Characterisation of Hepatitis C Virus antiviral resistance in the era of direct acting antivirals

David Smith

St Catherine's College

University of Oxford

A thesis presented for the degree of

Doctor of Philosophy of Clinical Medicine

Michaelmas Term 2019

I dedicate this thesis to my wife Kate.

Everyday your example inspires me to be a better man.

Characterisation of Hepatitis C Virus antiviral resistance in the era of direct acting antivirals

David Smith

St Catherine's College

Doctor of Philosophy in Clinical Medicine

Michaelmas Term, 2019

Hepatitis C Virus infects an estimated 71 million people world wide and accounts for approximately 399,000 deaths per year from end stage liver failure due to cirrhosis or hepatocellular carcinoma. The development of direct acting antivirals (DAAs) has led to the elimination of the HCV infection, known as a sustained viral response (SVR) in $>$95% of patients who are treated. However, resistance associated substitutions (RAS) have been described for all DAAs and the presence of these RAS have been shown by multiple studies to reduce SVR rates.

The aim of this work is to better characterise HCV RAS by studying the natural prevalence of RAS in a treatment naive cohort, to identify novel RAS and to study the prevalence of RAS in a cohort of patients who had not been successfully cured using first line DAA therapy and their effect on subsequent re-treatment.

RAS to NS5A inhibitors in treatment naive genotype 3 (gt3) HCV infected patients are common with $\sim$12% of patients in a treatment naive cohort having detectable RAS. All gt3b and gt3g viruses had the A30K + L31M RAS combination, which was shown to dramatically increase the EC50 of NS5A inhibitors.

Using viral genome wide association studies and whole genome HCV sequences the novel amino acid substitutions in NS2 I132V, NS2 119A NS3 67V and NS5B A150V were identified as being associated with sofosbuvir treatment failure in HCV gt3a infection.

SVR rates for the retreatment of patients with prior failure to direct acting antiviral therapy are high ($>$89%), despite RAS being present in 71% of patients. However, the SVR rate for gt3 infected patients is significantly lower only 81%.

This work demonstrates that despite SVR rates of $>$95% with DAA therapy, both innate resistance of understudied HCV subtypes and RAS present in treatment naive and experienced patients can effect treatment outcomes and should be viewed with concern particularly in patients who fail first-line therapy.

Declaration

I declare that the work presented in this thesis is my own and has not been submitted for any other degree in this or any other University or Institute of Learning.

Work completed by others: Clinical Data from the BOSON study used in chapters 2 and 3 was collected by BOSON Study team. RNA extraction was carried out by Anthony Brown at the Peter Medawