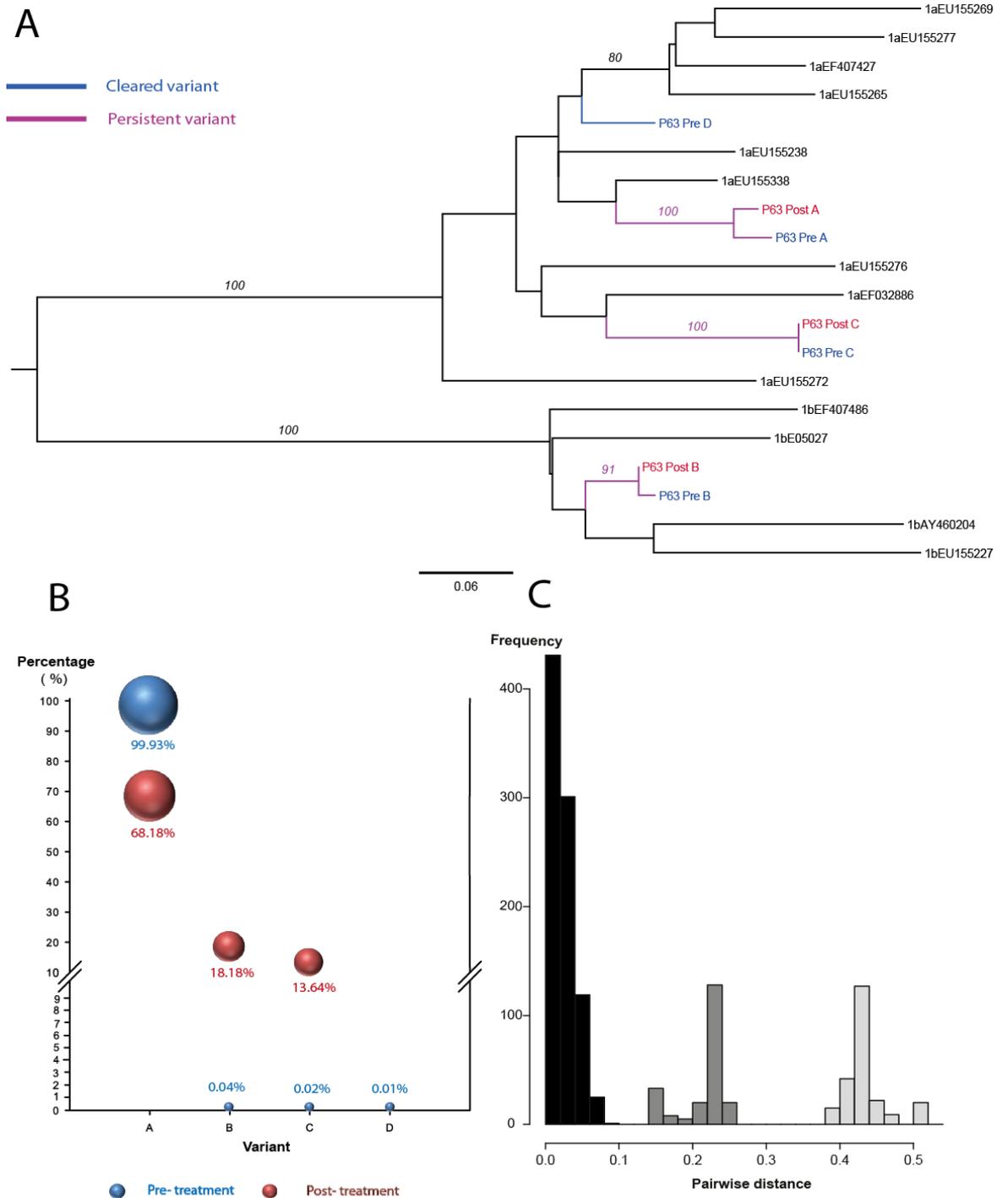
Outcome is determined by comparing consensus sequence of pre- and post-treatment samples using Sanger sequencing, 2- Pairwise distance is the pairwise distance between the similar variants in paired samples where a new dominance of pre-existing minority strain was noticed, 3- New dominance is the frequency of the new dominant variant of the post-treatment sample detected in the pre-treatment sample. 4- Number of new variants detected in the post-treatment sample, 5- Number of variants that cleared under treatment pressure.

Figure 4-7: Comparison of viral complexity in paired serum samples in P38.



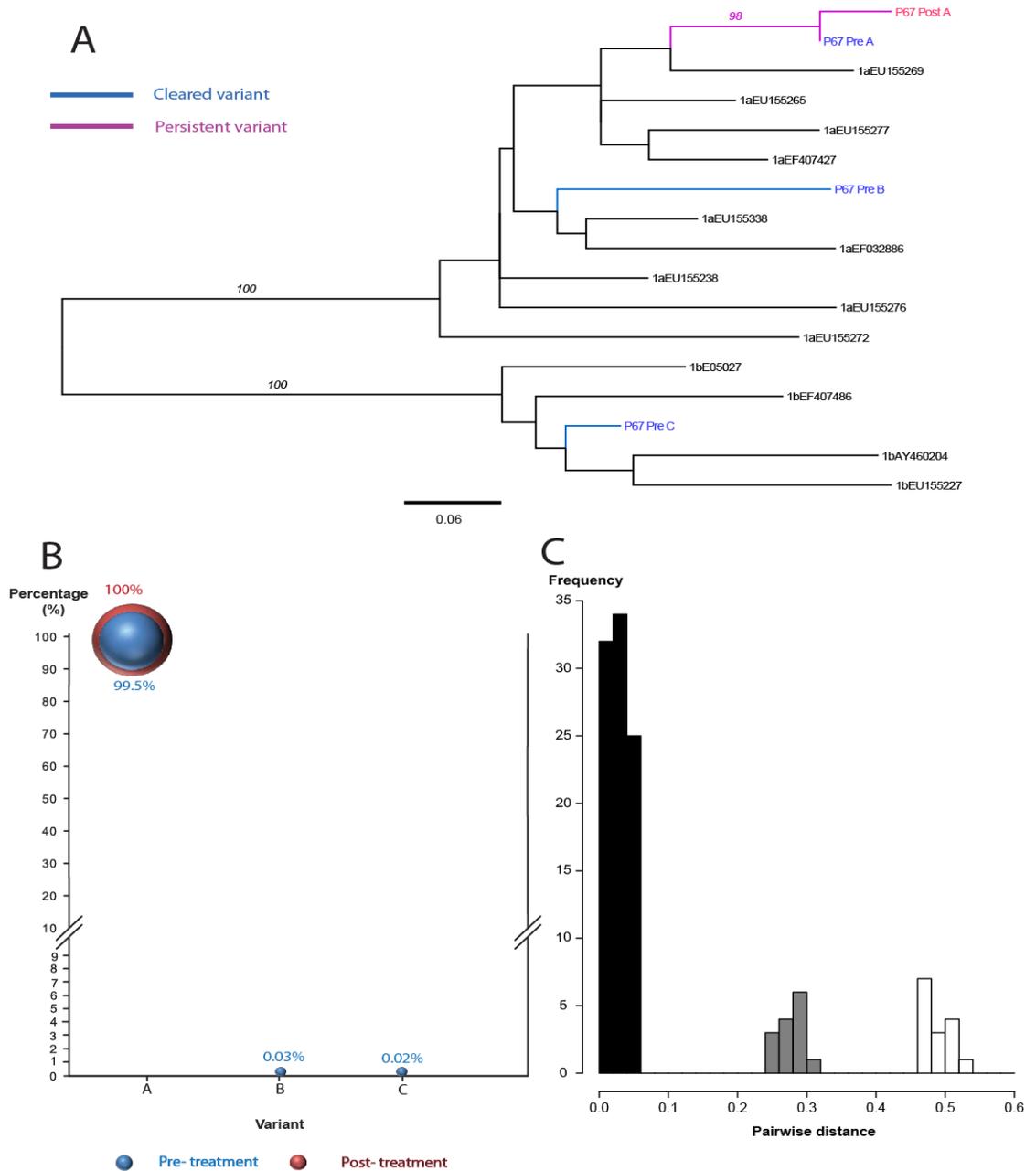
(A) ML tree was constructed using sequences from pre- and post-treatment samples in P38 (Null response, Persistent infection with new post-treatment variant detected), and selected HCV reference sequences for the Los Alamos HCV database. A total of 4(A-D) HCV variants detected. The analysis included; 25 clonal sequences (post-treatment) and 46755 reads derived from 454 pyrosequencing (pre-treatment). There was a total of 183 positions in the final dataset. (B) Bubble chart of the frequency of each variant (A-D) in pre- and post-treatment samples. (C) The pairwise distance between the most similar variants in the pre- and post-treatment samples (p-distance).

Figure 4-8: Comparison of viral complexity in paired serum samples in P63.



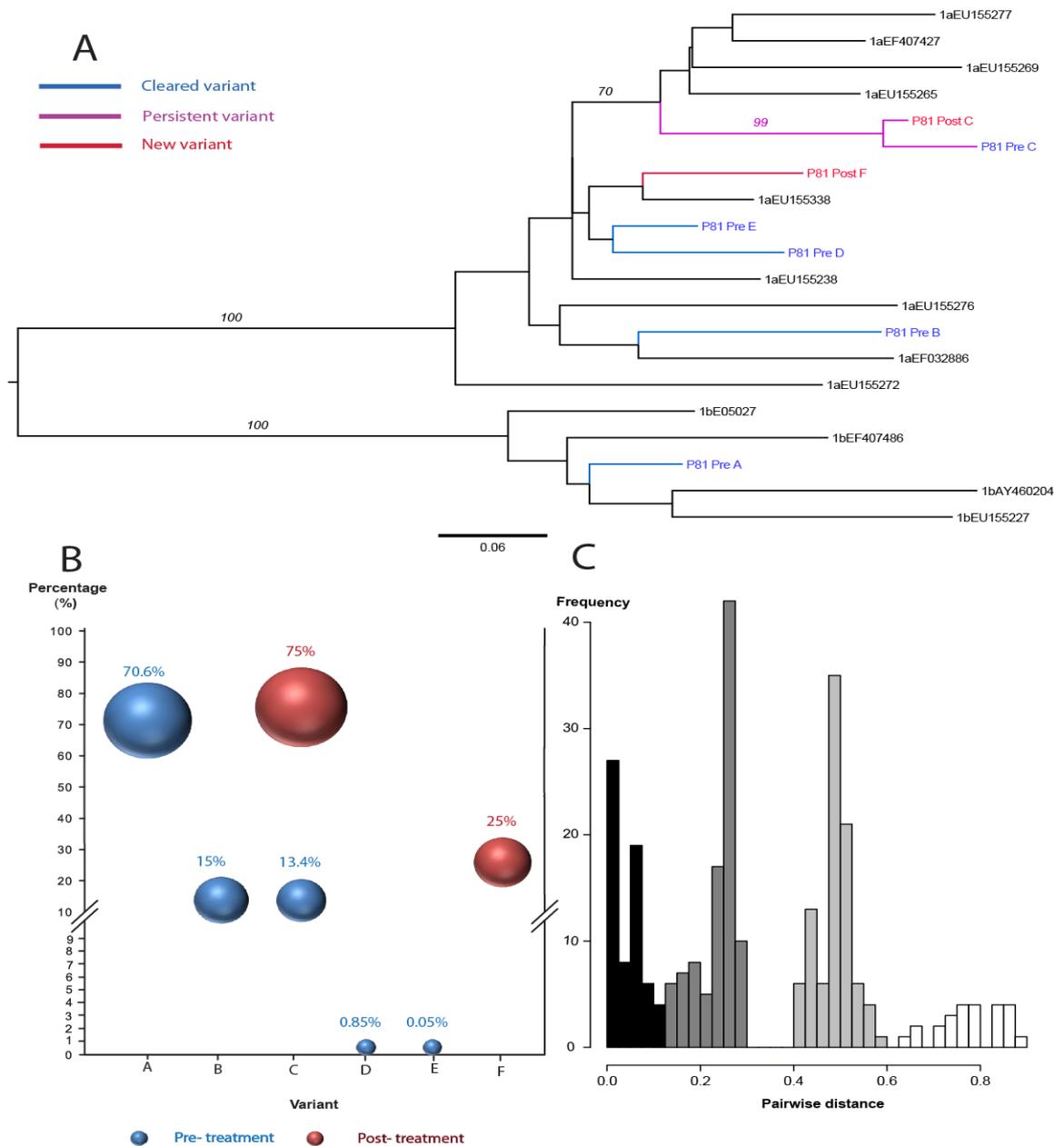
(A) ML tree was constructed using sequences from Pre- and post-treatment samples in P63 (Null response, Persistent infection) and selected HCV reference sequences for the Los Alamos HCV database. A total of 4 (A-D) HCV variants detected. The analysis included; 25 clonal sequences (post-treatment) and 46156 reads derived from 454 pyrosequencing (pre-treatment). There was a total of 183 positions in the final dataset. (B) Bubble chart of the frequency of each variant (A-D) in paired samples. (C) The pairwise distance between the most similar variants in the paired samples (p-distance).

Figure 4-9: Comparison of viral complexity in paired serum in P67.



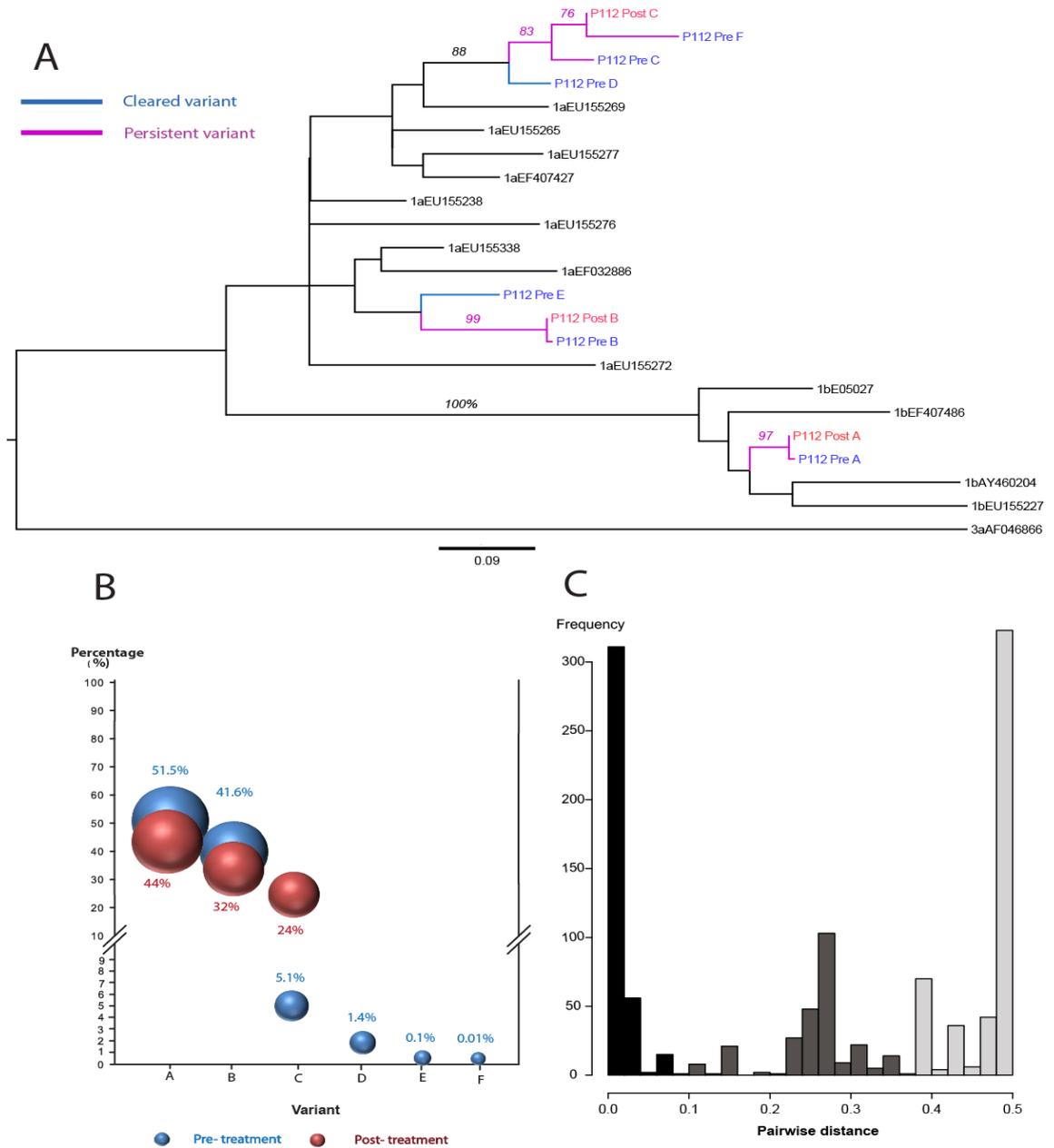
(A) ML tree was constructed using sequences from paired samples- P67 (Null response, Persistent infection), and selected HCV reference sequences for the Los Alamos HCV database. A total of 3 (A-C) HCV variants detected. The analysis included; 8 clonal sequences (post-treatment) and 10742 reads derived from 454 pyrosequencing (pre-treatment). There was a total of 183 positions in the final dataset. (B) Bubble chart of the frequency of each variant (A-C) in pre- and post-treatment samples. (C) The pairwise distance between the most similar variants in the pre- and post-treatment samples (p-distance).

Figure 4-10: Comparison of viral complexity in paired serum samples in P81



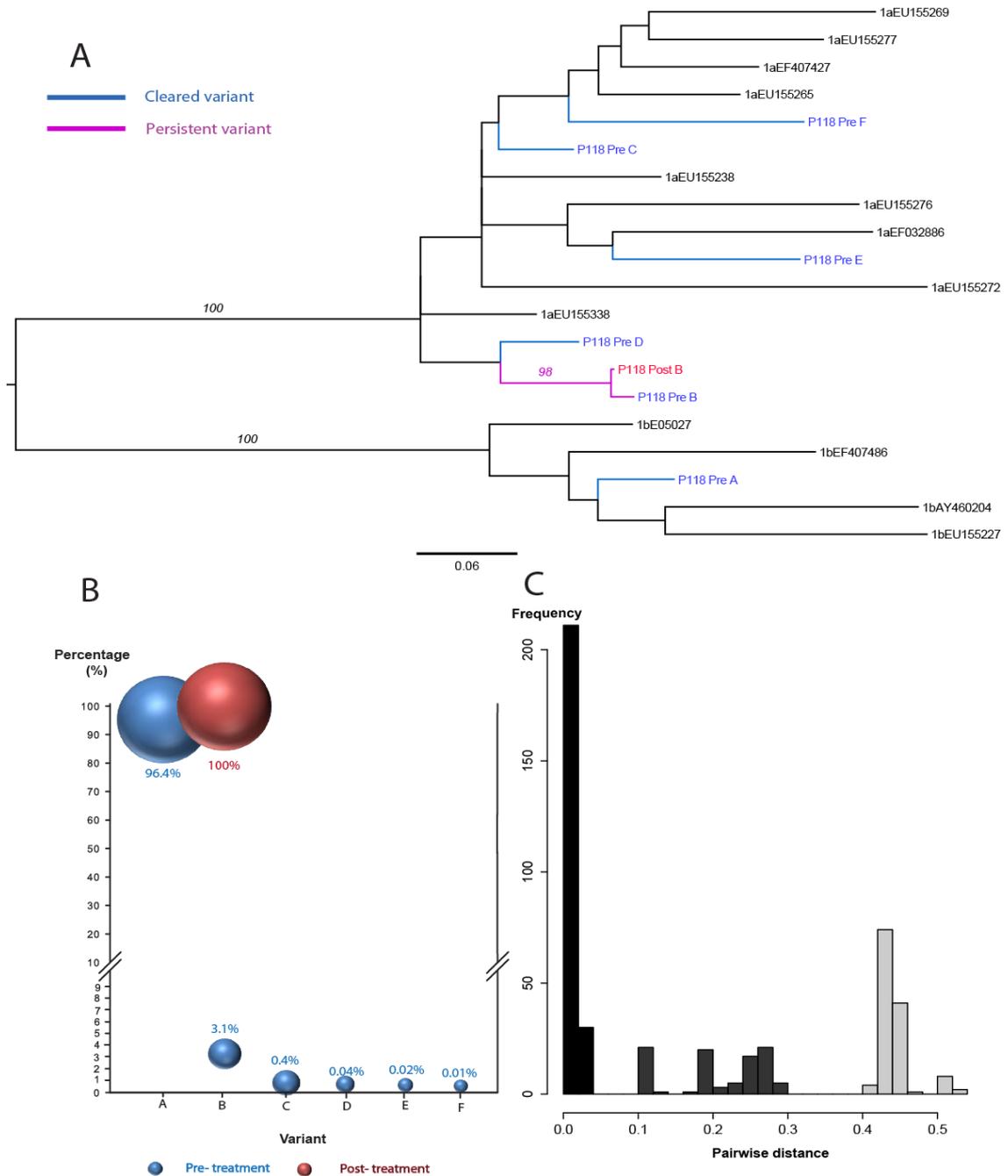
(A) ML tree was constructed using sequences from paired samples- P81 (Null response, Persistent infection with a new post-treatment variant detected), and selected HCV reference sequences for the Los Alamos HCV database. A total of 6 (A-F) HCV variants detected. The analysis included; 12 clonal sequences (post-treatment) and 19610 reads derived from 454 pyrosequencing (pre-treatment). There was a total of 183 positions in the final dataset. (B) Bubble chart of the frequency of each variant (A-F) in pre- and post-treatment samples. (C) The pairwise distance between the most similar variants in the pre- and post-treatment samples (p-distance).

Figure 4-11: Comparison of viral complexity in paired serum samples in P112.



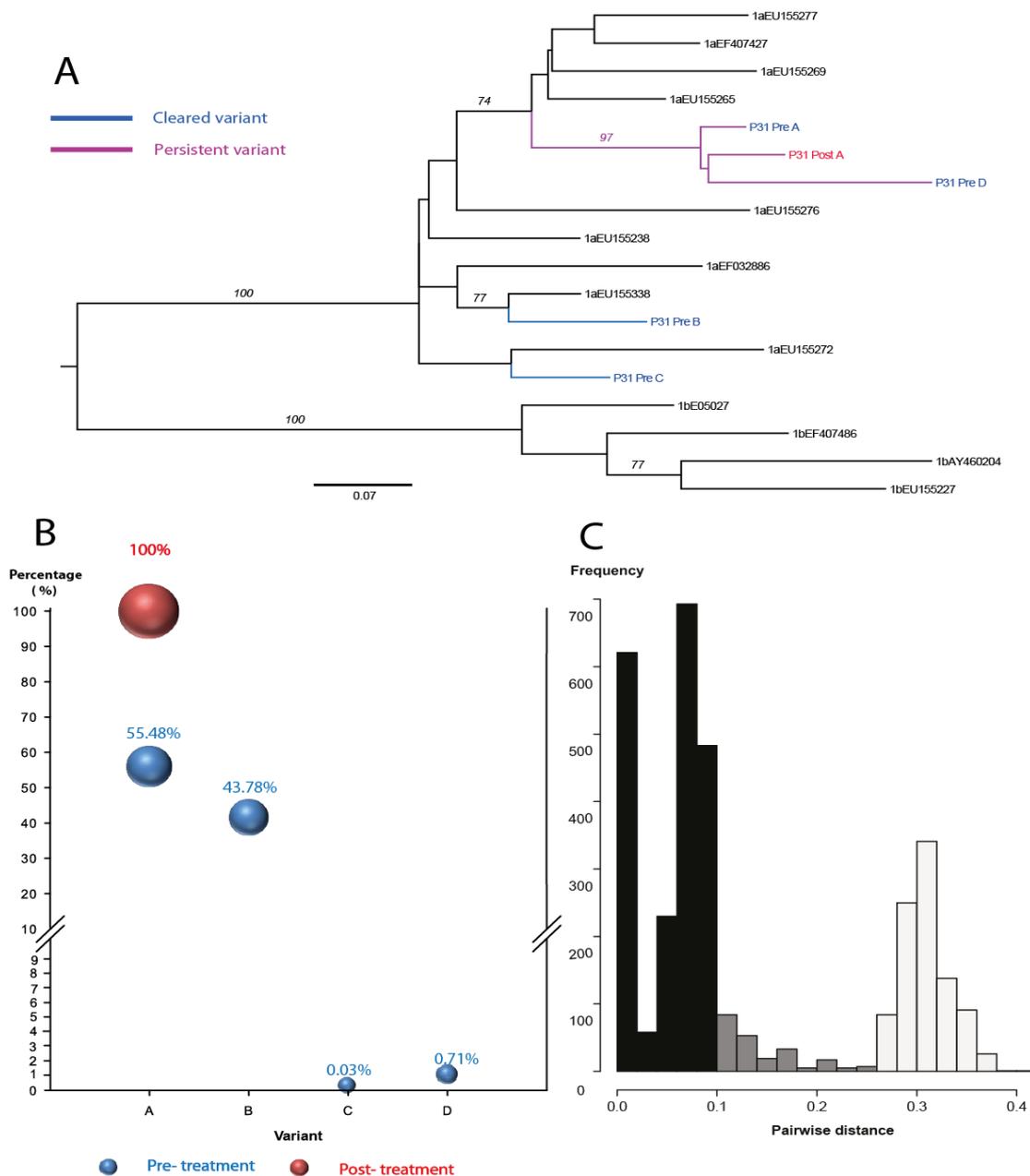
(A) ML tree was constructed using nucleotide sequences from paired samples from P112 (Null response, Persistent infection), and selected HCV reference sequences for the Los Alamos HCV database. A total of 6 (A-F) HCV variants detected. The analysis included; 25 clonal sequences (post-treatment) and 21246 reads derived from 454 pyrosequencing (pre-treatment). There was a total of 183 positions in the final dataset. (B) Bubble chart of the frequency of each variant (A-F) in pre- and post-treatment samples. (C) The pairwise distance between the most similar variants in the pre- and post-treatment samples (p-distance).

Figure 4-12: Comparison of viral complexity in paired serum samples in P118.



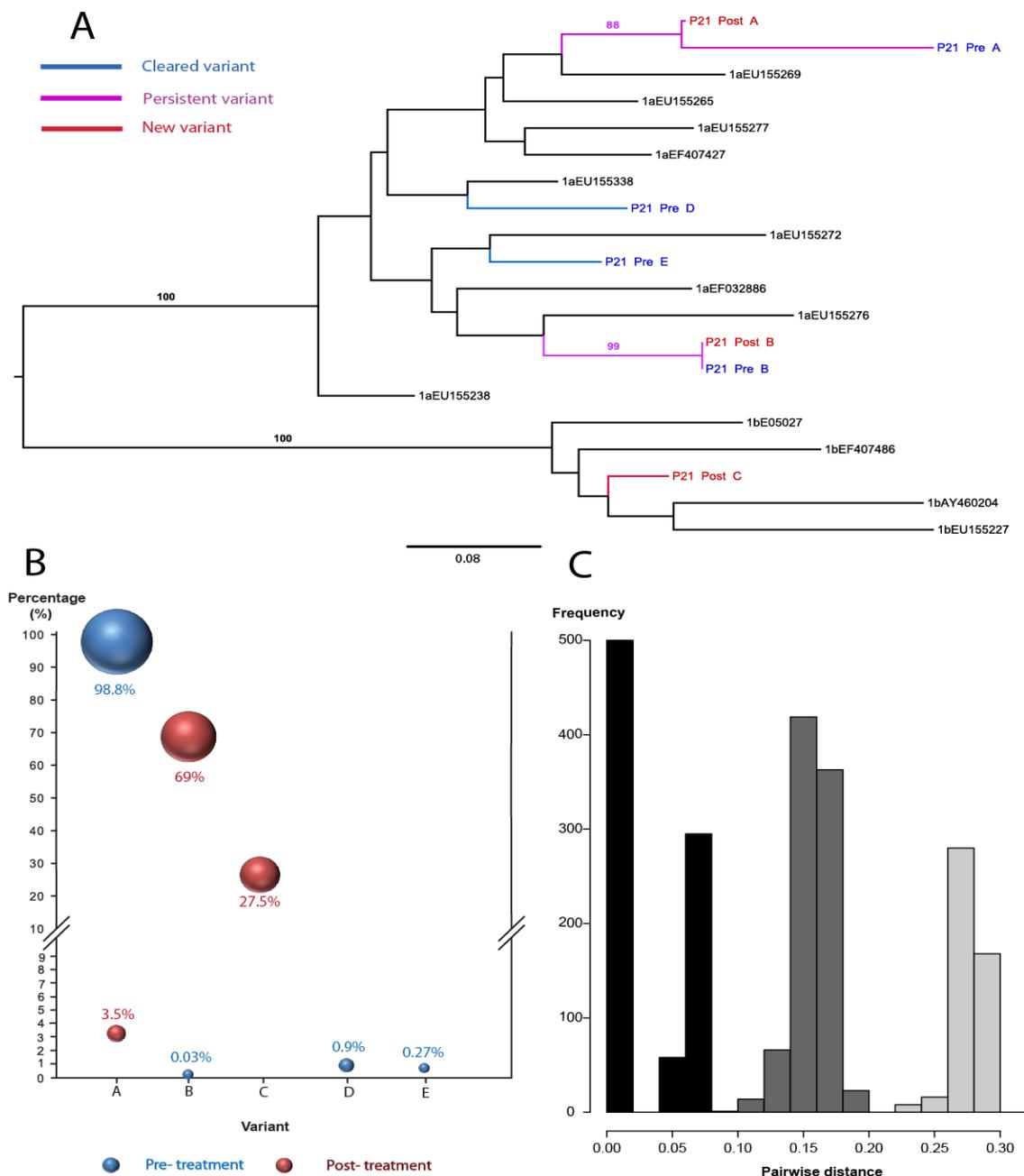
(A) A maximum likelihood tree was constructed using sequences from paired samples-P118 (Null response, Persistent infection with new dominance of a pre-existing minority variant), and selected HCV reference sequences for the Los Alamos HCV database. A total of 6 (A-F) HCV variants detected. The analysis included; 20 clonal sequences (post-treatment) and 28131 reads derived from 454 pyrosequencing (pre-treatment). There was a total of 183 positions in the final dataset. (B) Bubble chart of the frequency of each variant (A-F) in pre- and post-treatment samples. (C) The pairwise distance between the most similar variants in the pre- and post-treatment samples (p-distance).

Figure 4-13: Comparison of viral complexity in paired serum samples in P31.



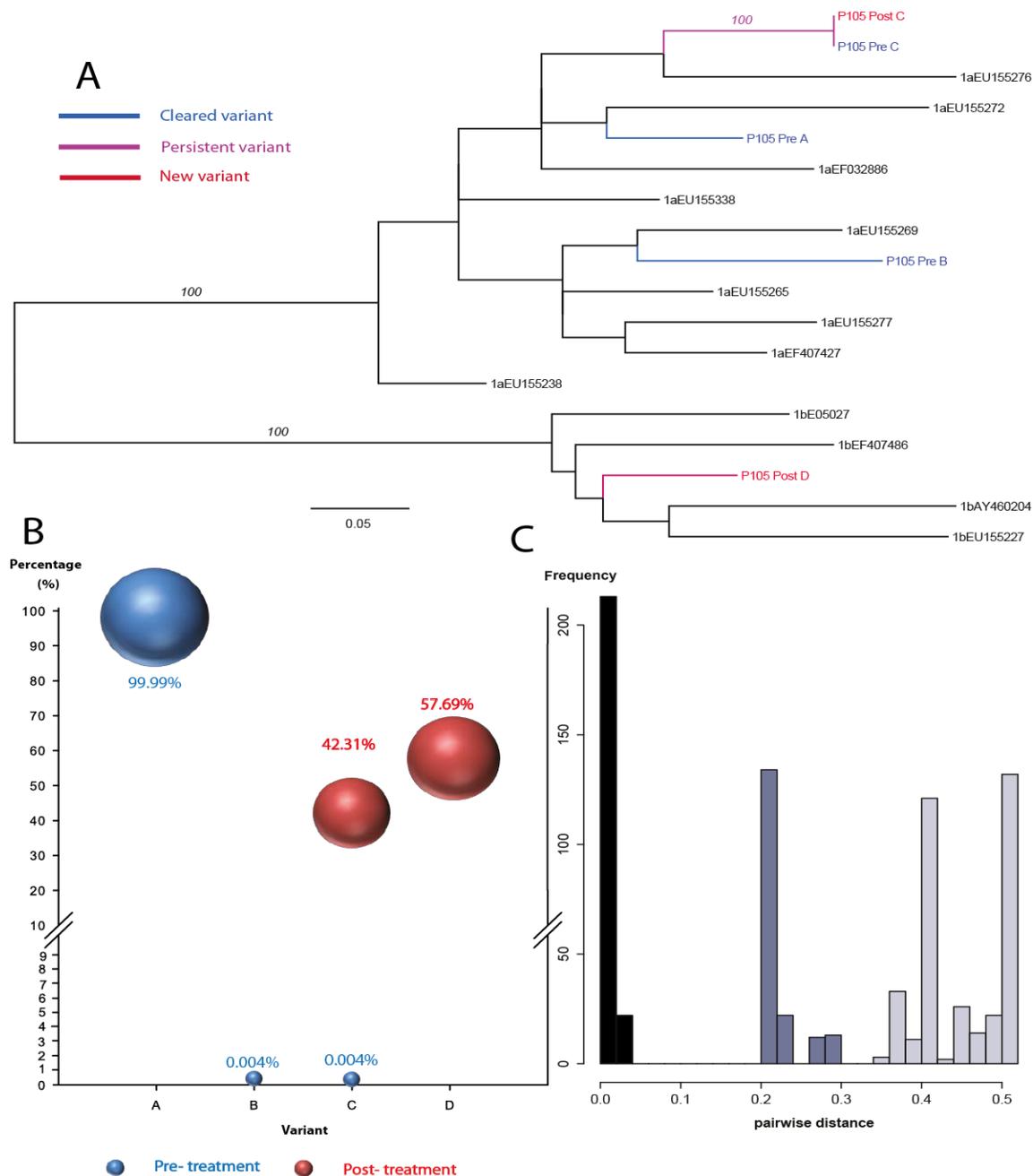
(A) ML tree was constructed using sequences from paired samples-P31 (Partial response, Persistent infection), and selected HCV reference sequences for the Los Alamos HCV database. A total of 4 (A-D) HCV variants detected. The analysis included; 35 clonal sequences (post-treatment) and 36422 reads derived from 454 pyrosequencing (pre-treatment). There was a total of 183 positions in the final dataset. (B) Bubble chart of the frequency of each variant (A-D) in pre- and post-treatment samples. (C) The pairwise distance between the most similar variants in the pre- and post-treatment samples (p-distance).

Figure 4-14: Comparison of viral complexity in paired serum samples in P21.



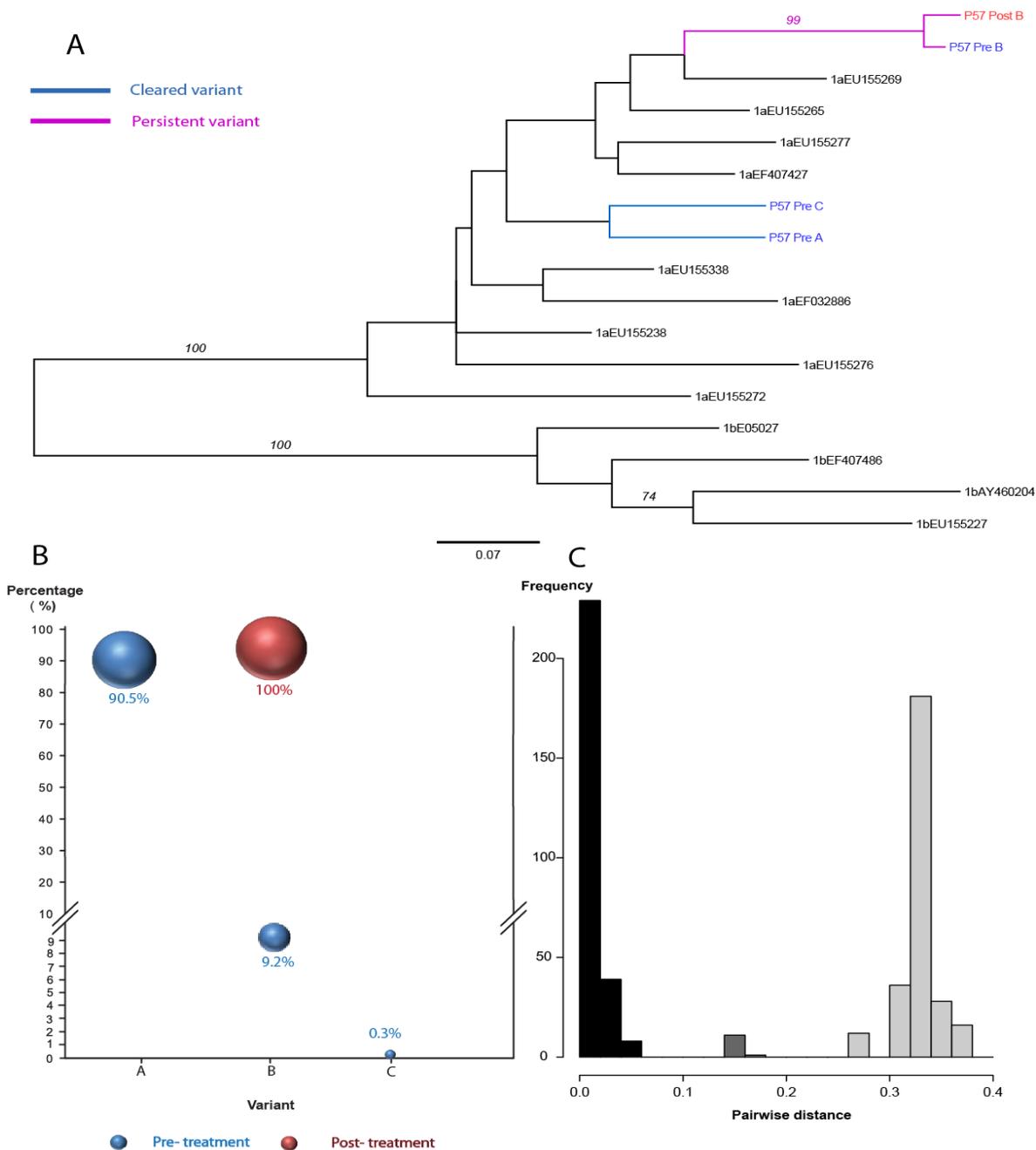
A) ML tree was constructed using sequences from paired samples-P21 (Partial response, Persistent infection with new post-treatment variant detected) and selected HCV reference sequences for the Los Alamos HCV database. A total of 5 (A-E) HCV variants detected. The analysis included; 29 clonal sequences (post-treatment) and 46755 reads derived from 454 pyrosequencing (pre-treatment). There was a total of 183 positions in the final dataset. (B) Bubble chart of the frequency of each variant (A-E) in pre- and post-treatment samples. (C) The pairwise distance between the most similar variants in the pre- and post-treatment samples (p-distance).

Figure 4-15: Comparison of viral complexity in paired serum samples in P105.



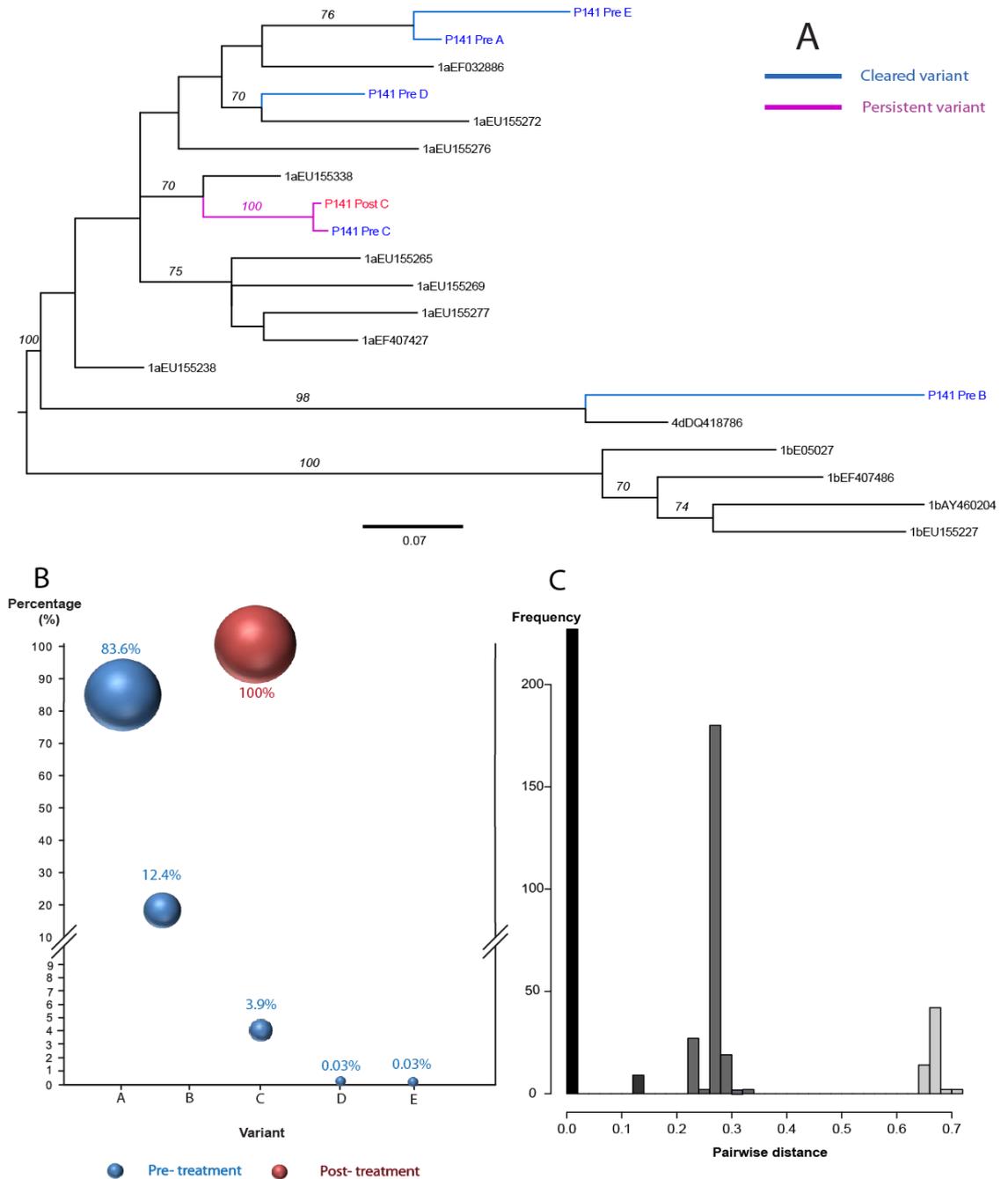
A) ML tree was constructed using sequences from paired samples-P105 (Partial response, Persistent infection with expansion of a pre-existing minority variant with a new post-treatment variant detected) and selected HCV reference sequences for the Los Alamos HCV database. A total of 4 (A-D) HCV variants detected. The analysis included; 26 clonal sequences (post-treatment) and 44296 reads derived from 454 pyrosequencing (pre-treatment). There was a total of 183 positions in the final dataset. (B) Bubble chart of the frequency of each variant (A-D) in pre- and post-treatment samples. (C) The pairwise distance between the most similar variants in the pre- and post-treatment samples (p-distance).

Figure 4-16: Comparison of viral complexity in paired serum samples in P57.



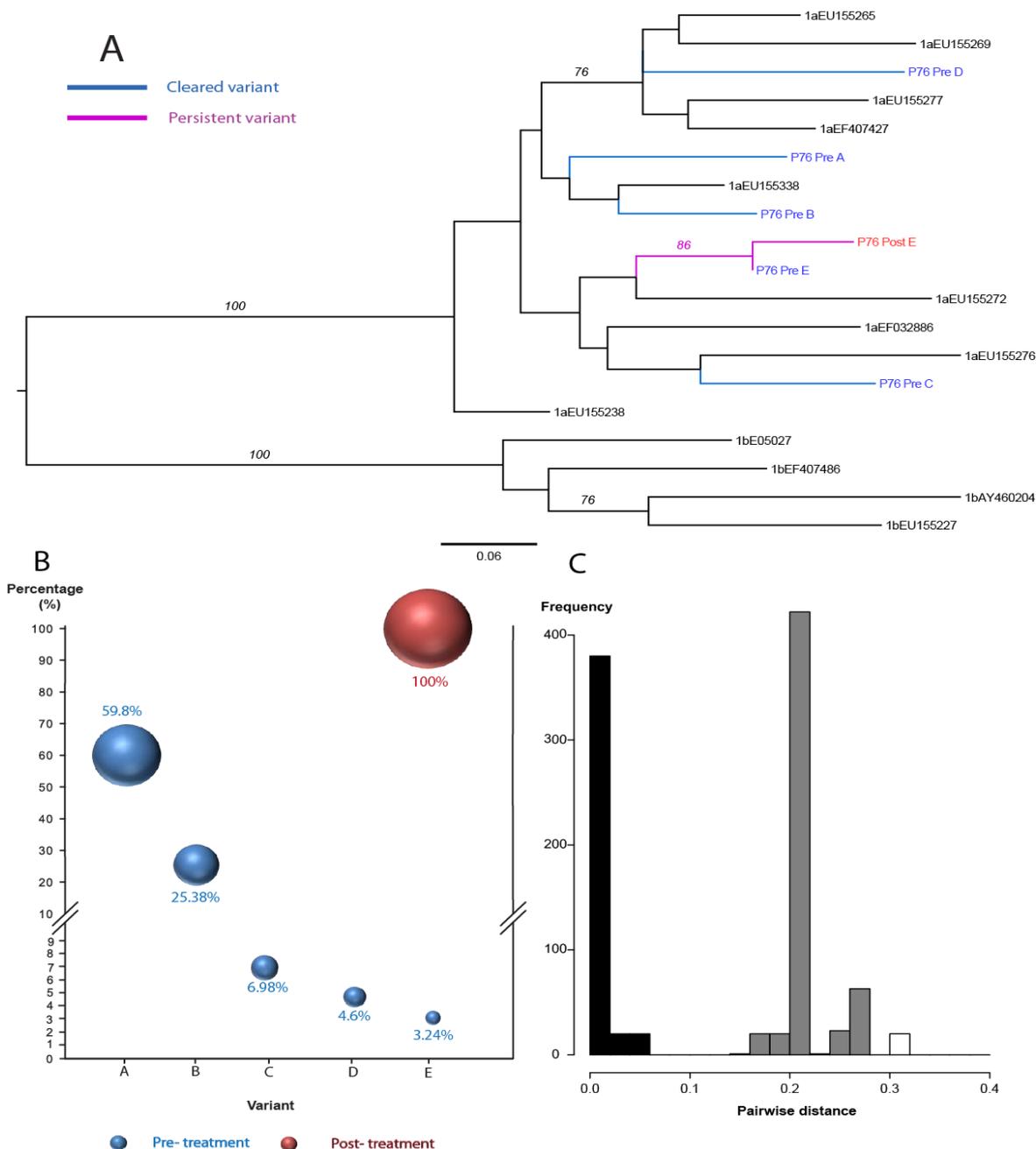
ML tree was constructed using sequences from paired samples-P57 (Relapse, Persistent infection with new dominance of pre-existing minority variant), and selected HCV reference sequences for the Los Alamos HCV database. A total of 3(A-C) HCV variants detected. The analysis included; 20 clonal sequences (post-treatment) and 23042 reads derived from 454 pyrosequencing (pre-treatment). There was a total of 183 positions in the final dataset. (B) Bubble chart of the frequency of each variant (A-C) in pre- and post-treatment samples. (C) The pairwise distance between the most similar variants in the pre- and post-treatment samples (p-distance).

Figure 4-17: Comparison of viral complexity in paired serum samples in P141.



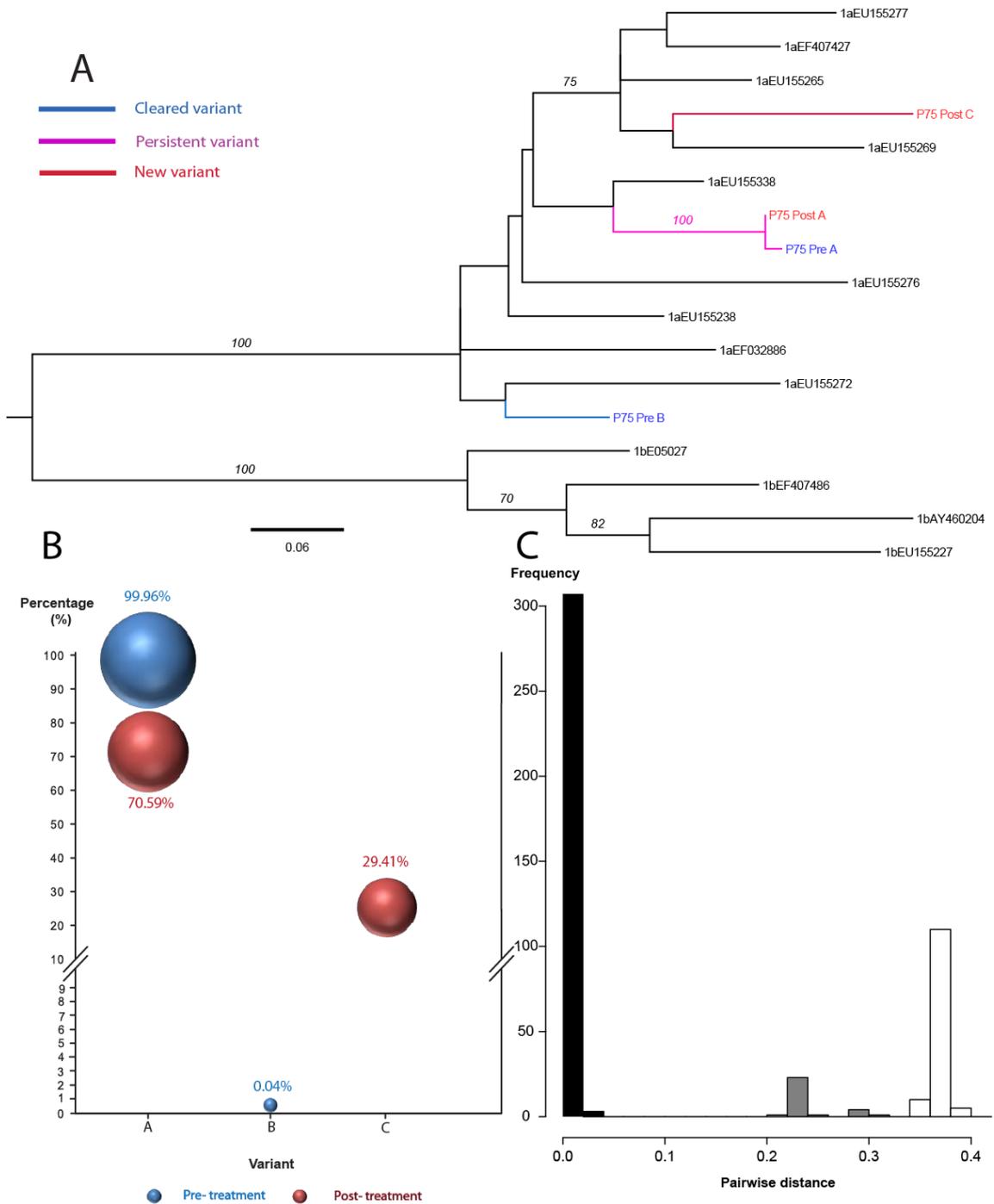
(A) ML tree was constructed using sequences from paired samples-P141 (Relapse, Persistent infection with new dominance of pre-existing minority variant) and selected HCV reference sequences for the Los Alamos HCV database. A total of 5 (A-E) HCV variants detected. The analysis included; 18 clonal sequences (post-treatment) and 23588 reads derived from 454 pyrosequencing (pre-treatment). There was a total of 183 positions in the final dataset. (B) Bubble chart of the frequency of each variant (A-E) in pre- and post-treatment samples. (C) The pairwise distance between the most similar variants in the pre- and post-treatment samples (p-distance).

Figure 4-18: Comparison of viral complexity in paired serum samples in P76.



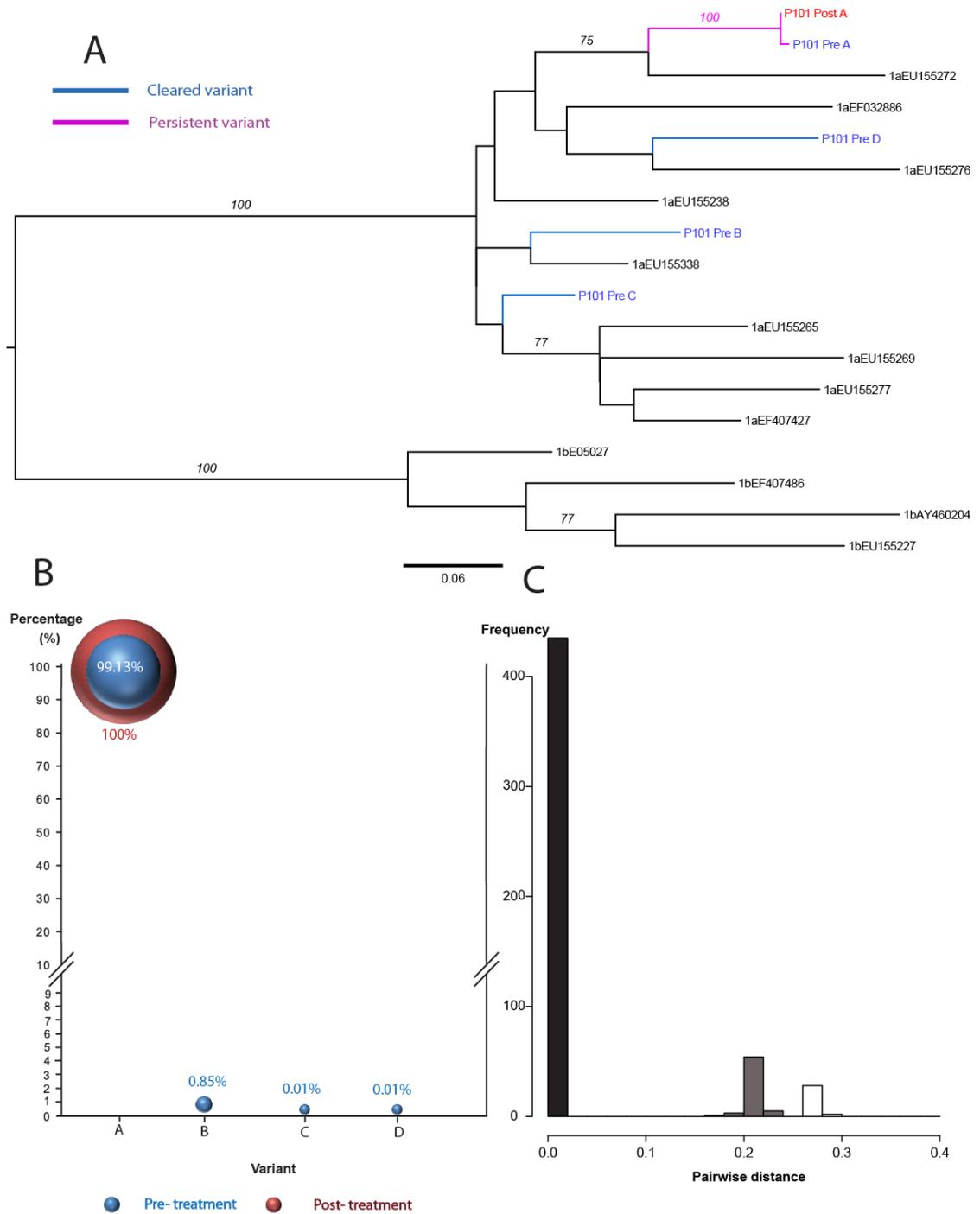
(A) ML tree was constructed using sequences from paired samples-P76 (Relapse, Persistent infection with new dominance of pre-existing minority variant) and selected HCV reference sequences for the Los Alamos HCV database. A total of 4(A-E) HCV variants detected. The analysis included; 20 clonal sequences (post-treatment) and 587 reads derived from 454 pyrosequencing (pre-treatment). There was a total of 183 positions in the final dataset. (B) Bubble chart of the frequency of each variant (A-E) in pre- and post-treatment samples. (C) The pairwise distance between the most similar variants in the pre- and post-treatment samples (p-distance).

Figure 4-19: Comparison of viral complexity in paired serum samples in P75.



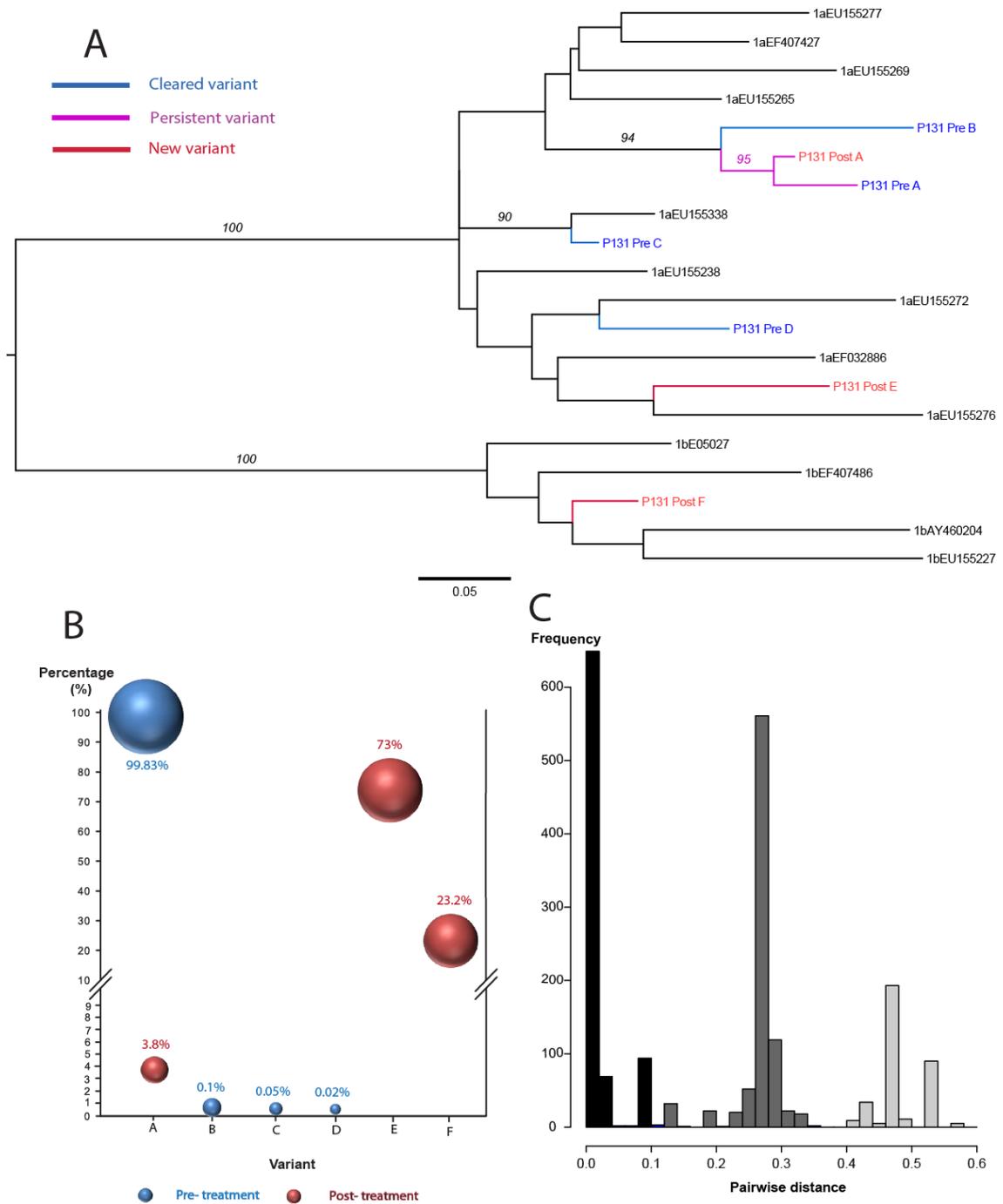
(A) A Maximum likelihood tree was constructed using sequences from paired samples-P75 (Relapse, Persistent infection with new post-treatment variant detected) and selected HCV reference sequences for the Los Alamos HCV database. A total of 3(A-C) HCV variants detected. The analysis included; 17 clonal sequences (post-treatment) and 23639 reads derived from 454 pyrosequencing (pre-treatment). There was a total of 183 positions in the final dataset. (B) Bubble chart of the frequency of each variant (A-C) in pre- and post-treatment samples. (C) The pairwise distance between the most similar variants in the pre- and post-treatment samples (p-distance).

Figure 4-20: Comparison of viral complexity in paired serum samples in P101.



(A) ML tree was constructed using sequences from paired samples-P101 (Relapse, Persistent infection) and selected HCV reference sequences for the Los Alamos HCV database. A total of 4(A-D) HCV variants detected. The analysis included; 24 clonal sequences (post-treatment) and 21265 reads derived from 454 pyrosequencing (pre-treatment). There was a total of 183 positions in the final dataset. (B) Bubble chart of the frequency of each variant (A-D) in pre- and post-treatment samples. (C) The pairwise distance between the most similar variants in the pre- and post-treatment samples (p-distance).

Figure 4-21: Comparison of viral complexity in paired serum samples in P131.



(A) ML tree was constructed using sequences from paired samples-P131 (Relapse, Persistent infection with new post-treatment variant detected) and selected HCV reference sequences for the Los Alamos HCV database. A total of 6 (A-F) HCV variants detected. The analysis included; 26 clonal sequences (post-treatment), and 19739 reads derived from 454 pyrosequencing (pre-treatment). There was a total of 183 positions in the final dataset. (B) Bubble chart of the frequency of each variant (A-F) in pre- and post-treatment samples. (C) The pairwise distance between the most similar variants in the pre- and post-treatment samples (p-distance).

4.3 Discussion

We hypothesized that high quasispecies diversity is a predictive factor of negative treatment outcome while treatment response might be predicted by low viral diversity. We further hypothesized that deep sequencing will give an accurate estimate of diversity providing a better predictor than the use of less sensitive techniques such as Sanger sequencing.

4.3.1 Predictors of outcome in patients treated with PegIFN α /RBV.

The response to antiviral therapy is influenced by several viral and host factors. Several viral factors have been reported to play a role in the response to treatment in HCV-infected patients including viral load and molecular profile of quasispecies at baseline.

Although patients with low baseline viral load have a favourable treatment response, with an arbitrary value to differentiate between high and low HC viral load commonly defined at 800,000 IU/ml, the influence of baseline viral load on response to treatment is unlikely to be only due to a high absolute number of copies of the RNA (Sallie, 2007).

Several studies have been carried out to assess variation within the HCV quasispecies and its effect on the response to treatment of chronic HCV infection, mainly in HIV-negative populations, but more recently in HIV-positive patients.

Payan *et al.* proposed an algorithm for prediction of treatment response in HIV/HCV co-infected patients using rapid virological response (RVR; at week 2 and week 4 after starting treatment) and HCV RNA level. The algorithm allowed the prediction of non-SVR as early as week 4 (Payan *et al.*, 2007). Another model integrated both host and viral variables including viral complexity as the main variable. It had a high positive predictive value in predicting the outcome of therapy in HCV-genotype 1b treatment-naïve patients (Saludes *et al.*, 2013).

Pre-treatment quasispecies complexity within the HVR1 region is negatively associated with SVR during chronic HCV infection using heteroduplex and clonal sequence analysis in both HIV-positive and negative patients (Shire et al., 2006, Ueda et al., 2004) (Yeh et al., 2002, Moreau et al., 2008). However, Abbate *et al.* did not find any significant difference in Shannon entropy between the patients who failed treatment and those who achieved SVR in a homogeneous genotype population (Abbate et al., 2004).

Another study in HCV genotype 1a and 1b infected patients by Chambers *et al.*, concluded that viral diversity could not significantly predict treatment responders from non-responders (Chambers et al., 2005). Meanwhile, Fan *et al.* reported that early viral response is correlated with a high genetic diversity at base line, although the end point was at 12 weeks of treatment, so a direct comparison is not possible as SVR was not investigated in that study. Moreover, The difference between two groups did not reach statistical significance (Fan et al., 2009).

These contradictory results could be attributed to techniques used to assess HCV quasispecies diversity which may be another source of data discrepancy (Fan et al., 2009). Clonal analysis bias towards missing low-frequency variants could affect the analysis of viral complexity. Furthermore, until now, the presence or absence of low-frequency variants that may contribute to quasispecies complexity may have been undetected; the introduction of NGS as a sensitive tool will shed more light on the viral complexity due to its ability to detect minor variants.

In the present study, we retrospectively investigated the diversity profile of two groups of acute HCV/HIV co-infected patients using NGS; a group of ten patients who failed to respond to treatment (TF group), and another group of ten patients who achieved SVR after standard treatment (SVR group).

When we compared the genetic diversity between the two groups, we observed that the patients in the TF group had a high degree of inpatient viral diversity in addition to a higher frequency of variable sites in the HVR1 when

compared to the SVR group. Consistent with these observations, the entropy analysis between the two groups also showed statistically significant differences in the variability of amino acids in the TF group ($p < 0.05$) when compared with the SVR group.

Our data suggest that it may be possible to predict treatment outcome using viral diversity measurement at baseline. This is in keeping with previous studies that have investigated viral diversity as a potential predictor of the treatment outcome (Farci et al., 2002, Chambers et al., 2005, Abbate et al., 2004, Yeh et al., 2002).

The two groups of our study are well defined and matched (e.g., treatment naïve, all infected with genotype 1, concomitant infections rigorously excluded), and quality genetic data were obtained. However, the small sample size of our study makes it difficult to draw conclusions on the correlation between viral diversity and response to treatment in HCV. Therefore, larger studies are needed to confirm the present findings.

4.3.2 Transmission diversity

In order to distinguish infections initiated by a single variant from those where multiple variants entered the host, we used Poisson Fitter to estimate Hamming distance (HD) frequencies. Although the correlation between high viral diversity and low SVR rate has been reported, the time at which viral diversity reaches a threshold that causes failure of treatment is unknown.

It remains to be determined whether the diversity level at transmission defines the outcome, or if there is a gradual expansion of the quasispecies over time offering higher possibility of achieving SVR if treatment is initiated during acute disease. If transmission diversity is a factor, it might be expected that patients with haemophilia as a route of transmission would have greater diversity and lower SVR rates than intravenous drug users due to the larger inocula introduced; further studies is needed to explore this hypothesis.

In our study, the samples in the TF group did not follow star phylogeny, with different variants detected, compared to samples following star phylogeny in the responder group where one or two variants only were identified as founder variants. This may be attributed to higher transmission diversity in the TF group, but this study was not set up to investigate this hypothesis. Another possibility is that the quasispecies changed after transmission because of adaptation to immune pressure.

If the time at which the quasispecies diversity reaches a threshold level defines the time of treatment, this could potentially be used to target treatment appropriately in individual patients.

4.3.3 Definition of relapse and re-infection in HCV

In the setting of antiviral treatment in HCV/HIV co-infected MSM, the reappearance of viral RNA after treatment is often assumed secondary to re-infection. This assumption was based on behavioural studies that have shown that HIV-infected patients with acute HCV are likely to be at high risk of re-exposure (Grebely et al., Danta et al., 2007).

Multiple variant infections in HCV/HIV co-infected individuals have been reported in up to 40% of those infected (Thomson et al., 2011). However, the screening methods may lack sensitivity for the detection of low-frequency variants. Thus this prevalence may be underestimated. Moreover, the frequency of sampling may affect estimates as infrequent sampling may miss transient infections. In this study, pre- and post-treatment samples were sequenced to investigate whether treatment failure occurred due to viral relapse or re-infection.

Our findings indicate that multiple infections are common in early HCV infection, reaching 100% in our cohort, with a mean of 3.8 variants present prior to treatment. Other studies may not have detected multiple strains because of limited sampling, primer selection bias, or reduced sensitivity of the method used (Wang et al., 2010).

The samples used in this study were collected early in the infection timeline. Hence, the presence of multiple viral strains at a single time-point is attributed mainly to simultaneous transmission or superinfection within a short timeframe as the time required for strain, sub-genotype or genotype evolution would be longer than the length of infection in each individual.

Mixed HCV infection is known to be transient due to immune selective pressure or competitive growth between variants leading to the outgrowth of the fitter variants (Pham et al., 2010). In contrast, we found that mixed infection was present at both time points examined in the majority of our patients; variation in quasispecies composition was common suggesting that certain strains may have been positively selected during treatment. The presence of multiple variants in all post-treatment samples in this cohort suggests that re-infection is not the only mechanism of treatment failure in this cohort. The presence of previously undetected variants in 40% of post-treatment samples could represent superinfection, but could also represent variants present below the detectability limit of NGS analysis.

In null responders, all patients had a persistent strain pre- and post-treatment and 50% of patients had a persistent variant that remained dominant throughout. Emerging dominance of a pre-existing minority variant occurred in two patients, one of whom had evidence of a previously undetected minority variant following treatment. In the remaining patient, a new majority variant emerged after treatment in addition to a persistent minority variant. This could represent superinfection or emergence of a strain below the limit of detection of pyrosequencing (in this case, 46,755 sequence reads were analysed pre-treatment).

In partial responders, we found evidence of persistent variants in all three patients. In two patients (P21 and P105), the dominant variant cleared following treatment while the third patient (P31) cleared a variant representing 44% of the pre-treatment viral population. In all cases, clearance was concurrent with a fall in viral load during treatment. A previously undetected variant was present in

two of three patients in post-treatment samples. Such variants may have been positively selected from minority variants undetected in the pre-treatment sample or could represent superinfection during treatment.

In those with viral relapse, all patients had evidence of a pre-existing variant present in post-treatment samples. Three patients (P57, P76, and P141) had evidence of emerging dominance of pre-existing minority strains (rising from 3-9% of in pre-treatment samples to 100% in post-treatment samples). Two patients (P101 and P75) had the same majority variant present pre- and post-treatment. In P75, a previously undetected minority variant was also detected post-treatment. One patient (P131) is the most likely case of superinfection in the cohort as he had evidence of a persisting minority variant and in addition, two new variants were detected.

In this study, NGS revealed that all patients who failed to respond to treatment had at least one persistent variant. If Sanger sequencing had been used alone, persisting variants would have been detected in only 34% of cases. The new variants detected by direct Sanger sequencing were found to be due to the emergence of a minority variant already present in the pre-treatment sample. It is likely that such emergent variants represent viral strains with reduced sensitivity to antiviral medications.

There is no evidence of re-infection in this cohort. However, view the ongoing behavioural risk of this group, we cannot rule out the possibility of re-infection from the same source. A new variant has been detected in as many as 6/15 (40%) of the patient cohort in whom a previously undetected (new) variant was found. In three of these cases (20%), the new variant represented the majority variant and in three cases (20%), the new variant was a minority variant. The presence of new variants in post-treatment samples could be due to superinfection, but it also may represent previously undetected minority variants selected by treatment or compartmentalised variants within different regions of the liver, lymphocytes or the central nervous system (Sobesky et al., 2007, Forton et al., 2004a, Thomson et al., 2011, Blackard et al., 2007).

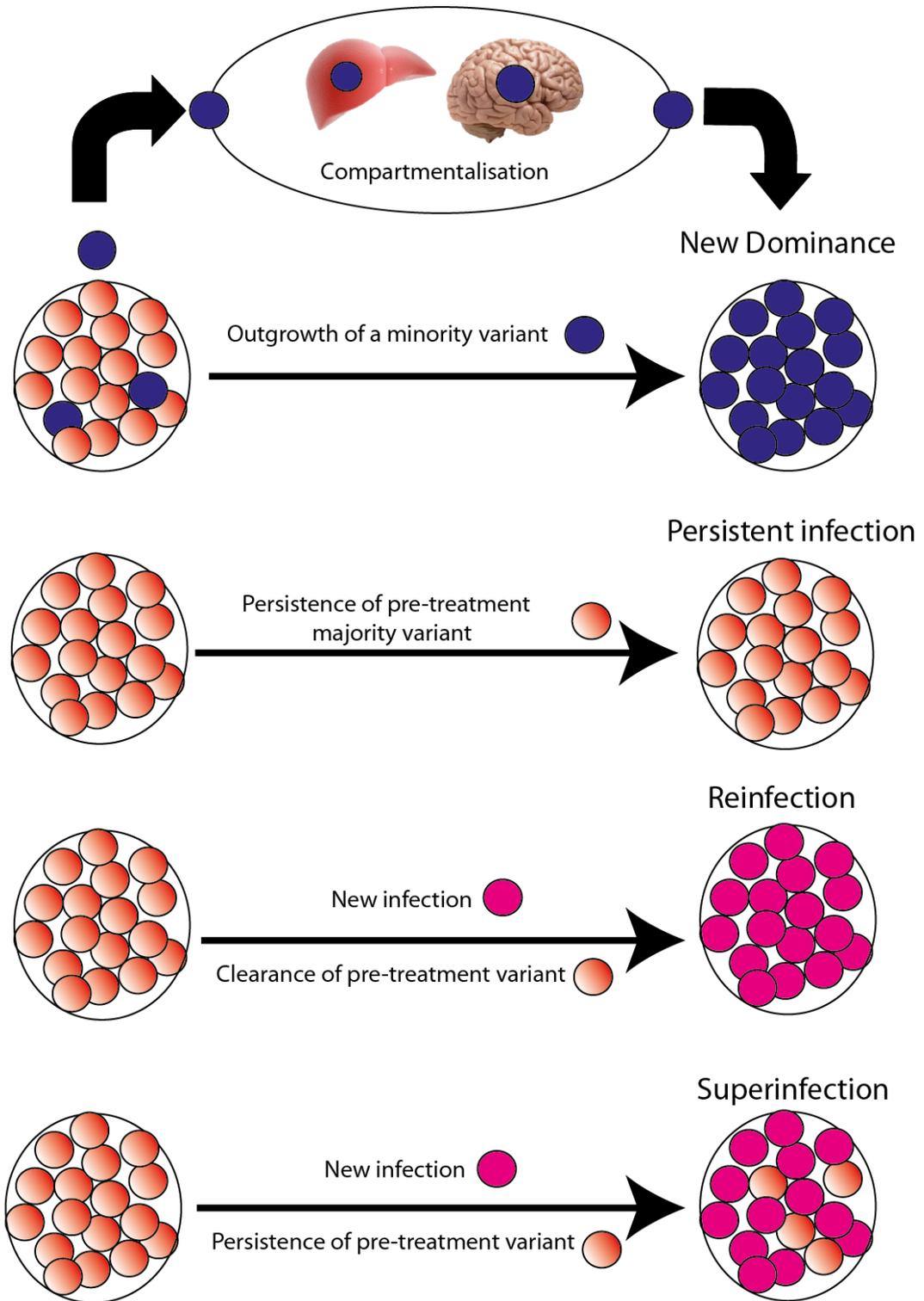
Several studies of HCV re-infection have been limited by the lack of sensitivity of detection method or even the absence of phylogenetic support. Lambers *et al.* reported a high incidence rate of HCV re-infection (15.2 per 100 person years) among HIV-infected MSM, who previously cleared HCV after treatment (Lambers *et al.*, 2011). Martin *et al.* also described a high risk of HCV re-infection among HIV-positive MSM who were either treated for or who spontaneously cleared initial HCV infection (Martin *et al.*, 2013). In German PWID cohort, a re-infection rate of 0-4.1/100 person-years has been reported (Grady *et al.*, 2012). These studies were not designed to identify multiple variant infections prior to treatment, nor the emergence of minority variants following treatment. We propose therefore that the definition of re-infection, persisting infection or superinfection should always be based on rigorous viral sequencing techniques (Figure 4-22).

4.3.4 HCV compartmentalisation

Detection of HCV RNA in extrahepatic compartments is reported. However, the role of compartmentalised virus acting as a reservoir for future recrudescence is, as yet, relatively unexplored (Sobesky *et al.*, 2007, Forton *et al.*, 2004a, Forton *et al.*, 2004b, Thomson *et al.*, 2011, Blackard *et al.*, 2007). It has been considered an independent predictor of treatment outcome (Di Liberto *et al.*, 2006). Hara *et al.* demonstrated that in late relapsers, HCV variants could be detected in liver biopsies during the aviraemic phase highlighting the possibility of compartmentalisation in patients with viral relapse (Hara *et al.*, 2013).

HCV compartmentalization is relatively common among patients with HCV/HIV co-infection. Hence, measuring viral diversity in the serum/plasma alone may not represent virus replicating within the liver or extrahepatic compartments. Therefore, the higher sensitivity offered by NGS in the detection of minority variant may not resolve the complexity of intra-host viral population in other compartments. This issue needs to be considered in future studies (Blackard *et al.*, 2014).

Figure 4-22: Viral dynamics during treatment failure.



Four different outcomes are expected after applying drug-selective pressure.

4.4 Conclusion

Deep sequencing has a potential role in laboratory diagnosis of HCV as it offers a better understanding of viral populations than current techniques; however, further clinical studies are required to validate this technology in the clinical setting.

The complexity of HCV evolution affects our understanding of the natural history of disease. Hence, a phylogenetic analysis is required to understand the viral diversity in any clinical samples, while considering that variants detected in peripheral blood may not entirely reflect the viral dynamics due to lack of representation from other sites (e.g. liver biopsies).

Using NGS we were able to address two important questions in the field of HCV; SVR of acute HCV infection can be predicted by low viral diversity within the quasispecies population, and the emergence of new viral strains following treatment failure is most commonly associated with emerging dominance of pre-existing minority variants rather than re-infection. Super-infection may occur in this cohort, but re-infection is overestimated by current diagnostic techniques.

Chapter 5: HCV antiviral drug resistance

5.1 Background

As discussed in Chapter 1, HCV DAAs target several proteins and functions, namely the NS3 protease, NS5A and the NS5B RNA-dependent polymerase (Heim, 2013, Schneider and Sarrazin, 2014). IFN-free regimens are now recommended as first-line therapies in the latest HCV treatment guidelines issued by the European Association for the Study of the Liver (EASL) and the American Association for the Study of the Liver (AASLD) and will also be recommended in the forthcoming WHO guidelines in 2015/2016 (EASL, 2015, AASLD/IDSA, 2015). The question of whether the successful launch of DAAs will be limited by the emergence of drug resistance is a subject of intense debate.

The introduction of the first generation of protease inhibitors (PI) in combination with PegIFN α /RBV provided the first evidence of an increased likelihood of achieving SVR in both HCV mono-infected and HCV/HIV co-infected populations by directly targeting viral function. In genotype 1-infected patients, triple therapy with PegIFN α /RBV and the NS3 PIs, TVR or BOC, showed significantly improved SVR rates compared to PegIFN α /RBV alone (Poordad et al., 2011, Bacon and Khalid, Zeuzem et al., 2011, Enomoto et al., 2013). However, TVR and BOC have limited efficacy against non-genotype 1 NS3 proteases and so were restricted for use to genotype 1-infected patients.

Since 2011, additional PIs including SMV and PTV have demonstrated pan-genotypic activity. Newer DAAs now include inhibitors of the NS5B polymerase (SOF and DSV) and NS5A (DCV, LDV, and OBV). The efficacy profiles of the different DAAs are described in detail in Section 1.12.

Drug resistance information on DAAs used to treat HCV infection is accumulating rapidly (Sarrazin and Zeuzem, 2010, Welsch et al., 2012). Strains within the HCV quasispecies may carry mutations that confer resistance to DAAs by either preventing elongation of RNA synthesis or blocking other functions required for replication.

Naturally occurring RAVs are selected early in monotherapy (Sarrazin et al., 2007, Susser et al., 2009), and natural polymorphisms are prevalent in treatment naïve populations (Kuntzen et al., 2008).

The prevalence of RAVs is variable and depends on the domain involved in binding the DAAs, the exposure to a drug, the genetic barrier to resistance, and HCV genotype. Moreover, host immune responses exert powerful selection pressure that influence HCV genetic diversity and replication dynamics, thereby affecting the development of RAVs (Gaudieri et al., 2009). The ability of RAVs to persist and induce treatment failure is related to their fitness compared to the wild-type virus (Welsch et al., 2012).

As described in Chapter 3, the detection of RAVs depends primarily on the sensitivity of the method that is used. The impact of RAVs present at very low frequencies is not well understood, and further studies are needed. It is very likely, however, that most drug resistance variants are present at a low level before subsequent selection under treatment pressure rather than arising *de novo* following the start of treatment.

RAVs must be able to replicate efficiently in order to occupy replication space left by susceptible WT virus during drug exposure. Thus, a low-level resistant RAV that can propagate efficiently in the presence of the drug has more clinical significance than a highly resistant RAV with low replication fitness. The fitness of a resistant variant may be restored with compensatory mutations that allow it to replicate efficiently in the presence of the drug, and even to persist after drug withdrawal.

In the following sections, an exhaustive literature review was carried out to create a database of all reported *in vitro* and *in vivo* RAVs that confer resistance against current DAAs (SMV, PTV, LDV, OBV, DCV, DSV, or SOF). A total of 1070 journal hits were retrieved from a literature search and 140 reports fulfilled the inclusion criteria for our review.

5.1.1.1 NS3/4A protease inhibitors.

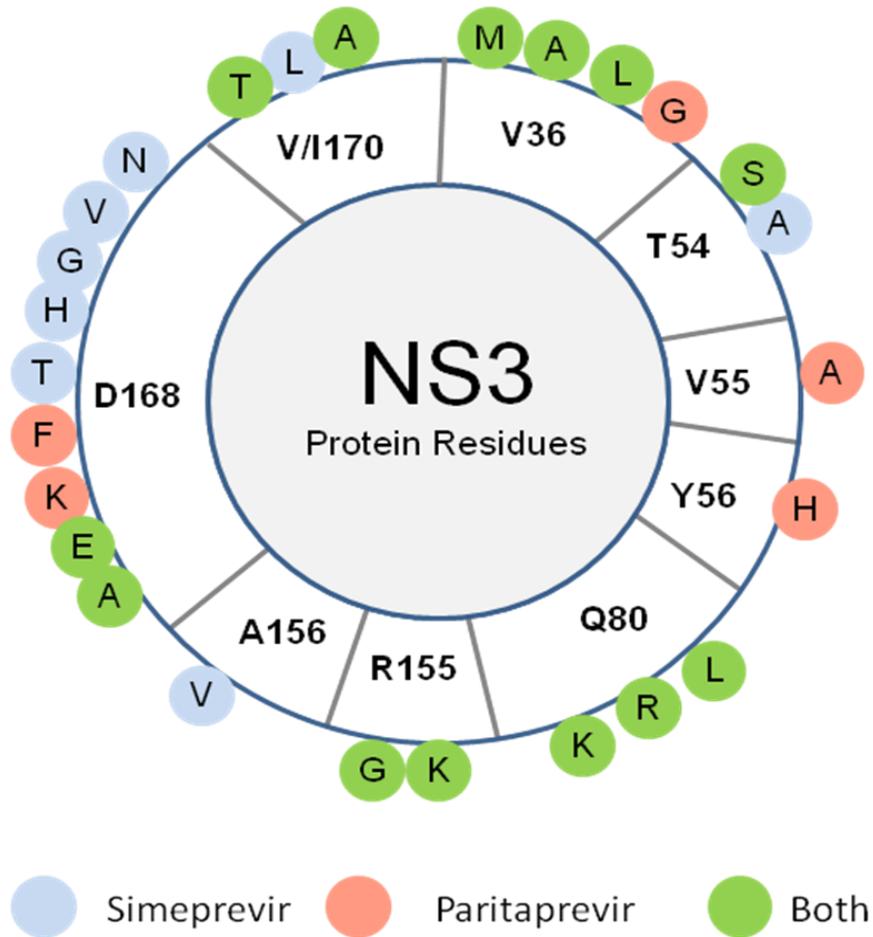
Currently, there are two types of NS3 PI. The first generation PIs; TVR and BOC, were designed to form covalent bonds within the active site of the viral protease, whereas second generation PIs, including SMV and PTV, are both non-covalent inhibitors of NS3. NS3 PIs act by inhibiting protease-mediated cleavage at boundaries between non-structural proteins encoded by the viral genome (Lin et al., 2006). This is described further in Section 1.12 (Romano et al., 2010).

In vivo RAMs are shown in Figure 5-1 and *in vitro* mutations conferring resistance to SMV are listed in the Appendix (7.21). The natural Q80K resistance polymorphism has a prevalence that varies geographically - in South America, it occurs in 9.1% of genotype 1a isolates, in Europe 19.4%, and in North America 48.1% but is rarely detected in genotype 1b (0.5%) (Schneider and Sarrazin, 2014). It can also emerge on therapy; the median time until loss of the Q80K mutation is between 24 and 36 months after stopping treatment (Schneider and Sarrazin, 2014).

PTV shows a mean viral load decline of 4 log₁₀ when administered for three days as monotherapy. In order to increase its bioavailability and prolong its half-life, it is boosted with ritonavir leading to a reduced dosage regimen requirement (Gentile et al., 2014a, Pilot-Matias et al., 2015). Ritonavir-boosted PTV/ OBV/ DSV (PrOD) ± RBV for 12 weeks is associated with SVR rates of 90-97.5% in naïve and treatment-experienced non-cirrhotic patients with HCV genotype 1 (Ferenci et al., 2014) (Zeuzem et al., 2014b).

Mutations that present the highest level of resistance to PTV are D168A/H/Y and R155K, which result in a x59-219 and x37-43 fold increase in resistance compared to wild-type respectively. RAMs reported *in vivo* are shown in Figure 5-1 and *in vitro* in the Appendix (7.19). Overlapping cross-resistance profiles for SMV and PTV as shown in Figure 5-1.

Figure 5-1: Schematic representation of Protease inhibitors resistance associated mutations detected *in vivo*.



Reported RAMs to SIM and PTV occur at nine residues. The color code indicates the agent against which the residue confers resistance.

5.1.1.2 NS5A inhibitors

Inhibitors of NS5A interfere with viral replication and assembly, but their mechanisms of action are unclear. A single dose of DCV, OBV or LDV results in a substantial reduction in viremia in patients with genotype 1 HCV (Nettles et al., 2011), but they have a low genetic barrier to resistance (Pawlotsky, 2013a, Nakamoto et al., 2014). RAVs reported with DCV, LDV, and OBV treatment are summarised in Figure 5-2.

5.1.1.3 NS5B inhibitors

NS5B inhibitors can be divided into two different groups; non-nucleoside inhibitor and nucleoside analogues (Asselah and Marcellin, 2011). SOF is a nucleoside analogue that serves as a chain terminator; hence, it hampers the elongation of RNA transcripts. In contrast, DSV is a non-nucleoside inhibitor that binds to a distinct site on the HCV RdRp leading to disruption of viral replication (Asselah and Marcellin, 2011).

Mutations that confer resistance to SOF and DSV are presented in Figure 5-3. RAVs reported *in vitro* are detailed in the Appendix (7.29 and 7.31). The first *in vitro* RAV reported against SOF was the substitution of serine 282 with threonine (S282T) (Wohnsland et al., 2007). However, as this mutation results in a large reduction in replicative capacity (up to 20%) compared with WT replicon, it is not prevalent as a natural polymorphism (Ludmerer et al., 2005). None of the 1292 patients reported in phase 3 studies harboured the S282T mutation (Schneider and Sarrazin, 2014).

Compensatory mutations may enhance replication capacity without altering the level of resistance conferred by the S282T substitution (Ali et al., 2008). SOF maintains activity against HCV variants harbouring mutations conferring resistance to other classes of DAAs, which present it as a potent option for retreatment in case of treatment failure (Abraham and Spooner, 2014). Residues associated with resistance to the non-nucleoside inhibitor DSV have been reported more frequently *in vitro* rather than *in vivo*.

The mutation with the highest level of resistance is C316Y, which results in 1472-fold resistance compared to wild-type *in vitro*.

5.1.2 Replicative fitness

Viral fitness is defined as the replication capacity of mutated variants in proportion to the replication capacity of the WT virus. It is a major determinant of the frequency of RAVs within the viral quasispecies, as the persistence of RAVs depends on their viral replication fitness (Welsch et al., 2012).

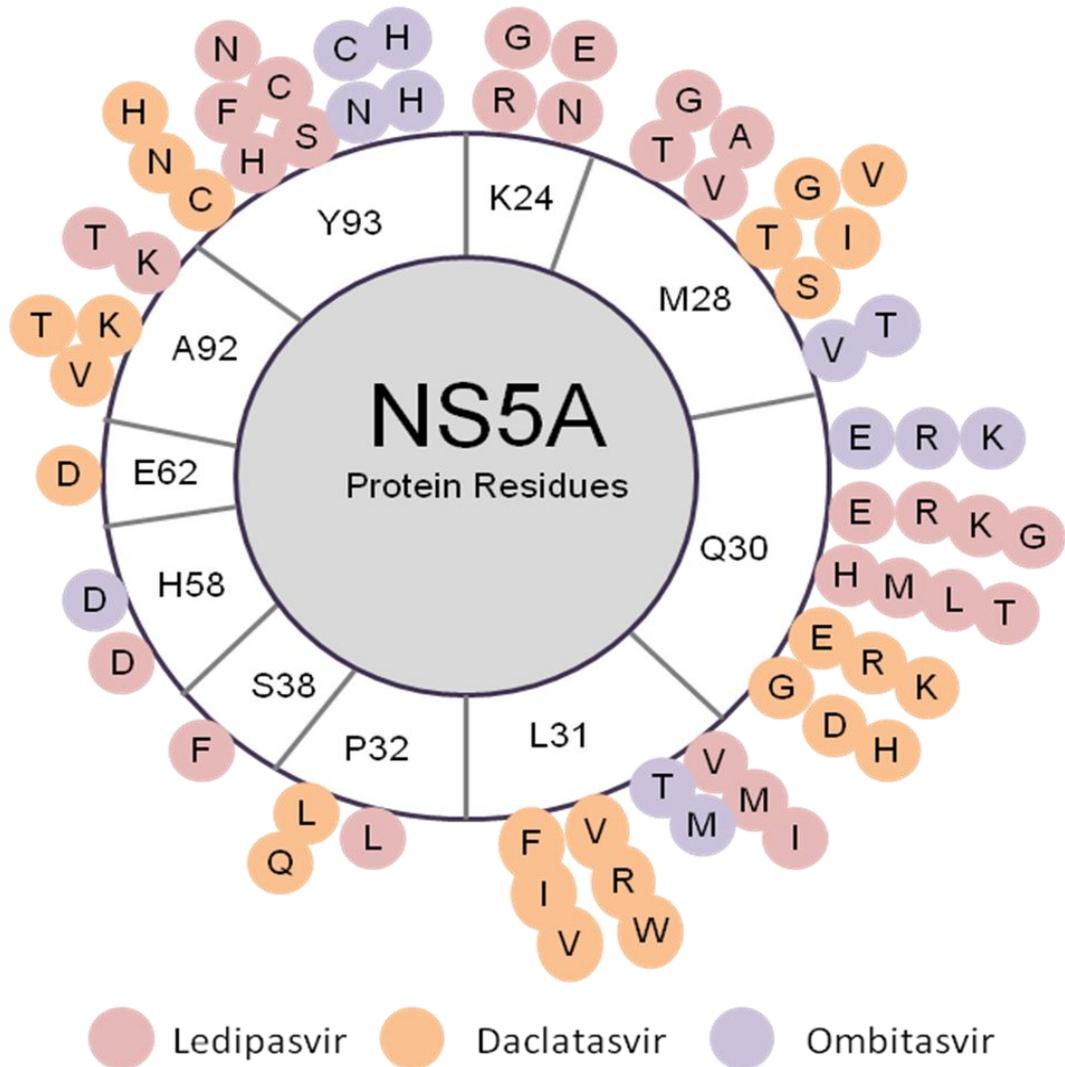
The HCV replicon system is extensively used to assess the replicative fitness of HCV variants *in vitro* (Bartenschlager and Lohmann, 2000). Viral fitness may also be estimated *in vivo* using viral load and clonal analysis at different time points after the end of therapy to assess the growth rate of RAVs compared to WT virus after the withdrawal of drug-selective pressure (Susser et al., 2009). RAVs regularly emerge, but they do not persist in the absence of selective pressure due to the intrinsic fitness cost of resistance mutations (Rong and Perelson, 2010, Adiwijaya et al., 2010).

In this chapter, we describe the construction of an HCV genotype 1a subgenomic transient replication system to evaluate the replication fitness of mutations in this genotype, and we describe a bioinformatics approach to predict putative SOF resistance mutations (designed in collaboration with Dr Sreenu Vattipally).

5.1.3 Importance of resistance testing

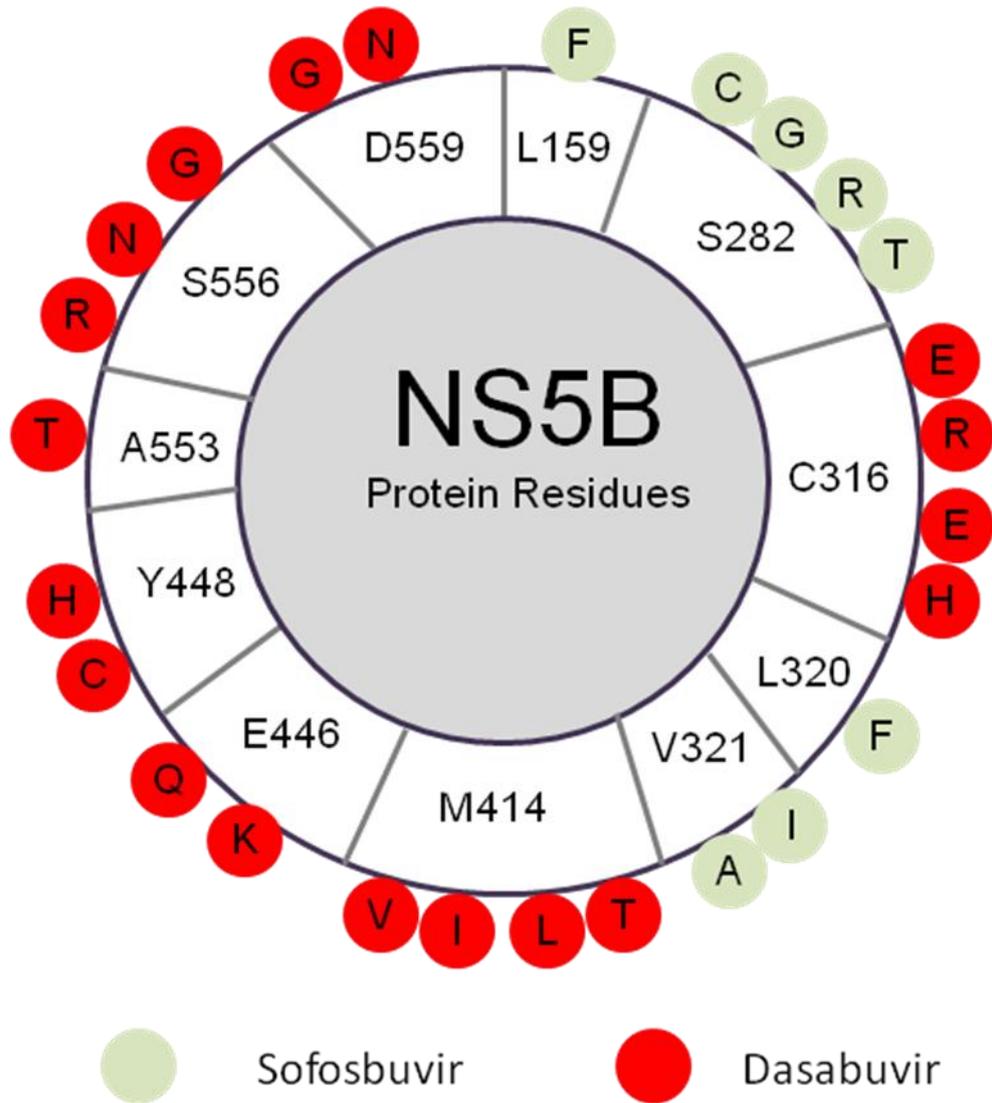
Several studies have identified mutations that confer resistance to DAAs (Welsch et al., 2012). Understanding resistance patterns is important to enable the use of optimum therapies. The emergence of RAVs under drug selection pressure leads to reduced DAA efficacy resulting in treatment failure (Pawlotsky, 2011, Schneider and Sarrazin, 2014). Almost all individuals that have failed DAA treatment have had a resistant mutation present in their sample (Kuntzen et al., 2008).

Figure 5-2: Schematic representation of NS5A inhibitors resistance associated mutations detected *in vivo*.



Reported RAMs to NS5A inhibitors occur at ten residues. The color code indicates the agent against which the residue confers resistance.

Figure 5-3: Schematic representation of NS5B inhibitors resistance associated mutations detected *in vivo*.



Reported RAMs to NS5B inhibitors occur at 11 residues. The color code indicates the agent against which the residue confers resistance.

The clinical significance of pre-existing minority RAVs at baseline and the impact of their presence on treatment outcome are still partially described (Thomson et al., 2009). A simple model explaining the emergence of HCV resistance to DAA drugs is shown in Figure 5-4.

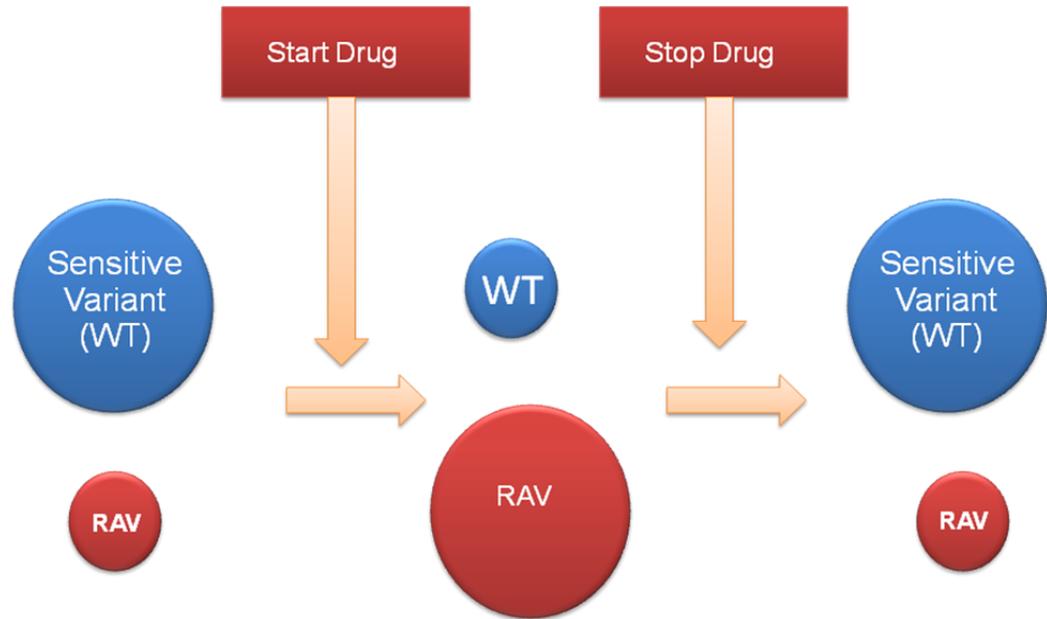
The rapid emergence of RAVs in patients who received TVR or BOC monotherapy (after <15 days) may be attributed to HCV high replication rate or the pre-existence of minority RAVs. The majority of PI RAVs are rarely detected by Sanger sequencing as they circulate at low frequency (0.1-3%) and typically their fitness is impaired (Schneider and Sarrazin, 2014). However, if their fitness reached that of WT virus, they would likely circulate as dominant variants and are detected more frequently (Halfon and Sarrazin, 2012).

The introduction of NGS enables the detection of minority variants and has other advantages; faster processing and larger scale sequencing (Hiraga et al., 2011, Nasu et al., 2011, Ninomiya et al., 2012). However, challenges of this technology include the need for sophisticated bioinformatic tools to enable reliable data analysis and exclude possible cross-contamination (NGS is highly prone to cross-contamination) (Gregori et al., 2013).

The overall aims of this series of experiments were:

1. To sequence the HCV genome from serum samples of acutely infected, treatment-naïve patients using Illumina deep sequencing, and to identify baseline polymorphisms associated with known resistance to DAAs.
2. To construct a genotype 1a subgenomic replicon to quantify viral fitness *in vitro* of selected resistance mutations introduced into the NS5B gene. These mutations included four mutations predicted *in silico* and two control mutations reported in the literature.

Figure 5-4: Model of HCV resistance in case of pre-existing minority RAV.



The figure illustrates the minority RAV that has poorly fitness before treatment. During treatment with DAAs, these pre-existing RAVs have a fitness advantage and can outgrow the dominant variant in the viral quasispecies, while the sensitive viral variants fail to replicate. After the end of therapy, the viral fitness returns to pre-treatment condition and the viral population revert to WT variant. Modified from (Pawlotsky, 2006).

5.2 Results

Sixteen samples from treatment-naïve acute HCV/HIV co-infected individuals were analysed in the following section; all patients were genotype 1a. Samples were processed using the metagenomic approach for HCV full genome sequencing as described in Section 2.10.2. Sequences were screened for substitutions at amino acid residues reported to confer resistance against currently recommended DAAs; PIs (SMV and PDV), NS5A inhibitors (DCV, LDV, and OBV) and NS5B inhibitors (SOF and DSV). A comparison group of sequences from 18 HCV mono-infected, genotype 1a treatment-naïve patients was used from the HCV Research UK cohort.

5.2.1 Natural polymorphisms in the NS3 protease region

A natural polymorphism at Q80 was the most prevalent RAM in the HCV/HIV cohort (56.25%). Other polymorphisms were present in approximately 25% of patients at residues V36, D168, and I/V170. As many as 12.5% of patients were identified to have a polymorphism at residues F43, T54, V55, A156, or F169 (Figure 5-5).

In the 18 HCV mono-infected patients, polymorphisms at baseline were found in fewer residues other than Q80; Q80 (83.33%), R155 (11%), D168 (11%), and F169 (5.6%). Unlike HIV/HCV co-infected patients, polymorphisms at V36, Q41, F43, T54, V55, A156, V163, and I/V170 were not detected in patients with mono-infection.

5.2.1.1 Resistance associated variants in the NS3 protease region.

In the HCV/HIV cohort, the prevalence of RAVs was calculated within each mutation site. All detected mutations were associated with PI resistance except two residues (V55 and V/I 170) (Figure 5-6). In the HCV mono-infection cohort, all detected mutations were known RAMs.

5.2.1.2 Distribution of resistance-associated variants in NS3 protein at intra-host Level

In the HCV/HIV cohort, RAVs that were identified in the NS3 region represented a minority variant except Q80 variants. Q80K was distributed as an intermediate (20-50%) and major variant (>50%), in three patients and one patient respectively (Figure 5-7). A list of individual mutations detected is shown in (Table 5-1).

For HCV mono-infected patients, RAVs were detected as minor variants at Q80, R155, D168, and F169 at low prevalence, which ranged from 6%-20% out of 18 samples and only 1 patient had Q80K as a dominant resistance variant. Only 3/18 subjects did not have natural occurring RAVs (Figure 5-8).

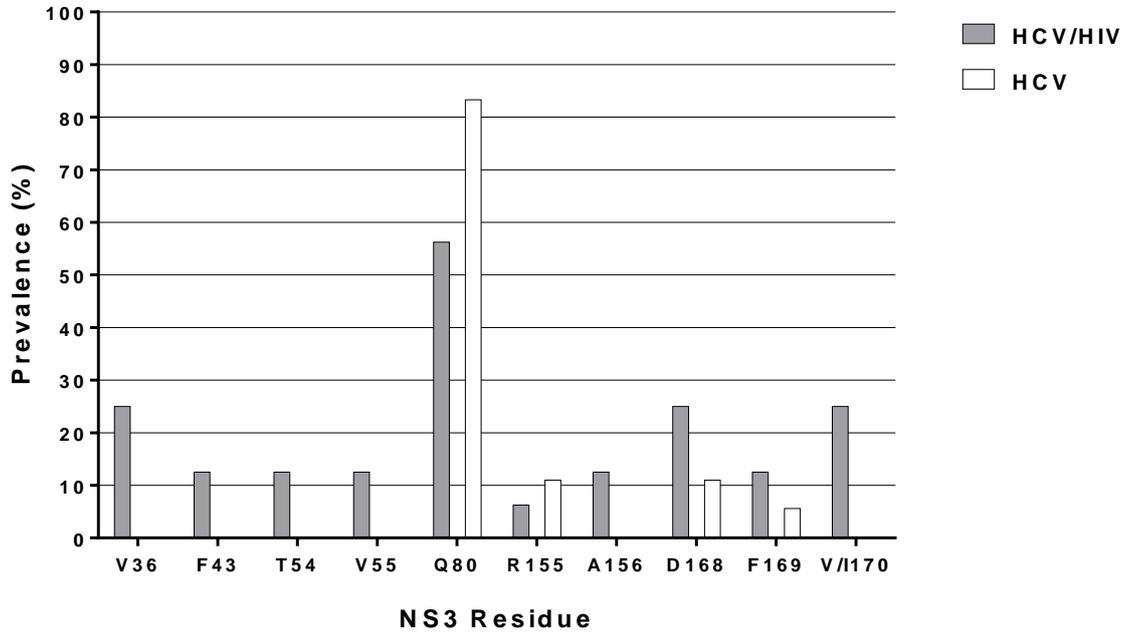
5.2.2 Natural polymorphism in the NS5A gene

Baseline polymorphisms within NS5A were found at residues M28 and M62 in half of the study cohort. In other reported residues, the percentage of mutated residues was Q30 (43.75%), H58 (37.5%), Y93 (37.5%), K24 (31.25%), A92 (18.75%), L31 (12.5%), S38 (12.5%) and mutations in P32 were the lowest detected at only 6.25% of total population (Figure 5-9). Amongst mono-infected samples, two residues, H58 and E62 were detected in 55% and 66% of total samples respectively. K24 and Y93 were detected in as many as 50%, whereas M28 and Q30 were slightly lower at 28%. P32 and A92 had a low prevalence level at 5.56%. Variants at residues L31 and S38 were not detected in any sample.

5.2.2.1 Resistance associated variants in NS5A

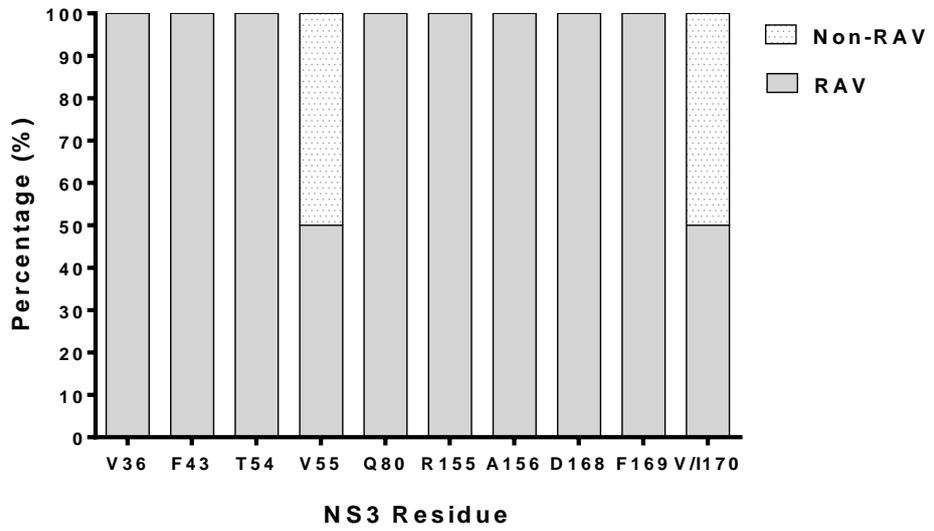
In the HCV/HIV cohort, the mutations detected in NS5A residues were not exclusively RAVs. All mutations detected at S38 and H58 were not reported in the literature to confer resistance. Only 37.5%, 40%, and 62% of detected mutation at residues E62, K24, and M28 respectively were known RAVs. However, all mutations detected in residues Q30, L31, P32, A92, Y93 were known RAVs (Figure 5-10).

Figure 5-5: Prevalence of natural polymorphisms in NS3 of HIV/HCV co-infected patients and HCV mono-infected patients.



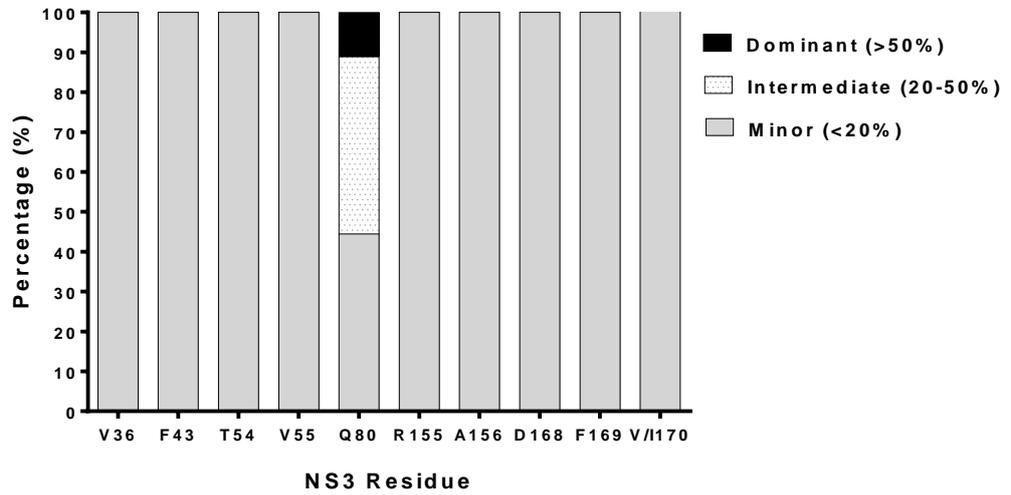
Samples were sequenced using the Illumina MiSeq. The prevalence of RAVs was measured as a percentage (%) of the total number of variants present at baseline in each cohort.

Figure 5-6: Frequency of RAVs within NS3 in HIV/HCV co-infected patients.



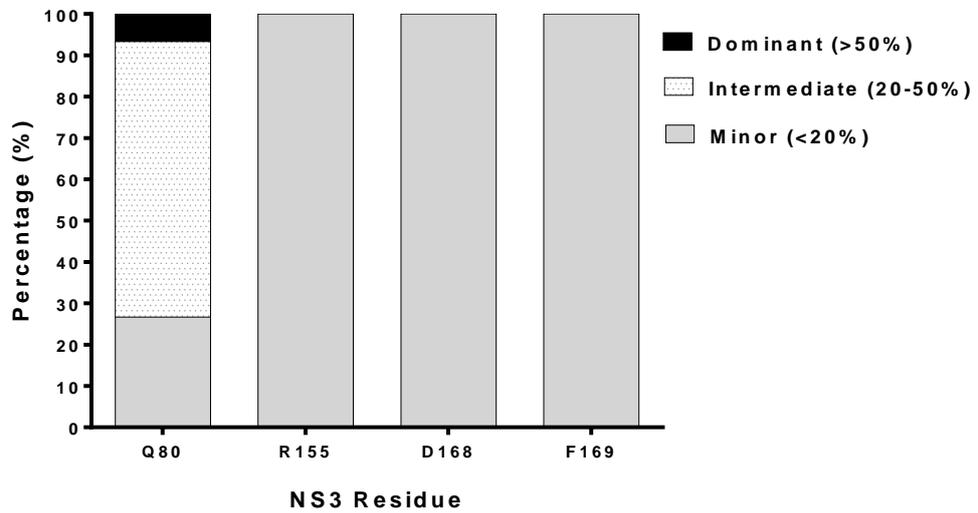
Baseline polymorphisms were stratified into resistance associated (RAV) and non-resistance associated variants (Non-RAV). Prevalence is shown as the percentage (%) of total HIV/HCV co-infected samples (N=16).

Figure 5-7: Distribution of baseline RAVs within NS3 of HIV/HCV co-infected patients.



The frequency of RAVs within the HCV intra-host quasispecies was categorized as minor (<20%), intermediate (20-50%) and dominant (>50%) variants. Prevalence is the percentage (%) of the total number of RAVs in all HIV/HCV co-infected patients.

Figure 5-8: Distribution of baseline RAVs in NS3 region of HCV mono-infected patients.



The frequency of RAVs within HCV intra-host quasispecies was categorized as minor (<20%), intermediate (20-50%) and dominant (>50%) variants. Prevalence was shown as the percentage (%) of total number RAVs in all HCV mono-infected patients.

Table 5-1: Prevalence of resistance-associated variants within the NS3 gene in HIV/HCV co-infected patients and distribution at the intra-host level.

NS3 Protein residues	Mutation	Prevalence in total population (%) N=16	Prevalence within individual (% range)
V36	A	12.5%	1.00%
	G	12.5%	3% (1-5%)
	L	18.75%	1.3% (1-2%)
T54	A	6.25%	2.00%
	S	12.5%	1.5% (1-2%)
V55	A	6.25%	4.00%
Q80	K	50%	29% (1-85%)
	L	6.25%	1%
	H	6.25%	1%
	R	6.25%	4%
R155	K	6.25%	1%
	T	6.25%	1%
	S	6.25%	3%
A156	S	12.5%	1%
D168	Y	12.5%	1%
	A	6.25%	1%
	V	6.25%	1%
	G	12.5%	1%
	E	6.25%	4%
	N	6.25%	2%
F169	L	12.5%	1.5% (1-2%)
V/I 170	L	12.5%	3%
	T	6.25%	1%

Meanwhile, in samples with HCV mono-infection, all variants at position Q30 and Y93 are associated with resistance. Furthermore, as many as 80% of patients with mutations at M28 were reported RAVs, while only 11% of detected mutations at K24 were RAVs. No HCV mono-infected patient was harbouring RAVs at P32, H58, E62 and A92 (Figure 5-11).

5.2.2.2 Distribution of RAVs in NS5A at the intra-host level

In the HCV/HIV cohort, only one patient harboured the M28V RAV as a dominant variant (90%), while other minor RAVs were distributed at residues K24, L31, P32, Q54, H58, E62, A92, and Y93 by prevalence from 6.25% up to 50% out of 16 patients. Intermediate RAVs were found in only one patient at Q30 (Figure 5-12). A detailed distribution of individual mutations detected is listed in Table 5-2.

In the NS5A region of HCV mono-infected patients, Y93H was a dominant variant in only one sample. In addition, minor variants were identified at residues K24, M28, Q30, and Y93 (Figure 5-13).

5.2.3 Natural polymorphisms in the NS5B gene

In the HCV/HIV cohort, polymorphisms were found mostly in the NS5B region compared to the other non-structural genes. The most prevalent mutation sites were L159, C316, A553, and S556 with a mutation detected in 37.5% of total samples and 31.25% at residue E446. The prevalence of baseline polymorphisms at other sites was as follows, M414 (18.75%), S282 (12.5%), L320 (12.5%), V321 (12.5%), Y448 (6.25%), and D559 (6.25%) (Figure 5-14).

Polymorphisms were fewer in HCV mono-infected patients. The most prevalent amino acid substitution was at C316 and detected in up to 61% of samples, (C316N was not reported as an RAV for genotype 1a). Natural polymorphisms were detected at residues E446 (55.5%), L159 (50%), A553 (44.4%), and S556 (44.4%).

Substitutions at S282 and Y448 (11.1%), L320 and V321 (5.5%), and M414 (11.1%) had lower prevalences. No mutations were identified at the D559 site.

5.2.3.1 Resistance associated variants within NS5B

In the HCV/HIV cohort, all substitutions detected at residues L159, S282, M414, A553, S556, and D559 have previously been reported as RAVs. The frequency of resistance polymorphisms at position E446 was 80% while 50% of the mutations at L320 and V321 were reported RAVs. Amino acid substitutions at C316 have also been identified as RAVs and were present at a frequency of 16.6%. Finally, none of the mutations detected at Y448 were identified to be RAVs (Figure 5-15).

The NS5B S282T mutation was detected in one patient with a frequency of 3% while two other substitutions were detected in the same patient, S282R and S282G at frequencies of 12% and 5% respectively.

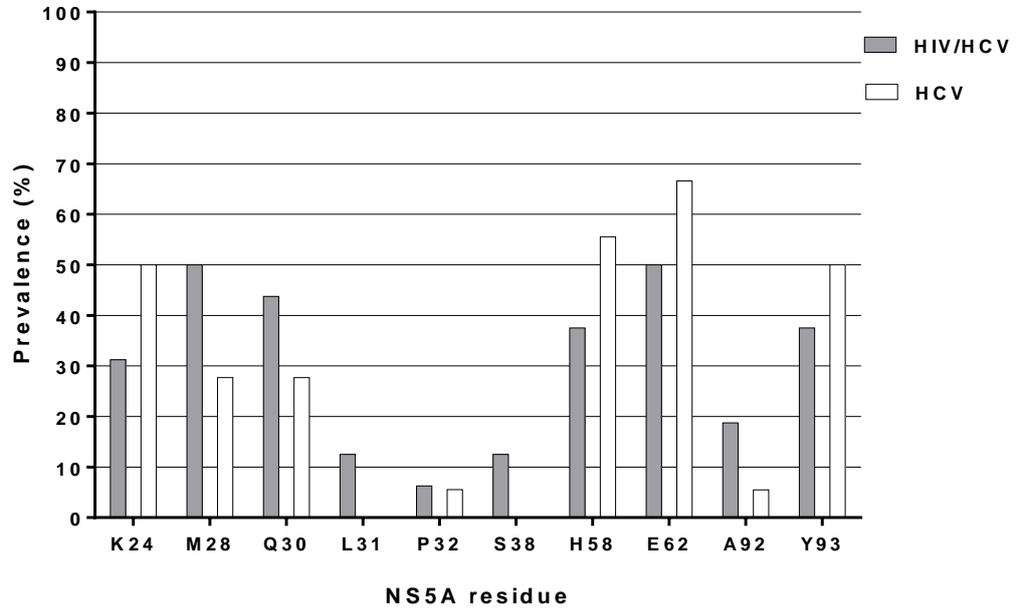
In the HCV mono-infected cohort, all substitutions detected at residues L320, M414, E446, A553, and S556 were known to confer resistance. Half of the mutations at S282, compared to 33.3% of mutations detected at L159 were found to be RAVs. Meanwhile, none of the mutations detected at residues C316, V321, Y448 was identified as an RAV (Figure 5-16).

5.2.3.2 Distribution of RAVs in NS5B at the intra-host level

In the HCV/HIV cohort, all RAVs were detected as minority variants. In the NS5B region, apart from the mutations reported *in vivo* that are shown above, many mutations are reported to confer resistance to SOF and DSV *in vitro* as listed in the appendix (7.29 and 7.31). Below is a detailed list of all mutations detected in this cohort including RAVs reported both *in vivo* and *in-vitro*

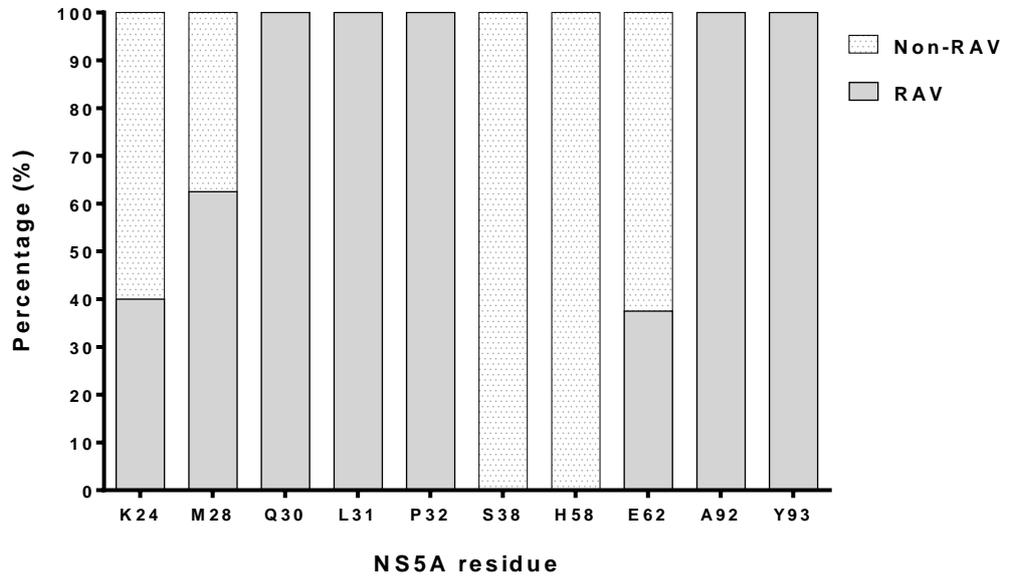
Dominant resistance-associated variants were detected in the HCV mono-infected cohort at residues L159 and E446Q. Intermediate variants were detected in 37.5% of samples at two residues, A553 and S556. Other detected RAVs at other residues were minority variants (Figure 5-17).

Figure 5-9: Prevalence of natural polymorphisms within NS5A in HIV/HCV coinfecting and HCV mono-infected patients.



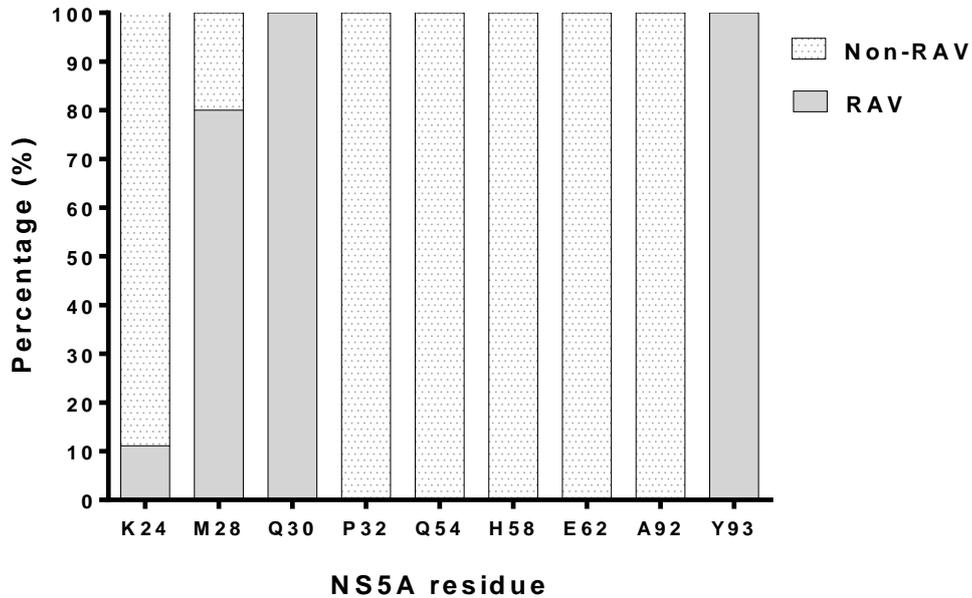
Samples were sequenced using the Illumina MiSeq. The prevalence of RAVs was measured as a percentage (%) of the total number of variants present at baseline in each cohort.

Figure 5-10: Frequencies of resistant and non-resistant variants within NS5A in HIV/HCV Co-infected Patients.



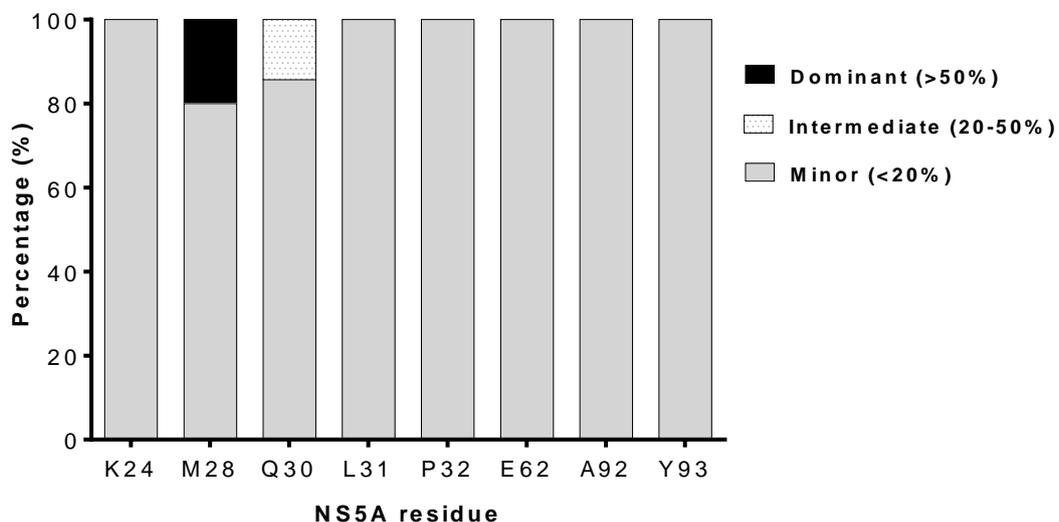
Baseline polymorphisms were stratified into resistance associated (RAV) and non-resistance associated variants (Non-RAV). Prevalence is shown as the percentage (%) of total HIV/HCV co-infected samples (N=16).

Figure 5-11: Frequencies of resistant and non-resistant variants within NS5A region in HCV mono-infected patients.



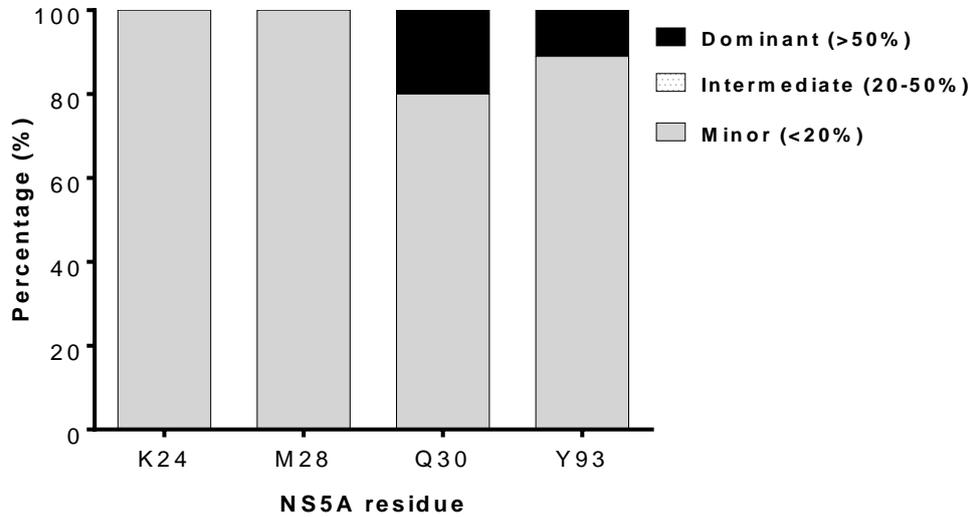
Baseline polymorphisms were stratified into resistance associated (RAV) and non-resistance associated variants (Non-RAV). Prevalence is shown as the percentage (%) of total HCV mono-infected samples (N=18).

Figure 5-12: Distribution of baseline RAVs in NS5A of HIV/HCV co-infected patients.



The frequency of RAVs within the HCV intra-host quasispecies was categorized as minor (<20%), intermediate (20-50%) and dominant (>50%) variants. Prevalence is the percentage (%) of the total number of RAVs in all HIV/HCV co-infected patients.

Figure 5-13: Distribution of baseline RAVs in NS5A of HCV mono-infected patients.

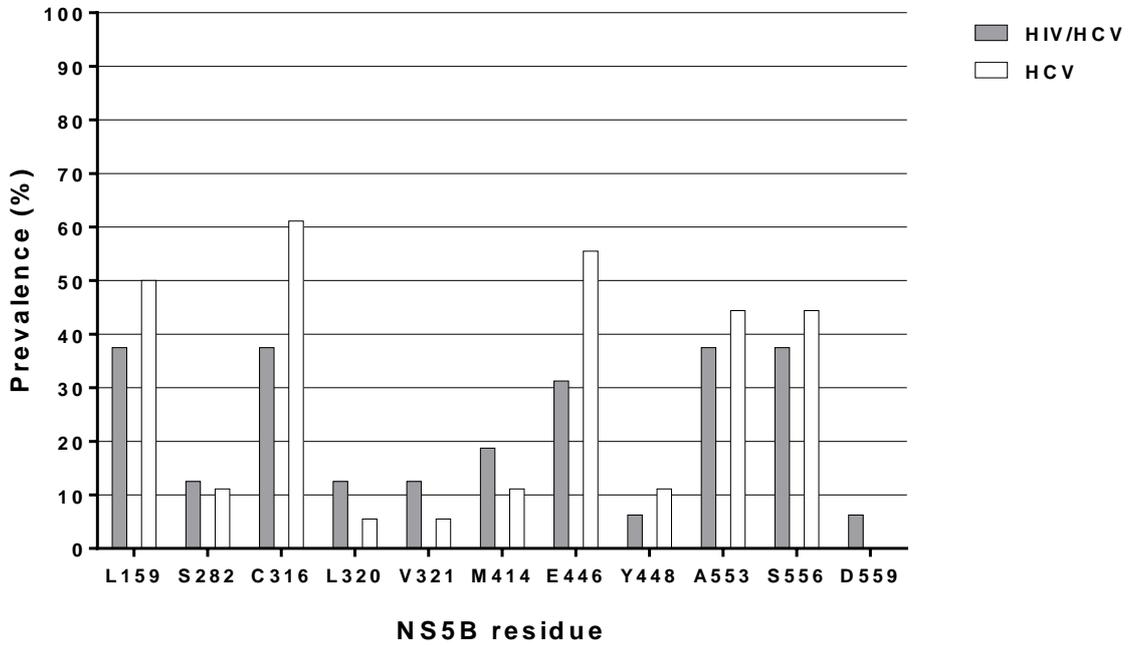


The frequency of RAVs within HCV intra-host quasispecies was categorized as minor (<20%), intermediate (20-50%) and dominant (>50%) variants. Prevalence was shown as the percentage (%) of total number RAVs in all HCV mono-infected patients.

Table 5-2: Prevalence of RAVs within NS5A in HIV/HCV co-infected patients and distribution at the intra-host level.

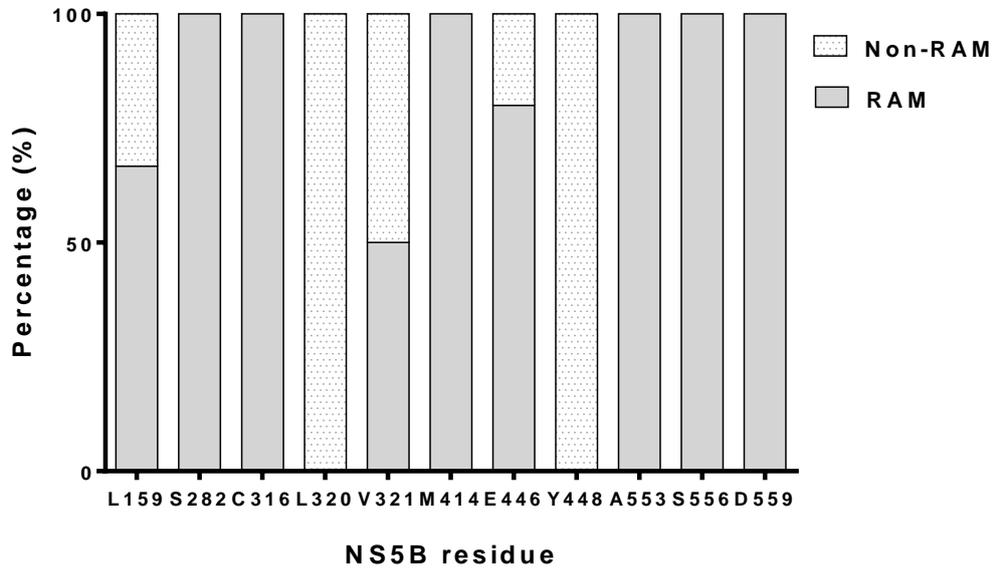
NS5A Protein residues	Mutation	Prevalence in total population (%) N=16	Prevalence within individual (% range)
K24	N	12.50%	3% (2-4%)
M28	I	18.75%	3% (1-5%)
	T	6.25%	1%
	V	6.25%	90%
Q30	H	25.00%	5.25% (1-13%)
	R	18.75%	1%
	K	6.25%	4%
	P	6.25%	4%
L31	M	12.5%	2.5% (2-3%)
	R	12.5%	2.5%(1-4%)
P32	L	6.25%	1%
E62	D	18.75%	2% (1-3%)
A92	T	18.75%	1%
Y93	C	6.25%	1%
	H	25.00%	2% (1-4%)
	S	12.50%	2.5% (2-3%)

Figure 5-14: Prevalence of Natural Polymorphisms in NS5B Region of HIV/HCV Co-infected Patients (HIV/HCV) and HCV mono-infected patients (HCV).



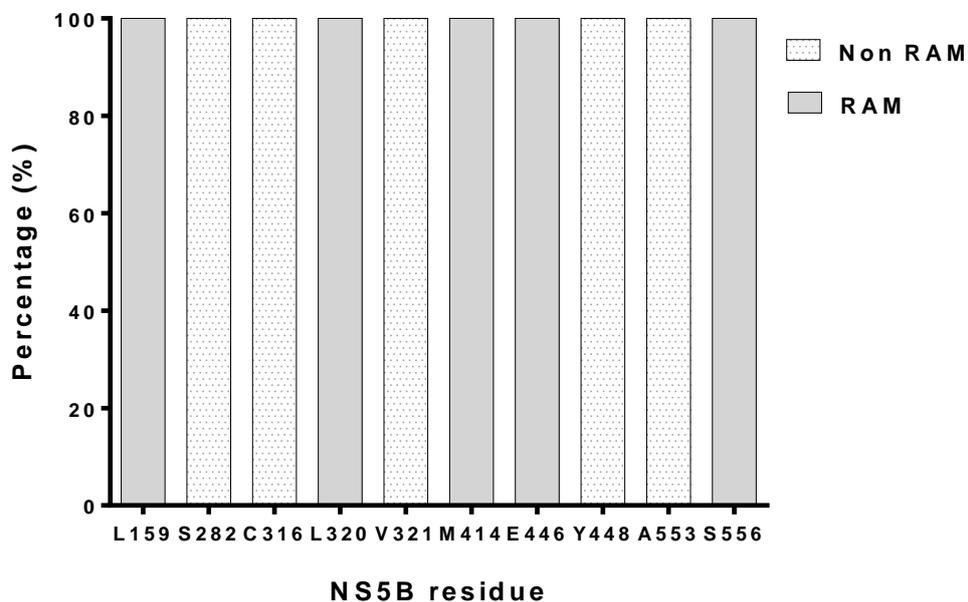
Samples were sequenced using Illumina MiSeq. Prevalence was measured as a percentage (%) of the total number variants at baseline at a particular site in all samples.

Figure 5-15: Frequencies of resistant and non-resistant variants within NS5B in HIV/HCV Co-infected Patients.



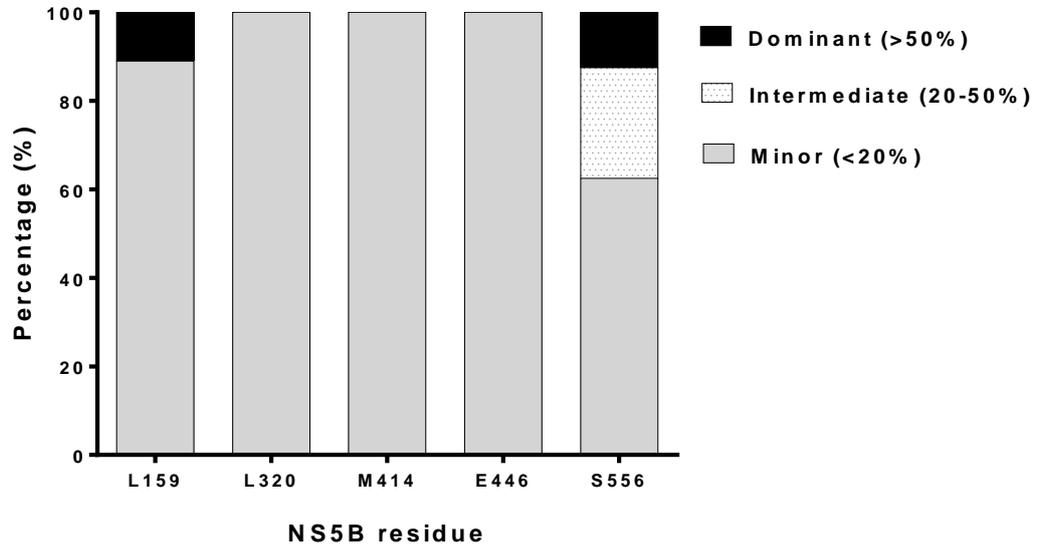
Baseline polymorphisms were stratified into resistance associated (RAM) and non-resistance associated variants (Non-RAM). Prevalence is shown as the percentage (%) of total HIV/HCV co-infected samples (N=16).

Figure 5-16: Frequencies of Resistant and Non-resistant Variants Within NS5A Region in HCV mono-infected Patients



Baseline polymorphisms were stratified into resistance associated (RAV) and non-resistance associated variants (Non-RAV). Prevalence is shown as the percentage (%) of total HCV mono-infected samples.

Figure 5-17: Distribution of baseline RAVs in NS3 region of HCV mono-infected patients.



The frequency of RAVs within HCV intra-host quasispecies was categorized as minor (<20%), intermediate (20-50%) and dominant (>50%) variants. Prevalence was shown as the percentage (%) of total number RAVs in all HCV mono-infected patients.

Table 5-3: Prevalence of resistance-associated variants within NS5B protein in HIV/HCV co-infected patients and distribution at intra-host level.

NS5B Protein residues	Mutation	Prevalence in total population (%) N=16	Prevalence within individual (% range)
C110	S	37.50%	21.7% (2-35%)
N117	S	12.50%	2.00%
L159	F	25.00%	1.25% (1-2%)
S282	G	6.25%	5.00%
	R	6.25%	1.00%
	T	6.25%	3.00%
C316	W	12.50%	4% (3-5%)
	N	25.00%	7.5% (4-18%)
	K	6.25%	1.00%
L320	F	6.25%	4.00%
V321	A	6.25%	2.00%
	I	6.25%	2.00%
A376	C	12.50%	1.5% (1-2%)
T390	I	6.25%	1.00%
A395	G	6.25%	1.00%
V405	I	43.75%	24.7% (1-99%)
N411	S	6.25%	1.00%
M414	I	18.75%	3% (1-7%)
	L	12.50%	3.5% (2-5%)
	V	6.25%	1.00%
F415	Y	31.25%	29.2% (1-39%)
S431	G		
I434	M	62.50%	27.1% (1-99%)
C445	F	6.25%	4.00%
E446	Q	25.00%	1.00%
C451	G	12.50%	3% (2-4%)
	S	12.50%	1.5% (1-2%)
A553	T	6.25%	2%
S556	G	37.50%	4.5% (1-12%)
	R	6.25%	6.00%
D559	G	6.25%	1.00%
	N	6.25%	1.00%
I585	V	6.25%	89.00%

5.2.3.3 Prevalence of predicted mutations in the HCV/HIV cohort

A prediction model of mutations that confer resistance to SOF was designed as described in Section 2.18.8. It predicts key residues that could interact with the SOF binding site at R32, G493, P495, and P496 (Figure 5-18). The cohort used above was analysed for the presence of any of these residues. Minority variants only were detected in only two samples as listed in Table 5-4.

5.2.4 Design of a genotype 1a replicon-based luciferase assay

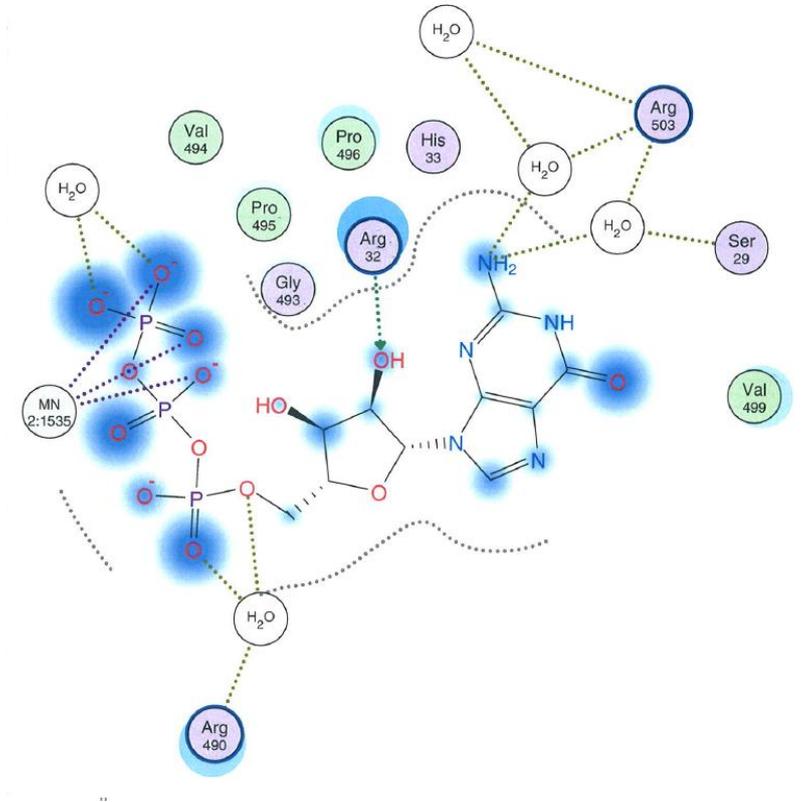
5.2.4.1 Strategy for the construction of a replication-competent genotype 1a subgenomic replicon.

The genotype 1a HCV replicon APP238 pH/SG-Neo (L+) was modified by introducing a luciferase firefly cassette and two adaptive mutations; NS4A-K1691R and NS4B-E1726G through a cloning strategy involving an intermediate cloning vector pGFP-C1 (Figure 5-19). This replicon was used as a backbone for mutations within the NS5B region, introduced to assess replicative fitness.

5.2.4.2 Introduction of the luciferase reporter gene

To construct subgenomic TA/SG/Luc/HCV-1a replicons containing the luciferase gene, a fragment composed of the firefly luciferase gene was amplified from pSGR-JFH1 using forward and reverse primers (Kato et al., 2003). The luciferase gene was amplified with primers flanked by restriction sites for *Apa*LI and *As*cl enzymes to produce a DNA fragment with *Apa*LI and *As*cl sites at the 5` and 3` termini, respectively with an expected product size of 1650 bp. The primers were designed using CLC genomics®: 5`-*Apa*LI-luc: Gtgcac c ATG GAA GAC GCC AAA AAC, 3- luc(stop)-*As*cl: “ GGC GGA AAG ATC GCC GTG TAA gGcgcgcc“.

Figure 5-18: Predicted NS5B resistance associated mutations based on HCV structure and interaction sites.



Residues interacting with the SOF active site on NS5B within 5 Å are shown. Halos around residues represent the degree of interaction, four residues (R32, G493, P495, and P496) are predicted as potential resistance mutations (Reproduced with permission from Dr Sreenu Vatipally).

Table 5-4: List of mutations detected in predicted residues.

	Sample 1	Sample 2
R32	P (2%), L (2%), G (2%), S (3%)	C (1%), L(2%),P(1%),S(3%)
G493	E (1%), L(1%)	E(1%),R(2%), V(6%),W(3%)
P495	Q(2%)	T(5%), Q(4%), L(2%)
P496	T(3%)	L(2%),H(4%),T(5%)

The four residues harboured minority variants in 12.5% of the study cohort.

The genotype 1a HCV replicon APP238 pH/SG-Neo (L+I) was digested using *AgeI* and *KpnI* as illustrated in (Figure 5-20), the schematic representation of this process is shown in (Figure 5-21). The resulting fragment was ligated into the pGFP-C1 plasmid as shown in (Figure 5-22), to allow the introduction of the luciferase cassette.

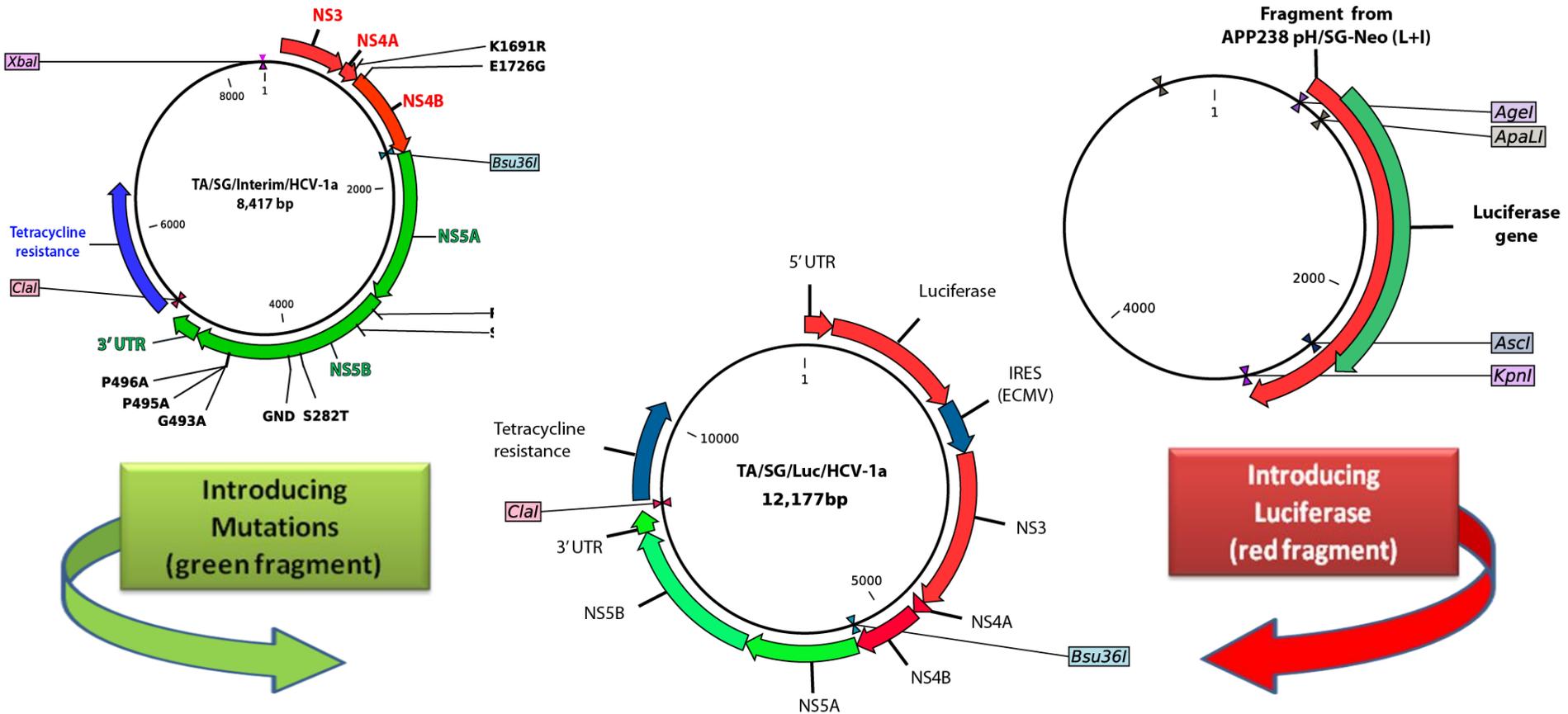
After introducing the luciferase cassette into the *AgeI-KpnI* fragment, another digest was carried out to re-introduce the fragment back into the subgenomic replicon backbone (shown in the red circle, 1 Figure 5-20). The ligation product (TA/SG/Luc/HCV-1a) was then transformed into *E. coli* followed by selection and culturing of tetracycline-resistant colonies. The resulting constructs were then sequenced to confirm the presence and correct orientation of the fragment carrying the luciferase cassette.

The two adaptive mutations, NS4A (K1691R) and NS4B (E1726G) were introduced into TA/SG/Interim/HCV-1a replicon using site-directed mutagenesis. Subcloning of the digested red fragment shown in Figure 5-23 using *NsiI* and *Bsu36I* in the TA/SG/Luc/HCV-1a backbone, the resulting replicon was sequenced using Sanger sequencing to confirm the orientation of the fragment and the presence of mutations as shown in Figure 5-24 and Figure 5-25.

Meanwhile, an intermediate plasmid TA/SG/Interim/HCV-1a was produced (Figure 5-23) by digesting TA/SG/Luc/HCV-1a using *AgeI* and *PmlI* to create a shorter plasmid of 8417 bp as shown in Figure 5-20 (green square). A shorter replicon was used to decrease the error rate in future subcloning.

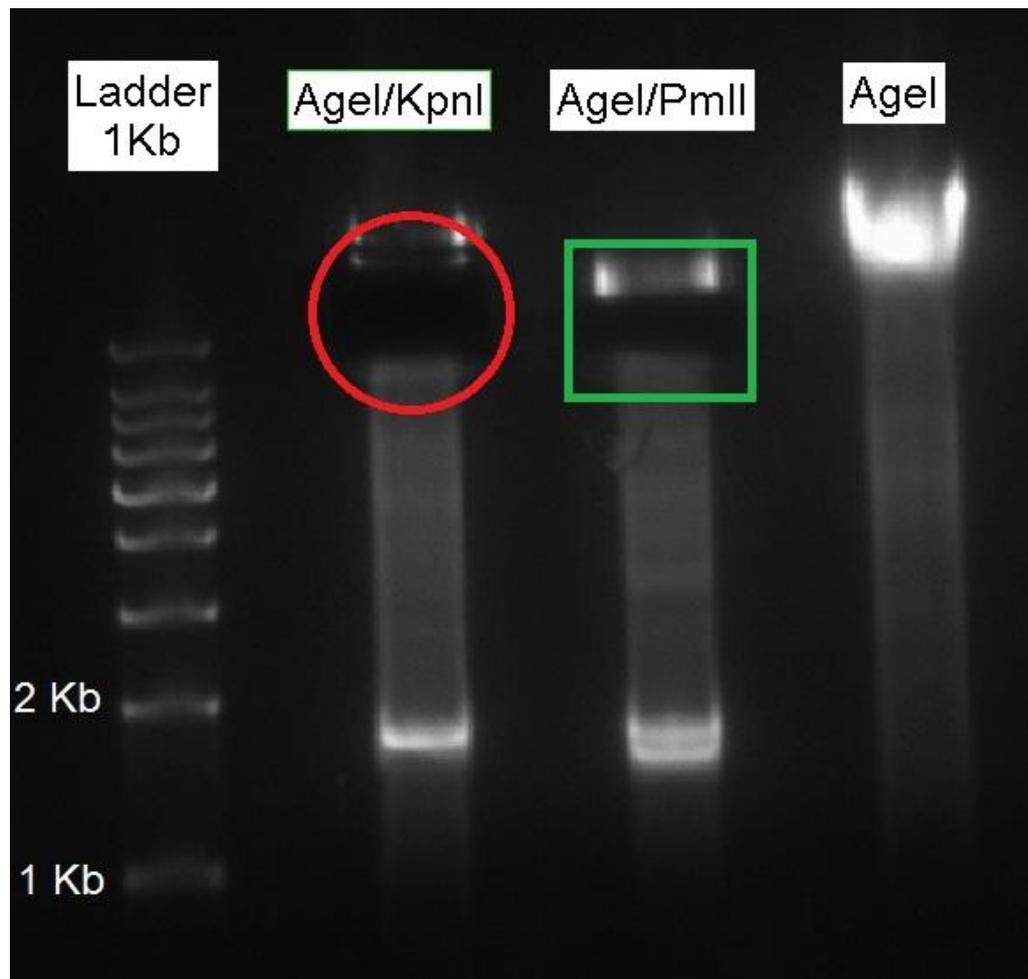
The predicted resistance mutation in our model was introduced to the intermediate replicon via site-directed mutagenesis. Both TA/SG/Interim/HCV-1a and wild-type TA/SG/Luc/HCV-1a were digested using the restriction enzymes *Clal* and *Bsu36I*, to produce a final mutated subgenomic replicon as shown in Figure 5-26.

Figure 5-19 Cloning strategy.



The cloning strategy involved 3 different replicons and 3 steps; 1) Replacing the neomycin gene with a luciferase cassette (pGFP-C1), 2) Creating an interim replicon with shorter length to introduce the mutations (TA/SG/Interim/HCV-1a), 3) Subcloning of the mutated fragment in the interim replicon into the final fit replicon (TA/SG/Luc/HCV-1a).

Figure 5-20: Restriction digest of replicon APP238 pH/SG-Neo (L+I).



The plasmid was digested using *AgeI* & *KpnI* in lane 1 with the excised larger fragment used as a vector for TA/SG/Luc/HCV-1a (red circle). In lane 2, APP238 pH/SG-Neo (L+I) was digested using *AgeI* & *PmlI* creating the TA/SG/Interim/HCV-1a of 8417bp (green square). Lane 3 shows linearised APP238 pH/SG-Neo (L+I) using the single cutter restriction enzyme *AgeI*.

Figure 5-21: Schematic illustration of a restriction digest of APP238 pH/SG-Neo (L+I) using *AgeI* and *KpnI*.

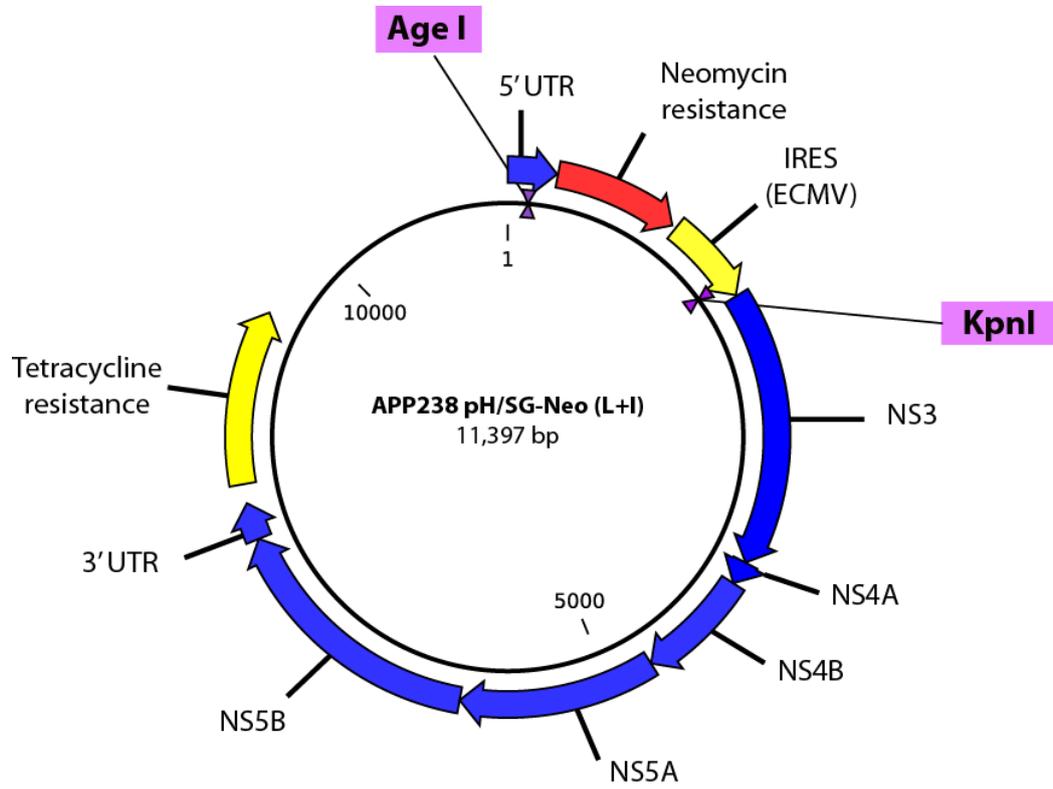


Figure 5-22: pGFP-C1 as intermediate vector to introduce the Luciferase firefly cassette

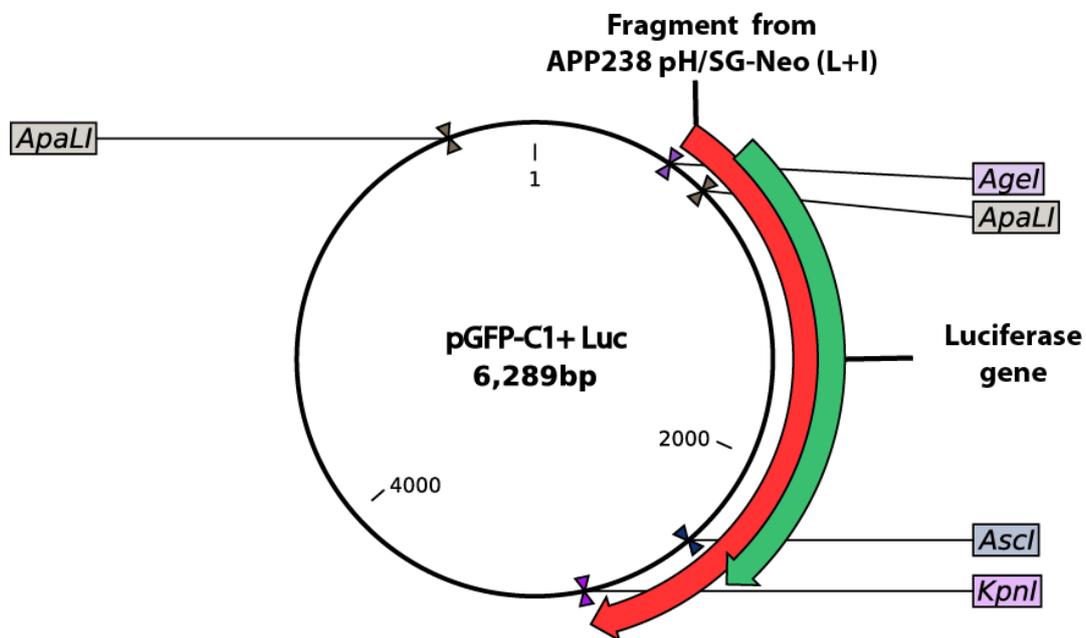


Figure 5-23: Schematic illustration of TA/SG/Interim/HCV-1a with different mutations to be introduced.

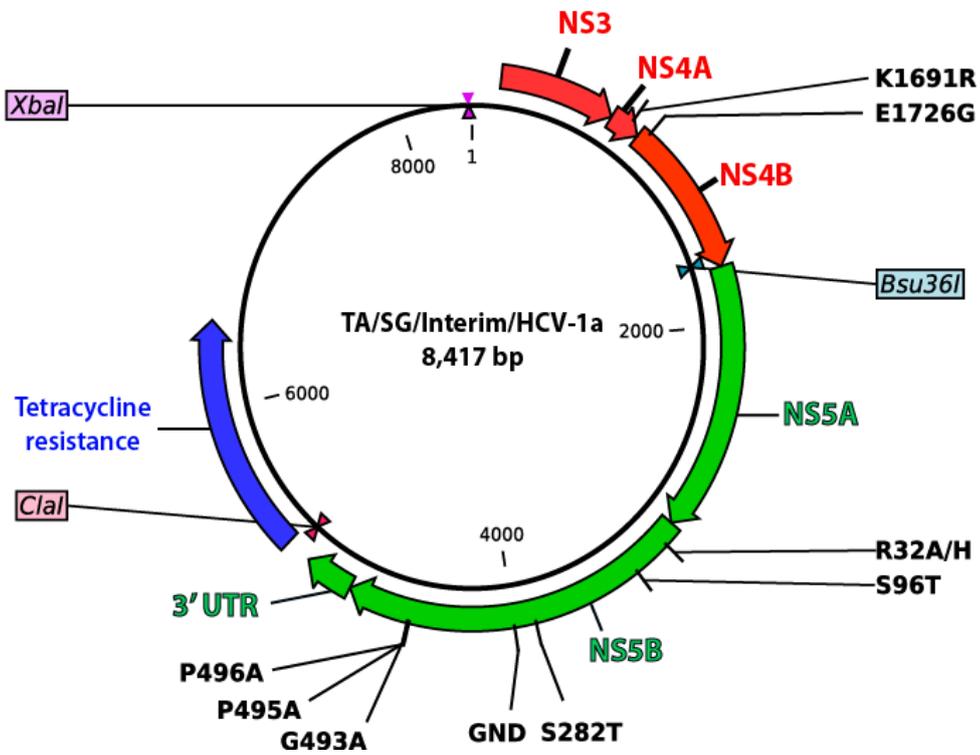


Figure 5-24: E1726R mutation (GAG)-(GGG)

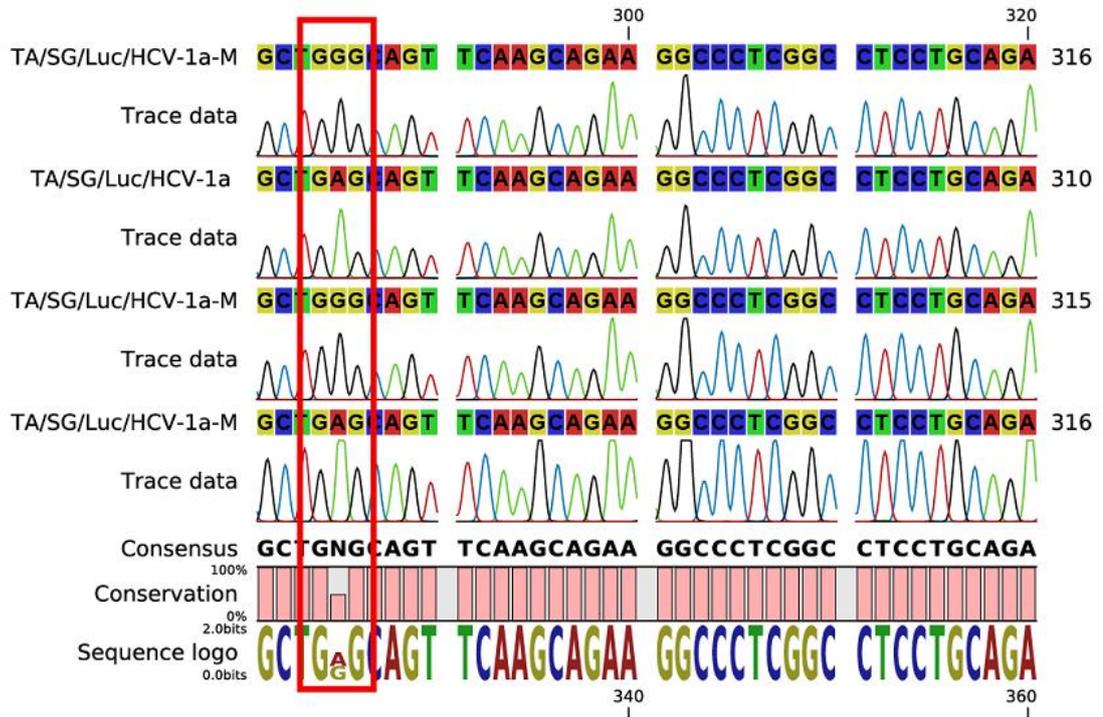


Figure 5-25: K1691R mutation (AAG) - (AGG)

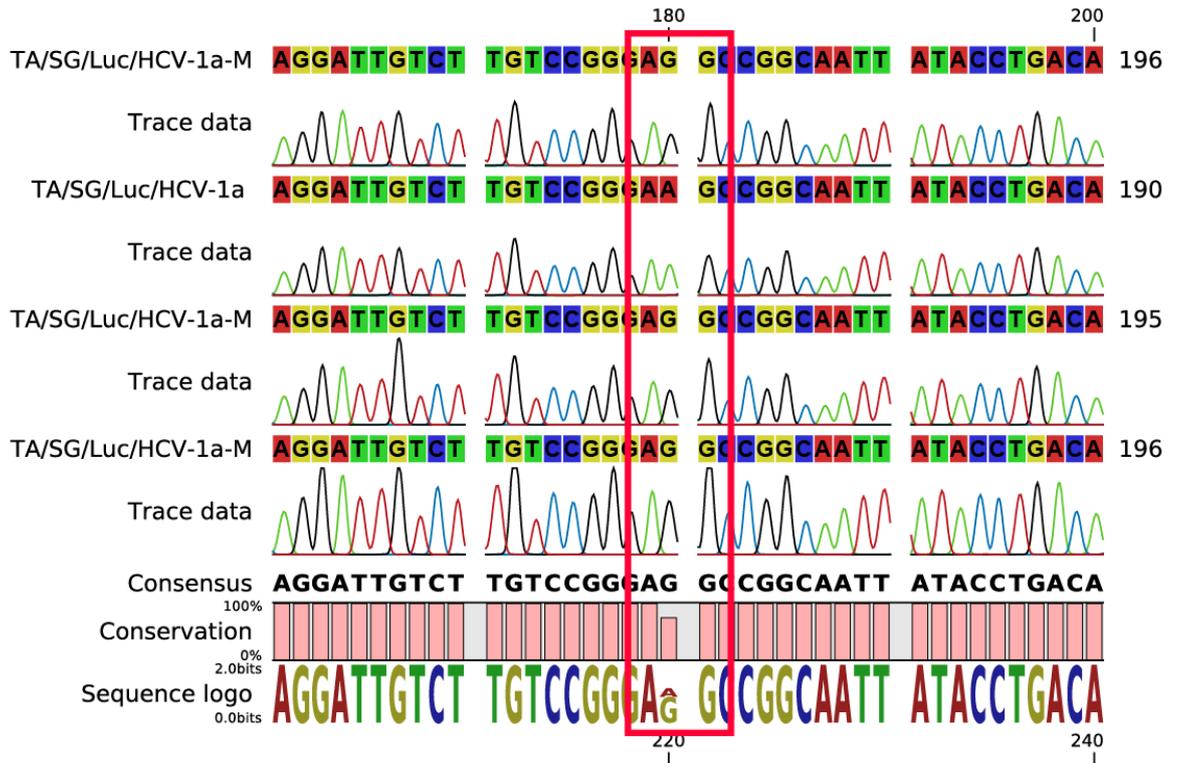
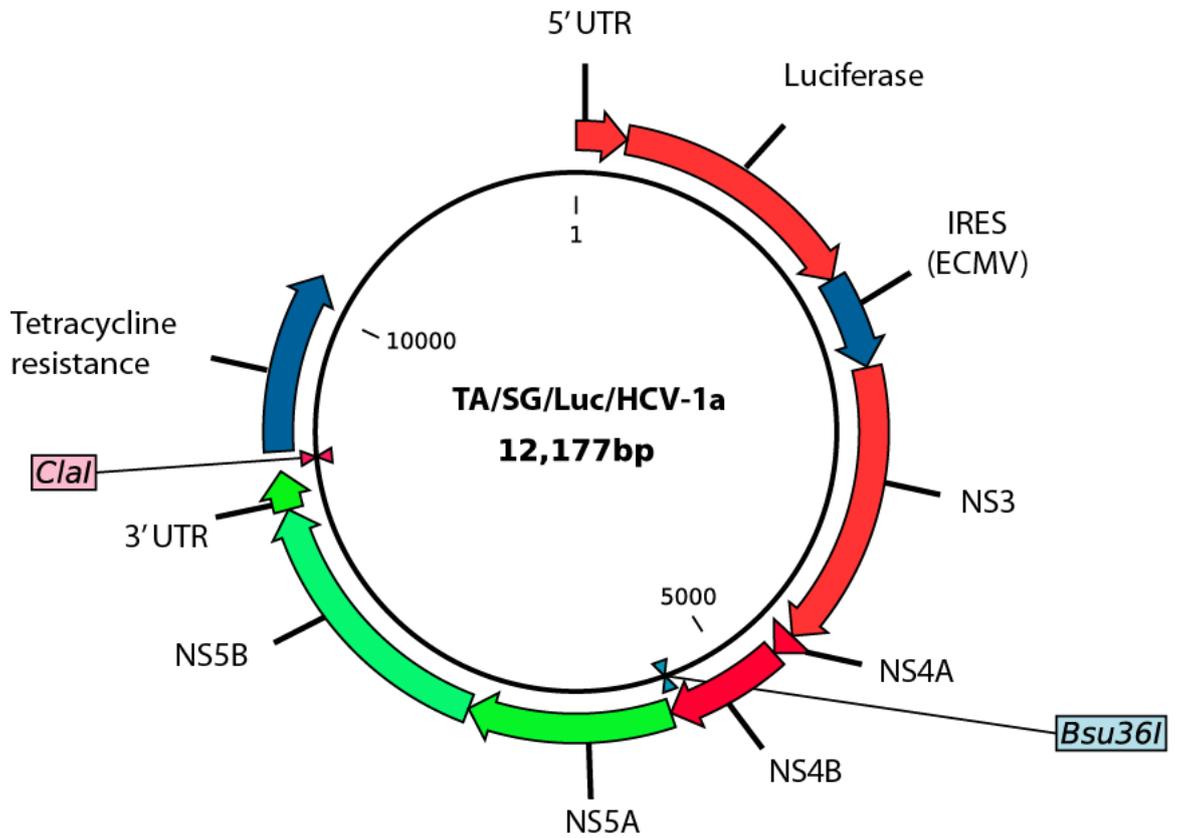


Figure 5-26: TA/SG/Luc/HCV-1a culture adapted replicon.



The green fragment of the TA/SG/Interim/HCV-1a contained the target mutation and was ligated with the TA/SG/Luc/HCV-1a vector fragment, forming the final mutated subgenomic replicon.

5.2.4.3 Transient replication assay

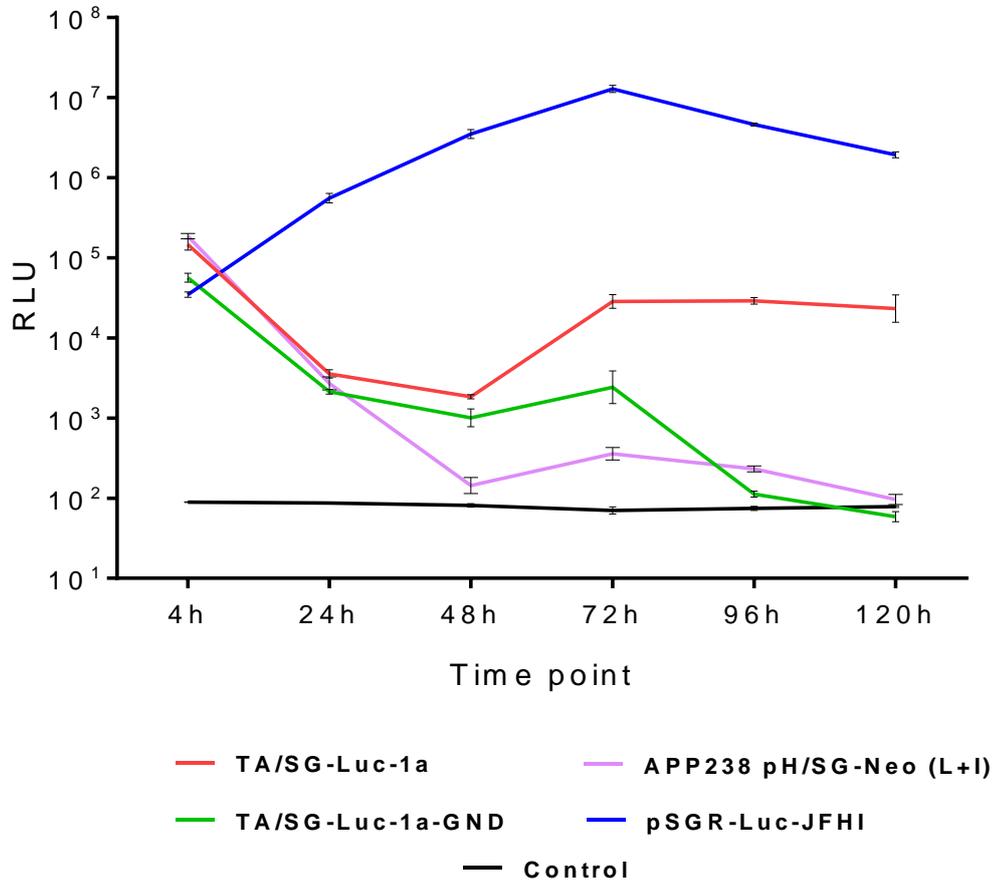
To investigate the replicative fitness of replicon variants containing RAVs, a transient replication assay was performed to measure replication capacity in Huh7.5 cells via luciferase reporter gene activity using pSGR-Luc-JFHI as a control. The pattern of luciferase activity displayed by the pSGR-Luc-JFHI was similar to previously published data (Targett-Adams and McLauchlan, 2005).

The number of Huh7.5 cells in each assay was adjusted to avoid confluence at 96h; 150,000 cells in each well of a 24 well plate was found to be optimum for the assay. For the TA/SG/Luc/HCV-1a replicon, luciferase activity gradually decreased after the initial time point (4h), but enzyme levels increased from 48 h to 72 h and then stabilized. The luciferase signal was 100 fold more than the activity displayed by the original APP238 pH/SG-Neo (L+I) replicon and the TA/SG/Luc/HCV-1a replicons encoding the GND mutation in the NS5B sequence, which abolishes the activity of the HCV RNA polymerase (Figure 5-27).

After optimisation of the transient replication assay using the TA/SG/Luc/HCV-1a replicon, Huh-7.5 cells were electroporated with 1, 2.5, 5 or 10 µg TA/SG/Luc/HCV-1a replicon RNA and the luciferase activity contained within cell extracts was monitored over 72 h to determine whether the amount of RNA used to electroporated cells influenced replication levels. By 4 h, luciferase activities were enhanced as the amount of RNA introduced into cells was increased. Enzyme levels reached at the end of the experiment were similar, irrespective of the amount of input RNA (data not shown). The overall luciferase reporter gene activity of the APP238 pH/SG-Neo (L+I) construct did not increase with time and remained at similar levels to the negative control TA/SG-Luc-1a-GND.

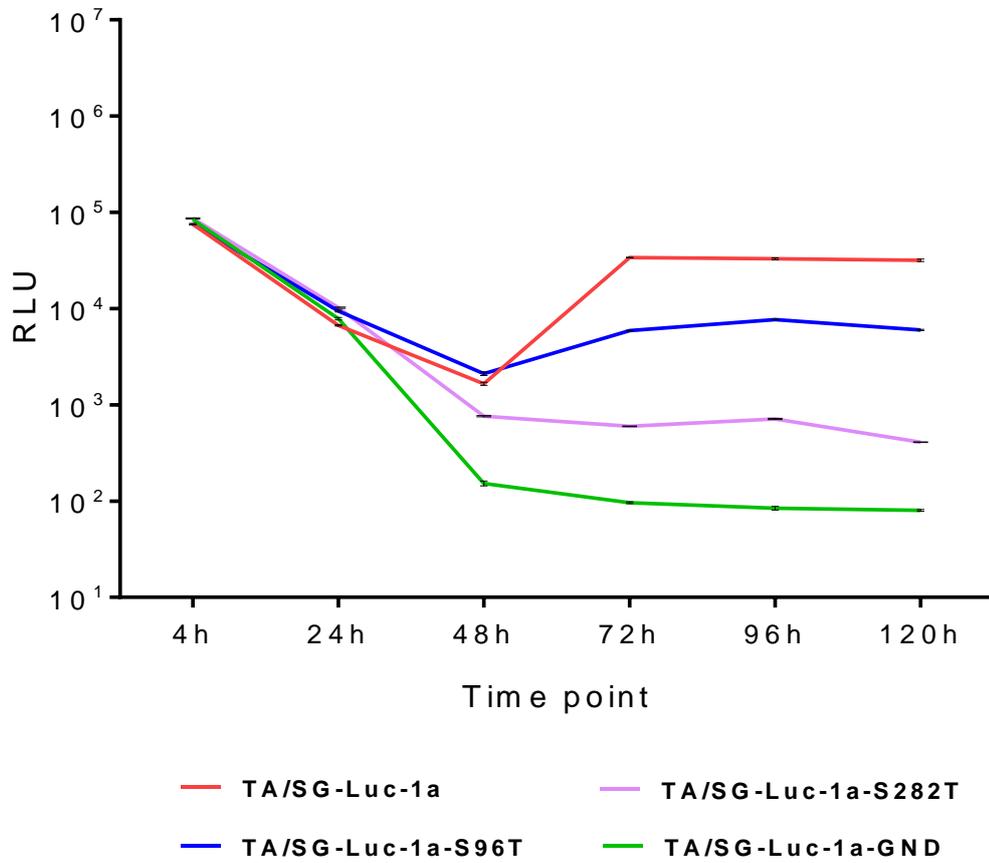
Different mutations introduced to the NS5B protein affected replication capacity when compared to the TA/SG/Luc/HCV-1a (WT) replicon (Figure 5-28, Figure 5-29, and Figure 5-30). Of these, only two maintained reduced replication fitness; S96T and P495A ($24.43 \pm 1.7\%$) and ($20.95 \pm 0.75\%$), respectively, while all other mutations showed less than 2% replicative fitness (Figure 5-31).

Figure 5-27: Optimisation of a transient replication assay for the subgenomic replicon TA/SG/Luc/HCV-1a.



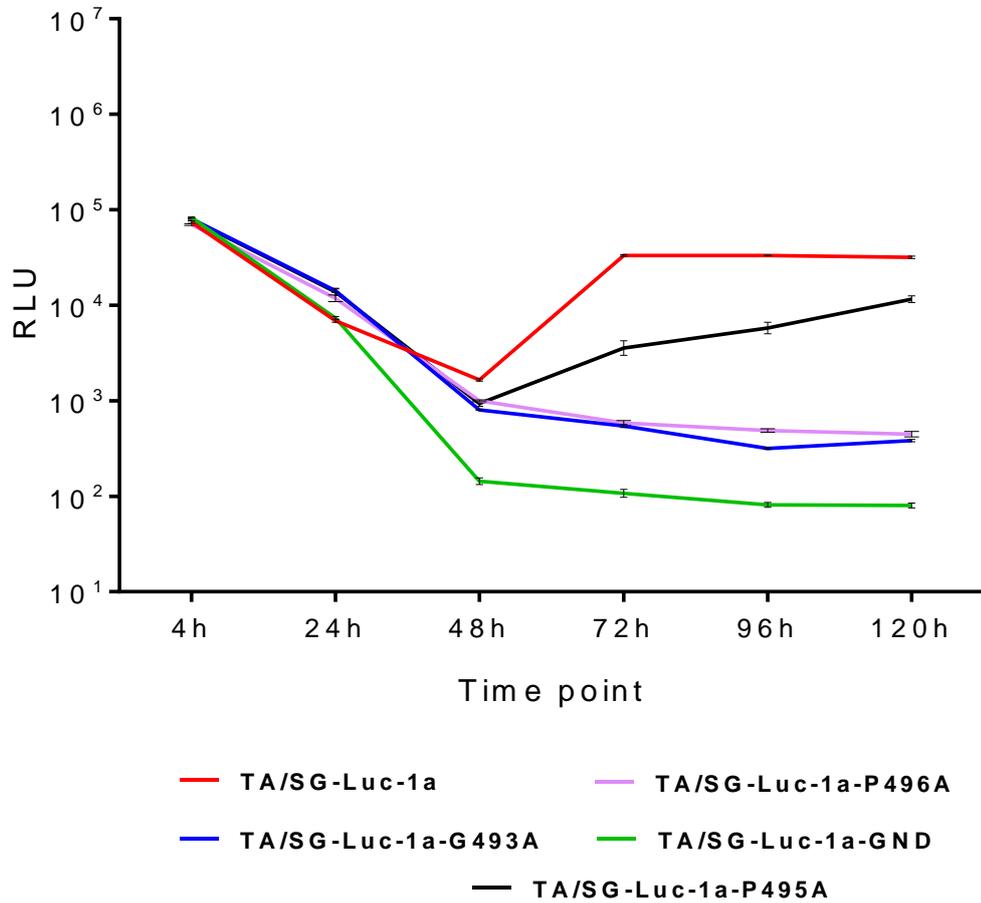
RNA was electroporated into Huh-7.5 cells and luciferase activities contained within cell extracts prepared at 4, 24, 48, 72 h, 96h, and 120h post-electroporation were determined. RLU; Relative light units. This figure represents one experiment carried out in triplicate.

Figure 5-28: Replication capacity of mutated replicon (S96T, S282T).



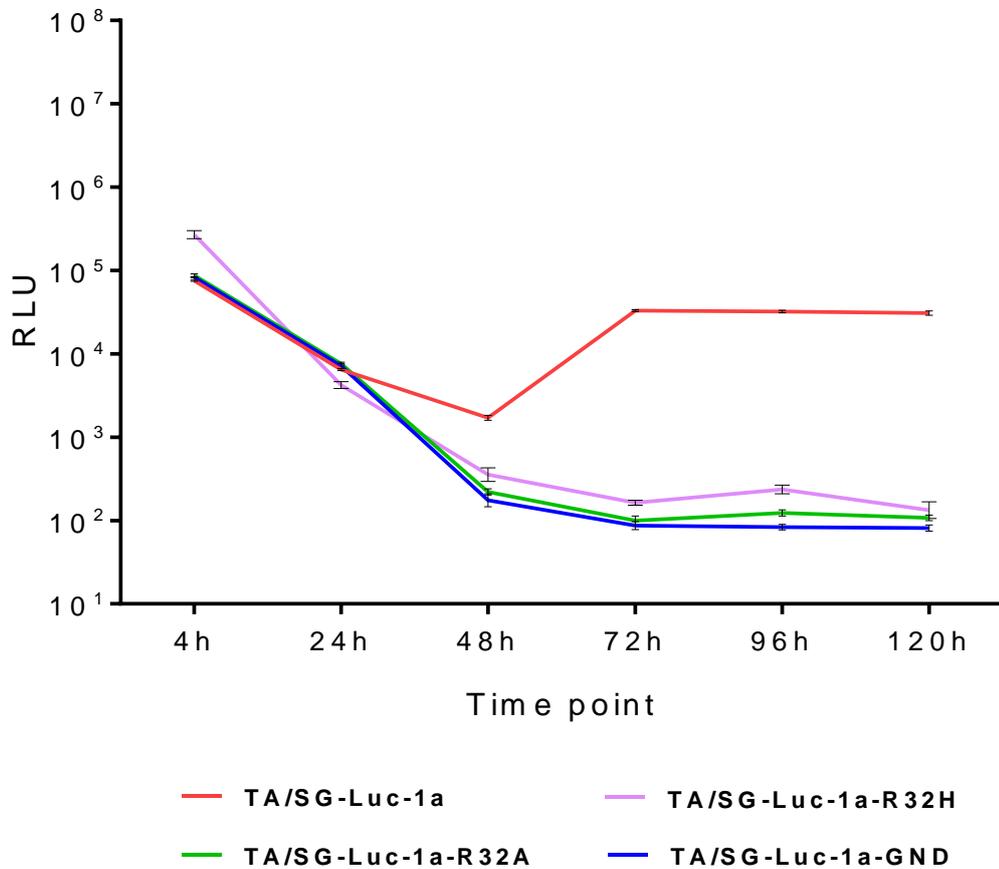
RNA was electroporated into Huh-7.5 cells and luciferase activities contained within cell extracts prepared at 4, 24, 48, 72 h, 96h, and 120h post-electroporation were determined. RLU, Relative light units. This represents one of four independent experiments with triplicates.

Figure 5-29: Replication capacity of the mutated replicon (G493A, P495A, P496A).



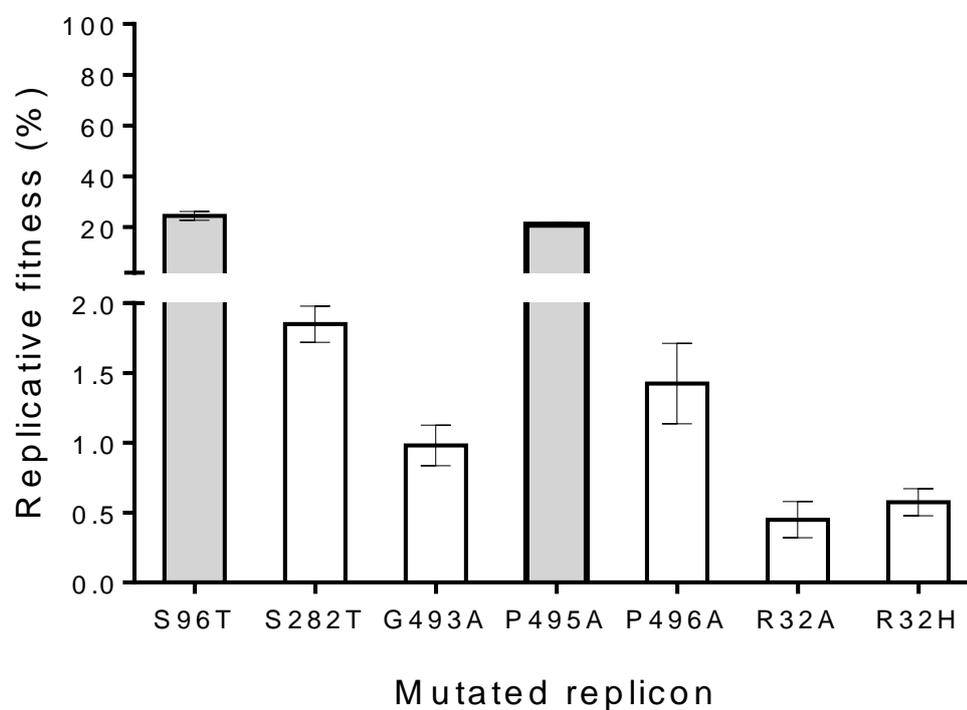
RNA was electroporated into Huh-7.5 cells and luciferase activities contained within cell extracts prepared at 4, 24, 48, 72 h, 96h, and 120h post-electroporation were determined. RLU, Relative light units. This represents one of four independent experiments with triplicates.

Figure 5-30: Replication capacity of mutated replicon (R32A, R32H).



RNA was electroporated into Huh-7.5 cells and luciferase activities contained within cell extracts prepared at 4h, 24h, 48h, 72h, 96h, and 120h post-electroporation were determined. RLU, Relative light units. This represents one of four independent experiments with triplicates.

Figure 5-31: Replicative fitness of replicons containing predicted RAVs.



Replicative fitness was calculated as a percentage= $[(\text{RLU mutant replicon@96h} / \text{RLU mutant replicon@4h}) / (\text{RLU WT replicon@96h} / \text{RLU WT replicon@4h}) \times 100]$. Where RLU (relative light units) is the absolute value given by Luminometer.

5.3 Discussion

In the new era of DAAs targeting HCV, there is a drive towards the routine use of IFN-free, all-oral combinations. Although the new regimens are effective, the clinical relevance of HCV drug resistance, the role of baseline natural polymorphism, and the potential need for resistance testing are important clinical questions that remain unanswered.

Although detectable at only relatively low percentage prevalence, the presence of RAVs in acutely infected, treatment-naïve patients points towards two significant phenomena. Firstly, resistance mutations are present within the quasispecies even prior to treatment. This can mainly be attributed to the dynamics of the viral quasispecies and the error-prone viral RNA polymerase (Rong and Perelson, 2010, Strahotin and Babich, 2012). Several studies have reported that all single RAVs and around 10% of all potential double RAVs pre-exist in infected individuals (Pawlotsky, 2009). Secondly, minority RAVs within the viral population may be transmissible among humans. As this is the first study involving a cohort of acutely infected, treatment-naïve patients, this phenomenon has not been described before. It suggests that despite the reduced viral fitness usually reported in RAVs, they can circulate within the quasispecies in enough numbers to enable transmission. However, the effect of transmission bottleneck in which the fittest variant becomes the dominant variant cannot be excluded.

The possibility that an agent will exert a selection pressure to allow outgrowth of RAVs is influenced by numerous factors including the DAA's genetic barrier to resistance, the level of drug exposure, and the replication fitness of RAVs (Lontok et al., 2015).

HCV genotype 1a has a lower genetic barrier to resistance than genotype 1b variants, hence patients with genotype 1a can acquire RAVs faster, for instance, a RAM at residue R155 requires only one nucleotide change in genotype 1a variants while two nucleotide changes are needed to confer resistance in HCV subtype 1b isolates (McCown et al., 2009, Pawlotsky, 2009). The prevalence of

RAVs at baseline in the studied cohort was higher in HCV/HIV co-infection group than in HCV mono-infection group, but the difference did not reach statistical significance ($p=0.25$). This is in line with reports that the prevalence of natural baseline polymorphisms varies in different populations, a higher prevalence was reported among hemophiliacs (OR 1.67, 95% CI [0.67-4.2]) and HCV/HIV co-infected patients (OR 3.8, CI [0.86-16.8], but the difference in prevalence was not statistically significant when compared with non-hemophiliacs and HCV groups respectively (Lin et al., 2014).

In HIV infection, numerous studies reported transmitted drug resistance with prevalence rates ranging from 3.4% to 26%, which prompted the recommendation of drug resistance testing before initiating drug therapy (Taiwo, 2009). HCV is not expected to have a similar resistance pattern due to two main features: i) it has a higher replication rate that consistently introduces errors into the circulating variants which may fasten its reversion to WT and ii) It does not integrate into genomic DNA to archive RAVs (Lontok et al., 2015). However reports of polymorphisms at certain positions that confer resistance to DAAs (e.g. Q80K) indicate drug resistance testing in HCV before treatment.

5.3.1 Prevalence of natural polymorphisms at resistance-associated residues.

There is an intense debate on the relevance of naturally occurring mutations with respect to the development of resistance and probability of success of DAAs.

5.3.1.1 Resistance to NS3 protease inhibitors.

The majority of RAVs to HCV PIs circulate as low-frequency variants at baseline. This is due to the inherent fitness cost caused by the decreased catalytic activity of the viral protease (Pawlotsky, 2011). The naturally occurring RAVs in the NS3 gene have been reported in several studies of HCV viral complexity using standard sequencing (Morsica et al., 2009, Paolucci et al., 2012, Bartolini et al., 2013).

In this study, a number of mutations were detected in the NS3 region. V36A/G/L, F43L/I/S/V, T54A/S, Q80K/L/H/R, V55A, R155K, A156S, D168A/V/Y/G/E/N, F169L, and I/V170L/T substitutions were observed in co-infected patients. In contrast, only mutations Q80K, R155S/W, D168G/E/N were detected in mono-infected patients. These substitutions have been reported to confer resistance to SMV and PTV and were found mostly as minority variants and at low prevalence except Q80K. Q80K had a high prevalence at baseline in both HIV/HCV coinfecting and HCV mono-infected patients (56.25% and 83.3% respectively). Other studies also reported Q80K as a common variant at baseline by Sanger sequencing and NGS. The prevalence of Q80K in this study was slightly higher than in previous reports, which reported 5-40% prevalence by Sanger and 40-57% by NGS (Shepherd et al., 2015, Jabara et al., 2014, McCormick et al., 2015, Ruggiero et al., 2015, Ogishi et al., 2015, Leggewie et al., 2013, Kirst et al., 2013).

Minority RAVs were detected at residues D168A/H/V/Y and R155K, which confer the highest level of resistance towards PTV and SMV, at 25% and 6.5% respectively. This result corresponds with earlier studies that high levels of resistance variants were present at a low level at baseline in treatment naïve patients or not at all in other studies (Bartels et al., 2013, Leggewie et al., 2013, Jabara et al., 2014, Shepherd et al., 2015).

Some minority variants were only observed in HIV/HCV co-infected patients including V36A/G/L, T54A/S, and V55A. These mutations are associated with low-level resistance to SMV and PTV (Lange and Zeuzem, 2013, Pilot-Matias et al., 2015).

In all analysed patients, V170I and Q86P were dominant variants. These substitutions represent natural polymorphisms that frequently occurs in genotype 1a sequences (data from the NCBI database shows that 96.71% of recorded HCV genotype 1a sequences contain the V170I variant and 99.34% contain Q86P). These two residues do not produce resistance to PIs.

Only one patient had a dominant Q80K mutation in HCV mono-infected (1/18) and HCV/HIV co-infection groups (1/16). The dominant RAVs are rarely reported in NS3 gene at baseline as explained above. The frequency of dominant RAVs reported in 507 treatment-naïve HCV genotype 1 patients was 0.3%-2.8%; most patients harbouring these dominant RAVs had a high viral load indicating RAVs may reach replicative capacities similar to non-resistant variants *in vivo* (Kuntzen et al., 2008). In a similar study using population sequencing of NS3 region, a low prevalence of dominant RAVs was detected as V36M was detected in only 0.9% of patients at baseline and R155K was detected in 0.7% of patients at baseline (Bartels et al., 2008).

The low prevalence of dominant R155K is in line with both *in vitro* and *in vivo* reports that have indicated that the R155K variant displays impaired fitness compared with WT virus. Thus, these detected dominant variants are most probably associated with compensatory mutations that improve the replication capacity of this RAV (Sarrazin et al., 2007).

5.3.1.2 Resistance to NS5A inhibitors

In this study, baseline resistance mutations within the NS5A region were detected in all samples using NGS. This result is significantly greater than that previously detected by population-based sequencing studies in which the prevalence of NS5A RAVs has been reported at around 12.5% (Paolucci et al., 2013).

Substitutions at residues Q30 and Y93 produce the highest level of resistance of all NS5A inhibitors (>1000 fold change) (Wong et al., 2013, Lawitz et al., 2012, Krishnan et al., 2015a). Minority variants Y93H and Q30H/R were detected in patients with HIV/HCV and HCV alone, at 25% (Y93H) and 25/18.75% (Q30H/R) respectively. Previous studies have also shown variant Y93H at a similar prevalence of 7-33% in DAA-treatment naïve HCV genotype 1 patients (McCormick et al., 2015, Yoshimi et al., 2015, Plaza et al., 2012).

The other common baseline polymorphisms within NS5A are L31M and M28T. However, L31M/R mutations were only detected in HIV/HCV co-infected patients in this study at low prevalence (12.5%). The prevalence of L31M was noticeably higher than in previous studies using direct sequencing, where L31M was detected in only 3% of patients (Paolucci et al., 2013, Yoshimi et al., 2015, Wong et al., 2013). Apart from the difference in technology used, the small sample size in our study could be a reason for this discrepancy.

The H58P mutation was also found at high prevalence in our cohort; 50% in mono-infected and 31% in co-infected subjects. The result is in accordance with an earlier NGS study that identified H58P at high frequency, reaching 87% compared to only 6.2% by direct sequencing (Paolucci et al., 2013). This variant is a natural polymorphism that does not confer resistance in genotype 1a. The H/P58D variants that do confer resistance against DCV, LDV and OBV were not detected in any sample (Lontok et al., 2015).

5.3.1.3 Resistance to NS5B inhibitors

In our study cohorts, the following resistance mutations were observed at high prevalence, C316N, L159F, V405I, F415Y, I434M, E446Q, A553V, and S556G, which were above 40% prevalence for both co-infected and mono-infected samples. L159F and F415Y have been reported to confer low-level resistance to SOF of 1.9 and 1.3 fold respectively (Donaldson et al., 2015).

Mutation S282T was only identified at low prevalence in HIV/HCV co-infected patients (6.25%). In previous clinical trials, the S282T mutation was not detected in treatment naïve individuals at baseline using deep sequencing (McCormick et al., 2015, Franco et al., 2013). SOF is characterized by a high resistance barrier. The presence of double mutations, L159F and L320F increases fold resistance to SOF, and this resistance can be enhanced with an additional S282T variant (Tong et al., 2014, Poveda et al., 2014). Meanwhile, residues S96, S282 and P495 were previously reported as being 100% conserved in treatment-naïve HCV genotype 1 infected patients (Kuntzen et al., 2008).

Using clonal analysis, no S282 mutation was detected in viral quasispecies at baseline in genotype 1 HCV infected individuals who did not receive prior treatment (Le Pogam et al., 2012); similar results were obtained after screening 16 HCV/HIV treatment-naïve patients at baseline using NGS (Franco et al., 2013). Moreover, the S282T mutation has been extremely difficult to detect *in vivo* even in patients with failure to SOF (Wyles, 2013).

A systematic review of the literature showed that RAVs to DSV were more prevalent than those for SOF. This may be related to the mechanism of action of DSV as an allosteric inhibitor within NS5B, resulting in a low barrier of resistance (Powdrill et al., 2010). RAVs to DSV were detected in all patients in this study.

The NS5B amino acid substitutions most commonly observed in genotype 1a-infected patients who did not achieve SVR were M414T and S556G (Lontok et al., 2015). The C316N and S556G variants were present at high prevalence in the HIV/HCV co-infected cohort at 25% and 37.5% respectively. These two RAMs have been reported to confer low-level resistance of 5 and 11-fold respectively. The presence of the double mutation C316N + S556G increases fold resistance by up to 38-fold (Koev et al., 2009, Krishnan et al., 2015b). Both M414T and C316Y were not detected at baseline in any of the patients.

5.3.2 Prediction of *in vitro* resistance mutations within NS5B

Numerous clinical studies have suggested that acquisition of RAMs is associated with viral fitness cost, characterised by an inverse relationship between the level of resistance conferred by a mutant variant and its replicative capacity. V36M represent an example of this relationship. It has been characterised as a low-level resistance mutant while it exhibits one of the best *in vivo* fitness rates (95% of wild-type fitness) of all reported PI-RAMs (Sarrazin et al., 2007, Tong et al., 2008). It has previously been reported as a dominant variant in acute HCV/HIV co-infected patients (Leggewie et al., 2013).

We used a prediction model to identify residues in the NS5B region that interact directly with the SOF binding site. Using the model described above, we hypothesized that mutations at residues R32, G493, P495, and P496 could confer resistance to SOF. These predicted mutations were introduced into a genotype 1a subgenomic replicon using site-directed mutagenesis. The previously described NS5B resistance mutations S282T and S96T were also introduced (Powdrill et al., 2010).

A simplified assay to examine transient replication of the TA/SG/Luc/HCV-1a subgenomic replicon in Huh-7.5 cells was developed. The introduction of two mutations in NS4A and NS4B enhanced replication fitness around 100-fold compared with the H77 genotype 1a replicon (Voitenleitner et al., 2012). This adapted replicon construct provided a rapid assay through measuring luciferase readout, used as a reporter for the reliable determination of replicative capacity. Because of the relative ease of introduction of mutations into the replicon, this system could be used to quickly profile a large number of compounds on a panel of resistance mutations against various HCV targets.

Using this system, only two constructs containing the mutations S96T and P495A were replication competent. Replication occurred at low level only, 24% and 21% respectively. In keeping with the *in vitro* data, only two clinical samples harboured mutations at residues S96 and P495 at a frequency of 1-3%, but none of the detected mutations have been reported to confer resistance to SOF or DSV. Higher prevalence of P495A/L/T was reported when a group of 27 treatment naïve genotype 1b patients was screened for RAVs at base line using NGS, 33% of patients were reported to harbour P495A/L/T mutations at a frequency of less than 1%, but these substitutions are not reported to be associated with resistance against either SOF or DSV; and no patient samples contained S282T mutants (Nasu et al., 2011).

All other constructs did not replicate and markedly reduced replication of the S282T construct was in agreement with previous reports (Ludmerer et al., 2005).

These sites were highly conserved both in our samples and in all published genotype 1a full genome sequences on NCBI.

These findings extend and confirm previous studies that showed a low prevalence of the S282T mutation *in vivo*. It has also been reported that S282T reverts to wild type after cessation of treatment (Ludmerer et al., 2005). These results may explain the absence of breakthroughs in patients failing therapy with NS5B nucleos(t)ide analogues due to their high *in vivo* barrier for developing resistance and likely rapid reversion following cessation of treatment.

5.3.3 Resistance profile in HIV/HCV co-infected patients

The effects of HIV co-infection and HAART on HCV quasispecies variability have not been firmly established. Some studies suggested that greater diversity occurred in HCV/HIV co-infected compared to HCV mono-infected subjects (Tanaka et al., 2007, Blackard and Sherman, 2007), while others studies have reported the opposite (Jabara et al., 2014, Shuhart et al., 2006).

T-cell responses may play a vital role in determining genetic diversity in both patient groups (Jabara et al., 2014). Heterogeneity and replication dynamics could expand under selective forces provided by host immune response and antiviral agents. HIV/HCV co-infection will reduce immunological pressure and could limit the escape mutants, therefore, reducing diversity (Nunez et al., 2006). Nevertheless, there is evidence that the differences are generally not statistically significant, similar to the data obtained here (Morsica et al., 2009, Paolucci et al., 2013, Trimoulet et al., 2011).

In this study, HCV/HIV co-infected patients had a higher number of RAVs compared to mono-infected patients, but the results did not achieve statistical significance. However, the prevalence of few RAMs has been detected at a higher level in patients with HCV alone (e.g. Q80K). The main limitations of our study were the low number of patients analysed and a lack of direct comparability between HCV-mono-infected and HIV co-infected patient groups.

Thus, it is not possible to fully establish whether HIV co-infection is an independent factor associated with a higher prevalence of baseline RAMs.

5.3.4 The potential role of NGS in antiviral resistance testing

Sanger sequencing remains the standard method to detect RAVs in clinical samples. Nevertheless, the Sanger sequencing method is not able to reveal mutations that have a frequency below 20-30% in a viral population (Palmer et al., 2005). Although clinical impact of minority variants on DAAs resistance is still under investigation, several studies have suggested that minor variants are clinically relevant to drug resistance. Thus, there is a need for an improved diagnostic tool that provides a better sensitivity in detecting the circulating minority HCV variants (Barzon et al., 2011).

NGS detected low-frequency RAVs in all our samples. A major challenge in SNP recalling is to eliminate different sources of errors that may occur during sample preparation before sequencing, during reverse transcription, in PCR reactions and during sequencing itself. These errors are likely to reduce the validity of variant detection at low frequency (Jabara et al., 2014, Thys et al., 2015). For an accurate SNP calling, a frequency at 1% was used and a depth at 100 reads were used as cut-offs to exclude false variants.

The limited read length of Illumina® sequencing technology to 200-300 bp was addressed by powerful alignment and assembly tools (Loman et al., 2012). However, this limited the ability to reconstruct the haplotypes to understand the synergistic effect of RAMs or the presence of compensatory mutations. In order to overcome this limitation different haplotype reconstruction tools (e.g. QuRe) were introduced but validation of these tools is still needed.

Despite these limitations, NGS is a potential tool for detecting minority variants due to its ability to produce large amounts of data in a timely manner (Ninomiya et al., 2012, Beerenwinkel and Zagordi, 2011, Mardis, 2011).

5.3.5 Clinical utility of resistance testing

The presence of naturally occurring RAVs may hinder the efforts to eradicate HCV, and there is an ongoing debate as to whether resistance testing should be performed prior to DAA treatment in some patient groups (e.g. genotype 3 and cirrhotic patients). In the era of DAAs, the importance of pre-existing as well as selected variants during antiviral therapies may become increasingly important. However, treatment-naïve patients with TVR-RAVs achieved similar SVR rates compared to patients without RAVs (Bartels et al., 2013, Halfon and Sarrazin, 2012).

The possibility of the selection for highly replication competent RAVs urged the implementation of stop rules in case of viral breakthrough. These variants probably persist for a considerable period leading to increased possibility of transmission to other individuals (Buhler and Bartenschlager, 2012).

Different RAVs are reported to be clinically relevant when present at baseline. Q80K was shown to reduce SVR rates when patients are treated with SMV in combination with IFN-based regimens (Jacobson et al., 2014). Common RAVs in NS5A at baseline are associated with reduced efficacy of DCV (McPhee et al., 2013, Dore et al., 2015). Moreover, PTV and LDV monotherapy reported a lower response rate in the presence of baseline RAVs (Lawitz et al., 2015a). Despite these reported links with treatment failure, the rates of virological failure are low when treatment include another agent such as SOF, which rendered the resistance testing unnecessary in the case of combination therapy (Sulkowski et al., 2014a).

The resistance testing before treatment with NS5A inhibitors may be of value in genotype 1a due to the low genetic barrier to resistance of the current agents; for example at M/L28T, this change requires two nucleotide changes in the genotype 1b sequences (L28T; CTC to ACC) but only one change in the genotype 1a replicon (M28T; ATG to ACG). The second generation agents may have a higher genetic barrier which may render baseline testing irrelevant (Nakamoto et al., 2014).

There is contradictory evidence on the value of baseline resistance testing. A combination of SOF and DCV ± RBV in genotype 1-3 patients was successfully used to treat a cohort of patients in which 16% of patients had baseline DCV-RAVs. However only one patient had a relapse (Sulkowski et al., 2014a). Thus, regardless of the pre-existing baseline RAMs, high SVR rates can still be achieved in most patients (Lawitz et al., 2014b). A similar outcome was reported after treatment by PrOD combination +/- RBV (Sulkowski et al., 2015). These results confirmed the value of combining different classes of DAA to reduce the effect of pre-existing RAVs on achieving SVR.

The emergence of *de novo* TVR-RAVs and SMV-RAVs after the start of triple therapy was not predicted at baseline (Akuta et al., 2013). Over time, the majority of *de novo* resistant variants become undetectable (Akuta et al., 2014). These results confirm that baseline resistance testing could be of limited value in prediction of resistance to DAAs.

Recent EASL guidelines recommended limited HCV baseline resistance testing in patients infected with subtype 1a who receive a combination of PegIFN α , RBV and SMV because SVR rates are very high both in patients without and with detectable amounts of RAVs using population sequencing at baseline. In special groups (e.g. cirrhotic patients and those infected with genotype 3) where the effectiveness of SOF shows some limitations, and before retreatment after NS5A inhibitors treatment failure the need for resistance testing of NS5A inhibitors may be of value.

5.4 Conclusions

RAVs occurring as natural polymorphisms are frequently detected in HIV-positive patients with acute HCV infection. The significance of prevalence thresholds is not yet established and further studies to investigate the impact of RAVs are needed. Further large-scale studies using deep sequencing should be performed to investigate the effects of RAVs on the response to treatment using new drugs, including DAAs, monitoring the evolution of RAVs during treatment and after stopping treatment in cases of treatment failure.

The replicon model is potentially a useful tool to monitor the evolution of resistance variants arising from DAA therapy.

NS5B inhibitors have a high genetic barrier to resistance. In this chapter, we have shown that the TA/SG/Luc/HCV-1a replicon is a useful model for investigating the replication fitness of any potential RAV.

Antiviral resistance testing is an important subject for future research in the era of DAAs; NGS is a superior tool for detecting minority RAVs although the clinical utility of such testing needs to be investigated further. The introduction of new agents with higher genetic barriers of resistance should be recommended in combination regimens to limit the impact of RAVs in the future.

Chapter 6: Conclusions and further work

6.1 Conclusions

6.1.1 Sequencing of the HCV genotype 1a open reading frame

HCV full genome sequencing has been a main challenge in the field for many years. The main obstacles are the high diversity of HCV and the presence of secondary structures in different regions of the genome.

In the series of experiments described in Chapter 3, the complete ORF was assembled from four amplicons produced by nested PCR using high fidelity DNA polymerase to achieve an efficient HCV amplification with minimal bias. The primer design and optimisation of PCR conditions produced a robust, rapid and sensitive method for amplification and sequencing of HCV genotype 1a with a detectability limit of 1000 copies/ml.

Another full genome sequencing method was validated; employing a metagenomic sequencing approach. It is an improved tool for detecting SNPs across the genome with limited systematic PCR error and without any primer selection bias. The latter enables the detection of all HCV genotypes as the metagenomic approach is not limited to genotype-specific primers. However, metagenomic sequencing is limited by lack of depth and potential under-representation of minority variant in low viral load samples. This approach was used in the detection of RAVs as described in chapter 5.

6.1.2 The role of NGS in studying HCV quasispecies

The development of NGS platforms has significantly enhanced our understanding of intra-host viral populations, providing a far better understanding of viral complexity with the potential to improve on conventional methodologies in both research and clinical settings.

Pyrosequencing of amplicons covering the HVR1 region revealed the presence of multiple variant infections which were not detected by direct Sanger sequencing or clonal analysis. Pyrosequencing had a similar performance to the Illumina platform in detecting minority variants but had an inherently higher error rate.

An artificial quasispecies was created (Mock community) to assess the performance of different haplotype reconstruction programs in detecting circulating variants. Two analysis software packages were assessed and (QuRe and PredictHaplo) were comparable and detected the most frequent variants in the mock community. To date, this is the first study of HCV quasispecies using these programs.

6.1.3 The effect of viral dynamics on treatment outcome

In chapter 4, I described how 454 pyrosequencing can be used to address HCV viral dynamics and impact on treatment outcome. Low viral diversity at baseline is a predictor of achieving SVR in HIV-infected patients with acute HCV. We based this observation on comparing ten patients who responded to a PegIFN α /RBV course with a matched group of ten patients who failed the same treatment.

In patients who failed PegIFN α /RBV treatment (15 patients; null responders, partial responders, and relapsers), all patients had persistent variants that were present in both pre- and post-treatment samples. The emergence of new viral variants detected by Sanger sequencing following treatment failure was associated with emerging dominance of pre-existing minority variants rather than reinfection. These results indicate that relapse and re-infection should be defined by phylogenetic support using sequences obtained using an NGS platform.

6.1.4 Antiviral resistance in HCV

NGS has a potential role in screening for known antiviral resistance mutations and is superior to conventional sequencing methods in the detection of minority variants. However, the role of resistance testing in the HCV field remains controversial, and the impact of minority RAVs is yet to be understood. High SVR rates and high genetic barriers to resistance are likely to limit the requirement for testing in the majority of HCV patients.

The prevalence of naturally occurring RAVs within NS3, NS5A and NS5B was higher at baseline in HIV/HCV co-infected compared with HCV mono-infected treatment-naïve genotype 1a patients. Several RAVs mutations were detected predominantly as minority variants (<10%) within all of these genes.

A trend towards more RAVs detected at baseline occurred while the prevalence of few RAVs (e.g. Q80K) was higher in patients with HCV mono-infection. Lack of statistical support for a significant difference in the prevalence of RAVs at baseline may be attributed to small sample size and the fact that both groups were not matched for other confounding factors such as viral load.

During this thesis, an extensive literature review of RAMs in the HCV genome that are reported to confer resistance to DAAs was carried out (SMV, PTV, LDV, OBV, DCV, DSV, and SOF). Chapter 5 provides a comprehensive summary of RAMs for currently recommended HCV DAAs. Differences between our findings and published resistance reviews may relate to different methodologies for the compilation of clinically relevant substitutions. In this rapidly evolving field, regular updates of RAV databases will be required. This database will contribute to the set up a national HCV Research UK database that will enable clinicians to tailor treatment strategies.

6.1.5 Replicon based assay for RAVs replication capacity

In chapter 5, I showed that the TA/SG/Luc/HCV-1a replicon provided a useful

model for investigating the replication capacity of any potential RAV. Resistance mutations at residues that have an interaction with the binding site of SOF were introduced by site-directed mutagenesis. It was shown that all of the introduced mutations had a viral fitness cost resulting in inefficient replication. These results explain the reported high genetic barrier to resistance of SOF. This also matches clinical trials experiences and the current recommendation that resistance testing at baseline is not indicated if any combination including SOF is used for HCV treatment. New RAVs could be tested using the same protocol in this genotype 1a system.

6.2 Future work

6.2.1 Sequencing of the HCV genotype 1a open reading frame

HCV full genome sequencing has evolved during the period of my thesis and the target enrichment approach has now become a routine method in our lab. Future plans include exploration of other platforms for full genome sequencing, such as SMRT sequencing (Pacific Bioscience). These alternative platforms allow longer read lengths but require approaches to decrease higher error rates.

6.2.2 The effect of viral dynamics on treatment outcome

After exploring the role of viral diversity in predicting treatment outcome in this unique cohort, further studies are needed to investigate the role of viral diversity in spontaneous clearance to determine whether it could be a predictor of spontaneous clearance in this population.

A group of spontaneous clearers will be sequenced using NGS in parallel with a group of progressors from the same cohort. Both groups will be matched for age, viral load, nadir CD4, duration of infection and genotype. We hypothesise that HCV diversification and evolution predicts the outcome of acute infection in this cohort.

6.2.3 Mock communities

Further mock community analysis will enable benchmarking of current haplotype reconstruction programs (e.g. QuRe). The mock community used during this thesis represented a low diversity sample. Further communities are designed to have higher diversity and larger genetic distance between different variants and include a different genotype with sequence divergence 0-45%. Fifteen genotype 1a samples have already been amplified using the amplicon 1 primer set shown earlier with an average size 2.5kb and an amplicon with an average size of 3kb from one genotype 3 patient (Figure 6-1).

The frequency of variants within the mock communities are as follow:

Mock 1; 40 clones with uniform abundance

Mock 2; 40 clones with log-normal abundance

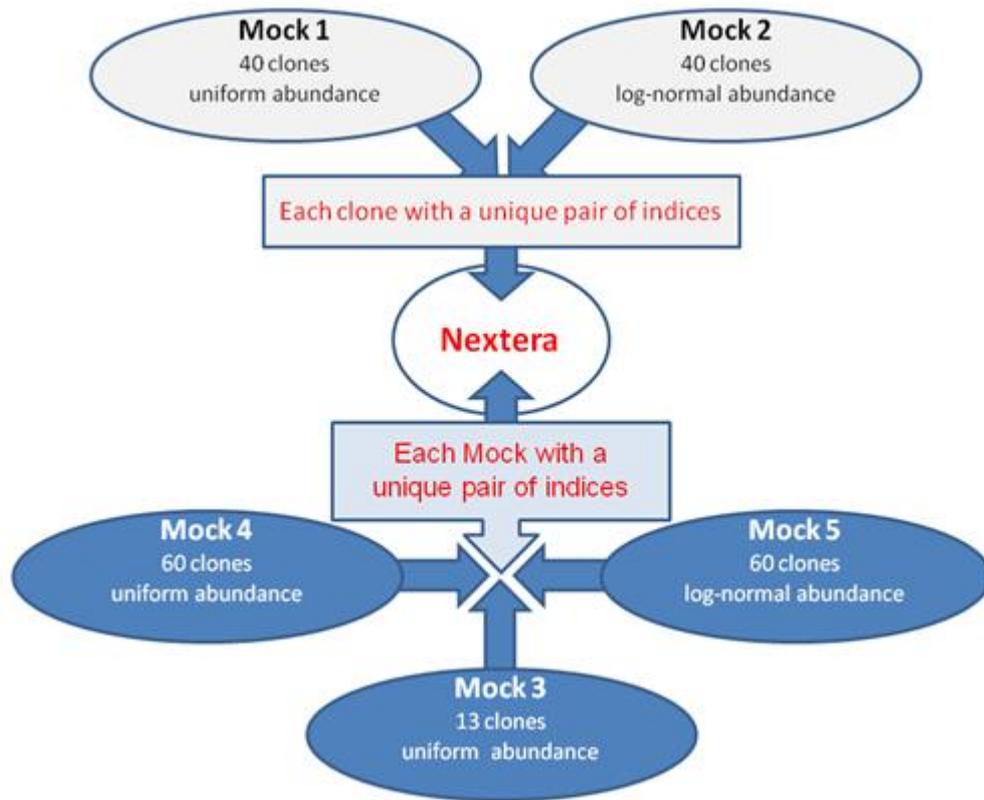
Mock 3; 13 clones with log-normal abundance

Mock 4; 60 clones with uniform abundance

Mock 5; 60 clones with log-normal abundance

NGS data will be analysed using the optimised bioinformatics pipeline, and different haplotype reconstruction programs will be evaluated to find the best available software to offer a precise and sensitive tool that can give a nearly accurate estimation of the circulating variants and permit linkage studies of different mutations and its linkage on the same haplotypes. This project will be in collaboration with Dr Melanie Schirmer, Broad Institute, USA.

Figure 6-1: The structure of mock communities



6.2.4 HCV antiviral resistance

Further large-scale studies using NGS should be performed to investigate the effects of RAVs on response to treatment using DAAs. These studies must target larger populations of HCV-infected patients in particular and patients with difficult to treat genotypes (genotype 3), patients with cirrhosis and those with previous DAA treatment failure.

The sequencing platform used will be Illumina using optimised bioinformatics pipelines. In order to ensure accurate analysis of antiviral resistance, the RAV database created during this thesis will be updated regularly (4 monthly) to include any new resistance mutations or DAAs.

Patients who fail DAA treatments will provide the opportunity to study resistance further, including the role of frequency threshold of baseline RAVs and the role of minority variants in predicting SVR. Currently, all samples from patients who failed DAA treatment referred to West of Scotland Specialist Virology Centre in Glasgow are stored pending ethical approval to be included in an antiviral resistance study including monitoring the evolution of RAVs after the end of therapy. This study will offer new insights into the evolution of viral populations pre- and post-treatment.

6.2.5 Cost - effectiveness of NGS in the diagnostic setting

Further experiments are required to validate the use of NGS in diagnostic virology laboratories and define an error threshold specific to platforms in order to improve the specificity of NGS and to exclude any artefactual variants that may have an implication on planning the treatment strategy.

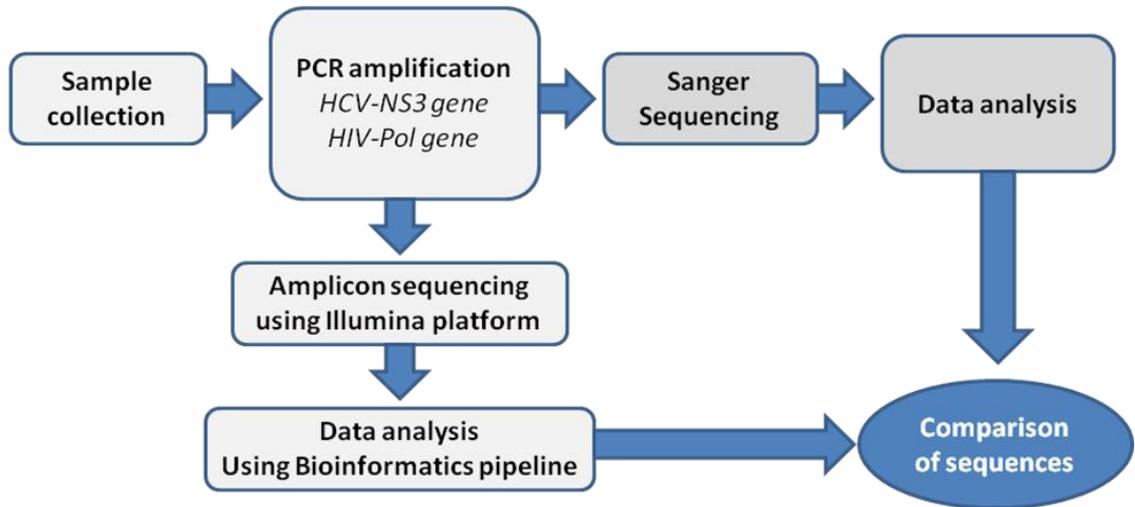
Once NGS is upscaled in the laboratory setting, there will be a reduction in the running costs. For this reason, major clinical virology laboratories in the USA and the UK (Public Health England) are currently optimising this technology.

During this thesis, as well as identifying multiple strain infections, we employed NGS to identify major and minor resistance associated variants that were not detected by Sanger sequencing.

A study at the West of Scotland Specialist Virology Centre (WOSSVC) is currently submitted for funding; it aims to compare the performance, turn-around time, and costs of NGS against Sanger sequencing.

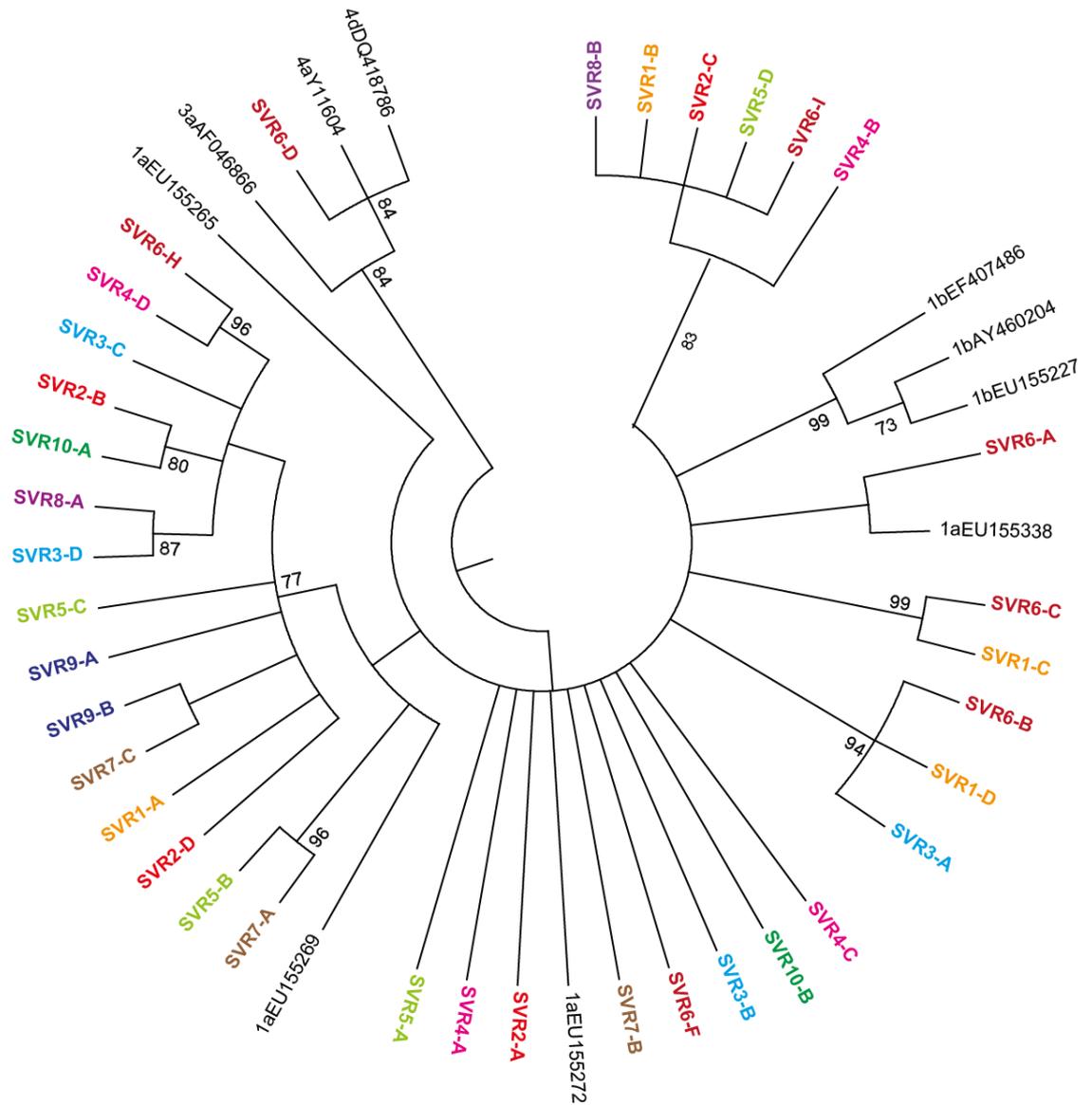
The study will validate the use of NGS as a diagnostic virology tool for genotyping and the detection and monitoring of HIV and HCV high and low abundance RAVs. The samples used in this study will be amplicons generated as part of the routine diagnostic service (targeting the NS3 region of HCV genotype 1a and Pol gene of HIV genome) will be anonymised and attributed a study number. Based on a study power of 80%, a sample size of 100 samples tested by both techniques was estimated to detect a significant difference between both techniques in detecting minor variants. The project workflow is shown in (Figure 6-2).

Figure 6-2: Workflow of cost-effectiveness study of NGS in diagnostic setting



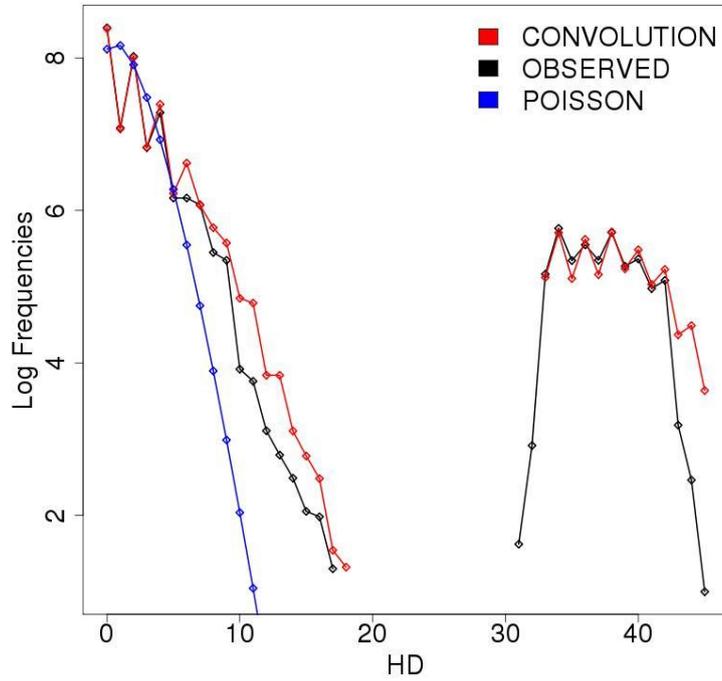
Appendices

7.1 The viral complexity in SVR group

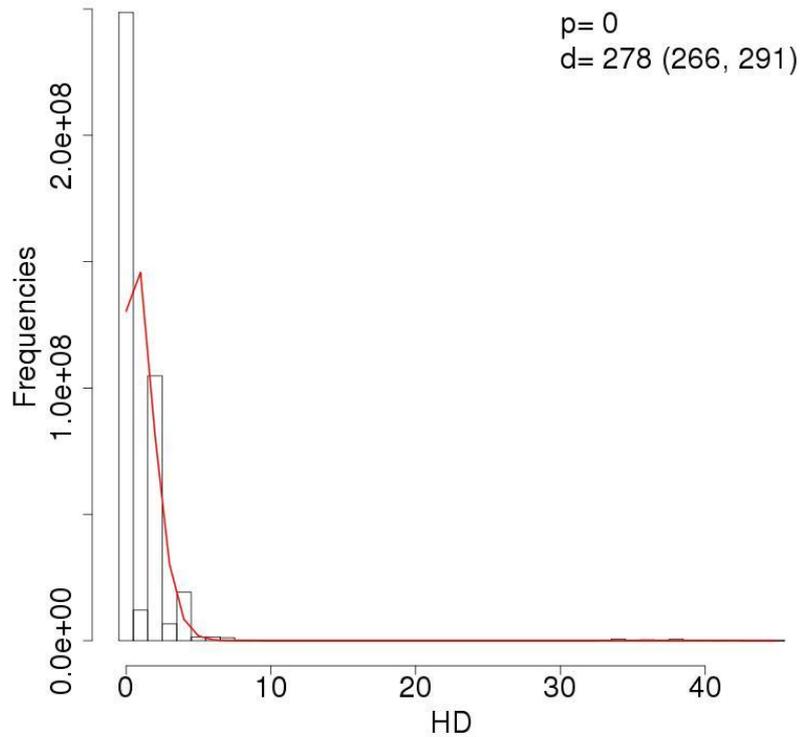


Maximum likelihood tree showing the variants the SVR group detailed in Section 4.2.1.4. The bootstrap values are shown next to the branch points (1,000 replicates); only values higher than 70% are shown.

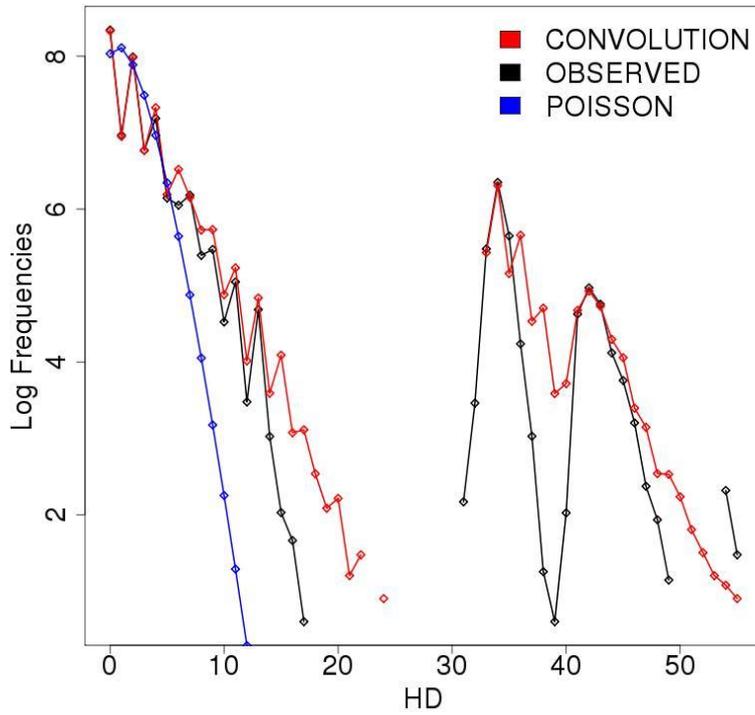
7.3 Poisson Fitter of variants detected in P6



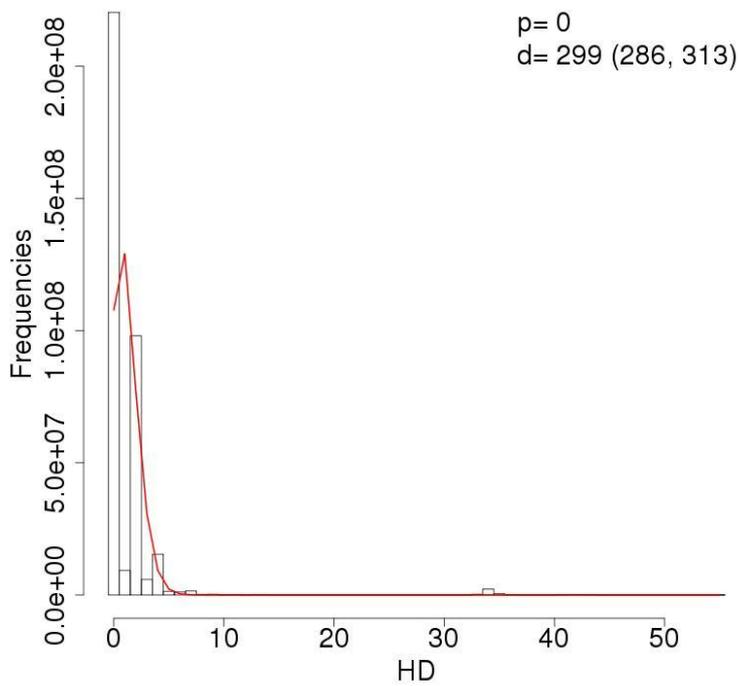
P6_D16_04_2004.



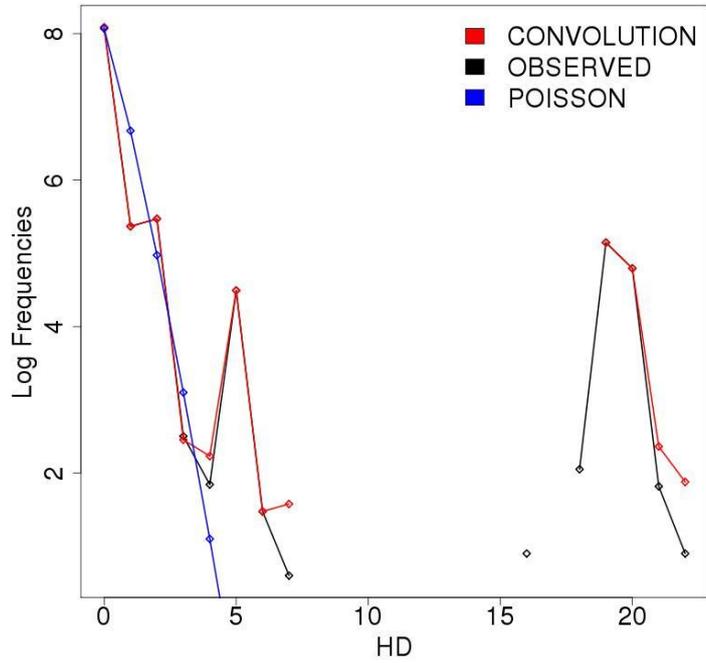
7.4 Poisson Fitter of variants detected in P9



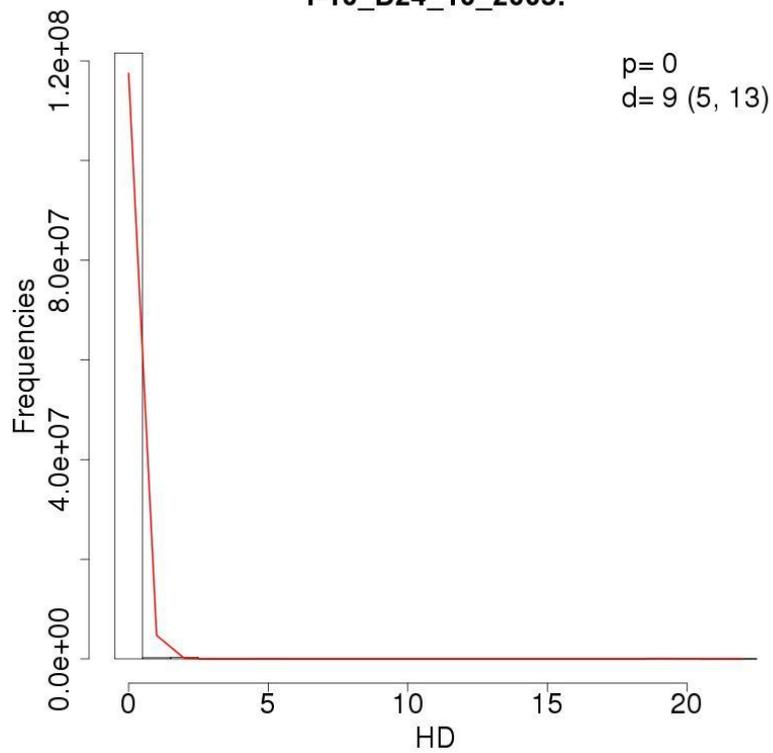
P9_D21_03_2005.



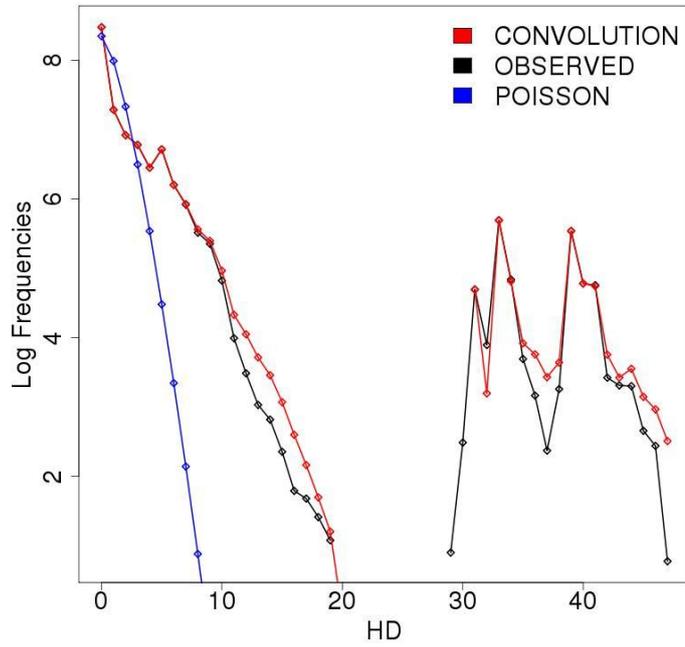
7.5 Poisson Fitter of variants detected in P10



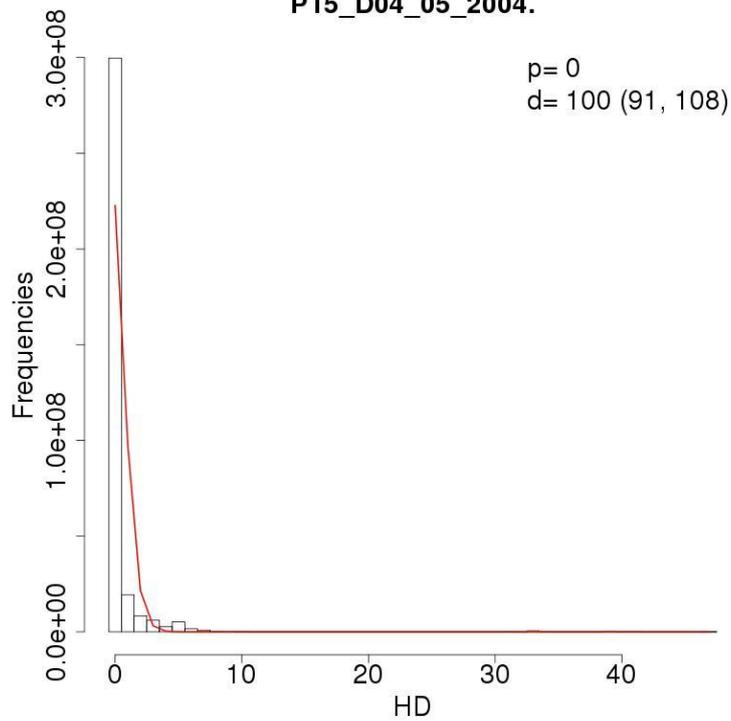
P10_D24_10_2005.



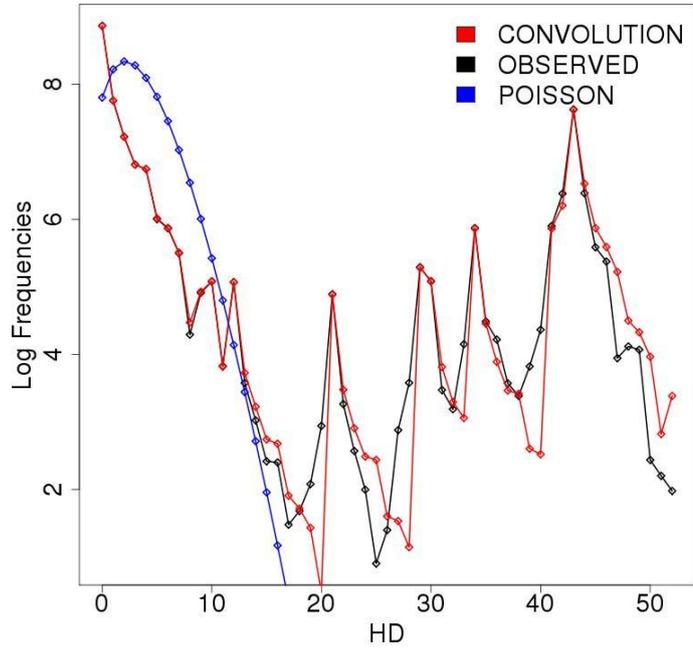
7.6 Poisson Fitter of variants detected in P15



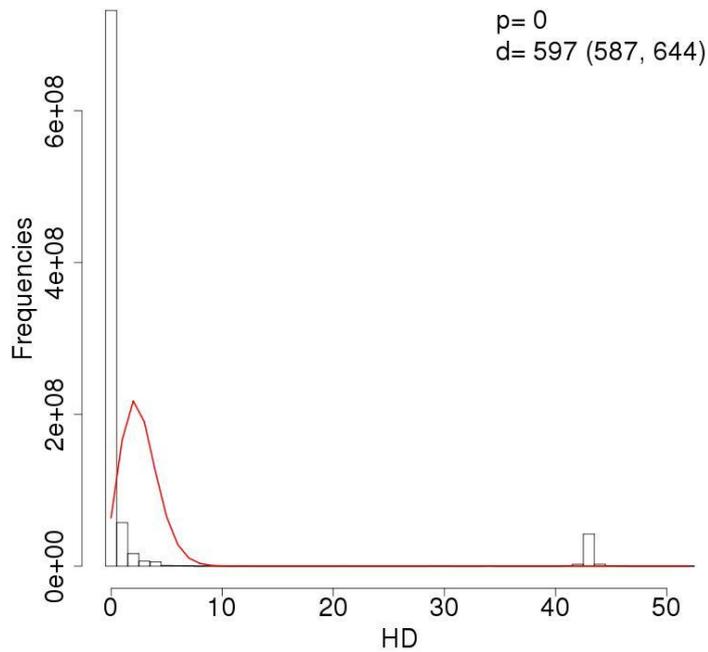
P15_D04_05_2004.



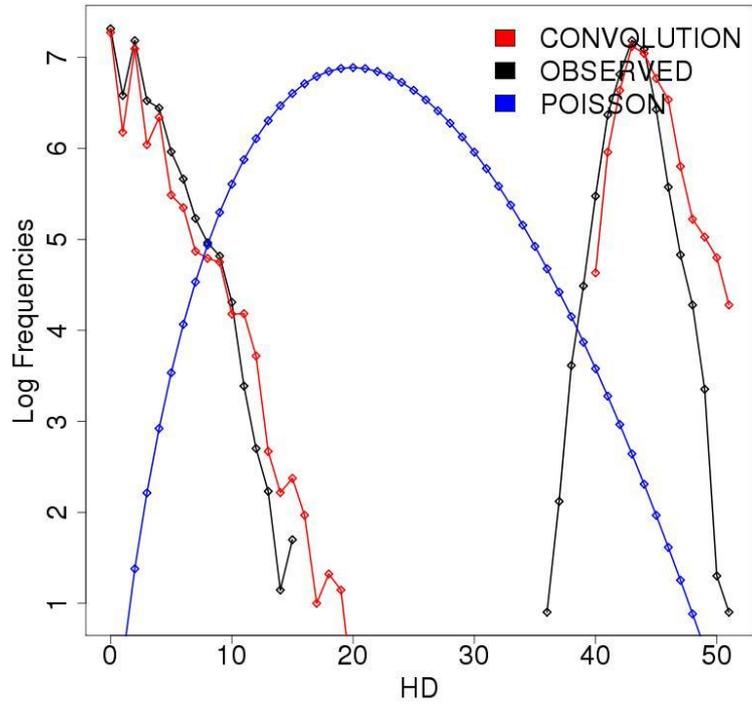
7.7 Poisson Fitter of variants detected in P26



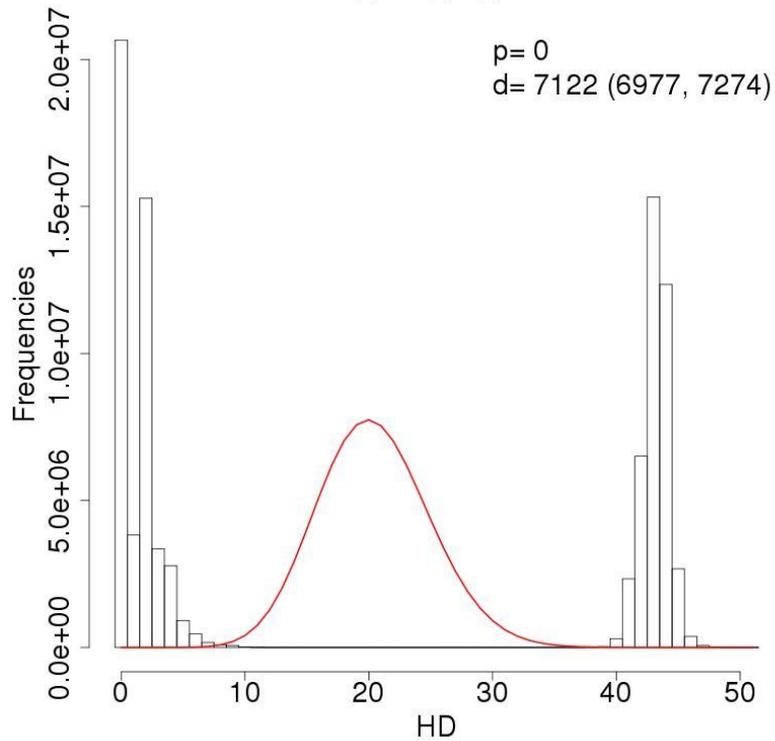
P26_D05_11_2004.



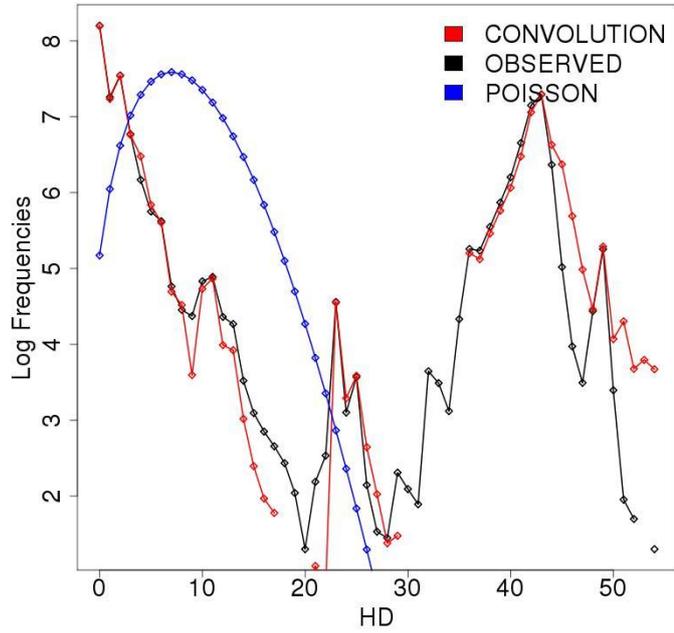
7.8 Poisson Fitter of variants detected in P31(TF group)



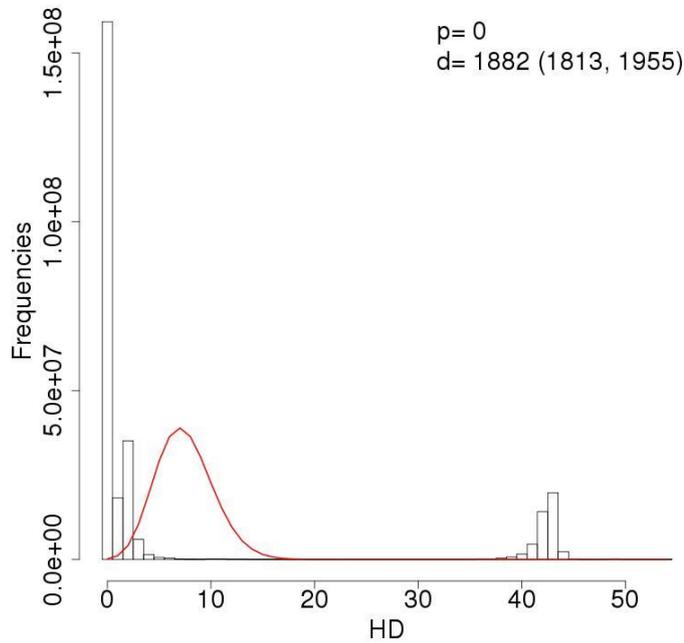
P31_D01_09_2003.



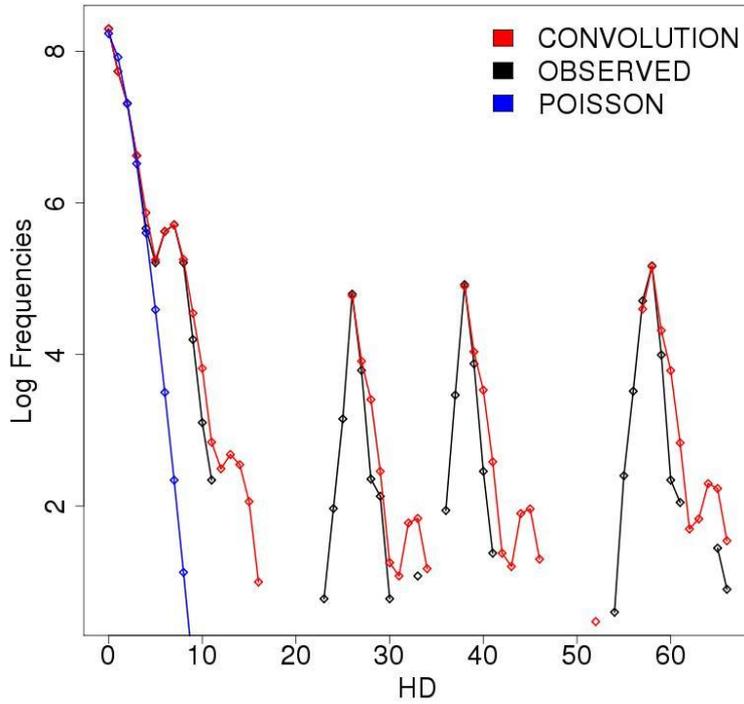
7.9 Poisson Fitter of variants detected in P57



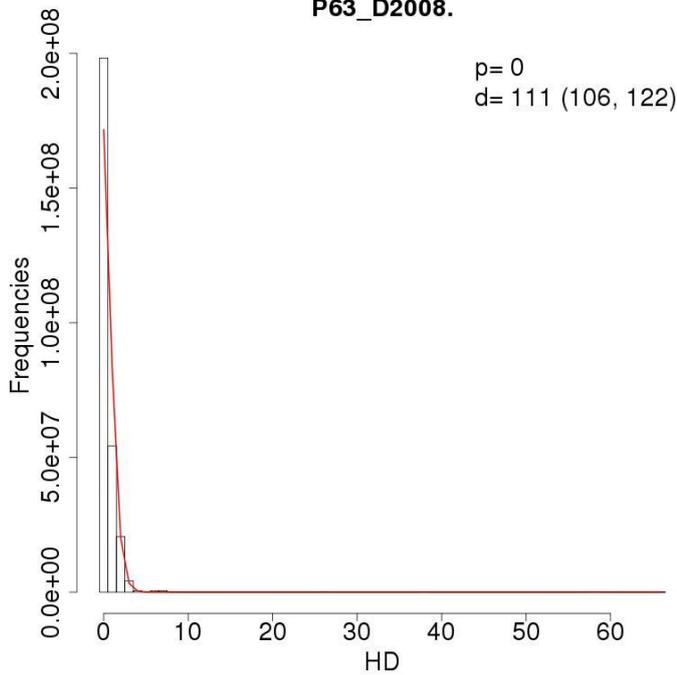
P57_D21_02_2008.



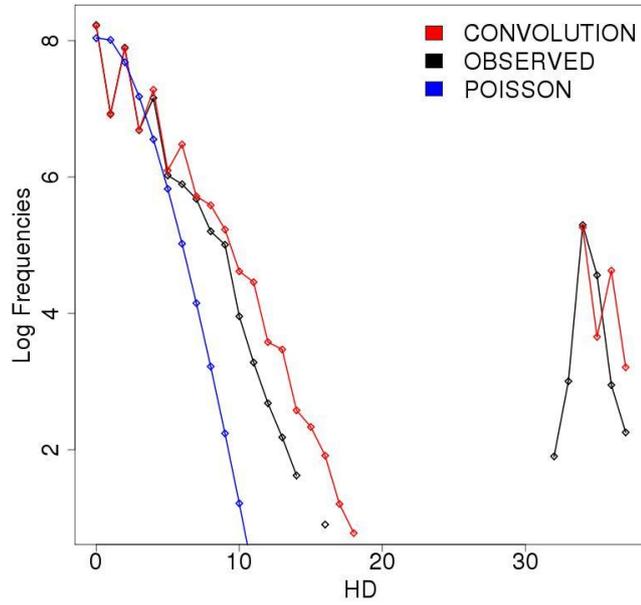
7.10 Poisson Fitter of variants detected in P63



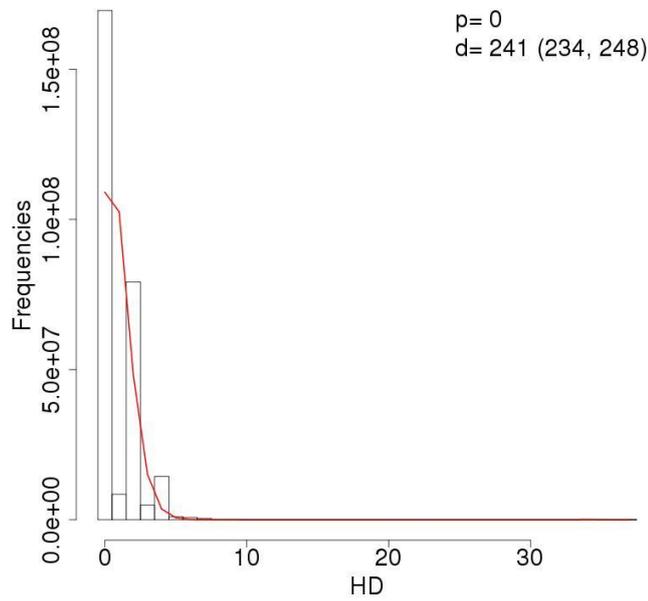
P63_D2008.



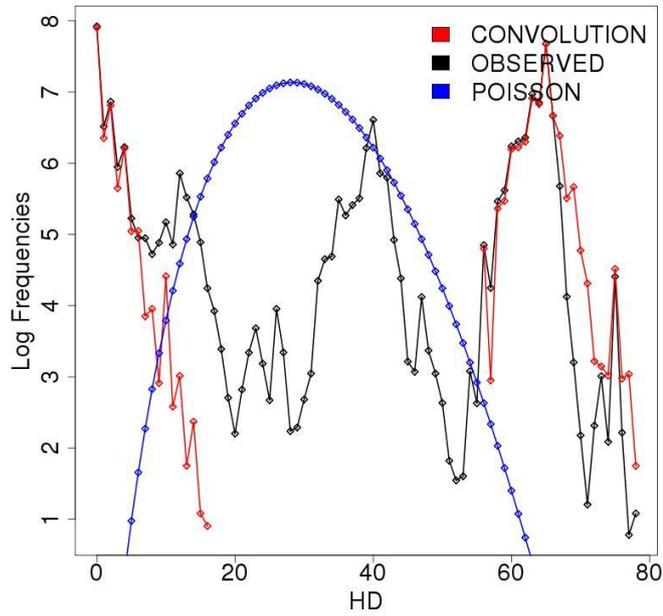
7.11 Poisson Fitter of variants detected in P75



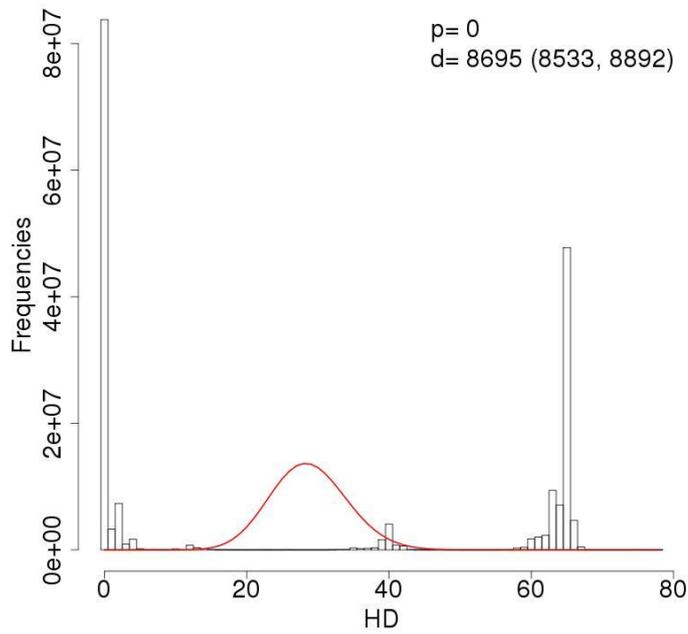
P75_D25_09_2008.



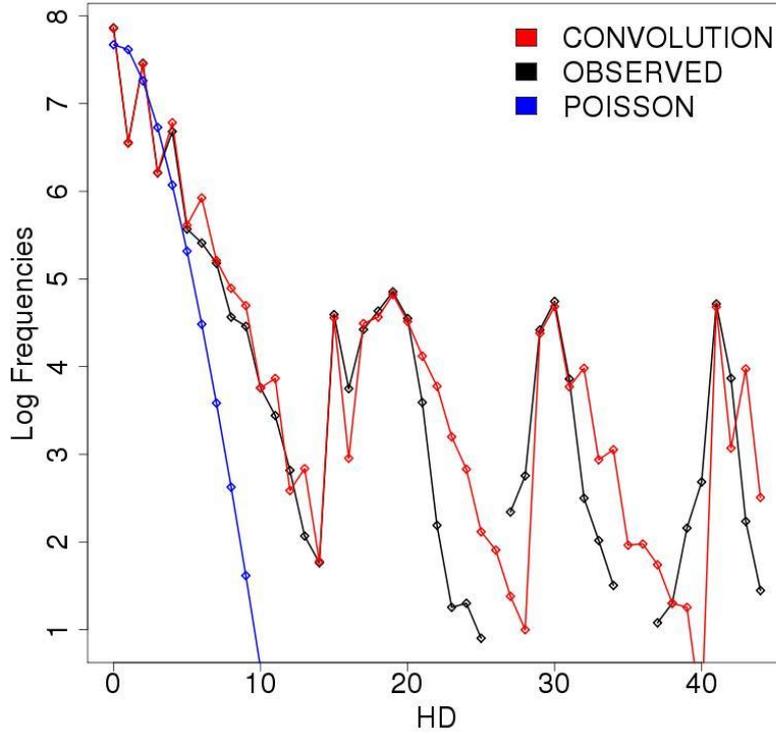
7.12 Poisson Fitter of variants detected in P81



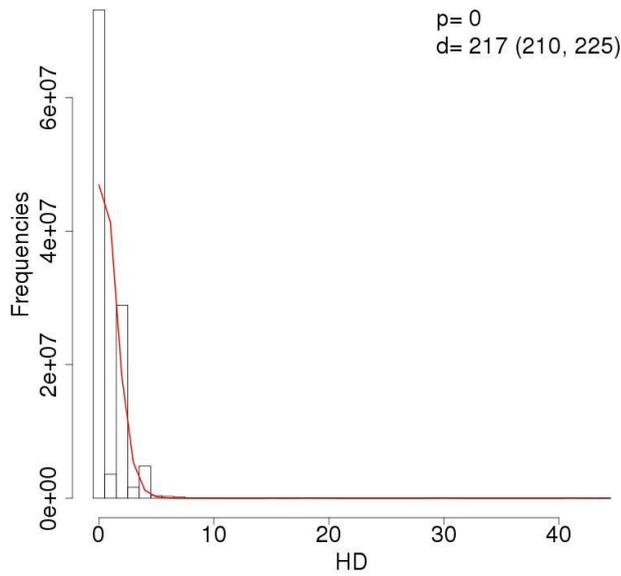
P81_D2005.



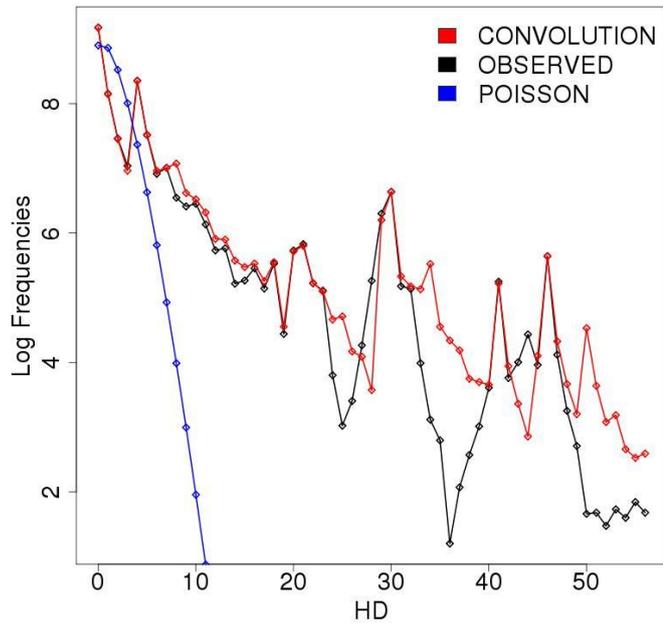
7.13 Poisson Fitter of variants detected in P103



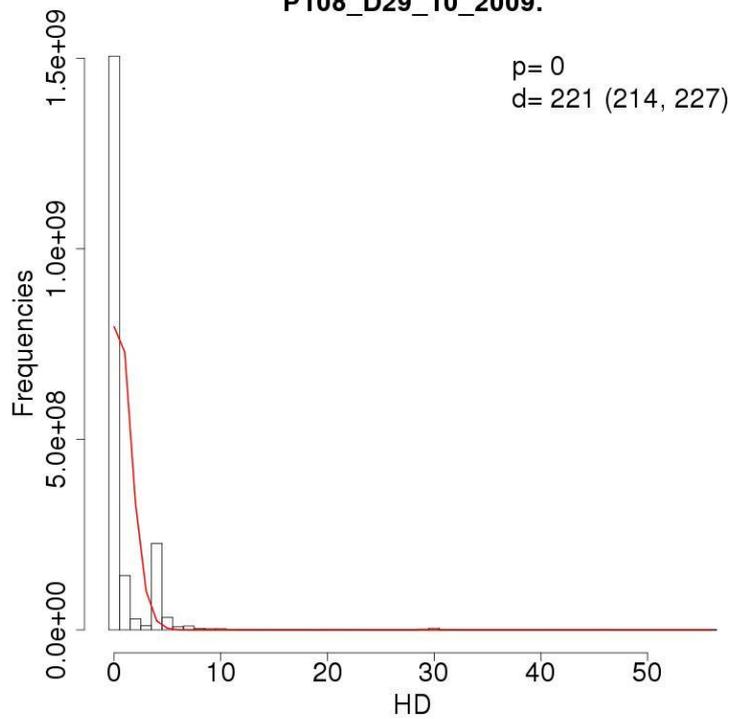
P103_D15_01_2009.



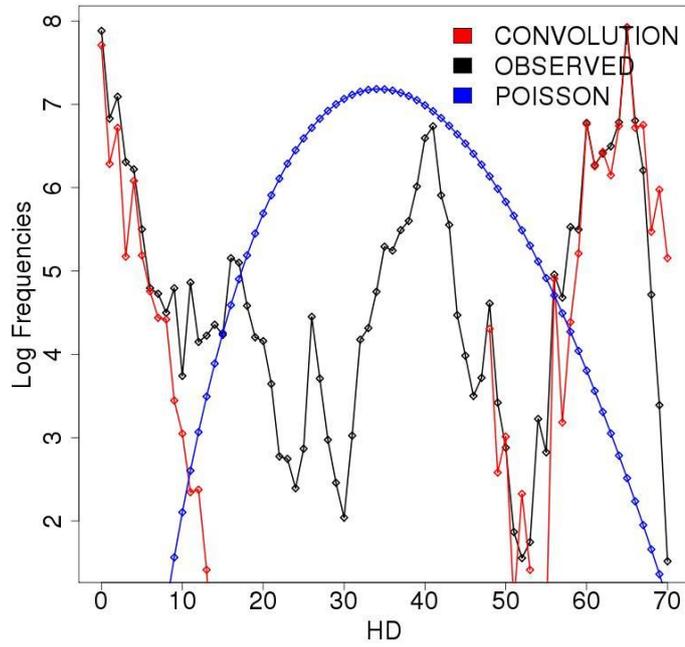
7.14 Poisson Fitter of variants detected in P108



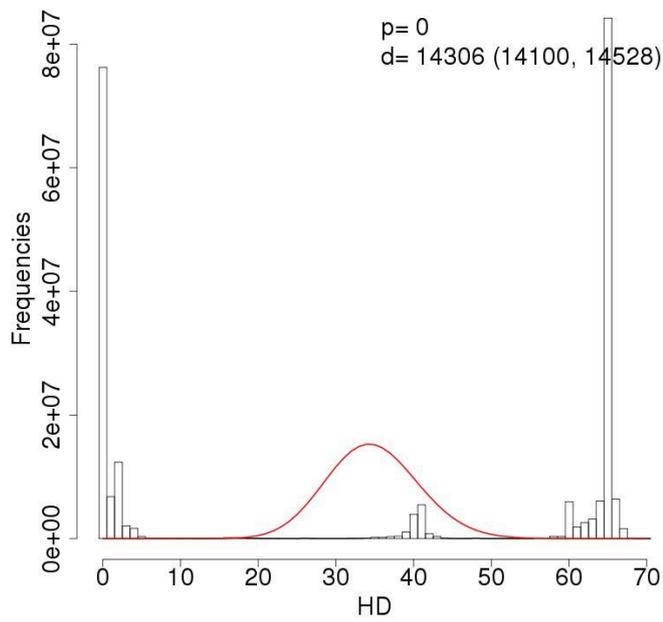
P108_D29_10_2009.



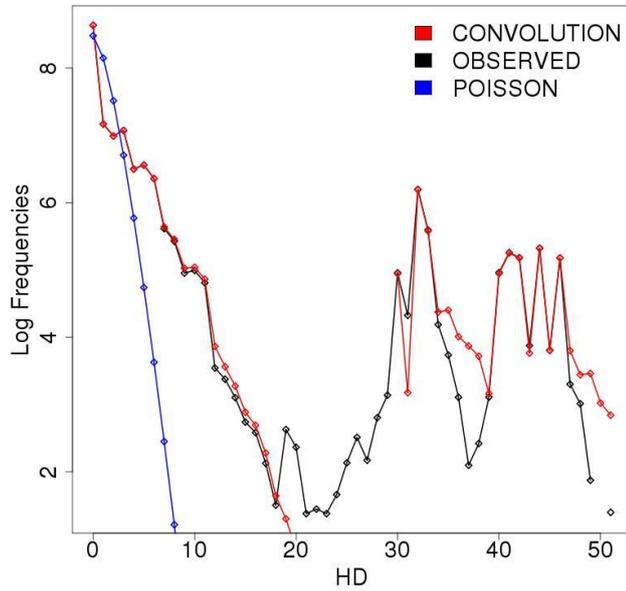
7.15 Poisson Fitter of variants detected in P112



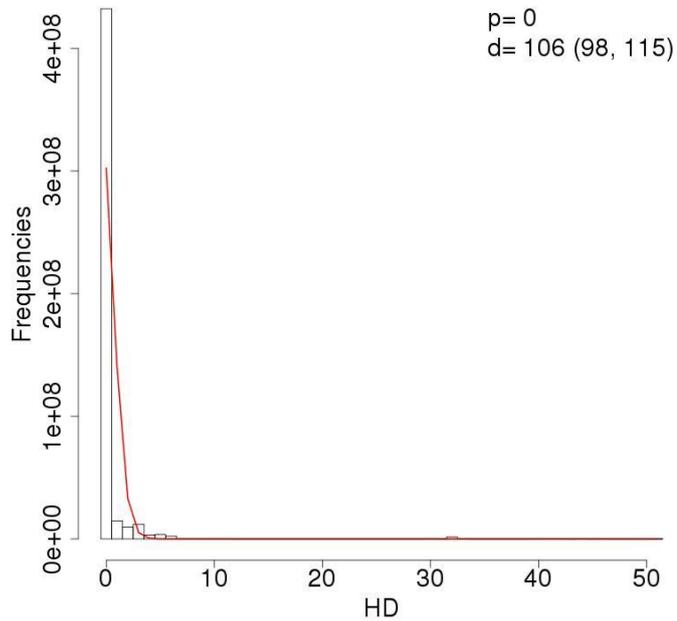
P112_D2008.



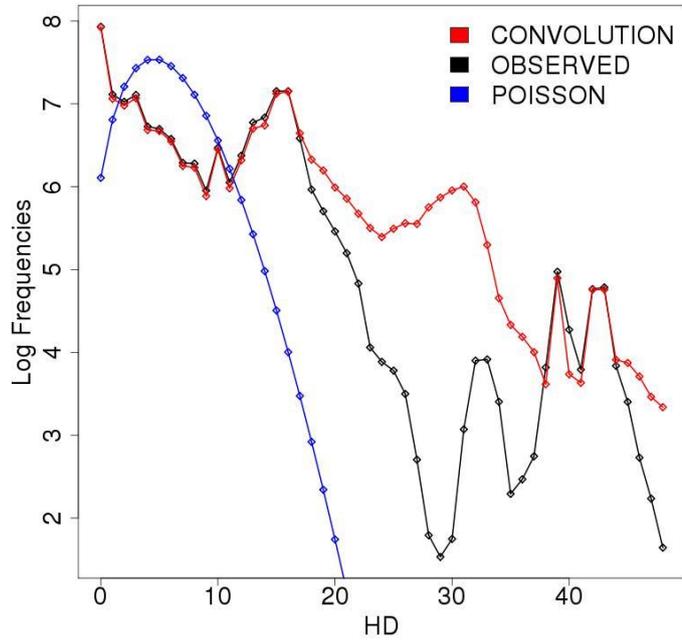
7.16 Poisson Fitter of variants detected in P114



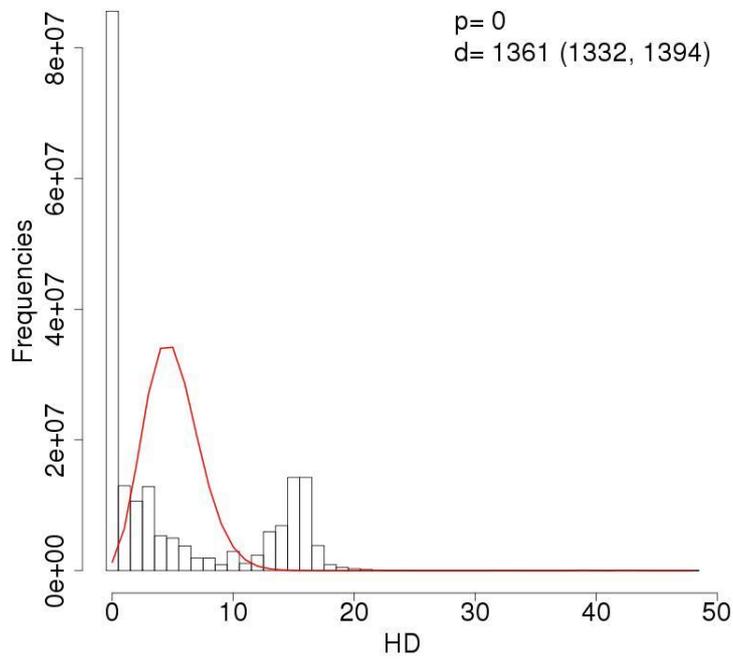
P114_D23_07_2009.



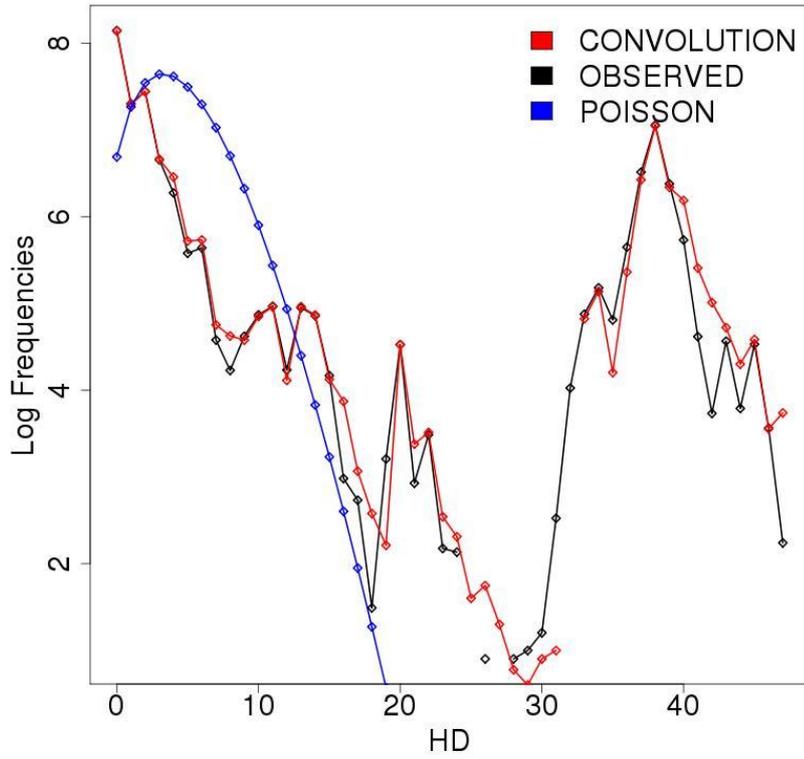
7.17 Poisson Fitter of variants detected in P131



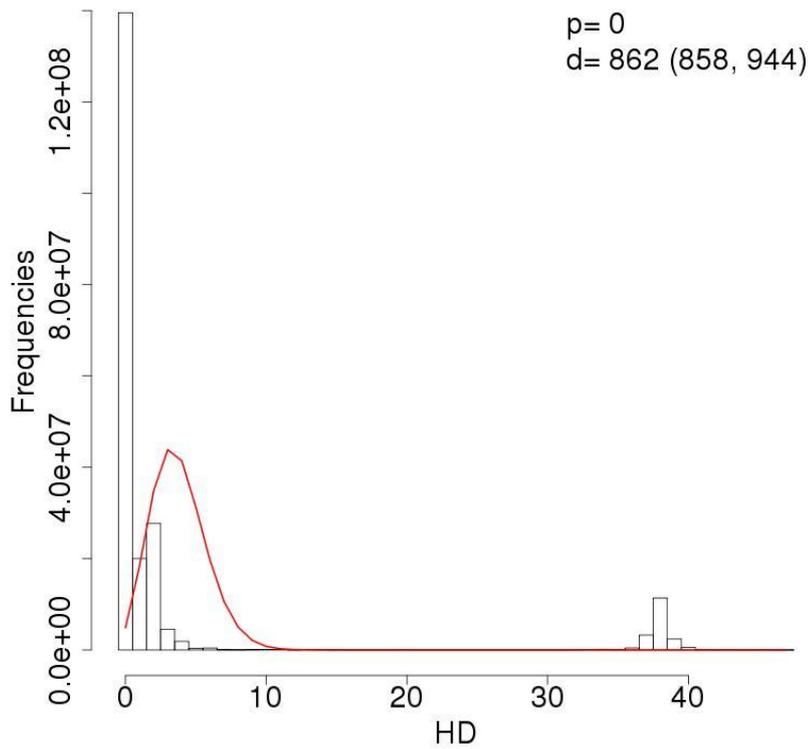
P131_D2009.



7.18 Poisson Fitter of variants detected in P141



P141_D2009.



7.19 Reported resistance mutations to PTV.

Mutation	Data source	Resistance	Sequencing	References
V36A	<i>In vitro</i>	3-fold	Sanger	5
V36L	Clinical	2-fold	Sanger/NGS	5,6
V36M	Clinical/ <i>in vitro</i>	2-fold	Sanger/NGS	5,6,8
V36G	Clinical		Sanger/NGS	7
F43L	<i>In vitro</i>	20-fold	Sanger	5,11
T54S	Clinical/ <i>in vitro</i>	0.4-fold	Sanger/NGS	5,6
T54A	<i>In vitro</i>	1-fold	Sanger	5,8
V55I	<i>In vitro</i>	1-fold	Sanger	5,7
V55A	Clinical/ <i>in vitro</i>	≤ 4-fold	Sanger/NGS	6,8
Y56H	Clinical	NA		12
Q80K	Clinical/ <i>in vitro</i>	3-fold	Sanger/NGS	6,8,11
Q80L	Clinical/ <i>in vitro</i>	2-fold	Sanger/NGS	6
Q80R	Clinical/ <i>in vitro</i>	2-fold	Sanger/NGS	6,8
R155K	Clinical/ <i>in vitro</i>	37-43 fold (1a)	Sanger/NGS	1,2,3,5,7,8,9,10,11,12
R155G	Clinical/ <i>in vitro</i>	16-fold	Sanger/NGS	2,11
R155S	<i>In vitro</i>	7-fold		11
R155T	<i>In vitro</i>	7-fold		8
R155W	<i>In vitro</i>	11-fold	Sanger	5
R155Q	<i>In vitro</i>	≥ 4-fold	Sanger	5,8
A156T	<i>In vitro</i>	17-fold (1a); 7-fold (1b)	Sanger	5,11,12
A156S	<i>In vitro</i>	0.5-fold	Sanger	5,8
D168A	Clinical/ <i>in vitro</i>	59-fold(1a); 27-fold	Sanger	2,8,11,12
D168Y	<i>In vitro</i>	219-fold (1a); 337-fold (1b)	Sanger	5,11,12
D168N	<i>In vitro</i>	13-fold	Sanger	11
D168E	Clinical/ <i>in vitro</i>	14-fold (1a); 4-fold (1b)	Sanger/NGS	5,6,11
D168H	<i>In vitro</i>	62-fold (1a); 76-fold (1b)	Sanger	5,8,11
D168T	<i>In vitro</i>	49-fold	Sanger	5,8
D168F	Clinical/ <i>in vitro</i>	NA	Sanger	9,11
D168K	Clinical	NA		12
V/I170A	Clinical/ <i>in vitro</i>	1-fold	Sanger/NGS	5,6,8
V/I170T	Clinical	NA	Sanger/NGS	6
V36M+R155K	<i>In vitro</i>	79-fold	Sanger	5
Q80K+R155K	<i>In vitro</i>	19-fold	Sanger	5

7.20 References for reported PTV resistance associated variants.

- 1 Abstract #779 AASLD 2012
- 2 Pilot-Matias, T *et al.*, 2011 J Hepatology
- 3 Abstract #1936 AASLD 2014
- 4 Pilot-Matias, T *et al.* 2011 J Hepatology
- 5 Pilot-Matias, T *et al.*, 2015 J Antimicrobial Agents and Chemotherapy
- 6 Beloukas A *et al.*, 2015
- 7 Wyles, D 2013 JID Antiviral Resistance and HCV Therapy
- 8 Halfon P *et al.*, 2011 J Hepatology
- 9 Krishnan P *et al.*, 2015 J Hepatology
- 10 Krishnan P *et al.*, 2015 J Antibacterial Agents and Chemotherapy
- 11 www.accessdata.fda.gov/drugsatfda_docs/label/2014/206619lbl.pdf
- 12 Gentile I *et al.*, Current Medicinal Chemistry 2014

7.21 Reported resistance mutations to SMV.

Mutation	Data source	Resistance	Sequencing	References
V36A	Clinical/ <i>in vitro</i>	2.8 fold	Pyrosequencing	1,2
V36M	Clinical/ <i>in vitro</i>	2 fold	Pyrosequencing	1,2
V36L	Clinical/ <i>in vitro</i>	1.7 fold	Direct	1,8,9
F43I	<i>In vitro</i>	89 fold	RT-PCR	1
F43S	<i>In vitro</i>	12 fold	Direct	1,6
F43V	<i>In vitro</i>	99 fold	RT-PCR	1
T54A	Clinical/ <i>in vitro</i>	0.6 fold	Direct	8
T54S	Clinical/ <i>in vitro</i>	1.2 fold	Direct	8
Q80G	Clinical/ <i>in vitro</i>	1.8 fold	Direct	8
Q80H	<i>In vitro</i>	3.6 fold	RT-PCR	1
Q80K	Clinical/ <i>in vitro</i>	7.7 folds/8 fold	Direct/Deep	1,3-5,8-10
Q80L	Clinical		454 Deep	10
Q80R	Clinical/ <i>in vitro</i>	6.9 folds	Direct/Deep	1,6,8,10
R155G	Clinical/ <i>in vitro</i>	20 folds	Deep	1,10
R155K	Clinical/ <i>in vitro</i>	30 folds	Deep	1,10,11
R155R	<i>In vitro</i>		454 Deep	10
R155T	<i>In vitro</i>	24 folds	RT-PCR	1
A156G	<i>In vitro</i>	16 folds	RT-PCR	1,6
A156T	<i>In vitro</i>	44 folds	RT-PCR	1,6
A156V	Clinical/ <i>in vitro</i>	177 folds	Pyrosequencing	1,2,11
D168A	Clinical/ <i>in vitro</i>	594 folds	Deep	1,3,10,11
D168D	<i>In vitro</i>		454 Deep	10
D168E	Clinical/ <i>in vitro</i>	40 folds	Deep	1,6,9,10
D168G	Clinical/ <i>in vitro</i>	4.4 folds	Deep	1,10
D168H	Clinical/ <i>in vitro</i>	368 folds	Deep	1,3,10
D168I	<i>In vitro</i>	1,807 folds	RT-PCR	1
D168N	Clinical/ <i>in vitro</i>	6.6 folds	Deep	1,10
D168T	Clinical/ <i>in vitro</i>	308 folds	RT-PCR	1,3
D168V	Clinical/ <i>in vitro</i>	2,591 folds	Deep	1,3,9,10
D168Q	Clinical/ <i>in vitro</i>	384/700 folds	Direct	3, 8, 9
D168Y	<i>In vitro</i>	666 folds	RT-PCR	1
V/I170A	Clinical/ <i>in vitro</i>	1.8 folds	RT-PCR	1,3
V/I170L	Clinical/ <i>in vitro</i>			3
V/I170T	Clinical/ <i>in vitro</i>	5.4 folds	RT-PCR	1,3

7.22 References for reported SMV resistance associated variants.

- 1 Lenz *et al.*, 2010 Antimicrobial agents and chemotherapy
- 2 Jabara *et al.*, 2014 Antimicrobial agents
- 3 Proveda *et al.*, 2014 Antiviral Research
- 4 Bichoupan & Dieterich 2014 Drugs
- 5 Pawlotsky J.M., 2014 Gastroenterology
- 6 Schneider & Sarrazin 2014 Antiviral Research
- 8 Palanisamy *et al.*, 2013 Antiviral Research
- 9 Lenz *et al.*, 2013 Journal of Hepatology
- 10 Lenz *et al.*, 2010 Gastroenterology
- 11 Xue *et al.*, 2012 Antiviral Research
- 12 Lawitz *et al.*, 2013 Gastroenterology

7.23 Reported resistance mutations to DCV.

Mutation	Data source	Resistance	Sequencing	Reference
M/F28S	Clinical/ <i>In vitro</i>		Sanger	2,4,7
M28A	<i>In vitro</i>	4,591 fold	Sanger	21
M28G	Clinical		Sanger	13
M28I	Clinical/ <i>In vitro</i>	1.2 fold	Replicon	4,13
M28T	Clinical/ <i>In vitro</i>	390 fold, 683 fold	Sanger	4,11,13,21,22
M28V	Clinical/ <i>In vitro</i>	1.3 fold	Sanger	4,11-13,21
Q30E	Clinical/ <i>In vitro</i>	6217 fold	Sanger	2,4,7,11-13,18,19,21,22
Q30G	Clinical/ <i>In vitro</i>	2,055 fold	Sanger	13,19
Q30H	Clinical/ <i>In vitro</i>	400 folds	Sanger	4,10,11,13,19,21,22
Q/A30K	Clinical/ <i>In vitro</i>	3,732 fold (1a)	Sanger	11,13,14,18,19,21,22
Q30L	<i>In vitro</i>	3.6 folds	Sanger	19
Q30P	<i>In vitro</i>		Sanger	19
Q30R	Clinical/ <i>In vitro</i>	252 fold (1a)	Sanger/NGS	4,8-13,18,19,21,22
L31M	Clinical/ <i>In vitro</i>	878 fold (1a)	Sanger/NGS	1,2,4,10-16,19,21,22
L31V	Clinical/ <i>In vitro</i>	458/710 fold	Sanger/NGS	1,2,4,7,10,12-16,19,21,22
L31R	Clinical		Sanger	13
P32L	Clinical/ <i>In vitro</i>	16 fold	Sanger	2,4,8,10,19,22
H58D	Clinical/ <i>In vitro</i>	483 fold	Sanger	11,13,19,21
H58P	<i>In vitro</i>	2 fold	Sanger	21
E62D	<i>In vitro</i>	2 fold	Replicon	13
Y93C	Clinical/ <i>In vitro</i>	520 fold (1a)	Sanger	4,10-13,19,21,22
Y93H	Clinical/ <i>In vitro</i>	24 fold	Sanger/NGS	1,2,4,7,8,10,11,13-16,18,19,21,22
Y93N	Clinical/ <i>In vitro</i>	28 fold	Sanger	2,4,11-13,18,19,21,22
M28T + Q30H	<i>In vitro</i>	103,767 folds	Replicon system	22
M28T + Q30R	<i>In vitro</i>		Sanger	19
Q30H + Y93H	<i>In vitro</i>	92,217 folds	Replicon system	22
Q30R + L31M	<i>In vitro</i>	9400 folds	Sanger	12
Q30R + L31V	<i>In vitro</i>	>33, 333 folds	Sanger	12
Q30R + H58D	<i>In vitro</i>		Sanger	11
Q30R + E62D	<i>In vitro</i>	High level resistance	Sanger	18
Q30R + Y93C	<i>In vitro</i>	High level resistance	Sanger	11
L31V + H58P	<i>In vitro</i>	High level resistance	Sanger	12
L31V + Y93C	<i>In vitro</i>	High level resistance	Sanger	12

7.24 References for reported DCV resistance associated variants.

- 1 Kosako *et al* 2014 Journal of Viral Hepatitis
- 2 Wang *et al.*, 2014 Antimicrobial Agents Chemotherapy
- 4 Nakamoto *et al.*, 2014 World J Gastroenterol
- 7 Schneider & Sarrazin 2014 Antiviral Research
- 8 Murakami *et al.*, 2014 Antimicrobial Agents Chemotherapy
- 10 Lee C 2013 Drug Des Devel Ther
- 11 Wong *et al.*, 2013 Antimicrob Agents Chemotherapy
- 12 McPhee *et al.*, 2013 Hepatology
- 13 Wang *et al.*, 2013 Antimicrobial Agents Chemotherapy
- 14 Hernandez *et al.*, 2013 J.Clin Virol
- 16 Suzuki *et al.*, 2012 J Clin Virology
- 18 Sun *et al.*, 2012 Hepatology
- 19 Wang *et al.*, 2012 Antimicrobial Agents Chemotherapy
- 20 Nettles *et al.*, 2011 Hepatology
- 21 Fridell *et al.*, 2011 Hepatology
- 22 Fridell *et al.*, 2010 Hepatology
- 23 Bunchorntavakul & Reddy 2015 Aliment Pharmacol Ther
- 24 Yoshimi *et al.*, 2015 J Med Virol
- 25 Bartolini *et al.*, 2015 J Clin Virology
- 26 Miura *et al.*, 2014 Hepatology Research
- 27 Dore *et al.*, 2015 Gastroenterology
- 28 Peres-da-Silva *et al.*, 2014 J Antimicrob Chemother
- 29 Hirotsu Y *et al.*, 2015 Hepatol Int

7.25 Reported resistance associated mutations to LDV

Mutation	Data Source	Resistance	Sequencing	Reference
K24G	Clinical	10-50 fold	Sanger/NGS	3,12
K24N	Clinical	10-50 fold	Sanger/NGS	3,12
K24R	Clinical	2.5-10 fold	Sanger/NGS	3,12,28
K26E	Clinical		Sanger	27
M28A	Clinical	>1000 fold	NGS	12
M28G	Clinical	>1000 fold	NGS	12
M28T	<i>In vitro</i> / Clinical	>30-fold	Sanger/NGS	3,5,11,12,21,32,34
M28V	Clinical		Sanger/NGS	3
Q30E	<i>In vitro</i> / Clinical	5458 fold	Sanger/NGS	3,5,12,21
Q30G	Clinical	100-1000 fold	NGS	12
Q30H	<i>In vitro</i> / Clinical	100-1000 fold	Sanger/NGS	5,8,11,12,21,28,32,34
Q30K	Clinical	>1000 fold	NGS	12
Q30L	Clinical	2.5-10 fold	NGS	12
Q30M	Clinical	183 fold	Sanger	3
Q30R	<i>In vitro</i> / Clinical	>100-fold	Sanger/NGS	3,5,12,21,25,31,32,34
Q30T	Clinical	2.5-10 fold	NGS	12
L31I	Clinical	100-1000 fold	Sanger/NGS	12, 25, 28
L31M	<i>In vitro</i> / Clinical	>100-fold	Sanger/NGS	3,5,7,11,12,16,21,22,23,24,25,28,34,32,35
L31V	Clinical	100-1000 fold	NGS	12,34
P32L	Clinical	100-1000 fold	NGS	12
S38F	Clinical	50-100 fold	Sanger/NGS	12, 25
H58D	Clinical	1177 fold	Sanger/NGS	3,12
P58D	Clinical	100-1000 fold	NGS	12
A92K	Clinical	>1000 fold	NGS	12
A92T	Clinical	10-50 fold	NGS	12
Y93C	<i>In vitro</i> / Clinical	>100-fold	Sanger/NGS	3,5,11,12,21,33
Y93F	Clinical	10-50 fold	NGS	12
Y93H	<i>In vitro</i> / Clinical	≥1000-fold	Sanger/NGS	1,3,5,6,8,11,12,16,21,28,31,32,34,35
Y93N	Clinical	14706 fold	Sanger/NGS	2,3,11,12,31
Y93S	Clinical	>1000 fold	NGS	12,31

7.26 References for reported LDV resistance associated variants.

- 1 Osinusi, A, *et al* 2015 JAMA
- 2 Nakamoto, Shingo *et al.*, 2014 World J Gastroenterol
- 3 Wong KA *et al.*, 2013 JASM
- 4 Kwong HJ *et al.*, 2015 Plos One
- 5 Lawitz EJ *et al.*, 2012 J Hepatol
- 6 Hernandez D *et al.*, 2013 J Clinical Virology
- 7 Lim PJ & Gallay A 2014 Current Opinion in Virology
- 8 Gentile I, *et al.*, 2014 Expert Opinion
- 9 S Sierrei *et al.*, Antiviral Therapy an Infectious Disease 2015
- 10 E. Poveda *et al.*, 2014 Antiviral Research
- 11 Cook, J *et al.*, 2015 CROI 2015
- 12 Dvory-Sobol, H *et al.*, EASL 2015
- 13 Afdhal N *et al.*, 2014 N ENGL J MED
- 14 Link JO *et al.*, 2013 ACS Publications
- 15 Lindström *et al.*, 2015 Infectious Diseases
- 16 Afdhal N *et al.*, 2014 N Engl J Med
- 17 Kowdley KV *et al.*, 2014 N Engl J Med
- 18 Kohler *et al.*, 2014 Infection and Drug Resistance Dovepress
- 19 Wyles DL *et al.*, 2014 Hepatology
- 20 Noel B *et al.*, 2015 Dovepress drug design, development and therapy
- 21 Cheng G *et al.*, 2012 EASL
- 22 Sarrazin C *et al.*, 2014 AASLD
- 23 Larousse JA *et al.*, 2015 Virology Journal
- 24 Paolucci S *et al.*, 2013 Virology Journal
- 25 Wong KA *et al.*, 2012 J Hepatology
- 26 Hebner C *et al.*, 2013 AASLD
- 27 Kitrinis *et al.*, 2014 AASLD
- 28 Svarovskaia E *et al.*, 2015 EASL
- 29 Sarrazin C *et al.*, 2015 EASL P0773
- 30 Mizokami M *et al.*, 2015 The Lancet Infect Dis
- 31 Wyles D *et al.*, 2015 EASL
- 32 Lawitz E *et al.*, 2015 EASL
- 33 Abergel A *et al.*, 2015 EASL
- 34 Bourliere M *et al.*, 2015 Lancet Infect Dis
- 35 Gane E *et al.*, 2015 ISVHLD
- 36 Lawitz E *et al.*, 2014 APASL

7.27 Reported resistance associated mutations to OBV

Mutation	Data Source	Resistance	Sequencing	References
L28M	<i>Clinical/In vitro</i>	2-fold (1b)	Sanger	1,11,25
L28T	<i>In vitro</i>	661-fold	Sanger	1
M28T	<i>Clinical/In vitro</i>	8965-fold	Sanger	1,3,6,9,12,19,22,24,34
M28V	<i>Clinical/In vitro</i>	58-fold	Sanger	1,3,7,12,22,24,34
Q30E	Clinical		Sanger	1,12
Q30H	<i>In vitro</i>	3-fold	Sanger	1
Q30K	Clinical		Sanger	12
Q30R	<i>Clinical/In vitro</i>	800-fold	Sanger	1,3,6,7,8,9,12,16,18,22,24,34
L31F	<i>In vitro</i>	10-fold	Sanger	1
L31M	<i>Clinical/In vitro</i>	2-fold	Sanger	1,3,6,24,25
L31V	<i>Clinical/In vitro</i>	8-fold (1b)	Sanger	1,18,24
H58D	<i>In vitro</i>	243-fold	Sanger	1
Y93C	<i>Clinical/In vitro</i>	1675-fold	Sanger	1,18,24,25
Y93H	<i>Clinical/In vitro</i>	41383-fold	Sanger	1,3,6,7,8,11,13,16,18,19,22,24,25,34
Y93L	<i>In vitro</i>		Sanger	14
Y93N	<i>Clinical/In vitro</i>	66740-fold	Sanger	1,18,21,24
Y93S	<i>Clinical/In vitro</i>	1013-fold	Sanger	1
L28M+L31F	<i>In vitro</i>	569-fold	Sanger	1
L28V+L31F	<i>In vitro</i>	2170-fold	Sanger	1

7.28 References for reported OBV resistance associated variants.

- 1 Krishnan P *et al.*, 2015 J Antimicrobial Agents and Chemotherapy
- 2 De Goey *et al.*, 2014 J Med Chem
- 3 Stirnimann G 2014 Expert Opin Pharmacother
- 4 Hezode C *et al.*, 2014 Liver Congr. Oral Present
- 5 Kwo P *et al.*, 2014 Liver Congr. Oral Present
- 6 Feld JJ, *et al.*, 2014 N Engl J Med
- 7 Zeuzem S, *et al.*, 2014 N Engl J Med
- 8 Poordad F, *et al.*, 2014 N Engl J Med
- 9 Ferenci P, *et al.*, 2014 N Engl J Med
- 10 Krishnan P *et al.*, 2014 J Hepatology
- 11 Schnell G *et al.*, 2014 AASLD
- 12 Krishnan P *et al.*, 2015 J Hepatology EASL Conference Oral Presentation
- 13 Andreone P *et al.*, 2014 J gastroenterology
- 14 http://www.accessdata.fda.gov/drugsatfda_docs/label/2014/206619lbl.pdf
- 15 Aisso Larousse *et al.*, 2015 Virology journal
- 16 Linstrom Ida *et al.*, 2015 Infectious Diseases
- 17 Gentile I *et al.*, 2014 Expert Rev. Anti Infect. Ther
- 18 Krishnan P *et al.*, 2015 J Antimicrobial Agents and Chemotherapy
- 19 Sulkowski M.S *et al.*, 2015 J American Medical Association
- 20 Krishnan P *et al.*, 2014 AASLD
- 21 Leenheer D *et al.*, 2014 Antiviral Therapy International Workshop
- 22 Lawitz E *et al.*, 2012 J Hepatology
- 23 Pilot-Matias T *et al.*, 2012 J Hepatology
- 24 Krishnan P *et al.*, 2014 J Hepatology
- 26 Stanislas Pol *et al.*, 2014 AASLD
- 27 Chayama K *et al.*, 2015 APASL
- 28 Wyles D *et al.*, 2014 AASLD
Eron JJ *et al.*, 2014 ICAAC; Sulkowski MS *et al.* 2014 International AIDS
29 Conference
- 30 Bernstein *et al.*, 2014 ICAAC
- 31 Poordad F *et al.*, 2014 EASL
- 32 Feld JJ *et al.*, 2014 EASL
- 33 Sullivan JG *et al.*, 2012 EASL
- 34 Klibanov OM *et al.*, 2015 Annals of Pharmacotherapy

7.29 Reported resistance-associated mutations to SOF

Mutation	Data source	Resistance	Sequencing	Reference
L159F	Clinical/ <i>In vitro</i>	1.9 fold (1a)	Sanger/NGS	1,2,3,4,7
S282R	Clinical		Sanger/NGS	4
S282T	Clinical/ <i>In vitro</i>	13.5 fold	Sanger/NGS	1,2,3,4,5,6,7,8
L320F	Clinical/ <i>In vitro</i>	2.7 fold	Sanger/NGS	4,7
V321A	Clinical		Sanger/NGS	1,2
V321I	Clinical		Sanger/NGS	3,4
F415Y	<i>In vitro</i>	1.3 fold	Sanger	8
L159F + S282T	Clinical/ <i>In vitro</i>	30 fold (1a)	Sanger	7
L159F + L316N	<i>In vitro</i>		Sanger/NGS	1,3,4
L159F + L320F	<i>In vitro</i>	4.3 fold (1a)	Sanger/NGS	1,3,7
S282T + L320F	<i>In vitro</i>	41 fold (1a)	Sanger	7
S282T + I434M	<i>In vitro</i>		Sanger	8

7.30 References for reported SOF resistance associated variants.

- 1 Tong & Kwong 2014 Clin Infect Disease
- 2 Svarovskaia *et al.*, 2014 Clin Infect Dis
- 3 Margeridon-Thermet *et al.*, 2014 PLoSOne
- 4 Donaldson *et al.*, 2014 Hepatology
- 5 Di Maio *et al.*, 2014 Antimicrobial Agents Chemotherapy
- 6 Schneider & Sarrazin 2014 Antiviral Research
- 7 Tong *et al.*, 2014 J. Infect Disease
- 8 Lam *et al.*, 2012 Antimicrobial Agents Chemotherapy

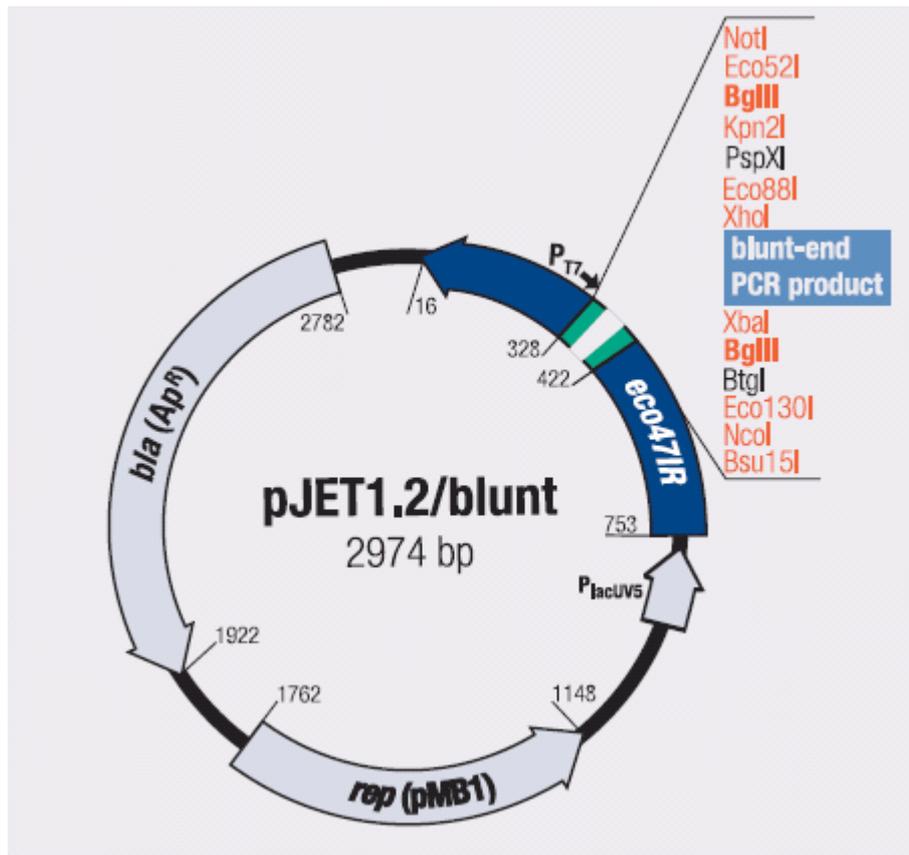
7.31 Reported resistance-associated mutations to DSV

Mutation	Data source	Resistance	Sequencing	Reference
S556G	Clinical/ <i>In vitro</i>	30-84-fold	Sanger	1,2,3,4,5,6,8,12,13,14,15,17, 18,19,20,21,22,23,25
S556N	Clinical	29-fold	Sanger	1,2
S556R	Clinical/ <i>In vitro</i>		Sanger	1,6,15
S556G	<i>In vitro</i>		Sanger	2,6
S565F	<i>In vitro</i>	17-fold	Sanger	2
M414T	Clinical/ <i>In vitro</i>	32-61-fold	Sanger	1,2,3,4,5,6,12,14,15
M414I	Clinical/ <i>In vitro</i>	8-fold	Sanger	1,6,7,14,15
M414L	Clinical		Sanger	1,14
M414V	Clinical/ <i>In vitro</i>	6-fold	Sanger	3,5,12,14
D559G	Clinical/ <i>In vitro</i>		Sanger	1,2,3,6,12,14,15
D559N	Clinical		Sanger	15,18
C316Y	Clinical/ <i>In vitro</i>	1472 fold	Sanger	1,2,3,4,5,6,7,10,12,14,15
C316K	Clinical		Sanger	1
C316N	Clinical/ <i>In vitro</i>	5-fold (1b)	Sanger	1,2,10,19,21,25
C316W	clinical		Sanger	1
C316H	<i>In vitro</i>		Sanger	6
E446K/Q	Clinical/ <i>In vitro</i>		Sanger	1,6,15
A553T	Clinical/ <i>In vitro</i>	152-fold	Sanger	1,6,15
A553V	<i>In vitro</i>	24-120-fold	Sanger	2,5,6
A553D	Clinical		Sanger	15
G554S	Clinical/ <i>In vitro</i>	198-fold	Sanger	1,3,6,14,15,16,24
G558R	Clinical/ <i>In vitro</i>		Sanger	1,15
S368A	<i>In vitro</i>		Sanger	1
S368T	<i>In vitro</i>	65-fold (1b)	Sanger	2,3,5,6
N411S	<i>In vitro</i>	84-fold (1b)	Sanger	1,2,5,6
C445F	<i>In vitro</i>	16-fold	Sanger	1,2
S556G	<i>In vitro</i>	11-fold	Sanger	1,2,5
A395G	<i>In vitro</i>	10-fold (1a)	Sanger	2,5
N444K	<i>In vitro</i>	23-fold (1a)	Sanger	2,5
Y448C	Clinical/ <i>In vitro</i>	400-940- fold	Sanger	1,2,3,4,5,6,12
Y448H	Clinical/ <i>In vitro</i>	250-975- fold	Sanger	1,2,3,4,5,6,12
C451R	<i>In vitro</i>	1-fold	Sanger	1,2,5
C451S	<i>In vitro</i>	16-fold	Sanger	2
C451G/T	<i>In vitro</i>		Sanger	2
I585V	<i>In vitro</i>	16-fold	Sanger	2
Y561H	Clinical/ <i>In vitro</i>		Sanger	6,18
P495A	<i>In vitro</i>	2.4-fold	Sanger	2
P495S	<i>In vitro</i>	1.1-fold	Sanger	2

7.32 References for reported DSV resistance associated variants.

- 1 Krishnan P *et al.*, 2015 J AAC
- 2 Kati W *et al.*, 2014 J Antimicrobial Agents and Chemotherapy
- 3 Trivella JP 2015 Expert Opin. Pharmacother
- 4 Rodriguez-Torres, M *et al.*, 2009 AASLD 2009
- 5 Koev G *et al.*, 2009 EASL 44th annual meeting
- 6 http://www.accessdata.fda.gov/drugsatfda_docs/label/2014/206619lbl.pdf
- 7 Poordad F *et al.*, 2014 NEJM
- 8 Ferenci P *et al.*, 2014 NEJM
- 9 Pilot-Matias T *et al.*, 2012 Poster J Hepatology
- 10 Plaza Z *et al.*, 2011 J Antimicrob Chemother
- 11 Middleton T *et al.*, 2011 J Hepatology
- 12 Middleton T *et al.*, 2010 J Hepatology
- 13 Krishnan P 2014 AASLD
- 14 Pilot-Matias T *et al.*, 2012 AASLD
- 15 Krishnan P EASL 2015
- 16 Mantry P *et al.*, AASLD 2014
- 17 Wyles D *et al.*, 2014 AASLD
- 18 Feld J.J *et al.*, 2014 EASL
- 19 Zeuzem S *et al.*, 2014 EASL
- 20 Feld J.J *et al.*, 2014 EASL
- 21 Zeuzem S *et al.*, 2014 EASL
- 22 Sulkowski MS *et al.*, 2014 International AIDS Conference
- 23 Sulkowski MS *et al.*, 2015 JAMA
- 24 Kwo P *et al.*, 2014 N Engl J Med
- 25 Andreone P *et al.*, 2014 Gastroenterology

7.33 pJET1.2/blunt plasmid map



Bibliography

- EASL Recommendations on Treatment of Hepatitis C 2015. *Journal of Hepatology*, 63, 199-236.
- AASLD/IDSA 2015. Hepatitis C guidance: AASLD-IDSA recommendations for testing, managing, and treating adults infected with hepatitis C virus. *Hepatology*, 62, 932-54.
- ABBATE, I., LO IACONO, O., DI STEFANO, R., CAPIELLO, G., GIRARDI, E., LONGO, R., FERRARO, D., ANTONUCCI, G., DI MARCO, V., SOLMONE, M., CRAXI, A., IPPOLITO, G. & CAPOBIANCHI, M. R. 2004. HVR-1 quasispecies modifications occur early and are correlated to initial but not sustained response in HCV-infected patients treated with pegylated- or standard-interferon and ribavirin. *J Hepatol*, 40, 831-6.
- ABDELRAHMAN, T., HUGHES, J., MAIN, J., MCLAUCHLAN, J., THURSZ, M. & THOMSON, E. 2015. Next-generation sequencing sheds light on the natural history of hepatitis C infection in patients who fail treatment. *Hepatology*, 61, 88-97.
- ABRAHAM, G. M. & SPOONER, L. M. 2014. Sofosbuvir in the Treatment of Chronic Hepatitis C: New Dog, New Tricks. *Clin Infect Dis*.
- ADIWIJAYA, B. S., HERRMANN, E., HARE, B., KIEFFER, T., LIN, C., KWONG, A. D., GARG, V., RANDLE, J. C., SARRAZIN, C., ZEUZEM, S. & CARON, P. R. 2010. A multi-variant, viral dynamic model of genotype 1 HCV to assess the in vivo evolution of protease-inhibitor resistant variants. *PLoS Comput Biol*, 6, e1000745.
- AFDHAL, N., REDDY, K. R., NELSON, D. R., LAWITZ, E., GORDON, S. C., SCHIFF, E., NAHASS, R., GHALIB, R., GITLIN, N., HERRING, R., LALEZARI, J., YOUNES, Z. H., POCKROS, P. J., DI BISCEGLIE, A. M., ARORA, S., SUBRAMANIAN, G. M., ZHU, Y., DVORY-SOBOL, H., YANG, J. C., PANG, P. S., SYMONDS, W. T., MCHUTCHISON, J. G., MUIR, A. J., SULKOWSKI, M., KWO, P. & INVESTIGATORS, I. O. N. 2014a. Ledipasvir and sofosbuvir for previously treated HCV genotype 1 infection. *N Engl J Med*, 370, 1483-93.
- AFDHAL, N., ZEUZEM, S., KWO, P., CHOJKIER, M., GITLIN, N., PUOTI, M., ROMERO-GOMEZ, M., ZARSKI, J.-P., AGARWAL, K., BUGGISCH, P., FOSTER, G. R., BRAEU, N., BUTI, M., JACOBSON, I. M., SUBRAMANIAN, G. M., DING, X., MO, H., YANG, J. C., PANG, P. S., SYMONDS, W. T., MCHUTCHISON, J. G., MUIR, A. J., MANGIA, A., MARCELLIN, P. & INVESTIGATORS, I. O. N. 2014b. Ledipasvir and Sofosbuvir for

Untreated HCV Genotype 1 Infection. *New England Journal of Medicine*, 370, 1889-1898.

- AKUTA, N., SUZUKI, F., SEKO, Y., KAWAMURA, Y., SEZAKI, H., SUZUKI, Y., HOSAKA, T., KOBAYASHI, M., HARA, T., KOBAYASHI, M., SAITOH, S., ARASE, Y., IKEDA, K. & KUMADA, H. 2013. Emergence of telaprevir-resistant variants detected by ultra-deep sequencing after triple therapy in patients infected with HCV genotype 1. *J Med Virol*, 85, 1028-36.
- AKUTA, N., SUZUKI, F., SEZAKI, H., SUZUKI, Y., HOSAKA, T., KOBAYASHI, M., KOBAYASHI, M., SAITOH, S., IKEDA, K. & KUMADA, H. 2014. Evolution of simeprevir-resistant variants over time by ultra-deep sequencing in HCV genotype 1b. *J Med Virol*, 86, 1314-22.
- ALI, S., LEVEQUE, V., LE POGAM, S., MA, H., PHILIPP, F., INOCENCIO, N., SMITH, M., ALKER, A., KANG, H., NAJERA, I., KLUMPP, K., SYMONS, J., CAMMACK, N. & JIANG, W. R. 2008. Selected replicon variants with low-level in vitro resistance to the hepatitis C virus NS5B polymerase inhibitor PSI-6130 lack cross-resistance with R1479. *Antimicrob Agents Chemother*, 52, 4356-69.
- ALLEN, B., KON, M. & BAR-YAM, Y. 2009. A new phylogenetic diversity measure generalizing the shannon index and its application to phyllostomid bats. *Am Nat*, 174, 236-43.
- ALTER, M. J. 2006. Epidemiology of viral hepatitis and HIV co-infection. *J Hepatol*, 44, S6-9.
- ANDREONE, P., COLOMBO, M. G., ENEJOSA, J. V., KOKSAL, I., FERENCI, P., MAIERON, A., MULLHAUPT, B., HORSMANS, Y., WEILAND, O., REESINK, H. W., RODRIGUES, L., JR., HU, Y. B., PODSADECKI, T. & BERNSTEIN, B. 2014. ABT-450, ritonavir, ombitasvir, and dasabuvir achieves 97% and 100% sustained virologic response with or without ribavirin in treatment-experienced patients with HCV genotype 1b infection. *Gastroenterology*, 147, 359-365 e1.
- APPEL, N., PIETSCHMANN, T. & BARTENSCHLAGER, R. 2005. Mutational analysis of hepatitis C virus nonstructural protein 5A: potential role of differential phosphorylation in RNA replication and identification of a genetically flexible domain. *J Virol*, 79, 3187-94.
- ASSELAH, T. & MARCELLIN, P. 2011. New direct-acting antivirals' combination for the treatment of chronic hepatitis C. *Liver Int*, 31 Suppl 1, 68-77.

- ASTROVSKAYA, I., TORK, B., MANGUL, S., WESTBROOKS, K., MANDOIU, I., BALFE, P. & ZELIKOVSKY, A. 2011. Inferring viral quasispecies spectra from 454 pyrosequencing reads. *BMC Bioinformatics*, 12 Suppl 6, S1.
- BACON, B. R. & KHALID, O. 2012. Triple therapy with boceprevir for HCV genotype 1 infection: phase III results in relapsers and nonresponders. *Liver Int*, 32 Suppl 1, 51-3.
- BAO, S., JIANG, R., KWAN, W., WANG, B., MA, X. & SONG, Y. Q. 2011. Evaluation of next-generation sequencing software in mapping and assembly. *J Hum Genet*, 56, 406-14.
- BARIL, M. & BRAKIER-GINGRAS, L. 2005. Translation of the F protein of hepatitis C virus is initiated at a non-AUG codon in a +1 reading frame relative to the polyprotein. *Nucleic Acids Res*, 33, 1474-86.
- BARNARD, R. J., HOWE, J. A., OGERT, R. A., ZEUZEM, S., POORDAD, F., GORDON, S. C., RALSTON, R., TONG, X., SNIUKIENE, V., STRIZKI, J., RYAN, D., LONG, J., QIU, P., BRASS, C. A., ALBRECHT, J., BURROUGHS, M., VUOCOLO, S. & HAZUDA, D. J. 2013. Analysis of boceprevir resistance associated amino acid variants (RAVs) in two phase 3 boceprevir clinical studies. *Virology*, 444, 329-36.
- BARTELS, D. J., SULLIVAN, J. C., ZHANG, E. Z., TIGGES, A. M., DORRIAN, J. L., DE MEYER, S., TAKEMOTO, D., DONDERO, E., KWONG, A. D., PICCHIO, G. & KIEFFER, T. L. 2013. Hepatitis C virus variants with decreased sensitivity to direct-acting antivirals (DAAs) were rarely observed in DAA-naive patients prior to treatment. *J Virol*, 87, 1544-53.
- BARTELS, D. J., ZHOU, Y., ZHANG, E. Z., MARCIAL, M., BYRN, R. A., PFEIFFER, T., TIGGES, A. M., ADIWIJAYA, B. S., LIN, C., KWONG, A. D. & KIEFFER, T. L. 2008. Natural prevalence of hepatitis C virus variants with decreased sensitivity to NS3.4A protease inhibitors in treatment-naive subjects. *J Infect Dis*, 198, 800-7.
- BARTENSCHLAGER, R., FRESE, M. & PIETSCHMANN, T. 2004. Novel insights into hepatitis C virus replication and persistence. *Adv Virus Res*, 63, 71-180.
- BARTENSCHLAGER, R. & LOHMANN, V. 2000. Replication of hepatitis C virus. *J Gen Virol*, 81, 1631-48.
- BARTENSCHLAGER, R. & LOHMANN, V. 2001. Novel cell culture systems for the hepatitis C virus. *Antiviral Res*, 52, 1-17.

- BARTENSCHLAGER, R., LOHMANN, V., WILKINSON, T. & KOCH, J. O. 1995. Complex formation between the NS3 serine-type proteinase of the hepatitis C virus and NS4A and its importance for polyprotein maturation. *J Virol*, 69, 7519-28.
- BARTOLINI, B., GIOMBINI, E., ZACCARO, P., SELLERI, M., ROZERA, G., ABBATE, I., COMANDINI, U. V., IPPOLITO, G., SOLMONE, M. & CAPOBIANCHI, M. R. 2013. Extent of HCV NS3 protease variability and resistance-associated mutations assessed by next generation sequencing in HCV monoinfected and HIV/HCV coinfecting patients. *Virus Res*, 177, 205-8.
- BARTOSCH, B., DUBUISSON, J. & COSSET, F. L. 2003a. Infectious hepatitis C virus pseudo-particles containing functional E1-E2 envelope protein complexes. *J Exp Med*, 197, 633-42.
- BARTOSCH, B., VITELLI, A., GRANIER, C., GOUJON, C., DUBUISSON, J., PASCALE, S., SCARSELLI, E., CORTESE, R., NICOSIA, A. & COSSET, F. L. 2003b. Cell entry of hepatitis C virus requires a set of co-receptors that include the CD81 tetraspanin and the SR-B1 scavenger receptor. *J Biol Chem*, 278, 41624-30.
- BARZON, L., LAVEZZO, E., MILITELLO, V., TOPPO, S. & PALU, G. 2011. Applications of next-generation sequencing technologies to diagnostic virology. *Int J Mol Sci*, 12, 7861-84.
- BEERENWINKEL, N. & ZAGORDI, O. 2011. Ultra-deep sequencing for the analysis of viral populations. *Curr Opin Virol*, 1, 413-8.
- BENEDET, M., ADACHI, D., WONG, A., WONG, S., PABBARAJU, K., TELLIER, R. & TANG, J. W. 2014. The need for a sequencing-based assay to supplement the Abbott m2000 RealTime HCV Genotype II assay: A 1 year analysis. *J Clin Virol*.
- BENTLEY, D. R., BALASUBRAMANIAN, S., SWERDLOW, H. P., SMITH, G. P., MILTON, J., BROWN, C. G., HALL, K. P., EVERS, D. J., BARNES, C. L., BIGNELL, H. R., BOUTELL, J. M., BRYANT, J., CARTER, R. J., CHEETHAM, R. K., COX, A. J., ELLIS, D. J., FLATBUSH, M. R., GORMLEY, N. A., HUMPHRAY, S. J., IRVING, L. J., KARBELASHVILI, M. S., KIRK, S. M., LI, H., LIU, X. H., MAISINGER, K. S., MURRAY, L. J., OBRADOVIC, B., OST, T., PARKINSON, M. L., PRATT, M. R., RASOLONJATOVO, I. M. J., REED, M. T., RIGATTI, R., RODIGHIERO, C., ROSS, M. T., SABOT, A., SANKAR, S. V., SCALLY, A., SCHROTH, G. P., SMITH, M. E., SMITH, V. P., SPIRIDOU, A., TORRANCE, P. E., TZONEV, S. S., VERMAAS, E. H., WALTER, K., WU, X. L., ZHANG, L., ALAM, M. D., ANASTASI, C., ANIEBO, I. C., BAILEY, D. M. D., BANCARZ, I. R., BANERJEE, S., BARBOUR, S. G., BAYBAYAN, P. A., BENOIT, V. A.,

- BENSON, K. F., BEVIS, C., BLACK, P. J., BOODHUN, A., BRENNAN, J. S., BRIDGHAM, J. A., BROWN, R. C., BROWN, A. A., BUERMANN, D. H., BUNDU, A. A., BURROWS, J. C., CARTER, N. P., CASTILLO, N., CATENAZZI, M. C. E., CHANG, S., COOLEY, R. N., CRAKE, N. R., DADA, O. O., DIAKOU MAKOS, K. D., DOMINGUEZ-FERNANDEZ, B., EARNSHAW, D. J., EGBUJOR, U. C., ELMORE, D. W., ETCHIN, S. S., EWAN, M. R., FEDURCO, M., FRASER, L. J., FAJARDO, K. V. F., FUREY, W. S., GEORGE, D., GIETZEN, K. J., GODDARD, C. P., GOLDA, G. S., GRANIERI, P. A., GREEN, D. E., GUSTAFSON, D. L., HANSEN, N. F., HARNISH, K., HAUDENSCHILD, C. D., HEYER, N. I., HIMMS, M. M., HO, J. T., HORGAN, A. M., et al. 2008. Accurate whole human genome sequencing using reversible terminator chemistry. *Nature*, 456, 53-59.
- BICA, I., MCGOVERN, B., DHAR, R., STONE, D., MCGOWAN, K., SCHEIB, R. & SNYDMAN, D. R. 2001. Increasing mortality due to end-stage liver disease in patients with human immunodeficiency virus infection. *Clin Infect Dis*, 32, 492-7.
- BIEBRICHER, C. K. & EIGEN, M. 2006. What is a quasispecies? *Curr Top Microbiol Immunol*, 299, 1-31.
- BLACKARD, J. T., HIASA, Y., SMEATON, L., JAMIESON, D. J., RODRIGUEZ, I., MAYER, K. H. & CHUNG, R. T. 2007. Compartmentalization of hepatitis C virus (HCV) during HCV/HIV coinfection. *J Infect Dis*, 195, 1765-73.
- BLACKARD, J. T., MA, G., SENGUPTA, S., MARTIN, C. M., POWELL, E. A., SHATA, M. T. & SHERMAN, K. E. 2014. Evidence of distinct populations of hepatitis C virus in the liver and plasma of patients co-infected with HIV and HCV. *J Med Virol*, 86, 1332-41.
- BLACKARD, J. T. & SHERMAN, K. E. 2007. Hepatitis C virus coinfection and superinfection. *J Infect Dis*, 195, 519-24.
- BLIGHT, K. J., KOLYKHALOV, A. A. & RICE, C. M. 2000. Efficient initiation of HCV RNA replication in cell culture. *Science*, 290, 1972-4.
- BLIGHT, K. J., MCKEATING, J. A., MARCOTRIGIANO, J. & RICE, C. M. 2003. Efficient replication of hepatitis C virus genotype 1a RNAs in cell culture. *J Virol*, 77, 3181-90.
- BLIGHT, K. J., MCKEATING, J. A. & RICE, C. M. 2002. Highly permissive cell lines for subgenomic and genomic hepatitis C virus RNA replication. *J Virol*, 76, 13001-14.
- BLIGHT, K. J. & NORGARD, E. A. 2006. HCV Replicon Systems. In: TAN, S. L. (ed.) *Hepatitis C Viruses: Genomes and Molecular Biology*. Norfolk (UK).

- BOURLIERE, M., BRONOWICKI, J. P., DE LEDINGHEN, V., HEZODE, C., ZOULIM, F., MATHURIN, P., TRAN, A., LARREY, D. G., RATZIU, V., ALRIC, L., HYLAND, R. H., JIANG, D., DOEHLE, B., PANG, P. S., SYMONDS, W. T., SUBRAMANIAN, G. M., MCHUTCHISON, J. G., MARCELLIN, P., HABERSETZER, F., GUYADER, D., GRANGE, J. D., LOUSTAUD-RATTI, V., SERFATY, L., METIVIER, S., LEROY, V., ABERGEL, A. & POL, S. 2015. Ledipasvir-sofosbuvir with or without ribavirin to treat patients with HCV genotype 1 infection and cirrhosis non-responsive to previous protease-inhibitor therapy: a randomised, double-blind, phase 2 trial (SIRIUS). *Lancet Infect Dis*, 15, 397-404.
- BOURLIERE, M., BRONOWICKI, J. P., DE LEDINGHEN, V., HEZODE, C., ZOULIM, F., MATHURIN, P., TRAN, A., LARREY, D. G., RATZIU, V., ALRIC, L., HYLAND, R. H., JIANG, D. Y., DOEHLE, B., PANG, P. S., SYMONDS, W. T., SUBRAMANIAN, M., MCHUTCHISON, J. G., MARCELLIN, P., HABERSETZER, F., GUYADER, D., GRANGE, J. D., LOUSTAUD-RATTI, V., SERFATY, L., METIVIER, S., LEROY, V., ABERGEL, A. & POL, S. 2014a. Ledipasvir/Sofosbuvir Fixed Dose Combination is Safe and Efficacious in Cirrhotic Patients Who Have Previously Failed Protease-Inhibitor Based Triple Therapy. *Hepatology*, 60, 1270a-1271a.
- BOURLIERE, M., SULKOWSKI, M. S., OMATA, M., ZEUZEM, S., FELD, J. J., LAWITZ, E., MARCELLIN, P., HYLAND, R. H., DING, X., YANG, J. C., KNOX, S. J., PANG, P. S., SUBRAMANIAN, M., SYMONDS, W. T., MCHUTCHISON, J. G., MANGIA, A., GANE, E. J., REDDY, K. R., MIZOKAMI, M., POL, S. & AFDHAL, N. H. 2014b. An Integrated Safety and Efficacy Analysis of > 500 Patients with Compensated Cirrhosis Treated with Ledipasvir/Sofosbuvir with or without Ribavirin. *Hepatology*, 60, 239a-239a.
- BRAITSTEIN, P., PALEPU, A., DIETERICH, D., BENHAMOU, Y. & MONTANER, J. S. 2004. Special considerations in the initiation and management of antiretroviral therapy in individuals coinfecting with HIV and hepatitis C. *AIDS*, 18, 2221-34.
- BROOK, G., MAIN, J., NELSON, M., BHAGANI, S., WILKINS, E., LEEN, C., FISHER, M., GILLEECE, Y., GILSON, R., FREEDMAN, A., KULASEGARAM, R., AGARWAL, K., SABIN, C. & DEACON-ADAMS, C. 2010. British HIV Association guidelines for the management of coinfection with HIV-1 and hepatitis B or C virus 2010. *HIV Med*, 11, 1-30.
- BROWN, E. A., ZHANG, H., PING, L. H. & LEMON, S. M. 1992. Secondary structure of the 5' nontranslated regions of hepatitis C virus and pestivirus genomic RNAs. *Nucleic Acids Res*, 20, 5041-5.
- BUHLER, S. & BARTENSCHLAGER, R. 2012. New targets for antiviral therapy of chronic hepatitis C. *Liver Int*, 32 Suppl 1, 9-16.

- BUNCHORNTAVAKUL, C. & REDDY, K. R. 2015. Review article: the efficacy and safety of daclatasvir in the treatment of chronic hepatitis C virus infection. *Alimentary Pharmacology & Therapeutics*, n/a-n/a.
- CARABALLO CORTES, K., ZAGORDI, O., LASKUS, T., PLOSKI, R., BUKOWSKA-OSKO, I., PAWELCZYK, A., BERAK, H. & RADKOWSKI, M. 2013. Ultradeep pyrosequencing of hepatitis C virus hypervariable region 1 in quasispecies analysis. *Biomed Res Int*, 2013, 626083.
- CHAMBERS, T. J., FAN, X., DROLL, D. A., HEMBRADOR, E., SLATER, T., NICKELLS, M. W., DUSTIN, L. B. & DIBISCEGLIE, A. M. 2005. Quasispecies heterogeneity within the E1/E2 region as a pretreatment variable during pegylated interferon therapy of chronic hepatitis C virus infection. *J Virol*, 79, 3071-83.
- CHO, H. S., HA, N. C., KANG, L. W., CHUNG, K. M., BACK, S. H., JANG, S. K. & OH, B. H. 1998. Crystal structure of RNA helicase from genotype 1b hepatitis C virus. A feasible mechanism of unwinding duplex RNA. *J Biol Chem*, 273, 15045-52.
- CHOO, Q. L., KUO, G., WEINER, A. J., OVERBY, L. R., BRADLEY, D. W. & HOUGHTON, M. 1989. Isolation of a cDNA clone derived from a blood-borne non-A, non-B viral hepatitis genome. *Science*, 244, 359-62.
- COCQUEREL, L., WYCHOWSKI, C., MINNER, F., PENIN, F. & DUBUISSON, J. 2000. Charged residues in the transmembrane domains of hepatitis C virus glycoproteins play a major role in the processing, subcellular localization, and assembly of these envelope proteins. *J Virol*, 74, 3623-33.
- COMAS, I., MOYA, A. & GONZALEZ-CANDELAS, F. 2005. Validating viral quasispecies with digital organisms: a re-examination of the critical mutation rate. *BMC Evol Biol*, 5, 5.
- COOPER, C. & KLEIN, M. 2014. HIV/hepatitis C virus coinfection management: changing guidelines and changing paradigms. *HIV Med*.
- CROOKS, G. E., HON, G., CHANDONIA, J. M. & BRENNER, S. E. 2004. WebLogo: a sequence logo generator. *Genome Res*, 14, 1188-90.
- CRUZ-RIVERA, M., CARPIO-PEDROZA, J. C., ESCOBAR-GUTIERREZ, A., LOZANO, D., VERGARA-CASTANEDA, A., RIVERA-OSORIO, P., MARTINEZ-GUARNEROS, A., CHACON, C. A., FONSECA-CORONADO, S. & VAUGHAN, G. 2013. Rapid hepatitis C virus divergence among chronically infected individuals. *J Clin Microbiol*, 51, 629-32.

- D'OLIVEIRA, A., JR., VOIRIN, N., ALLARD, R., PEYRAMOND, D., CHIDIAC, C., TOURAINE, J. L., FABRY, J., TREPO, C. & VANHEMS, P. 2005. Prevalence and sexual risk of hepatitis C virus infection when human immunodeficiency virus was acquired through sexual intercourse among patients of the Lyon University Hospitals, France, 1992-2002. *J Viral Hepat*, 12, 330-2.
- DANTA, M., BROWN, D., BHAGANI, S., PYBUS, O. G., SABIN, C. A., NELSON, M., FISHER, M., JOHNSON, A. M. & DUSHEIKO, G. M. 2007. Recent epidemic of acute hepatitis C virus in HIV-positive men who have sex with men linked to high-risk sexual behaviours. *Aids*, 21, 983-91.
- DANTA, M., SEMMO, N., FABRIS, P., BROWN, D., PYBUS, O. G., SABIN, C. A., BHAGANI, S., EMERY, V. C., DUSHEIKO, G. M. & KLENERMAN, P. 2008. Impact of HIV on host-virus interactions during early hepatitis C virus infection. *J Infect Dis*, 197, 1558-66.
- DE AMORIM, R. M., COELHO, A., LAMPE, E., RAIOL, T. & MARTINS, R. M. 2014. Genetic diversity of hepatitis C virus quasispecies in chronic renal failure patients in Midwest Brazil. *Arch Virol*, 159, 1917-25.
- DE MEYER, S., DIERYNCK, I., GHYS, A., BEUMONT, M., DAEMS, B., VAN BAELEN, B., SULLIVAN, J. C., BARTELS, D. J., KIEFFER, T. L., ZEUZEM, S. & PICCHIO, G. 2012. Characterization of telaprevir treatment outcomes and resistance in patients with prior treatment failure: results from the REALIZE trial. *Hepatology*, 56, 2106-15.
- DEGASPERI, E. & AGHEMO, A. 2014. Sofosbuvir for the treatment of chronic hepatitis C: between current evidence and future perspectives. *Hepat Med*, 6, 25-33.
- DELANG, L., NEYTS, J., VLIEGEN, I., ABRIGNANI, S., NEDDERMANN, P. & DE FRANCESCO, R. 2013. Hepatitis C virus-specific directly acting antiviral drugs. *Curr Top Microbiol Immunol*, 369, 289-320.
- DI LIBERTO, G., ROQUE-AFONSO, A. M., KARA, R., DUCOULOMBIER, D., FALLOT, G., SAMUEL, D. & FERAY, C. 2006. Clinical and therapeutic implications of hepatitis C virus compartmentalization. *Gastroenterology*, 131, 76-84.
- DIETERICH, D., BACON, B. R., FLAMM, S. L., KOWDLEY, K. V., MILLIGAN, S., TSAI, N., YOUNOSSI, Z. & LAWITZ, E. 2014a. Evaluation of sofosbuvir and simeprevir-based regimens in the TRIO network: academic and community treatment of a real-world, heterogeneous population. *Hepatology*, 60, 220a-220a.

- DIETERICH, D., ROCKSTROH, J. K., ORKIN, C., GUTIERREZ, F., KLEIN, M. B., REYNES, J., SHUKLA, U., JENKINS, A., LENZ, O., OUWERKERK-MAHADEVAN, S., PEETERS, M., DE LA ROSA, G., TAMBUYZER, L. & JESSNER, W. 2014b. Simeprevir (TMC435) With Pegylated Interferon/Ribavirin in Patients Coinfected With HCV Genotype 1 and HIV-1: A Phase 3 Study. *Clinical Infectious Diseases*, 59, 1579-1587.
- DIONNE-ODOM, J., OSBORN, M. K., RADZIEWICZ, H., GRAKOU, A. & WORKOWSKI, K. 2009. Acute hepatitis C and HIV coinfection. *Lancet Infect Dis*, 9, 775-83.
- DOHM, J. C., LOTTAZ, C., BORODINA, T. & HIMMELBAUER, H. 2008. Substantial biases in ultra-short read data sets from high-throughput DNA sequencing. *Nucleic Acids Res*, 36, e105.
- DOMINGO, E., ESCARMIS, C., LAZARO, E. & MANRUBIA, S. C. 2005. Quasispecies dynamics and RNA virus extinction. *Virus Res*, 107, 129-39.
- DOMINGUEZ, S., GHOSN, J., VALANTIN, M. A., SCHRUNIGER, A., SIMON, A., BONNARD, P., CAUMES, E., PIALOUX, G., BENHAMOU, Y., THIBAUT, V. & KATLAMA, C. 2006. Efficacy of early treatment of acute hepatitis C infection with pegylated interferon and ribavirin in HIV-infected patients. *Aids*, 20, 1157-61.
- DONALDSON, E. F., HARRINGTON, P. R., O'REAR, J. J. & NAEGER, L. K. 2015. Clinical evidence and bioinformatics characterization of potential hepatitis C virus resistance pathways for sofosbuvir. *Hepatology*, 61, 56-65.
- DORE, G. J., LAWITZ, E., HEZODE, C., SHAFRAN, S. D., RAMJI, A., TATUM, H. A., TALIANI, G., TRAN, A., BRUNETTO, M. R., ZALTRON, S., STRASSER, S. I., WEIS, N., GHESQUIERE, W., LEE, S. S., LARREY, D., POL, S., HARLEY, H., GEORGE, J., FUNG, S. K., DE LEDINGHEN, V., HAGENS, P., MCPHEE, F., HERNANDEZ, D., COHEN, D., COONEY, E., NOVIELLO, S. & HUGHES, E. A. 2015. Daclatasvir plus peginterferon and ribavirin is noninferior to peginterferon and ribavirin alone, and reduces the duration of treatment for HCV genotype 2 or 3 infection. *Gastroenterology*, 148, 355-366 e1.
- DRESSMAN, D., YAN, H., TRAVERSO, G., KINZLER, K. W. & VOGELSTEIN, B. 2003. Transforming single DNA molecules into fluorescent magnetic particles for detection and enumeration of genetic variations. *Proceedings of the National Academy of Sciences of the United States of America*, 100, 8817-8822.

- DUFFY, S., SHACKELTON, L. A. & HOLMES, E. C. 2008. Rates of evolutionary change in viruses: patterns and determinants. *Nat Rev Genet*, 9, 267-76.
- DUNDAS, N., LEOS, N. K., MITUI, M., REVELL, P. & ROGERS, B. B. 2008. Comparison of automated nucleic acid extraction methods with manual extraction. *J Mol Diagn*, 10, 311-6.
- DUVOUX, C., PAWLITSKY, J. M., BASTIE, A., CHERQUI, D., SOUSSY, C. J. & DHUMEAUX, D. 1999. Low HCV replication levels in end-stage hepatitis C virus-related liver disease. *J Hepatol*, 31, 593-7.
- EASL 2015. EASL Recommendations on Treatment of Hepatitis C 2015. *J Hepatol*, 63, 199-236.
- EDGAR, R. C. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res*, 32, 1792-7.
- EID, J., FEHR, A., GRAY, J., LUONG, K., LYLE, J., OTTO, G., PELUSO, P., RANK, D., BAYBAYAN, P., BETTMAN, B., BIBILLO, A., BJORNSON, K., CHAUDHURI, B., CHRISTIANS, F., CICERO, R., CLARK, S., DALAL, R., DEWINTER, A., DIXON, J., FOQUET, M., GAERTNER, A., HARDENBOL, P., HEINER, C., HESTER, K., HOLDEN, D., KEARNS, G., KONG, X., KUSE, R., LACROIX, Y., LIN, S., LUNDQUIST, P., MA, C., MARKS, P., MAXHAM, M., MURPHY, D., PARK, I., PHAM, T., PHILLIPS, M., ROY, J., SEBRA, R., SHEN, G., SORENSON, J., TOMANEY, A., TRAVERS, K., TRULSON, M., VIECELI, J., WEGENER, J., WU, D., YANG, A., ZACCARIN, D., ZHAO, P., ZHONG, F., KORLACH, J. & TURNER, S. 2009. Real-time DNA sequencing from single polymerase molecules. *Science*, 323, 133-8.
- ELAZAR, M., CHEONG, K. H., LIU, P., GREENBERG, H. B., RICE, C. M. & GLENN, J. S. 2003. Amphipathic helix-dependent localization of NS5A mediates hepatitis C virus RNA replication. *J Virol*, 77, 6055-61.
- ELAZAR, M., LIU, P., RICE, C. M. & GLENN, J. S. 2004. An N-terminal amphipathic helix in hepatitis C virus (HCV) NS4B mediates membrane association, correct localization of replication complex proteins, and HCV RNA replication. *J Virol*, 78, 11393-400.
- ENOMOTO, M., TAMORI, A., KOBAYASHI, S., IWAI, S., MORIKAWA, H. & KAWADA, N. 2013. Treatment guidelines for HCV genotype 1: mono for low, triple for high, and dual for 'middle'? *J Gastroenterol*, 48, 555-6.
- ENOMOTO, N., SAKUMA, I., ASAHINA, Y., KUROSAKI, M., MURAKAMI, T., YAMAMOTO, C., IZUMI, N., MARUMO, F. & SATO, C. 1995. Comparison of full-length sequences of interferon-sensitive and resistant hepatitis

- C virus 1b. Sensitivity to interferon is conferred by amino acid substitutions in the NS5A region. *J Clin Invest*, 96, 224-30.
- ENOMOTO, N., SAKUMA, I., ASAHINA, Y., KUROSAKI, M., MURAKAMI, T., YAMAMOTO, C., OGURA, Y., IZUMI, N., MARUMO, F. & SATO, C. 1996. Mutations in the nonstructural protein 5A gene and response to interferon in patients with chronic hepatitis C virus 1b infection. *N Engl J Med*, 334, 77-81.
- ERIKSSON, N., PACHTER, L., MITSUYA, Y., RHEE, S. Y., WANG, C., GHARIZADEH, B., RONAGHI, M., SHAFER, R. W. & BEERENWINKEL, N. 2008. Viral population estimation using pyrosequencing. *PLoS Comput Biol*, 4, e1000074.
- ESCOBAR-GUTIERREZ, A., VAZQUEZ-PICHARDO, M., CRUZ-RIVERA, M., RIVERA-OSORIO, P., CARPIO-PEDROZA, J. C., RUIZ-PACHECO, J. A., RUIZ-TOVAR, K. & VAUGHAN, G. 2012. Identification of hepatitis C virus transmission using a next-generation sequencing approach. *J Clin Microbiol*, 50, 1461-3.
- EWING, B. & GREEN, P. 1998. Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res*, 8, 186-94.
- EWING, B., HILLIER, L., WENDL, M. C. & GREEN, P. 1998. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res*, 8, 175-85.
- FAN, X., MAO, Q., ZHOU, D., LU, Y., XING, J., XU, Y., RAY, S. C. & DI BISCEGLIE, A. M. 2009. High diversity of hepatitis C viral quasispecies is associated with early virological response in patients undergoing antiviral therapy. *Hepatology*, 50, 1765-72.
- FARCI, P., SHIMODA, A., WONG, D., CABEZON, T., DE GIOANNIS, D., STRAZZERA, A., SHIMIZU, Y., SHAPIRO, M., ALTER, H. J. & PURCELL, R. H. 1996. Prevention of hepatitis C virus infection in chimpanzees by hyperimmune serum against the hypervariable region 1 of the envelope 2 protein. *Proc Natl Acad Sci U S A*, 93, 15394-9.
- FARCI, P., STRAZZERA, R., ALTER, H. J., FARCI, S., DEGIOANNIS, D., COIANA, A., PEDDIS, G., USAI, F., SERRA, G., CHESSA, L., DIAZ, G., BALESTRIERI, A. & PURCELL, R. H. 2002. Early changes in hepatitis C viral quasispecies during interferon therapy predict the therapeutic outcome. *Proc Natl Acad Sci U S A*, 99, 3081-6.
- FELD, J. J., KOWDLEY, K. V., COAKLEY, E., SIGAL, S., NELSON, D. R., CRAWFORD, D., WEILAND, O., AGUILAR, H., XIONG, J., PILOT-MATIAS, T., DASILVA-TILLMANN, B., LARSEN, L., PODSADECKI, T. & BERNSTEIN,

- B. 2014. Treatment of HCV with ABT-450/r-ombitasvir and dasabuvir with ribavirin. *N Engl J Med*, 370, 1594-603.
- FENG, H., SHUDA, M., CHANG, Y. & MOORE, P. S. 2008. Clonal integration of a polyomavirus in human Merkel cell carcinoma. *Science*, 319, 1096-100.
- FERENCI, P., BERNSTEIN, D., LALEZARI, J., COHEN, D., LUO, Y., COOPER, C., TAM, E., MARINHO, R. T., TSAI, N., NYBERG, A., BOX, T. D., YOUNES, Z., ENAYATI, P., GREEN, S., BARUCH, Y., BHANDARI, B. R., CARUNTU, F. A., SEPE, T., CHULANOV, V., JANCZEWSKA, E., RIZZARDINI, G., GERVAIN, J., PLANAS, R., MORENO, C., HASSANEIN, T., XIE, W., KING, M., PODSADECKI, T. & REDDY, K. R. 2014. ABT-450/r-Ombitasvir and Dasabuvir with or without Ribavirin for HCV. *New England Journal of Medicine*, 370, 1983-1992.
- FIERER, D. S., URIEL, A. J., CARRIERO, D. C., KLEPPER, A., DIETERICH, D. T., MULLEN, M. P., THUNG, S. N., FIEL, M. I. & BRANCH, A. D. 2008. Liver fibrosis during an outbreak of acute hepatitis C virus infection in HIV-infected men: a prospective cohort study. *J Infect Dis*, 198, 683-6.
- FISHMAN, S. L. & BRANCH, A. D. 2009. The quasispecies nature and biological implications of the hepatitis C virus. *Infect Genet Evol*, 9, 1158-67.
- FLICEK, P. & BIRNEY, E. 2009. Sense from sequence reads: methods for alignment and assembly. *Nat Methods*, 6, S6-S12.
- FLINT, M. & MCKEATING, J. A. 2000. The role of the hepatitis C virus glycoproteins in infection. *Rev Med Virol*, 10, 101-17.
- FLORESE, R. H., NAGANO-FUJII, M., IWANAGA, Y., HIDAJAT, R. & HOTTA, H. 2002. Inhibition of protein synthesis by the nonstructural proteins NS4A and NS4B of hepatitis C virus. *Virus Res*, 90, 119-31.
- FORNS, X., LAWITZ, E., ZEUZEM, S., GANE, E., BRONOWICKI, J. P., ANDREONE, P., HORBAN, A., BROWN, A., PEETERS, M., LENZ, O., OUWERKERK-MAHADEVAN, S., SCOTT, J., DE LA ROSA, G., KALMEIJER, R., SINHA, R. & BEUMONT-MAUVIEL, M. 2014. Simeprevir with peginterferon and ribavirin leads to high rates of SVR in patients with HCV genotype 1 who relapsed after previous therapy: a phase 3 trial. *Gastroenterology*, 146, 1669-79 e3.
- FORTON, D. M., KARAYIANNIS, P., MAHMUD, N., TAYLOR-ROBINSON, S. D. & THOMAS, H. C. 2004a. Identification of unique hepatitis C virus quasispecies in the central nervous system and comparative analysis of internal translational efficiency of brain, liver, and serum variants. *J Virol*, 78, 5170-83.

- FORTON, D. M., THOMAS, H. C. & TAYLOR-ROBINSON, S. D. 2004b. Central nervous system involvement in hepatitis C virus infection. *Metab Brain Dis*, 19, 383-91.
- FOY, E., LI, K., WANG, C., SUMPTER, R., JR., IKEDA, M., LEMON, S. M. & GALE, M., JR. 2003. Regulation of interferon regulatory factor-3 by the hepatitis C virus serine protease. *Science*, 300, 1145-8.
- FRANCO, S., CASADELLA, M., NOGUERA-JULIAN, M., CLOTET, B., TURAL, C., PAREDES, R. & MARTINEZ, M. A. 2013. No detection of the NS5B S282T mutation in treatment-naïve genotype 1 HCV/HIV-1 coinfecting patients using deep sequencing. *J Clin Virol*, 58, 726-9.
- FRANK, C., MOHAMED, M. K., STRICKLAND, G. T., LAVANCHY, D., ARTHUR, R. R., MAGDER, L. S., EL KHOBY, T., ABDEL-WAHAB, Y., ALY OHN, E. S., ANWAR, W. & SALLAM, I. 2000. The role of parenteral antischistosomal therapy in the spread of hepatitis C virus in Egypt. *Lancet*, 355, 887-91.
- FRIDELL, R. A., QIU, D. K., WANG, C. F., VALERA, L. & GAO, M. 2010. Resistance Analysis of the Hepatitis C Virus NS5A Inhibitor BMS-790052 in an In Vitro Replicon System. *Antimicrobial Agents and Chemotherapy*, 54, 3641-3650.
- FRIDELL, R. A., WANG, C., SUN, J. H., O'BOYLE, D. R., 2ND, NOWER, P., VALERA, L., QIU, D., ROBERTS, S., HUANG, X., KIENZLE, B., BIFANO, M., NETTLES, R. E. & GAO, M. 2011. 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*, 54, 1924-35.
- FRIED, M. W., BUTI, M., DORE, G. J., FLISIAK, R., FERENCI, P., JACOBSON, I., MARCELLIN, P., MANNING, M., NIKITIN, I., POORDAD, F., SHERMAN, M., ZEUZEM, S., SCOTT, J., GILLES, L., LENZ, O., PEETERS, M., SEKAR, V., DE SMEDT, G. & BEUMONT-MAUVIEL, M. 2013. Once-daily simeprevir (TMC435) with pegylated interferon and ribavirin in treatment-naïve genotype 1 hepatitis C: the randomized PILLAR study. *Hepatology*, 58, 1918-29.
- FU, L., NIU, B., ZHU, Z., WU, S. & LI, W. 2010. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics*, 28, 3150-2.
- GANE, E. J., STEDMAN, C. A., HYLAND, R. H., DING, X., SVAROVSKAIA, E., SYMONDS, W. T., HINDES, R. G. & BERREY, M. M. 2013. Nucleotide polymerase inhibitor sofosbuvir plus ribavirin for hepatitis C. *N Engl J Med*, 368, 34-44.

- GAO, L., AIZAKI, H., HE, J. W. & LAI, M. M. 2004. Interactions between viral nonstructural proteins and host protein hVAP-33 mediate the formation of hepatitis C virus RNA replication complex on lipid raft. *J Virol*, 78, 3480-8.
- GAO, M. 2013. Antiviral activity and resistance of HCV NS5A replication complex inhibitors. *Curr Opin Virol*, 3, 514-20.
- GARCIA-SAMANIEGO, J., RODRIGUEZ, M., BERENQUER, J., RODRIGUEZ-ROSADO, R., CARBO, J., ASENSI, V. & SORIANO, V. 2001. Hepatocellular carcinoma in HIV-infected patients with chronic hepatitis C. *Am J Gastroenterol*, 96, 179-83.
- GAUDIERI, S., RAUCH, A., PFAFFEROTT, K., BARNES, E., CHENG, W., MCCAUGHAN, G., SHACKEL, N., JEFFREY, G. P., MOLLISON, L., BAKER, R., FURRER, H., GUNTARD, H. F., FREITAS, E., HUMPHREYS, I., KLENERMAN, P., MALLAL, S., JAMES, I., ROBERTS, S., NOLAN, D. & LUCAS, M. 2009. Hepatitis C virus drug resistance and immune-driven adaptations: relevance to new antiviral therapy. *Hepatology*, 49, 1069-82.
- GENTILE, I., BORGIA, F., BUONOMO, A. R., ZAPPULO, E., CASTALDO, G. & BORGIA, G. 2014a. ABT-450: A Novel Protease Inhibitor for the Treatment of Hepatitis C Virus Infection. *Current Medicinal Chemistry*, 21, 3261-3270.
- GENTILE, I., BUONOMO, A. R., ZAPPULO, E., COPPOLA, N. & BORGIA, G. 2014b. GS-9669: a novel non-nucleoside inhibitor of viral polymerase for the treatment of hepatitis C virus infection. *Expert Review of Anti-Infective Therapy*, 12, 1179-1186.
- GERBER, L., WELZEL, T. M. & ZEUZEM, S. 2013. New therapeutic strategies in HCV: polymerase inhibitors. *Liver Int*, 33 Suppl 1, 85-92.
- GIALLONARDO, F. D., TOPFER, A., REY, M., PRABHAKARAN, S., DUPORT, Y., LEEMANN, C., SCHMUTZ, S., CAMPBELL, N. K., JOOS, B., LECCA, M. R., PATRIGNANI, A., DAUMER, M., BEISEL, C., RUSERT, P., TRKOLA, A., GUNTARD, H. F., ROTH, V., BEERENWINKEL, N. & METZNER, K. J. 2014. Full-length haplotype reconstruction to infer the structure of heterogeneous virus populations. *Nucleic Acids Res*, 42, e115.
- GILBERT, P. B., ROSSINI, A. J. & SHANKARAPPA, R. 2005. Two-sample tests for comparing intra-individual genetic sequence diversity between populations. *Biometrics*, 61, 106-117.
- GILLEECE, Y. C., BROWNE, R. E., ASBOE, D., ATKINS, M., MANDALIA, S., BOWER, M., GAZZARD, B. G. & NELSON, M. R. 2005. Transmission of

hepatitis C virus among HIV-positive homosexual men and response to a 24-week course of pegylated interferon and ribavirin. *J Acquir Immune Defic Syndr*, 40, 41-6.

GIORGI, E. E., FUNKHOUSER, B., ATHREYA, G., PERELSON, A. S., KORBER, B. T. & BHATTACHARYA, T. 2010. Estimating time since infection in early homogeneous HIV-1 samples using a poisson model. *Bmc Bioinformatics*, 11.

GOTTWEIN, J. M., SCHEEL, T. K., JENSEN, T. B., LADEMANN, J. B., PRENTOE, J. C., KNUDSEN, M. L., HOEGH, A. M. & BUKH, J. 2009. Development and characterization of hepatitis C virus genotype 1-7 cell culture systems: role of CD81 and scavenger receptor class B type I and effect of antiviral drugs. *Hepatology*, 49, 364-77.

GOUTAGNY, N., FATMI, A., DE LEDINGHEN, V., PENIN, F., COUZIGOU, P., INCHAUSPE, G. & BAIN, C. 2003. Evidence of viral replication in circulating dendritic cells during hepatitis C virus infection. *J Infect Dis*, 187, 1951-8.

GRADY, B. P., VANHOMMERIG, J. W., SCHINKEL, J., WEEGINK, C. J., BRUISTEN, S. M., LINDENBURG, C. E. & PRINS, M. 2012. Low incidence of reinfection with the hepatitis C virus following treatment in active drug users in Amsterdam. *Eur J Gastroenterol Hepatol*, 24, 1302-7.

GRAKOU, A., MCCOURT, D. W., WYCHOWSKI, C., FEINSTONE, S. M. & RICE, C. M. 1993a. Characterization of the hepatitis C virus-encoded serine proteinase: determination of proteinase-dependent polyprotein cleavage sites. *J Virol*, 67, 2832-43.

GRAKOU, A., WYCHOWSKI, C., LIN, C., FEINSTONE, S. M. & RICE, C. M. 1993b. Expression and identification of hepatitis C virus polyprotein cleavage products. *J Virol*, 67, 1385-95.

GRAY, R. R., SALEMI, M., KLENERMAN, P. & PYBUS, O. G. 2012. A new evolutionary model for hepatitis C virus chronic infection. *PLoS Pathog*, 8, e1002656.

GRAY, R. R., TANAKA, Y., TAKEBE, Y., MAGIORKINIS, G., BUSKELL, Z., SEEFF, L., ALTER, H. J. & PYBUS, O. G. 2013. Evolutionary analysis of hepatitis C virus gene sequences from 1953. *Philos Trans R Soc Lond B Biol Sci*, 368, 20130168.

GREBELY, J., PHAM, S. T., MATTHEWS, G. V., PETOUMENOS, K., BULL, R. A., YEUNG, B., RAWLINSON, W., KALDOR, J., LLOYD, A., HELLARD, M., DORE, G. J. & WHITE, P. A. 2012. Hepatitis C virus reinfection and

superinfection among treated and untreated participants with recent infection. *Hepatology*, 55, 1058-69.

GREGORI, J., ESTEBAN, J. I., CUBERO, M., GARCIA-CEHIC, D., PERALES, C., CASILLAS, R., ALVAREZ-TEJADO, M., RODRIGUEZ-FRIAS, F., GUARDIA, J., DOMINGO, E. & QUER, J. 2013. Ultra-deep pyrosequencing (UDPS) data treatment to study amplicon HCV minor variants. *PLoS One*, 8, e83361.

GREGORI, J., SALICRU, M., DOMINGO, E., SANCHEZ, A., ESTEBAN, J. I., RODRIGUEZ-FRIAS, F. & QUER, J. 2014. Inference with viral quasispecies diversity indices: clonal and NGS approaches. *Bioinformatics*, 30, 1104-1111.

GREUB, G., LEDERGERBER, B., BATTEGAY, M., GROB, P., PERRIN, L., FURRER, H., BURGISSER, P., ERB, P., BOGGIAN, K., PIFFARETTI, J. C., HIRSCHL, B., JANIN, P., FRANCIOLI, P., FLEPP, M. & TELENTI, A. 2000. Clinical progression, survival, and immune recovery during antiretroviral therapy in patients with HIV-1 and hepatitis C virus coinfection: the Swiss HIV Cohort Study. *Lancet*, 356, 1800-5.

GROBLER, J. A., MARKEL, E. J., FAY, J. F., GRAHAM, D. J., SIMCOE, A. L., LUDMERER, S. W., MURRAY, E. M., MIGLIACCIO, G. & FLORES, O. A. 2003. Identification of a key determinant of hepatitis C virus cell culture adaptation in domain II of NS3 helicase. *J Biol Chem*, 278, 16741-6.

GU, B., GATES, A. T., ISKEN, O., BEHRENS, S. E. & SARISKY, R. T. 2003. Replication studies using genotype 1a subgenomic hepatitis C virus replicons. *J Virol*, 77, 5352-9.

GWACK, Y., KIM, D. W., HAN, J. H. & CHOE, J. 1997. DNA helicase activity of the hepatitis C virus nonstructural protein 3. *Eur J Biochem*, 250, 47-54.

HADZIYANNIS, S. J., SETTE, H., JR., MORGAN, T. R., BALAN, V., DIAGO, M., MARCELLIN, P., RAMADORI, G., BODENHEIMER, H., JR., BERNSTEIN, D., RIZZETTO, M., ZEUZEM, S., POCKROS, P. J., LIN, A., ACKRILL, A. M. & GROUP, P. I. S. 2004. Peginterferon-alpha2a and ribavirin combination therapy in chronic hepatitis C: a randomized study of treatment duration and ribavirin dose. *Ann Intern Med*, 140, 346-55.

HALFON, P. & LOCARNINI, S. 2011. Hepatitis C virus resistance to protease inhibitors. *J Hepatol*, 55, 192-206.

- HALFON, P. & SARRAZIN, C. 2012. Future treatment of chronic hepatitis C with direct acting antivirals: is resistance important? *Liver Int*, 32 Suppl 1, 79-87.
- HANG, J. Q., YANG, Y., HARRIS, S. F., LEVEQUE, V., WHITTINGTON, H. J., RAJYAGURU, S., AO-IEONG, G., MCCOWN, M. F., WONG, A., GIANNETTI, A. M., LE POGAM, S., TALAMAS, F., CAMMACK, N., NAJERA, I. & KLUMPP, K. 2009. Slow binding inhibition and mechanism of resistance of non-nucleoside polymerase inhibitors of hepatitis C virus. *J Biol Chem*, 284, 15517-29.
- HARA, K., RIVERA, M. M., KOH, C., DEMINO, M., PAGE, S., NAGABHYRU, P. R., REHERMANN, B., LIANG, T. J., HOOFNAGLE, J. H. & HELLER, T. 2013. Sequence analysis of hepatitis C virus from patients with relapse after a sustained virological response: relapse or reinfection? *J Infect Dis*, 209, 38-45.
- HAYASHI, N., IZUMI, N., KUMADA, H., OKANOUE, T., TSUBOUCHI, H., YATSUHASHI, H., KATO, M., KI, R., KOMADA, Y., SETO, C. & GOTO, S. 2014a. Simeprevir with peginterferon/ribavirin for treatment-naive hepatitis C genotype 1 patients in Japan: CONCERTO-1, a phase III trial. *J Hepatol*, 61, 219-27.
- HAYASHI, N., SETO, C., KATO, M., KOMADA, Y. & GOTO, S. 2014b. Once-daily simeprevir (TMC435) with peginterferon/ribavirin for treatment-naive hepatitis C genotype 1-infected patients in Japan: the DRAGON study. *J Gastroenterol*, 49, 138-47.
- HEIM, M. H. 2013. 25 years of interferon-based treatment of chronic hepatitis C: an epoch coming to an end. *Nat Rev Immunol*, 13, 535-42.
- HEZODE, C., HIRSCHFIELD, G. M., GHESQUIERE, W., SIEVERT, W., RODRIGUEZ-TORRES, M., SHAFRAN, S. D., THULUVATH, P. J., TATUM, H. A., WAKED, I., ESMAT, G., LAWITZ, E. J., RUSTGI, V. K., POL, S., WEIS, N., POCKROS, P. J., BOURLIERE, M., SERFATY, L., VIERLING, J. M., FRIED, M. W., WEILAND, O., BRUNETTO, M. R., EVERSON, G. T., ZEUZEM, S., KWO, P. Y., SULKOWSKI, M., BRAU, N., HERNANDEZ, D., MCPHEE, F., WIND-ROTOLO, M., LIU, Z. H., NOVIELLO, S., HUGHES, E. A., YIN, P. D. & SCHNITTMAN, S. 2015. Daclatasvir plus peginterferon alfa and ribavirin for treatment-naive chronic hepatitis C genotype 1 or 4 infection: a randomised study. *Gut*, 64, 948-956.
- HILLIER, L. W., MARTH, G. T., QUINLAN, A. R., DOOLING, D., FEWELL, G., BARNETT, D., FOX, P., GLASSCOCK, J. I., HICKENBOTHAM, M., HUANG, W., MAGRINI, V. J., RICHT, R. J., SANDER, S. N., STEWART, D. A., STROMBERG, M., TSUNG, E. F., WYLIE, T., SCHEDL, T., WILSON, R. K.

- & MARDIS, E. R. 2008. Whole-genome sequencing and variant discovery in *C. elegans*. *Nat Methods*, 5, 183-8.
- HIRAGA, N., IMAMURA, M., ABE, H., HAYES, C. N., KONO, T., ONISHI, M., TSUGE, M., TAKAHASHI, S., OCHI, H., IWAO, E., KAMIYA, N., YAMADA, I., TATENO, C., YOSHIKATO, K., MATSUI, H., KANAI, A., INABA, T., TANAKA, S. & CHAYAMA, K. 2011. Rapid emergence of telaprevir resistant hepatitis C virus strain from wildtype clone in vivo. *Hepatology*, 54, 781-8.
- HOLLAND, J., SPINDLER, K., HORODYSKI, F., GRABAU, E., NICHOL, S. & VANDEPOL, S. 1982. Rapid evolution of RNA genomes. *Science*, 215, 1577-85.
- HOLMES, E. C. 2010. Does hepatitis C virus really form quasispecies? *Infect Genet Evol*, 10, 431-2.
- HONDA, M., PING, L. H., RIJNBRAND, R. C., AMPHLETT, E., CLARKE, B., ROWLANDS, D. & LEMON, S. M. 1996. Structural requirements for initiation of translation by internal ribosome entry within genome-length hepatitis C virus RNA. *Virology*, 222, 31-42.
- HUELSENBECK, J. P. 1995. The robustness of two phylogenetic methods: four-taxon simulations reveal a slight superiority of maximum likelihood over neighbor joining. *Mol Biol Evol*, 12, 843-9.
- HUMAN MICROBIOME PROJECT, C. 2012. Structure, function and diversity of the healthy human microbiome. *Nature*, 486, 207-14.
- JABARA, C. B., HU, F., MOLLAN, K. R., WILLIFORD, S. E., MENEZES, P., YANG, Y., ERON, J. J., FRIED, M. W., HUDGENS, M. G., JONES, C. D., SWANSTROM, R. & LEMON, S. M. 2014. Hepatitis C Virus (HCV) NS3 sequence diversity and antiviral resistance-associated variant frequency in HCV/HIV coinfection. *Antimicrob Agents Chemother*, 58, 6079-92.
- JACOBSON, I. M., DORE, G. J., FOSTER, G. R., FRIED, M. W., RADU, M., RAFALSKY, V. V., MOROZ, L., CRAXI, A., PEETERS, M., LENZ, O., OUWERKERK-MAHADEVAN, S., DE LA ROSA, G., KALMEIJER, R., SCOTT, J., SINHA, R. & BEUMONT-MAUVIEL, M. 2014. Simeprevir with pegylated interferon alfa 2a plus ribavirin in treatment-naive patients with chronic hepatitis C virus genotype 1 infection (QUEST-1): a phase 3, randomised, double-blind, placebo-controlled trial. *Lancet*, 384, 403-13.
- JACOBSON, I. M., MCHUTCHISON, J. G., DUSHEIKO, G., DI BISCEGLIE, A. M., REDDY, K. R., BZOWEJ, N. H., MARCELLIN, P., MUIR, A. J., FERENCI,

- P., FLISIAK, R., GEORGE, J., RIZZETTO, M., SHOUVAL, D., SOLA, R., TERG, R. A., YOSHIDA, E. M., ADDA, N., BENGTSSON, L., SANKOH, A. J., KIEFFER, T. L., GEORGE, S., KAUFFMAN, R. S., ZEUZEM, S. & TEAM, A. S. 2011. Telaprevir for previously untreated chronic hepatitis C virus infection. *N Engl J Med*, 364, 2405-16.
- JENSEN, D. M., O'LEARY, J. G., POCKROS, P. J., SHERMAN, K. E., KWO, P. Y., MAILLIARD, M. E., KOWDLEY, K. V., MUIR, A. J., DICKSON, R. C., RAMANI, A., MANNS, M. P., LOK, A. S., AKUSKEVICH, L., NELSON, D. R. & FRIED, M. W. 2014. Safety and Efficacy of Sofosbuvir-Containing Regimens for Hepatitis C: Real-World Experience in a Diverse, Longitudinal Observational Cohort. *Hepatology*, 60, 219a-220a.
- JONES, C. T., MURRAY, C. L., EASTMAN, D. K., TASSELLO, J. & RICE, C. M. 2007. Hepatitis C virus p7 and NS2 proteins are essential for production of infectious virus. *J Virol*, 81, 8374-83.
- JONES, D. M., PATEL, A. H., TARGETT-ADAMS, P. & MCLAUCHLAN, J. 2009. The hepatitis C virus NS4B protein can trans-complement viral RNA replication and modulates production of infectious virus. *J Virol*, 83, 2163-77.
- KALITA, M. M., GRIFFIN, S., CHOU, J. J. & FISCHER, W. B. 2015. Genotype-specific differences in structural features of hepatitis C virus (HCV) p7 membrane protein. *Biochim Biophys Acta*, 1848, 1383-92.
- KATO, N., OOTSUYAMA, Y., TANAKA, T., NAKAGAWA, M., NAKAZAWA, T., MURAI, K., OHKOSHI, S., HIJIKATA, M. & SHIMOTOHNO, K. 1992. Marked sequence diversity in the putative envelope proteins of hepatitis C viruses. *Virus Res*, 22, 107-23.
- KATO, T., DATE, T., MIYAMOTO, M., FURUSAKA, A., TOKUSHIGE, K., MIZOKAMI, M. & WAKITA, T. 2003. Efficient replication of the genotype 2a hepatitis C virus subgenomic replicon. *Gastroenterology*, 125, 1808-17.
- KATOH, K. & STANDLEY, D. M. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol*, 30, 772-80.
- KIMURA, M. 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J Mol Evol*, 16, 111-20.
- KIRST, M. E., LI, E. C., WANG, C. X., DONG, H. J., LIU, C., FRIED, M. W., NELSON, D. R. & WANG, G. P. 2013. Deep sequencing analysis of HCV

NS3 resistance-associated variants and mutation linkage in liver transplant recipients. *PLoS One*, 8, e69698.

KOEV, G., MONDAL, R., BEYER, J., REISCH, T., MASSE, S., KATI, W., HUTCHINSON, D., FLENTGE, C., RANDOLPH, J., DONNER, P., KRUEGER, A., WAGNER, R., YAN, P., LIN, T., MARING, C. & MOLLA, A. 2009. Characterization of Resistance Mutations Selected *In Vitro* by the Non-Nucleoside Hcv Polymerase Inhibitors Abt-333 and Abt-072. *Journal of Hepatology*, 50, S346-S347.

KOLYKHALOV, A. A., FEINSTONE, S. M. & RICE, C. M. 1996. Identification of a highly conserved sequence element at the 3' terminus of hepatitis C virus genome RNA. *J Virol*, 70, 3363-71.

KORBEL, J. O., URBAN, A. E., AFFOURTIT, J. P., GODWIN, B., GRUBERT, F., SIMONS, J. F., KIM, P. M., PALEJEV, D., CARRIERO, N. J., DU, L., TAILLON, B. E., CHEN, Z., TANZER, A., SAUNDERS, A. C., CHI, J., YANG, F., CARTER, N. P., HURLES, M. E., WEISSMAN, S. M., HARKINS, T. T., GERSTEIN, M. B., EGHOLM, M. & SNYDER, M. 2007. Paired-end mapping reveals extensive structural variation in the human genome. *Science*, 318, 420-6.

KOWDLEY, K. V., GORDON, S. C., REDDY, K. R., ROSSARO, L., BERNSTEIN, D. E., LAWITZ, E., SHIFFMAN, M. L., SCHIFF, E., GHALIB, R., RYAN, M., RUSTGI, V., CHOJKIER, M., HERRING, R., DI BISCEGLIE, A. M., POCKROS, P. J., SUBRAMANIAN, G. M., AN, D., SVAROVSKAIA, E., HYLAND, R. H., PANG, P. S., SYMONDS, W. T., MCHUTCHISON, J. G., MUIR, A. J., POUND, D. & FRIED, M. W. 2014. Ledipasvir and sofosbuvir for 8 or 12 weeks for chronic HCV without cirrhosis. *N Engl J Med*, 370, 1879-88.

KOWDLEY, K. V., LAWITZ, E., CRESPO, I., HASSANEIN, T., DAVIS, M. N., DEMICCO, M., BERNSTEIN, D. E., AFDHAL, N., VIERLING, J. M., GORDON, S. C., ANDERSON, J. K., HYLAND, R. H., DVORY-SOBOL, H., AN, D., HINDES, R. G., ALBANIS, E., SYMONDS, W. T., BERREY, M. M., NELSON, D. R. & JACOBSON, I. M. 2013. Sofosbuvir with pegylated interferon alfa-2a and ribavirin for treatment-naïve patients with hepatitis C genotype-1 infection (ATOMIC): an open-label, randomised, multicentre phase 2 trial. *Lancet*, 381, 2100-2107.

KRIEGER, N., LOHMANN, V. & BARTENSCHLAGER, R. 2001. Enhancement of hepatitis C virus RNA replication by cell culture-adaptive mutations. *J Virol*, 75, 4614-24.

KRISHNAN, P., BEYER, J., MISTRY, N., KOEV, G., REISCH, T., DEGOEY, D., KATI, W., CAMPBELL, A., WILLIAMS, L., XIE, W., SETZE, C., MOLLA, A., COLLINS, C. & PILOT-MATIAS, T. 2015a. *In Vitro* and *In Vivo* Antiviral

Activity and Resistance Profile of Ombitasvir, an Inhibitor of Hepatitis C Virus NS5A. *Antimicrobial Agents and Chemotherapy*, 59, 979-987.

KRISHNAN, P., TRIPATHI, R., SCHNELL, G., REISCH, T., BEYER, J., IRVIN, M., XIE, W., LARSEN, L., COHEN, D., PODSADECKI, T., PILOT-MATIAS, T. & COLLINS, C. 2015b. Resistance Analysis of Baseline and Treatment-Emergent Variants in Hepatitis C Virus Genotype 1 in the AVIATOR Study with Paritaprevir-Ritonavir, Ombitasvir, and Dasabuvir. *Antimicrob Agents Chemother*, 59, 5445-54.

KRISHNAN, P., TRIPATHI, R., SCHNELL, G., REISCH, T., BEYER, J., IRVIN, M., XIE, W., LARSEN, L., PODSADECKI, T., PILOT-MATIAS, T. & COLLINS, C. 2014. Pooled analysis of resistance in patients treated with ombitasvir/ABT-450/r and dasabuvir with or without ribavirin in Phase 2 and Phase 3 clinical trials. *Hepatology*, 60, 1134A-1135A.

KUIKEN, C., HRABER, P., THURMOND, J. & YUSIM, K. 2008. The hepatitis C sequence database in Los Alamos. *Nucleic Acids Res*, 36, D512-6.

KUNTZEN, T., TIMM, J., BERICAL, A., LENNON, N., BERLIN, A. M., YOUNG, S. K., LEE, B., HECKERMAN, D., CARLSON, J., REYOR, L. L., KLEYMAN, M., MCMAHON, C. M., BIRCH, C., SCHULZE ZUR WIESCH, J., LEDLIE, T., KOEHRSEN, M., KODIRA, C., ROBERTS, A. D., LAUER, G. M., ROSEN, H. R., BIHL, F., CERNY, A., SPENGLER, U., LIU, Z., KIM, A. Y., XING, Y., SCHNEIDWIND, A., MADEY, M. A., FLECKENSTEIN, J. F., PARK, V. M., GALAGAN, J. E., NUSBAUM, C., WALKER, B. D., LAKE-BAKAAR, G. V., DAAR, E. S., JACOBSON, I. M., GOMPERTS, E. D., EDLIN, B. R., DONFIELD, S. M., CHUNG, R. T., TALAL, A. H., MARION, T., BIRREN, B. W., HENN, M. R. & ALLEN, T. M. 2008. Naturally occurring dominant resistance mutations to hepatitis C virus protease and polymerase inhibitors in treatment-naive patients. *Hepatology*, 48, 1769-78.

KWONG, A. D., NAJERA, I., BECHTEL, J., BOWDEN, S., FITZGIBBON, J., HARRINGTON, P., KEMPF, D., KIEFFER, T. L., KOLETZKI, D., KUKOLJ, G., LIM, S., PILOT-MATIAS, T., LIN, K., MANI, N., MO, H., O'REAR, J., OTTO, M., PARKIN, N., PAWLOTSKY, J. M., PETROPOULOS, C., PICCHIO, G., RALSTON, R., REEVES, J. D., SCHOOLEY, R. T., SEIWERT, S., STANDRING, D., STUYVER, L., SULLIVAN, J. & MILLER, V. 2011. Sequence and phenotypic analysis for resistance monitoring in hepatitis C virus drug development: recommendations from the HCV DRAG. *Gastroenterology*, 140, 755-60.

LAMBERS, F. A., PRINS, M., THOMAS, X., MOLENKAMP, R., KWA, D., BRINKMAN, K., VAN DER MEER, J. T. & SCHINKEL, J. 2011. Alarming incidence of hepatitis C virus re-infection after treatment of sexually acquired acute hepatitis C virus infection in HIV-infected MSM. *Aids*, 25, F21-7.

- LANFORD, R. E., GUERRA, B., LEE, H., AVERETT, D. R., PFEIFFER, B., CHAVEZ, D., NOTVALL, L. & BIGGER, C. 2003. Antiviral effect and virus-host interactions in response to alpha interferon, gamma interferon, poly(i)-poly(c), tumor necrosis factor alpha, and ribavirin in hepatitis C virus subgenomic replicons. *J Virol*, 77, 1092-104.
- LANGE, C. M. & ZEUZEM, S. 2013. Perspectives and challenges of interferon-free therapy for chronic hepatitis C. *J Hepatol*, 58, 583-92.
- LASKUS, T., RADKOWSKI, M., PIASEK, A., NOWICKI, M., HORBAN, A., CIANCIARA, J. & RAKELA, J. 2000. Hepatitis C virus in lymphoid cells of patients coinfectd with human immunodeficiency virus type 1: evidence of active replication in monocytes/macrophages and lymphocytes. *J Infect Dis*, 181, 442-8.
- LAWITZ, E., GHALIB, R., RODRIGUEZ-TORRES, M., YOUNOSSI, Z. M., CORREGIDOR, A., SULKOWSKI, M. S., DEJESUS, E., PEARLMAN, B., RABINOVITZ, M., GITLIN, N., LIM, J. K., POCKROS, P. J., FEVERY, B., LAMBRECHT, T., OUWERKERK-MAHADEVAN, S., CALLEWAERT, K., SYMONDS, W. T., PICCHIO, G., LINDSAY, K., BEUMONT-MAUVIEL, M. & JACOBSON, I. M. 2014a. Simeprevir Plus Sofosbuvir with/without Ribavirin in Hcv Genotype 1 Prior Null-Responder/Treatment-Naive Patients (Cosmos Study): Primary Endpoint (Svr12) Results in Patients with Metavir F3-4 (Cohort 2). *Journal of Hepatology*, 60, S524-S524.
- LAWITZ, E., MAKARA, M., AKARCA, U. S., THULUVATH, P. J., PREOTESCU, L. L., VARUNOK, P., MORILLAS, R. M., HALL, C., MOBASHERY, N., REDMAN, R., PILOT-MATIAS, T., VILCHEZ, R. A. & HEZODE, C. 2015a. Efficacy and Safety of Ombitasvir, Paritaprevir, and Ritonavir in an Open-Label Study of Patients With Genotype 1b Chronic Hepatitis C Virus Infection With and Without Cirrhosis. *Gastroenterology*, 149, 971-980 e1.
- LAWITZ, E., MANGIA, A., WYLES, D., RODRIGUEZ-TORRES, M., HASSANEIN, T., GORDON, S. C., SCHULTZ, M., DAVIS, M. N., KAYALI, Z., REDDY, K. R., JACOBSON, I. M., KOWDLEY, K. V., NYBERG, L., SUBRAMANIAN, G. M., HYLAND, R. H., ARTERBURN, S., JIANG, D., MCNALLY, J., BRAINARD, D., SYMONDS, W. T., MCHUTCHISON, J. G., SHEIKH, A. M., YOUNOSSI, Z. & GANE, E. J. 2013. Sofosbuvir for previously untreated chronic hepatitis C infection. *N Engl J Med*, 368, 1878-87.
- LAWITZ, E., POORDAD, F. F., PANG, P. S., HYLAND, R. H., DING, X., MO, H., SYMONDS, W. T., MCHUTCHISON, J. G. & MEMBRENO, F. E. 2014b. Sofosbuvir and ledipasvir fixed-dose combination with and without ribavirin in treatment-naive and previously treated patients with genotype 1 hepatitis C virus infection (LONESTAR): an open-label, randomised, phase 2 trial. *Lancet*, 383, 515-23.

- LAWITZ, E., SULKOWSKI, M. S., GHALIB, R., RODRIGUEZ-TORRES, M., YOUNOSSI, Z. M., CORREGIDOR, A., DEJESUS, E., PEARLMAN, B., RABINOVITZ, M., GITLIN, N., LIM, J. K., POCKROS, P. J., SCOTT, J. D., FEVERY, B., LAMBRECHT, T., OUWERKERK-MAHADEVAN, S., CALLEWAERT, K., SYMONDS, W. T., PICCHIO, G., LINDSAY, K. L., BEUMONT, M. & JACOBSON, I. M. 2014c. Simeprevir plus sofosbuvir, with or without ribavirin, to treat chronic infection with hepatitis C virus genotype 1 in non-responders to pegylated interferon and ribavirin and treatment-naive patients: the COSMOS randomised study. *Lancet*, 384, 1756-1765.
- LAWITZ, E., SULLIVAN, G., RODRIGUEZ-TORRES, M., BENNETT, M., POORDAD, F., KAPOOR, M., BADRI, P., CAMPBELL, A., RODRIGUES, L., HU, Y. R., PILOT-MATIAS, T. & VILCHEZ, R. A. 2015b. Exploratory trial of ombitasvir and ABT-450/r with or without ribavirin for HCV genotype 1, 2, and 3 infection. *Journal of Infection*, 70, 197-205.
- LAWITZ, E. J., GRUENER, D., HILL, J. M., MARBURY, T., MOOREHEAD, L., MATHIAS, A., CHENG, G., LINK, J. O., WONG, K. A., MO, H., MCHUTCHISON, J. G. & BRAINARD, D. M. 2012. A phase 1, randomized, placebo-controlled, 3-day, dose-ranging study of GS-5885, an NS5A inhibitor, in patients with genotype 1 hepatitis C. *Journal of Hepatology*, 57, 24-31.
- LE POGAM, S., YAN, J. M., CHHABRA, M., ILNICKA, M., KANG, H., KOSAKA, A., ALI, S., CHIN, D. J., SHULMAN, N. S., SMITH, P., KLUMPP, K. & NAJERA, I. 2012. Characterization of hepatitis C virus (HCV) quasispecies dynamics upon short-term dual therapy with the HCV NS5B nucleoside polymerase inhibitor mericitabine and the NS3/4 protease inhibitor danoprevir. *Antimicrob Agents Chemother*, 56, 5494-502.
- LEE, H., SHIN, H., WIMMER, E. & PAUL, A. V. 2004. cis-acting RNA signals in the NS5B C-terminal coding sequence of the hepatitis C virus genome. *J Virol*, 78, 10865-77.
- LEGGEWIE, M., SREENU, V. B., ABDELRAHMAN, T., LEITCH, E. C., WILKIE, G. S., KLYMENKO, T., MUIR, D., THURSZ, M., MAIN, J. & THOMSON, E. C. 2013. Natural NS3 resistance polymorphisms occur frequently prior to treatment in HIV-positive patients with acute hepatitis C. *AIDS*, 27, 2485-8.
- LENZ, O., VERBINNEN, T., FEVERY, B., TAMBUYZER, L., VIJGEN, L., PEETERS, M., BUELENS, A., CEULEMANS, H., BEUMONT, M., PICCHIO, G. & DE MEYER, S. 2015. Virology analyses of HCV isolates from genotype 1-infected patients treated with simeprevir plus peginterferon/ribavirin in Phase IIb/III studies. *J Hepatol*, 62, 1008-14.

- LENZ, O., VERBINNEN, T., LIN, T. I., VIJGEN, L., CUMMINGS, M. D., LINDBERG, J., BERKE, J. M., DEHERTOGH, P., FRANSEN, E., SCHOLLIERS, A., VERMEIREN, K., IVENS, T., RABOISSON, P., EDLUND, M., STORM, S., VRANG, L., DE KOCK, H., FANNING, G. C. & SIMMEN, K. A. 2010. In vitro resistance profile of the hepatitis C virus NS3/4A protease inhibitor TMC435. *Antimicrob Agents Chemother*, 54, 1878-87.
- LESBURG, C. A., CABLE, M. B., FERRARI, E., HONG, Z., MANNARINO, A. F. & WEBER, P. C. 1999. Crystal structure of the RNA-dependent RNA polymerase from hepatitis C virus reveals a fully encircled active site. *Nat Struct Biol*, 6, 937-43.
- LI, J. Z., CHAPMAN, B., CHARLEBOIS, P., HOFMANN, O., WEINER, B., PORTER, A. J., SAMUEL, R., VARDHANABHUTI, S., ZHENG, L., ERON, J., TAIWO, B., ZODY, M. C., HENN, M. R., KURITZKES, D. R., HIDE, W., TEAM, A. A. S., WILSON, C. C., BERZINS, B. I., ACOSTA, E. P., BASTOW, B., KIM, P. S., READ, S. W., JANIK, J., MERES, D. S., LEDERMAN, M. M., MONG-KRYSPIN, L., SHAW, K. E., ZIMMERMAN, L. G., LEAVITT, R., DE LA ROSA, G. & JENNINGS, A. 2014. Comparison of illumina and 454 deep sequencing in participants failing raltegravir-based antiretroviral therapy. *PLoS One*, 9, e90485.
- LI, K., FOY, E., FERREON, J. C., NAKAMURA, M., FERREON, A. C., IKEDA, M., RAY, S. C., GALE, M., JR. & LEMON, S. M. 2005. Immune evasion by hepatitis C virus NS3/4A protease-mediated cleavage of the Toll-like receptor 3 adaptor protein TRIF. *Proc Natl Acad Sci U S A*, 102, 2992-7.
- LIN, C., LINDENBACH, B. D., PRAGAI, B. M., MCCOURT, D. W. & RICE, C. M. 1994. Processing in the hepatitis C virus E2-NS2 region: identification of p7 and two distinct E2-specific products with different C termini. *J Virol*, 68, 5063-73.
- LIN, K., PERNI, R. B., KWONG, A. D. & LIN, C. 2006. VX-950, a novel hepatitis C virus (HCV) NS3-4A protease inhibitor, exhibits potent antiviral activities in HCV replicon cells. *Antimicrob Agents Chemother*, 50, 1813-22.
- LIN, M. V., CHARLTON, A. N., ROUSTER, S. D., ZAMOR, P. J. & SHERMAN, K. E. 2014. Hepatitis C virus NS3 mutations in haemophiliacs. *Haemophilia*, 20, 659-665.
- LIU, L., LI, Y., LI, S., HU, N., HE, Y., PONG, R., LIN, D., LU, L. & LAW, M. 2012. Comparison of next-generation sequencing systems. *J Biomed Biotechnol*, 2012, 251364.

- LOHMANN, V., HOFFMANN, S., HERIAN, U., PENIN, F. & BARTENSCHLAGER, R. 2003. Viral and cellular determinants of hepatitis C virus RNA replication in cell culture. *J Virol*, 77, 3007-19.
- LOHMANN, V., KORNER, F., KOCH, J., HERIAN, U., THEILMANN, L. & BARTENSCHLAGER, R. 1999. Replication of subgenomic hepatitis C virus RNAs in a hepatoma cell line. *Science*, 285, 110-3.
- LOMAN, N. J., MISRA, R. V., DALLMAN, T. J., CONSTANTINIDOU, C., GHARBIA, S. E., WAIN, J. & PALLEEN, M. J. 2012. Performance comparison of benchtop high-throughput sequencing platforms. *Nat Biotechnol*, 30, 434-9.
- LONTOK, E., HARRINGTON, P., HOWE, A., KIEFFER, T., LENNERSTRAND, J., LENZ, O., MCPHEE, F., MO, H., PARKIN, N., PILOT-MATIAS, T. & MILLER, V. 2015. Hepatitis C virus drug resistance-associated substitutions: State of the art summary. *Hepatology*, 62, 1623-32.
- LUDMERER, S. W., GRAHAM, D. J., BOOTS, E., MURRAY, E. M., SIMCOE, A., MARKEL, E. J., GROBLER, J. A., FLORES, O. A., OLSEN, D. B., HAZUDA, D. J. & LAFEMINA, R. L. 2005. Replication fitness and NS5B drug sensitivity of diverse hepatitis C virus isolates characterized by using a transient replication assay. *Antimicrob Agents Chemother*, 49, 2059-69.
- MANNIS, M., MARCELLIN, P., POORDAD, F., DE ARAUJO, E. S., BUTI, M., HORMANS, Y., JANCZEWSKA, E., VILLAMIL, F., SCOTT, J., PEETERS, M., LENZ, O., OUWERKERK-MAHADEVAN, S., DE LA ROSA, G., KALMEIJER, R., SINHA, R. & BEUMONT-MAUVIEL, M. 2014. Simeprevir with pegylated interferon alfa 2a or 2b plus ribavirin in treatment-naive patients with chronic hepatitis C virus genotype 1 infection (QUEST-2): a randomised, double-blind, placebo-controlled phase 3 trial. *Lancet*, 384, 414-26.
- MARDIS, E. R. 2008. Next-generation DNA sequencing methods. *Annu Rev Genomics Hum Genet*, 9, 387-402.
- MARDIS, E. R. 2011. A decade's perspective on DNA sequencing technology. *Nature*, 470, 198-203.
- MARDIS, E. R. 2013. Next-generation sequencing platforms. *Annu Rev Anal Chem (Palo Alto Calif)*, 6, 287-303.
- MARGERIDON, S., LE POGAM, S., LIU, T. F., HANCZARUK, B., SIMEN, B. B., SHULMAN, N., SHAFER, R. W. & NAJERA, I. 2011. No Detection of Variants Bearing Ns5b S282t Mericitabine (Mcb) Resistance Mutation in

Daa Treatment-Naive Hcv Genotype 1 (G1)-Infected Patients Using Ultra-Deep Pyrosequencing (Udps). *Hepatology*, 54, 532a-532a.

MARTELL, M., ESTEBAN, J. I., QUER, J., GENESCA, J., WEINER, A., ESTEBAN, R., GUARDIA, J. & GOMEZ, J. 1992. Hepatitis-C Virus (Hcv) Circulates as a Population of Different but Closely Related Genomes - Quasi-Species Nature of Hcv Genome Distribution. *Journal of Virology*, 66, 3225-3229.

MARTIN, T. C., MARTIN, N. K., HICKMAN, M., VICKERMAN, P., PAGE, E. E., EVERETT, R., GAZZARD, B. G. & NELSON, M. 2013. Hepatitis C virus reinfection incidence and treatment outcome among HIV-positive MSM. *Aids*, 27, 2551-7.

MATSEN, F. A., KODNER, R. B. & ARMBRUST, E. V. 2010. pplacer: linear time maximum-likelihood and Bayesian phylogenetic placement of sequences onto a fixed reference tree. *BMC Bioinformatics*, 11, 538.

MATTHEWS, G. V., HELLARD, M., HABER, P., YEUNG, B., MARKS, P., BAKER, D., MCCAUGHAN, G., SASADEUSZ, J., WHITE, P., RAWLINSON, W., LLOYD, A., KALDOR, J., DORE, G. J. & ACUTE, A. T. I. 2009. Characteristics and Treatment Outcomes among HIV-Infected Individuals in the Australian Trial in Acute Hepatitis C. *Clinical Infectious Diseases*, 48, 650-658.

MCCAUGHAN, G. W., LASKUS, T. & VARGAS, H. E. 2003. Hepatitis C virus quasispecies: misunderstood and mistreated? *Liver Transpl*, 9, 1048-52.

MCCORMICK, A. L., MOYNIHAN, L., MACARTNEY, M. J., GARCIA-DIAZ, A., SMITH, C., JOHNSON, M. A., RODGER, A. J., BHAGANI, S., HAQUE, T. & WEBSTER, D. P. 2015. Baseline drug resistance mutations are detectable in HCV genes NS3 and NS5A but not NS5B in acute and chronic HIV-coinfected patients. *Antivir Ther*, 20, 361-3.

MCCOWN, M. F., RAJYAGURU, S., KULAR, S., CAMMACK, N. & NAJERA, I. 2009. GT-1a or GT-1b subtype-specific resistance profiles for hepatitis C virus inhibitors telaprevir and HCV-796. *Antimicrob Agents Chemother*, 53, 2129-32.

MCLAUHLAN, J. 2000. Properties of the hepatitis C virus core protein: a structural protein that modulates cellular processes. *J Viral Hepat*, 7, 2-14.

MCNAUGHTON, A. L., THOMSON, E. C., TEMPLETON, K., GUNSON, R. N. & LEITCH, E. C. 2013. Mixed genotype hepatitis C infections and implications for treatment. *Hepatology*.

- MCPHEE, F., HERNANDEZ, D., YU, F., UELAND, J., MONIKOWSKI, A., CARIFA, A., FALK, P., WANG, C., FRIDELL, R., ELEY, T., ZHOU, N. & GARDINER, D. 2013. Resistance analysis of hepatitis C virus genotype 1 prior treatment null responders receiving daclatasvir and asunaprevir. *Hepatology*.
- METZKER, M. L. 2010. Sequencing technologies - the next generation. *Nat Rev Genet*, 11, 31-46.
- MILLER, J. R., KOREN, S. & SUTTON, G. 2010. Assembly algorithms for next-generation sequencing data. *Genomics*, 95, 315-27.
- MOHSEN, A. H., EASTERBROOK, P., TAYLOR, C. B. & NORRIS, S. 2002. Hepatitis C and HIV-1 coinfection. *Gut*, 51, 601-8.
- MOHSEN, A. H., EASTERBROOK, P. J., TAYLOR, C., PORTMANN, B., KULASEGARAM, R., MURAD, S., WISELKA, M. & NORRIS, S. 2003. Impact of human immunodeficiency virus (HIV) infection on the progression of liver fibrosis in hepatitis C virus infected patients. *Gut*, 52, 1035-40.
- MOREAU, I., LEVIS, J., CROSBIE, O., KENNY-WALSH, E. & FANNING, L. J. 2008. Correlation between pre-treatment quasispecies complexity and treatment outcome in chronic HCV genotype 3a. *Virology*, 5, 78.
- MORSICA, G., BAGAGLIO, S., UBERTI-FOPPA, C., GALLI, L. & LAZZARIN, A. 2009. Detection of hepatitis C mutants with natural resistance to NS3/4A protease inhibitors in HIV/HCV-coinfected individuals treated with antiretroviral therapy. *J Acquir Immune Defic Syndr*, 51, 106-8.
- MULLAN, B., KENNY-WALSH, E., COLLINS, J. K., SHANAHAN, F. & FANNING, L. J. 2001. Inferred hepatitis C virus quasispecies diversity is influenced by choice of DNA polymerase in reverse transcriptase-polymerase chain reactions. *Anal Biochem*, 289, 137-46.
- MURPHY, D. G., SABLON, E., CHAMBERLAND, J., FOURNIER, E., DANDAVINO, R. & TREMBLAY, C. L. 2015. Hepatitis C Virus Genotype 7, a New Genotype Originating from Central Africa. *Journal of Clinical Microbiology*, 53, 967-972.
- NAKAMOTO, S., KANDA, T., WU, S., SHIRASAWA, H. & YOKOSUKA, O. 2014. Hepatitis C virus NS5A inhibitors and drug resistance mutations. *World Journal of Gastroenterology*, 20, 2902-2912.
- NAKANO, T., LAU, G. M., LAU, G. M., SUGIYAMA, M. & MIZOKAMI, M. 2012. An updated analysis of hepatitis C virus genotypes and subtypes based on the complete coding region. *Liver Int*, 32, 339-45.

- NASU, A., MARUSAWA, H., UEDA, Y., NISHIJIMA, N., TAKAHASHI, K., OSAKI, Y., YAMASHITA, Y., INOKUMA, T., TAMADA, T., FUJIWARA, T., SATO, F., SHIMIZU, K. & CHIBA, T. 2011. Genetic heterogeneity of hepatitis C virus in association with antiviral therapy determined by ultra-deep sequencing. *PLoS One*, 6, e24907.
- NATTERMANN, J. & TACKE, F. 2009. The structure behind diversity: covariance networks in hepatitis C virus sequences are associated with treatment response. *Hepatology*, 49, 1767-9.
- NELSON, D. R., COOPER, J. N., LALEZARI, J. P., LAWITZ, E., POCKROS, P. J., GITLIN, N., FREILICH, B. F., YOUNES, Z. H., HARLAN, W., GHALIB, R., OGUCHI, G., THULUVATH, P. J., ORTIZ-LASANTA, G., RABINOVITZ, M., BERASTEIN, D., BENNETT, M., HAWKINS, T., RAVENDHRAN, N., SHEIKH, A. M., VARUNOK, P., KOWDLEY, K. V., HENNICKEN, D., MCPHEE, F., RANA, I., HUGHES, E. A. & TEAM, A.-S. 2015. All-Oral 12-Week Treatment With Daclatasvir Plus Sofosbuvir in Patients With Hepatitis C Virus Genotype 3 Infection: ALLY-3 Phase III Study. *Hepatology*, 61, 1127-1135.
- NETTLES, R. E., GAO, M., BIFANO, M., CHUNG, E., PERSSON, A., MARBURY, T. C., GOLDWATER, R., DEMICCO, M. P., RODRIGUEZ-TORRES, M., VUTIKULLIRD, A., FUENTES, E., LAWITZ, E., LOPEZ-TALAVERA, J. C. & GRASELA, D. M. 2011. Multiple ascending dose study of BMS-790052, a nonstructural protein 5A replication complex inhibitor, in patients infected with hepatitis C virus genotype 1. *Hepatology*, 54, 1956-65.
- NEUKAM, K., NATTERMANN, J., RALLON, N., RIVERO, A., CARUZ, A., MACIAS, J., VOGEL, M., BENITO, J., CAMACHO, A., MIRA, J., SCHWARZ-ZANDER, C., BARREIRO, P., MARTINEZ, A., ROCKSTROH, J., SORIANO, V. & PINEDA, J. 2011. Different distributions of hepatitis C virus genotypes among HIV-infected patients with acute and chronic hepatitis C according to interleukin-28B genotype. *HIV Med*, 12, 487-93.
- NEUMANN, A. U., LAM, N. P., DAHARI, H., GRETCH, D. R., WILEY, T. E., LAYDEN, T. J. & PERELSON, A. S. 1998. Hepatitis C viral dynamics in vivo and the antiviral efficacy of interferon-alpha therapy. *Science*, 282, 103-107.
- NIELSEN, S. U., BASSENDINE, M. F., BURT, A. D., BEVITT, D. J. & TOMS, G. L. 2004. Characterization of the genome and structural proteins of hepatitis C virus resolved from infected human liver. *J Gen Virol*, 85, 1497-507.
- NINOMIYA, M., UENO, Y., FUNAYAMA, R., NAGASHIMA, T., NISHIDA, Y., KONDO, Y., INOUE, J., KAKAZU, E., KIMURA, O., NAKAYAMA, K. &

- SHIMOSEGAWA, T. 2012. Use of illumina deep sequencing technology to differentiate hepatitis C virus variants. *J Clin Microbiol*, 50, 857-66.
- NUNEZ, M., SORIANO, V., LOPEZ, M., BALLESTEROS, C., CASCAJERO, A., GONZALEZ-LAHOZ, J. & BENITO, J. M. 2006. Coinfection with hepatitis C virus increases lymphocyte apoptosis in HIV-infected patients. *Clin Infect Dis*, 43, 1209-12.
- O'BRIEN, T. R., EVERHART, J. E., MORGAN, T. R., LOK, A. S., CHUNG, R. T., SHAO, Y., SHIFFMAN, M. L., DOTRANG, M., SNINSKY, J. J., BONKOVSKY, H. L. & PFEIFFER, R. M. 2011. An IL28B genotype-based clinical prediction model for treatment of chronic hepatitis C. *PLoS One*, 6, e20904.
- OGISHI, M., YOTSUYANAGI, H., TSUTSUMI, T., GATANAGA, H., ODE, H., SUGIURA, W., MORIYA, K., OKA, S., KIMURA, S. & KOIKE, K. 2015. Deconvoluting the Composition of Low-Frequency Hepatitis C Viral Quasispecies: Comparison of Genotypes and NS3 Resistance-Associated Variants between HCV/HIV Coinfected Hemophiliacs and HCV Monoinfected Patients in Japan. *PLoS ONE*, 10, e0119145.
- OSINUSI, A., TOWNSEND, K., KOHLI, A., NELSON, A., SEAMON, C., MEISSNER, E. G., BON, D., SILK, R., GROSS, C., PRICE, A., SAJADI, M., SIDHARTHAN, S., SIMS, Z., HERRMANN, E., HOGAN, J., TEFERI, G., TALWANI, R., PROSCHAN, M., JENKINS, V., KLEINER, D. E., WOOD, B. J., SUBRAMANIAN, G. M., PANG, P. S., MCHUTCHISON, J. G., POLIS, M. A., FAUCI, A. S., MASUR, H. & KOTTILIL, S. 2015. Virologic Response Following Combined Ledipasvir and Sofosbuvir Administration in Patients With HCV Genotype 1 and HIV Co-infection. *Jama-Journal of the American Medical Association*, 313, 1232-1239.
- OTA, T. & NEI, M. 1994. Variance and covariances of the numbers of synonymous and nonsynonymous substitutions per site. *Mol Biol Evol*, 11, 613-9.
- PAGE, K., HAHN, J. A., EVANS, J., SHIBOSKI, S., LUM, P., DELWART, E., TOBLER, L., ANDREWS, W., AVANESYAN, L., COOPER, S. & BUSCH, M. P. 2009. Acute hepatitis C virus infection in young adult injection drug users: a prospective study of incident infection, resolution, and reinfection. *J Infect Dis*, 200, 1216-26.
- PALMER, S., KEARNEY, M., MALDARELLI, F., HALVAS, E. K., BIXBY, C. J., BAZMI, H., ROCK, D., FALLOON, J., DAVEY, R. T., JR., DEWAR, R. L., METCALF, J. A., HAMMER, S., MELLORS, J. W. & COFFIN, J. M. 2005. Multiple, linked human immunodeficiency virus type 1 drug resistance mutations in treatment-experienced patients are missed by standard genotype analysis. *J Clin Microbiol*, 43, 406-13.

- PAOLUCCI, S., FIORINA, L., MARIANI, B., GULMINETTI, R., NOVATI, S., BARBARINI, G., BRUNO, R. & BALDANTI, F. 2013. Naturally occurring resistance mutations to inhibitors of HCV NS5A region and NS5B polymerase in DAA treatment-naive patients. *Virology*, 10, 355.
- PAOLUCCI, S., FIORINA, L., PIRALLA, A., GULMINETTI, R., NOVATI, S., BARBARINI, G., SACCHI, P., GATTI, M., DOSSENA, L. & BALDANTI, F. 2012. Naturally occurring mutations to HCV protease inhibitors in treatment-naive patients. *Virology*, 9, 245.
- PAWLOTSKY, J.-M. 2013a. NS5A inhibitors in the treatment of hepatitis C. *Journal of Hepatology*, 59, 375-382.
- PAWLOTSKY, J. M. 2006. Therapy of hepatitis C: from empiricism to eradication. *Hepatology*, 43, S207-20.
- PAWLOTSKY, J. M. 2009. Therapeutic implications of hepatitis C virus resistance to antiviral drugs. *Therapeutic Advances in Gastroenterology*, 2, 205-19.
- PAWLOTSKY, J. M. 2011. Treatment failure and resistance with direct-acting antiviral drugs against hepatitis C virus. *Hepatology*, 53, 1742-51.
- PAWLOTSKY, J. M. 2013b. NS5A inhibitors in the treatment of hepatitis C. *J Hepatology*, 59, 375-82.
- PAYAN, C., PIVERT, A., MORAND, P., FAFI-KREMER, S., CARRAT, F., POL, S., CACOUB, P., PERRONNE, C. & LUNEL, F. 2007. Rapid and early virological response to chronic hepatitis C treatment with IFN alpha2b or PEG-IFN alpha2b plus ribavirin in HIV/HCV co-infected patients. *Gut*, 56, 1111-6.
- PENIN, F., COMBET, C., GERMANIDIS, G., FRANAIS, P. O., DELEAGE, G. & PAWLOTSKY, J. M. 2001. Conservation of the conformation and positive charges of hepatitis C virus E2 envelope glycoprotein hypervariable region 1 points to a role in cell attachment. *J Virology*, 75, 5703-10.
- PHAM, S. T., BULL, R. A., BENNETT, J. M., RAWLINSON, W. D., DORE, G. J., LLOYD, A. R. & WHITE, P. A. 2010. Frequent multiple hepatitis C virus infections among injection drug users in a prison setting. *Hepatology*, 52, 1564-72.
- PICCININNI, S., VARAKLIOTI, A., NARDELLI, M., DAVE, B., RANEY, K. D. & MCCARTHY, J. E. 2002. Modulation of the hepatitis C virus RNA-dependent RNA polymerase activity by the non-structural (NS) 3 helicase and the NS4B membrane protein. *J Biol Chem*, 277, 45670-9.

- PILOT-MATIAS, T., TRIPATHI, R., COHEN, D., GAULTIER, I., DEKHTYAR, T., LU, L., REISCH, T., IRVIN, M., HOPKINS, T., PITHAWALLA, R., MIDDLETON, T., NG, T., MCDANIEL, K., OR, Y. S., MENON, R., KEMPF, D., MOLLA, A. & COLLINS, C. 2015. In Vitro and In Vivo Antiviral Activity and Resistance Profile of the Hepatitis C Virus NS3/4A Protease Inhibitor ABT-450. *Antimicrobial Agents and Chemotherapy*, 59, 988-997.
- PLAZA, Z., SORIANO, V., VISPO, E., DEL MAR GONZALEZ, M., BARREIRO, P., SECLÉN, E. & POVEDA, E. 2012. Prevalence of natural polymorphisms at the HCV NS5A gene associated with resistance to daclatasvir, an NS5A inhibitor. *Antivir Ther*, 17, 921-6.
- POH, W. T., XIA, E., CHIN-INMANU, K., WONG, L. P., CHENG, A. Y., MALASIT, P., SURIYAPHOL, P., TEO, Y. Y. & ONG, R. T. 2013. Viral quasispecies inference from 454 pyrosequencing. *BMC Bioinformatics*, 14, 355.
- POORDAD, F., HEZODE, C., TRINH, R., KOWDLEY, K. V., ZEUZEM, S., AGARWAL, K., SHIFFMAN, M. L., WEDEMEYER, H., BERG, T., YOSHIDA, E. M., FORNS, X., LOVELL, S. S., DA SILVA-TILLMANN, B., COLLINS, C. A., CAMPBELL, A. L., PODSADECKI, T. & BERNSTEIN, B. 2014. ABT-450/r-Ombitasvir and Dasabuvir with Ribavirin for Hepatitis C with Cirrhosis. *New England Journal of Medicine*, 370, 1973-1982.
- POORDAD, F., MCCONE, J., JR., BACON, B. R., BRUNO, S., MANNS, M. P., SULKOWSKI, M. S., JACOBSON, I. M., REDDY, K. R., GOODMAN, Z. D., BOPARAI, N., DINUBILE, M. J., SNIUKIENE, V., BRASS, C. A., ALBRECHT, J. K. & BRONOWICKI, J. P. 2011. Boceprevir for untreated chronic HCV genotype 1 infection. *N Engl J Med*, 364, 1195-206.
- POORDAD, F., SCHIFF, E., VIERLING, J., LANDIS, C., FONTANA, R., YANG, R., MCPHEE, F., HUGHES, E., NOVIELLO, S. & SWENSON, E. 2015. Daclatasvir, sofosbuvir, and ribavirin combination for HCV patients with advanced cirrhosis or post-transplant recurrence: phase 3 ALLY-1 study. *Journal of Viral Hepatitis*, 22, 30-31.
- POVEDA, E., WYLES, D. L., MENA, A., PEDREIRA, J. D., CASTRO-IGLESIAS, A. & CACHAY, E. 2014. Update on hepatitis C virus resistance to direct-acting antiviral agents. *Antiviral Res*.
- POWDRILL, M. H., BERNATCHEZ, J. A. & GOTTE, M. 2010. Inhibitors of the Hepatitis C Virus RNA-Dependent RNA Polymerase NS5B. *Viruses*, 2, 2169-95.
- PROSPERI, M. C., PROSPERI, L., BRUSELLES, A., ABBATE, I., ROZERA, G., VINCENTI, D., SOLMONE, M. C., CAPOBIANCHI, M. R. & ULIVI, G. 2011. Combinatorial analysis and algorithms for quasispecies reconstruction using next-generation sequencing. *BMC Bioinformatics*, 12, 5.

- PROSPERI, M. C. & SALEMI, M. 2012. QuRe: software for viral quasispecies reconstruction from next-generation sequencing data. *Bioinformatics*, 28, 132-3.
- PROSPERI, M. C., YIN, L., NOLAN, D. J., LOWE, A. D., GOODENOW, M. M. & SALEMI, M. 2013. Empirical validation of viral quasispecies assembly algorithms: state-of-the-art and challenges. *Sci Rep*, 3, 2837.
- RAO, C. R. 1982. Diversity and Dissimilarity Coefficients - a Unified Approach. *Theoretical Population Biology*, 21, 24-43.
- REDDY, K. R., ZEUZEM, S., ZOULIM, F., WEILAND, O., HORBAN, A., STANCIU, C., VILLAMIL, F. G., ANDREONE, P., GEORGE, J., DAMMERS, E., FU, M., KURLAND, D., LENZ, O., OUWERKERK-MAHADEVAN, S., VERBINNEN, T., SCOTT, J. & JESSNER, W. 2015. Simeprevir versus telaprevir with peginterferon and ribavirin in previous null or partial responders with chronic hepatitis C virus genotype 1 infection (ATTAIN): a randomised, double-blind, non-inferiority phase 3 trial. *Lancet Infectious Diseases*, 15, 27-35.
- RIBEIRO, R. M., LI, H., WANG, S., STODDARD, M. B., LEARN, G. H., KORBER, B. T., BHATTACHARYA, T., GUEDJ, J., PARRISH, E. H., HAHN, B. H., SHAW, G. M. & PERELSON, A. S. 2012. Quantifying the diversification of hepatitis C virus (HCV) during primary infection: estimates of the in vivo mutation rate. *PLoS Pathog*, 8, e1002881.
- RICHMAN, D. D. 2000. The impact of drug resistance on the effectiveness of chemotherapy for chronic hepatitis B. *Hepatology*, 32, 866-7.
- ROBERTS, R. J., CARNEIRO, M. O. & SCHATZ, M. C. 2013. The advantages of SMRT sequencing. *Genome Biol*, 14, 405.
- ROCKSTROH, J. K., MOCROFT, A., SORIANO, V., TURAL, C., LOSSO, M. H., HORBAN, A., KIRK, O., PHILLIPS, A., LEDERGERBER, B. & LUNDGREN, J. 2005. Influence of hepatitis C virus infection on HIV-1 disease progression and response to highly active antiretroviral therapy. *J Infect Dis*, 192, 992-1002.
- ROCKSTROH, J. K. & SPENGLER, U. 2004. HIV and hepatitis C virus co-infection. *Lancet Infect Dis*, 4, 437-44.
- ROMANO, K. P., ALI, A., AYDIN, C., SOUMANA, D., OZEN, A., DEVEAU, L. M., SILVER, C., CAO, H., NEWTON, A., PETROPOULOS, C. J., HUANG, W. & SCHIFFER, C. A. 2012. The molecular basis of drug resistance against hepatitis C virus NS3/4A protease inhibitors. *PLoS Pathog*, 8, e1002832.

- ROMANO, K. P., ALI, A., ROYER, W. E. & SCHIFFER, C. A. 2010. Drug resistance against HCV NS3/4A inhibitors is defined by the balance of substrate recognition versus inhibitor binding. *Proc Natl Acad Sci U S A*, 107, 20986-91.
- RONAGHI, M., UHLEN, M. & NYREN, P. 1998. A sequencing method based on real-time pyrophosphate. *Science*, 281, 363-+.
- RONG, L. & PERELSON, A. S. 2010. Treatment of hepatitis C virus infection with interferon and small molecule direct antivirals: viral kinetics and modeling. *Crit Rev Immunol*, 30, 131-48.
- ROQUEBERT, B., MALET, I., WIRDEN, M., TUBIANA, R., VALANTIN, M. A., SIMON, A., KATLAMA, C., PEYTAVIN, G., CALVEZ, V. & MARCELIN, A. G. 2006. Role of HIV-1 minority populations on resistance mutational pattern evolution and susceptibility to protease inhibitors. *AIDS*, 20, 287-9.
- ROTHBERG, J. M., HINZ, W., REARICK, T. M., SCHULTZ, J., MILESKI, W., DAVEY, M., LEAMON, J. H., JOHNSON, K., MILGREW, M. J., EDWARDS, M., HOON, J., SIMONS, J. F., MARRAN, D., MYERS, J. W., DAVIDSON, J. F., BRANTING, A., NOBILE, J. R., PUC, B. P., LIGHT, D., CLARK, T. A., HUBER, M., BRANCIFORTE, J. T., STONER, I. B., CAWLEY, S. E., LYONS, M., FU, Y., HOMER, N., SEDOVA, M., MIAO, X., REED, B., SABINA, J., FEIERSTEIN, E., SCHORN, M., ALANJARY, M., DIMALANTA, E., DRESSMAN, D., KASINSKAS, R., SOKOLSKY, T., FIDANZA, J. A., NAMSARAEV, E., MCKERNAN, K. J., WILLIAMS, A., ROTH, G. T. & BUSTILLO, J. 2011. An integrated semiconductor device enabling non-optical genome sequencing. *Nature*, 475, 348-52.
- RUGGIERO, T., PROIETTI, A., BOGLIONE, L., MILIA, M. G., ALLICE, T., BURDINO, E., OROFINO, G., BONORA, S., DI PERRI, G. & GHISSETTI, V. 2015. Predominance of hepatitis C virus Q80K among NS3 baseline-resistance-associated amino acid variants in direct-antiviral-agent-naïve patients with chronic hepatitis: single-centre experience. *Arch Virol*, 160, 2881-2885.
- RUPP, D. & BARTENSCHLAGER, R. 2014. Targets for antiviral therapy of hepatitis C. *Semin Liver Dis*, 34, 9-21.
- SALLIE, R. 2005. Replicative homeostasis: a fundamental mechanism mediating selective viral replication and escape mutation. *Virol J*, 2, 10.
- SALLIE, R. 2007. Replicative homeostasis III: implications for antiviral therapy and mechanisms of response and non-response. *Virol J*, 4, 29.

- SALUDES, V., BASCUNANA, E., JORDANA-LLUCH, E., CASANOVAS, S., ARDEVOL, M., SOLER, E., PLANAS, R., AUSINA, V. & MARTRO, E. 2013. Relevance of baseline viral genetic heterogeneity and host factors for treatment outcome prediction in hepatitis C virus 1b-infected patients. *PLoS One*, 8, e72600.
- SANDRES, K., DUBOIS, M., PASQUIER, C., PAYEN, J. L., ALRIC, L., DUFFAUT, M., VINEL, J. P., PASCAL, J. P., PUEL, J. & IZOPET, J. 2000. Genetic heterogeneity of hypervariable region 1 of the hepatitis C virus (HCV) genome and sensitivity of HCV to alpha interferon therapy. *J Virol*, 74, 661-8.
- SANTANTONIO, T., MEDDA, E., FERRARI, C., FABRIS, P., CARITI, G., MASSARI, M., BABUDIERI, S., TOTI, M., FRANCAVILLA, R., ANCARANI, F., ANTONUCCI, G., SCOTTO, G., DI MARCO, V., PASTORE, G. & STROFFOLINI, T. 2006. Risk factors and outcome among a large patient cohort with community-acquired acute hepatitis C in Italy. *Clin Infect Dis*, 43, 1154-9.
- SARRAZIN, C., DVORY-SOBOL, H., SVAROVSKAIA, E. S., DOEHLE, B., MCCARVILLE, J. F., PANG, P. S., AFDHAL, N. H., KOWDLEY, K. V., GANE, E. J., LAWITZ, E., MCHUTCHISON, J. G., MILLER, M. D. & MO, H. 2014. Baseline and Post-baseline Resistance Analyses of Phase 2/3 Studies of Ledipasvir/Sofosbuvir +/- RBV. *Hepatology*, 60, 1128A-1128A.
- SARRAZIN, C., KIEFFER, T. L., BARTELS, D., HANZELKA, B., MUH, U., WELKER, M., WINCHERINGER, D., ZHOU, Y., CHU, H. M., LIN, C., WEEGINK, C., REESINK, H., ZEUZEM, S. & KWONG, A. D. 2007. Dynamic hepatitis C virus genotypic and phenotypic changes in patients treated with the protease inhibitor telaprevir. *Gastroenterology*, 132, 1767-77.
- SARRAZIN, C. & ZEUZEM, S. 2010. Resistance to direct antiviral agents in patients with hepatitis C virus infection. *Gastroenterology*, 138, 447-62.
- SCHINAZI, R., HALFON, P., MARCELLIN, P. & ASSELAH, T. 2014. HCV direct-acting antiviral agents: the best interferon-free combinations. *Liver International*, 34, 69-78.
- SCHIRMER, M., SLOAN, W. T. & QUINCE, C. 2014. Benchmarking of viral haplotype reconstruction programmes: an overview of the capacities and limitations of currently available programmes. *Briefings in Bioinformatics*, 15, 431-442.
- SCHNEIDER, M. D. & SARRAZIN, C. 2014. Antiviral therapy of hepatitis C in 2014: do we need resistance testing? *Antiviral Res*, 105, 64-71.

- SCHWARTZ, S., OREN, R. & AST, G. 2011. Detection and removal of biases in the analysis of next-generation sequencing reads. *PLoS One*, 6, e16685.
- SERPAGGI, J., CHAIX, M. L., BATISSE, D., DUPONT, C., VALLET-PICHARD, A., FONTAINE, H., VIARD, J. P., PIKETTY, C., ROUVEIX, E., ROUZIOUX, C., WEISS, L. & POL, S. 2006. Sexually transmitted acute infection with a clustered genotype 4 hepatitis C virus in HIV-1-infected men and inefficacy of early antiviral therapy. *Aids*, 20, 233-40.
- SHEPARD, C. W., FINELLI, L. & ALTER, M. J. 2005. Global epidemiology of hepatitis C virus infection. *Lancet Infect Dis*, 5, 558-67.
- SHEPHERD, S. J., ABDELRAHMAN, T., MACLEAN, A. R., THOMSON, E. C., AITKEN, C. & GUNSON, R. N. 2015. Prevalence of HCV NS3 pre-treatment resistance associated amino acid variants within a Scottish cohort. *J Clin Virol*, 65, 50-3.
- SHERMAN, K. E., FLAMM, S. L., AFDHAL, N. H., NELSON, D. R., SULKOWSKI, M. S., EVERSON, G. T., FRIED, M. W., ADLER, M., REESINK, H. W., MARTIN, M., SANKOH, A. J., ADDA, N., KAUFFMAN, R. S., GEORGE, S., WRIGHT, C. I. & POORDAD, F. 2011. Response-guided telaprevir combination treatment for hepatitis C virus infection. *N Engl J Med*, 365, 1014-24.
- SHERMAN, K. E., ROUSTER, S. D., STANFORD, S., BLACKARD, J. T., SHIRE, N., KOZIEL, M., PETERS, M., CHUNG, R. T. & TEAM, A. C. T. G. S. 2010. Hepatitis C virus (HCV) quasispecies complexity and selection in HCV/HIV-coinfected subjects treated with interferon-based regimens. *J Infect Dis*, 201, 712-9.
- SHI, S. T., LEE, K. J., AIZAKI, H., HWANG, S. B. & LAI, M. M. 2003. Hepatitis C virus RNA replication occurs on a detergent-resistant membrane that cofractionates with caveolin-2. *J Virol*, 77, 4160-8.
- SHIMAKAMI, T., WELSCH, C., YAMANE, D., MCGIVERN, D. R., YI, M., ZEUZEM, S. & LEMON, S. M. 2011. Protease inhibitor-resistant hepatitis C virus mutants with reduced fitness from impaired production of infectious virus. *Gastroenterology*, 140, 667-75.
- SHIRE, N. J., HORN, P. S., ROUSTER, S. D., STANFORD, S., EYSTER, M. E. & SHERMAN, K. E. 2006. HCV kinetics, quasispecies, and clearance in treated HCV-infected and HCV/HIV-1-coinfected patients with hemophilia. *Hepatology*, 44, 1146-57.
- SHIRE, N. J. & SHERMAN, K. E. 2005. Clinical trials of treatment for hepatitis C virus infection in HIV-infected patients: past, present, and future. *Clin Infect Dis*, 41 Suppl 1, S63-8.

- SHORES, N. J., MAIDA, I., SORIANO, V. & NUNEZ, M. 2008. Sexual transmission is associated with spontaneous HCV clearance in HIV-infected patients. *J Hepatol*, 49, 323-8.
- SHUHART, M. C., SULLIVAN, D. G., BEKELE, K., HARRINGTON, R. D., KITAHATA, M. M., MATHISEN, T. L., THOMASSEN, L. V., EMERSON, S. S. & GRETCH, D. R. 2006. HIV infection and antiretroviral therapy: effect on hepatitis C virus quasispecies variability. *J Infect Dis*, 193, 1211-8.
- SIMMONDS, P. 2001. Reconstructing the origins of human hepatitis viruses. *Philos Trans R Soc Lond B Biol Sci*, 356, 1013-26.
- SIMMONDS, P. 2004. Genetic diversity and evolution of hepatitis C virus--15 years on. *J Gen Virol*, 85, 3173-88.
- SIMMONDS, P., SMITH, D. B., MCOMISH, F., YAP, P. L., KOLBERG, J., URDEA, M. S. & HOLMES, E. C. 1994. Identification of genotypes of hepatitis C virus by sequence comparisons in the core, E1 and NS-5 regions. *J Gen Virol*, 75 (Pt 5), 1053-61.
- SINGAL, A. K. & ANAND, B. S. 2009. Management of hepatitis C virus infection in HIV/HCV co-infected patients: clinical review. *World J Gastroenterol*, 15, 3713-24.
- SMITH, D. B., BUKH, J., KUIKEN, C., MUERHOFF, A. S., RICE, C. M., STAPLETON, J. T. & SIMMONDS, P. 2014. Expanded classification of hepatitis C virus into 7 genotypes and 67 subtypes: updated criteria and genotype assignment Web resource. *Hepatology*, 59, 318-27.
- SMITH, J. A., ABERLE, J. H., FLEMING, V. M., FERENCI, P., THOMSON, E. C., KARAYIANNIS, P., MCLEAN, A. R., HOLZMANN, H. & KLENERMAN, P. 2010. Dynamic coinfection with multiple viral subtypes in acute hepatitis C. *J Infect Dis*, 202, 1770-9.
- SOBESKY, R., FERAY, C., RIMLINGER, F., DERIAN, N., DOS SANTOS, A., ROQUE-AFONSO, A. M., SAMUEL, D., BRECHOT, C. & THIERS, V. 2007. Distinct hepatitis C virus core and F protein quasispecies in tumoral and nontumoral hepatocytes isolated via microdissection. *Hepatology*, 46, 1704-12.
- SORIANO, V., VISPO, E., POVEDA, E., LABARGA, P., MARTIN-CARBONERO, L., FERNANDEZ-MONTERO, J. V. & BARREIRO, P. 2011. Directly acting antivirals against hepatitis C virus. *J Antimicrob Chemother*, 66, 1673-86.

- STEBBING, J., WATERS, L., MANDALIA, S., BOWER, M., NELSON, M. & GAZZARD, B. 2005. Hepatitis C virus infection in HIV type 1-infected individuals does not accelerate a decrease in the CD4+ cell count but does increase the likelihood of AIDS-defining events. *Clin Infect Dis*, 41, 906-11.
- STEINKUHLER, C., BIASIOL, G., BRUNETTI, M., URBANI, A., KOCH, U., CORTESE, R., PESSI, A. & DE FRANCESCO, R. 1998. Product inhibition of the hepatitis C virus NS3 protease. *Biochemistry*, 37, 8899-905.
- STRAHOTIN, C. S. & BABICH, M. 2012. Hepatitis C variability, patterns of resistance, and impact on therapy. *Adv Virol*, 2012, 267483.
- SULKOWSKI, M. S. 2014. Interferon-Containing and Interferon-Free HCV Therapy for HIV-Infected Patients. *Semin Liver Dis*, 34, 72-8.
- SULKOWSKI, M. S., ERON, J. J., WYLES, D., TRINH, R., LALEZARI, J., WANG, C., SLIM, J., BHATTI, L., GATHE, J., RUANE, P. J., ELION, R., BREDEEK, F., BRENNAN, R., BLICK, G., KHATRI, A., GIBBONS, K., HU, Y. B., FREDRICK, L., SCHNELL, G., PILOT-MATIAS, T., TRIPATHI, R., DA SILVA-TILLMANN, B., MCGOVERN, B., CAMPBELL, A. L. & PODSADECKI, T. 2015. Ombitasvir, paritaprevir co-dosed with ritonavir, dasabuvir, and ribavirin for hepatitis C in patients co-infected with HIV-1: a randomized trial. *Jama*, 313, 1223-31.
- SULKOWSKI, M. S., GARDINER, D. F., RODRIGUEZ-TORRES, M., REDDY, K. R., HASSANEIN, T., JACOBSON, I., LAWITZ, E., LOK, A. S., HINESTROSA, F., THULUVATH, P. J., SCHWARTZ, H., NELSON, D. R., EVERSON, G. T., ELEY, T., WIND-ROTOLO, M., HUANG, S. P., GAO, M., HERNANDEZ, D., MCPHEE, F., SHERMAN, D., HINDES, R., SYMONDS, W., PASQUINELLI, C. & GRASELA, D. M. 2014a. Daclatasvir plus sofosbuvir for previously treated or untreated chronic HCV infection. *N Engl J Med*, 370, 211-21.
- SULKOWSKI, M. S., NAGGIE, S., LALEZARI, J., FESSEL, W. J., MOUNZER, K., SHUHART, M., LUETKEMEYER, A. F., ASMUTH, D., GAGGAR, A., NI, L., SVAROVSKAIA, E., BRAINARD, D. M., SYMONDS, W. T., SUBRAMANIAN, G. M., MCHUTCHISON, J. G., RODRIGUEZ-TORRES, M., DIETERICH, D. & INVESTIGATORS, P.-. 2014b. Sofosbuvir and ribavirin for hepatitis C in patients with HIV coinfection. *JAMA*, 312, 353-61.
- SULLIVAN, J. C., DE MEYER, S., BARTELS, D. J., DIERYNCK, I., ZHANG, E. Z., SPANKS, J., TIGGES, A. M., GHYS, A., DORRIAN, J., ADDA, N., MARTIN, E. C., BEUMONT, M., JACOBSON, I. M., SHERMAN, K. E., ZEUZEM, S., PICCHIO, G. & KIEFFER, T. L. 2013. Evolution of treatment-emergent resistant variants in telaprevir phase 3 clinical trials. *Clin Infect Dis*, 57, 221-9.

- SULLIVAN, P. S., HANSON, D. L., TESHALE, E. H., WOTRING, L. L. & BROOKS, J. T. 2006. Effect of hepatitis C infection on progression of HIV disease and early response to initial antiretroviral therapy. *Aids*, 20, 1171-9.
- SUMPTER, R., JR., LOO, Y. M., FOY, E., LI, K., YONEYAMA, M., FUJITA, T., LEMON, S. M. & GALE, M., JR. 2005. Regulating intracellular antiviral defense and permissiveness to hepatitis C virus RNA replication through a cellular RNA helicase, RIG-I. *J Virol*, 79, 2689-99.
- SUSSER, S., WELSCH, C., WANG, Y., ZETTLER, M., DOMINGUES, F. S., KAREY, U., HUGHES, E., RALSTON, R., TONG, X., HERRMANN, E., ZEUZEM, S. & SARRAZIN, C. 2009. Characterization of resistance to the protease inhibitor boceprevir in hepatitis C virus-infected patients. *Hepatology*, 50, 1709-18.
- SUZUKI, R., SAKAMOTO, S., TSUTSUMI, T., RIKIMARU, A., TANAKA, K., SHIMOIKE, T., MORIISHI, K., IWASAKI, T., MIZUMOTO, K., MATSUURA, Y., MIYAMURA, T. & SUZUKI, T. 2005. Molecular determinants for subcellular localization of hepatitis C virus core protein. *J Virol*, 79, 1271-81.
- SVAROVSKAIA, E. S., DVORY-SOBOL, H., PARKIN, N., HEBNER, C., GONTCHAROVA, V., MARTIN, R., OUYANG, W., HAN, B., XU, S. M., KU, K., CHIU, S., GANE, E., JACOBSON, I. M., NELSON, D. R., LAWITZ, E., WYLES, D. L., BEKELE, N., BRAINARD, D., SYMONDS, W. T., MCHUTCHISON, J. G., MILLER, M. D. & MO, H. M. 2014. Infrequent Development of Resistance in Genotype 1-6 Hepatitis C Virus-Infected Subjects Treated With Sofosbuvir in Phase 2 and 3 Clinical Trials. *Clinical Infectious Diseases*, 59, 1666-1674.
- TAIWO, B. 2009. Understanding transmitted HIV resistance through the experience in the USA. *International Journal of Infectious Diseases*, 13, 552-559.
- TAMURA K, PETERSON D, PETERSON N, STECHER G, NEI M & S., K. 2011. MEGA5 : Molecular Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. *Molecular Biology and Evolution* 28, 2731-2739.
- TANAKA, Y., HANADA, K., HANABUSA, H., KURBANOV, F., GOJOBORI, T. & MIZOKAMI, M. 2007. Increasing genetic diversity of hepatitis C virus in haemophiliacs with human immunodeficiency virus coinfection. *J Gen Virol*, 88, 2513-9.
- TARGETT-ADAMS, P. & MCLAUCHLAN, J. 2005. Development and characterization of a transient-replication assay for the genotype 2a hepatitis C virus subgenomic replicon. *J Gen Virol*, 86, 3075-80.

- THOMAS, D. L. 2013. Global control of hepatitis C: where challenge meets opportunity. *Nat Med*, 19, 850-8.
- THOMAS, D. L., THIO, C. L., MARTIN, M. P., QI, Y., GE, D., O'HUIGIN, C., KIDD, J., KIDD, K., KHAKOO, S. I., ALEXANDER, G., GOEDERT, J. J., KIRK, G. D., DONFIELD, S. M., ROSEN, H. R., TOBLER, L. H., BUSCH, M. P., MCHUTCHISON, J. G., GOLDSTEIN, D. B. & CARRINGTON, M. 2009. Genetic variation in IL28B and spontaneous clearance of hepatitis C virus. *Nature*, 461, 798-801.
- THOMPSON, A., PATEL, K., TILLMAN, H. & MCHUTCHISON, J. G. 2009. Directly acting antivirals for the treatment of patients with hepatitis C infection: a clinical development update addressing key future challenges. *J Hepatol*, 50, 184-94.
- THOMSON, B. J. 2009. Hepatitis C virus: the growing challenge. *Br Med Bull*, 89, 153-67.
- THOMSON, E. C., FLEMING, V. M., MAIN, J., KLENERMAN, P., WEBER, J., ELIAHOO, J., SMITH, J., MCCLURE, M. O. & KARAYIANNIS, P. 2011. Predicting spontaneous clearance of acute hepatitis C virus in a large cohort of HIV-1-infected men. *Gut*, 60, 837-45.
- THOMSON, E. C., NASTOULI, E., MAIN, J., KARAYIANNIS, P., ELIAHOO, J., MUIR, D. & MCCLURE, M. O. 2009. Delayed anti-HCV antibody response in HIV-positive men acutely infected with HCV. *Aids*, 23, 89-93.
- THURNER, C., WITWER, C., HOFACKER, I. L. & STADLER, P. F. 2004. Conserved RNA secondary structures in Flaviviridae genomes. *J Gen Virol*, 85, 1113-24.
- THYS, K., VERHASSELT, P., REUMERS, J., VERBIST, B. M., MAES, B. & AERSSSENS, J. 2015. Performance assessment of the Illumina massively parallel sequencing platform for deep sequencing analysis of viral minority variants. *J Virol Methods*, 221, 29-38.
- TONG, X., BOGEN, S., CHASE, R., GIRIJAVALLABHAN, V., GUO, Z., NJOROGI, F. G., PRONGAY, A., SAKSENA, A., SKELTON, A., XIA, E. & RALSTON, R. 2008. Characterization of resistance mutations against HCV ketoamide protease inhibitors. *Antiviral Res*, 77, 177-85.
- TONG, X., LE POGAM, S., LI, L., HAINES, K., PISO, K., BARONAS, V., YAN, J. M., SO, S. S., KLUMPP, K. & NAJERA, I. 2014. In vivo emergence of a novel mutant L159F/L320F in the NS5B polymerase confers low-level resistance to the HCV polymerase inhibitors mericitabine and sofosbuvir. *J Infect Dis*, 209, 668-75.

- TORRIANI, F. J., RODRIGUEZ-TORRES, M., ROCKSTROH, J. K., LISSEN, E., GONZALEZ-GARCIA, J., LAZZARIN, A., CAROSI, G., SASADEUSZ, J., KATLAMA, C., MONTANER, J., SETTE, H., JR., PASSE, S., DE PAMPHILIS, J., DUFF, F., SCHRENK, U. M. & DIETERICH, D. T. 2004. Peginterferon Alfa-2a plus ribavirin for chronic hepatitis C virus infection in HIV-infected patients. *N Engl J Med*, 351, 438-50.
- TOWNSEND, K. S., OSINUSI, A., NELSON, A. K., KOHLI, A., GROSS, C., POLIS, M. A., PANG, P. S., SAJADI, M. M., SUBRAMANIAN, M., MCHUTCHISON, J. G., MASUR, H. & KOTTILIL, S. 2014. High Efficacy of Sofosbuvir/Ledipasvir for the Treatment of HCV Genotype 1 in Patients Coinfected With HIV on or off Antiretroviral therapy: Results from The NIAID ERADICATE Trial. *Hepatology*, 60, 240a-241a.
- TRIMOULET, P., BELZUNCE, C., FAURE, M., WITTKOP, L., REIGADAS, S., DUPON, M., RAGNAUD, J. M., FLEURY, H. & NEAU, D. 2011. Hepatitis C virus (HCV) protease variability and anti-HCV protease inhibitor resistance in HIV/HCV-coinfected patients. *HIV Med*, 12, 506-9.
- UEDA, E., ENOMOTO, N., SAKAMOTO, N., HAMANO, K., SATO, C., IZUMI, N. & WATANABE, M. 2004. Changes of HCV quasispecies during combination therapy with interferon and ribavirin. *Hepatol Res*, 29, 89-96.
- VALLET-PICHARD, A. & POL, S. 2006. Natural history and predictors of severity of chronic hepatitis C virus (HCV) and human immunodeficiency virus (HIV) co-infection. *J Hepatol*, 44, S28-34.
- VAN DE LAAR, T., PYBUS, O., BRUISTEN, S., BROWN, D., NELSON, M., BHAGANI, S., VOGEL, M., BAUMGARTEN, A., CHAIX, M. L., FISHER, M., GOTZ, H., MATTHEWS, G. V., NEIFER, S., WHITE, P., RAWLINSON, W., POL, S., ROCKSTROH, J., COUTINHO, R., DORE, G. J., DUSHEIKO, G. M. & DANTA, M. 2009. Evidence of a large, international network of HCV transmission in HIV-positive men who have sex with men. *Gastroenterology*, 136, 1609-17.
- VAN DE LAAR, T. J., MATTHEWS, G. V., PRINS, M. & DANTA, M. 2010. Acute hepatitis C in HIV-infected men who have sex with men: an emerging sexually transmitted infection. *Aids*, 24, 1799-812.
- VATTIPALLY B. SREENU, G. N., SULTAN ALOTAIBI, TAMER ABDELRAHMAN, KIRSTYN BRUNKER, RICHARD ORTON, TETYANA KLYMENKO, GAVIN WILKIE, ANDREW DAVISON, EMMA C. THOMSON 2015. TANOTI: a BLAST guided divergent read mapper for small genomes.
- VERBEECK, J., MAES, P., LEMEY, P., PYBUS, O. G., WOLLANTS, E., SONG, E., NEVENS, F., FEVERY, J., DELPORT, W., VAN DER MERWE, S. & VAN

- RANST, M. 2006. Investigating the origin and spread of hepatitis C virus genotype 5a. *J Virol*, 80, 4220-6.
- VERBINNEN, T., VAN MARCK, H., VANDENBROUCKE, I., VIJGEN, L., CLAES, M., LIN, T. I., SIMMEN, K., NEYTS, J., FANNING, G. & LENZ, O. 2010. Tracking the evolution of multiple in vitro hepatitis C virus replicon variants under protease inhibitor selection pressure by 454 deep sequencing. *J Virol*, 84, 11124-33.
- VOELKERDING, K. V., DAMES, S. A. & DURTSCHI, J. D. 2009. Next-generation sequencing: from basic research to diagnostics. *Clin Chem*, 55, 641-58.
- VOGEL, M., BIENIEK, B., JESSEN, H., SCHEWE, C. K., HOFFMANN, C., BAUMGARTEN, A., KROIDL, A., BOGNER, J. R., SPENGLER, U. & ROCKSTROH, J. K. 2005. Treatment of acute hepatitis C infection in HIV-infected patients: a retrospective analysis of eleven cases. *J Viral Hepat*, 12, 207-11.
- VOITENLEITNER, C., BECHTEL, J., ARFSTEN, A. & HAMATAKE, R. 2012. Hepatitis C genotype 1a replicon improved through introduction of fitness mutations. *Biotechniques*, 52, 273-5.
- WAJID, B. & SERPEDIN, E. 2012. Review of general algorithmic features for genome assemblers for next generation sequencers. *Genomics Proteomics Bioinformatics*, 10, 58-73.
- WALEWSKI, J. L., KELLER, T. R., STUMP, D. D. & BRANCH, A. D. 2001. Evidence for a new hepatitis C virus antigen encoded in an overlapping reading frame. *RNA*, 7, 710-21.
- WANG, C. F., SUN, J. H., O'BOYLE, D. R., NOWER, P., VALERA, L., ROBERTS, S., FRIDELL, R. A. & GAO, M. 2013. Persistence of Resistant Variants in Hepatitis C Virus-Infected Patients Treated with the NS5A Replication Complex Inhibitor Daclatasvir. *Antimicrobial Agents and Chemotherapy*, 57, 2054-2065.
- WANG, G. P., SHERRILL-MIX, S. A., CHANG, K. M., QUINCE, C. & BUSHMAN, F. D. 2010. Hepatitis C virus transmission bottlenecks analyzed by deep sequencing. *J Virol*, 84, 6218-28.
- WELSCH, C., JESUDIAN, A., ZEUZEM, S. & JACOBSON, I. 2012. New direct-acting antiviral agents for the treatment of hepatitis C virus infection and perspectives. *Gut*, 61 Suppl 1, i36-46.
- WHO. 2009. *Global summary of HIV Epidemic-2009* [Online]. Geneva. Available: <http://www.who.int/hiv/data/en/> [Accessed 10/10/2011].

- WHO. 2011. *Hepatitis C Fact sheet N°164* [Online]. Available: <http://www.who.int/mediacentre/factsheets/fs164/en/> [Accessed 10/10/2011].
- WILKE, C. O., WANG, J. L., OFRIA, C., LENSKI, R. E. & ADAMI, C. 2001. Evolution of digital organisms at high mutation rates leads to survival of the flattest. *Nature*, 412, 331-3.
- WOHNSLAND, A., HOFMANN, W. P. & SARRAZIN, C. 2007. Viral determinants of resistance to treatment in patients with hepatitis C. *Clin Microbiol Rev*, 20, 23-38.
- WONG, K. A., WORTH, A., MARTIN, R., SVAROVSKAIA, E., BRAINARD, D. M., LAWITZ, E., MILLER, M. D. & MO, H. 2013. Characterization of Hepatitis C Virus Resistance from a Multiple-Dose Clinical Trial of the Novel NS5A Inhibitor GS-5885. *Antimicrobial Agents and Chemotherapy*, 57, 6333-6340.
- WYLES, D. L. 2013. Antiviral resistance and the future landscape of hepatitis C virus infection therapy. *J Infect Dis*, 207 Suppl 1, S33-9.
- WYLES, D. L., POCKROS, P. J., YANG, J. C., ZHU, Y. N., PANG, P. S., MCHUTCHISON, J. G., FLAMM, S. L. & LAWITZ, E. 2014a. Retreatment of Patients Who Failed Prior Sofosbuvir-Based Regimens with All Oral Fixed-Dose Combination Ledipasvir/Sofosbuvir Plus Ribavirin for 12 Weeks. *Hepatology*, 60, 317a-318a.
- WYLES, D. L., RUANE, P. J., SULKOWSKI, M. S., DIETERICH, D., LUETKEMEYER, A., MORGAN, T. R., SHERMAN, K. E., DRETTLER, R., FISHBEIN, D., GATHE, J. C., JR., HENN, S., HINESTROSA, F., HUYNH, C., MCDONALD, C., MILLS, A., OVERTON, E. T., RAMGOPAL, M., RASHBAUM, B., RAY, G., SCARSELLA, A., YOZVIAK, J., MCPHEE, F., LIU, Z., HUGHES, E., YIN, P. D., NOVIELLO, S. & ACKERMAN, P. 2015. Daclatasvir plus Sofosbuvir for HCV in Patients Coinfected with HIV-1. *N Engl J Med*, 373, 714-25.
- WYLES, D. L., SULKOWSKI, M. S., ERON, J. J., TRINH, R., LALEZARI, J., SLIM, J., GATHE, J. C., WANG, C. C., ELION, R., BREDEEK, F., BRENNAN, R. O., BLICK, G., KHATRI, A., GIBBONS, K., HU, Y. R., FREDRICK, L., PILOT-MATIAS, T., DA SILVA-TILLMANN, B., MCGOVERN, B. H., CAMPBELL, A. L. & PODSADECKI, T. 2014b. TURQUOISE-I: 94% SVR12 in HCV/HIV-1 Coinfected Patients Treated with ABT-450/r/Ombitasvir, Dasabuvir and Ribavirin. *Hepatology*, 60, 1136a-1137a.
- WYLIE, K. M., MIHINDUKULASURIYA, K. A., SODERGREN, E., WEINSTOCK, G. M. & STORCH, G. A. 2012. Sequence analysis of the human virome in febrile and afebrile children. *PLoS One*, 7, e27735.

- XU, B., LIU, L., HUANG, X., MA, H., ZHANG, Y., DU, Y., WANG, P., TANG, X., WANG, H., KANG, K., ZHANG, S., ZHAO, G., WU, W., YANG, Y., CHEN, H., MU, F. & CHEN, W. 2011. Metagenomic analysis of fever, thrombocytopenia and leukopenia syndrome (FTLS) in Henan Province, China: discovery of a new bunyavirus. *PLoS Pathog*, 7, e1002369.
- YAMAGA, A. K. & OU, J. H. 2002. Membrane topology of the hepatitis C virus NS2 protein. *J Biol Chem*, 277, 33228-34.
- YAO, E. & TAVIS, J. E. 2005. A general method for nested RT-PCR amplification and sequencing the complete HCV genotype 1 open reading frame. *Virology*, 2, 88.
- YEH, B. I., HAN, K. H., LEE, H. W., SOHN, J. H., RYU, W. S., YOON, D. J., YOON, J., KIM, H. W., KONG, I. D., CHANG, S. J. & CHOI, J. W. 2002. Factors predictive of response to interferon-alpha therapy in hepatitis C virus type 1b infection. *J Med Virol*, 66, 481-7.
- YI, M. & LEMON, S. M. 2004. Adaptive mutations producing efficient replication of genotype 1a hepatitis C virus RNA in normal Huh7 cells. *J Virol*, 78, 7904-15.
- YOSHIMI, S., IMAMURA, M., MURAKAMI, E., HIRAGA, N., TSUGE, M., KAWAKAMI, Y., AIKATA, H., ABE, H., HAYES, C. N., SASAKI, T., OCHI, H. & CHAYAMA, K. 2015. Long term persistence of NS5A inhibitor-resistant hepatitis C virus in patients who failed daclatasvir and asunaprevir therapy. *Journal of Medical Virology*, 87, 1913-1920.
- ZAGORDI, O., GEYRHOFER, L., ROTH, V. & BEERENWINKEL, N. 2010a. Deep sequencing of a genetically heterogeneous sample: local haplotype reconstruction and read error correction. *J Comput Biol*, 17, 417-28.
- ZAGORDI, O., KLEIN, R., DAUMER, M. & BEERENWINKEL, N. 2010b. Error correction of next-generation sequencing data and reliable estimation of HIV quasispecies. *Nucleic Acids Res*, 38, 7400-9.
- ZEUZEM, S., ANDREONE, P., POL, S., LAWITZ, E., DIAGO, M., ROBERTS, S., FOCACCIA, R., YOUNOSSI, Z., FOSTER, G. R., HORBAN, A., FERENCI, P., NEVENS, F., MULLHAUPT, B., POCKROS, P., TERG, R., SHOUVAL, D., VAN HOEK, B., WEILAND, O., VAN HEESWIJK, R., DE MEYER, S., LUO, D., BOOGAERTS, G., POLO, R., PICCHIO, G. & BEUMONT, M. 2011. Telaprevir for retreatment of HCV infection. *N Engl J Med*, 364, 2417-28.
- ZEUZEM, S., DUSHEIKO, G. M., SALUPERE, R., MANGIA, A., FLISIAK, R., HYLAND, R. H., ILLEPERUMA, A., SVAROVSKAIA, E., BRAINARD, D. M., SYMONDS, W. T., SUBRAMANIAN, M., MCHUTCHISON, J. G., WEILAND,

O., REESINK, H. W., FERENCI, P., HEZODE, C., ESTEBAN, R. & INVESTIGATORS, V. 2014a. Sofosbuvir and Ribavirin in HCV Genotypes 2 and 3. *New England Journal of Medicine*, 370, 1993-2001.

ZEUZEM, S., JACOBSON, I. M., BAYKAL, T., MARINHO, R. T., POORDAD, F., BOURLIERE, M., SULKOWSKI, M. S., WEDEMEYER, H., TAM, E., DESMOND, P., JENSEN, D. M., DI BISCEGLIE, A. M., VARUNOK, P., HASSANEIN, T., XIONG, J., PILOT-MATIAS, T., DASILVA-TILLMANN, B., LARSEN, L., PODSADECKI, T. & BERNSTEIN, B. 2014b. Retreatment of HCV with ABT-450/r-ombitasvir and dasabuvir with ribavirin. *N Engl J Med*, 370, 1604-14.

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. 2013. Development of a sensitive RT-PCR method for amplifying and sequencing near full-length HCV genotype 1 RNA from patient samples. *Virology*, 53, 53.

ZHENG, Z. L., ADVANI, A., MELEFORS, O., GLAVAS, S., NORDSTROM, H., YE, W. M., ENGSTRAND, L. & ANDERSSON, A. F. 2010. Titration-free massively parallel pyrosequencing using trace amounts of starting material. *Nucleic Acids Research*, 38.

ZHONG, W., USS, A. S., FERRARI, E., LAU, J. Y. & HONG, Z. 2000. De novo initiation of RNA synthesis by hepatitis C virus nonstructural protein 5B polymerase. *J Virol*, 74, 2017-22.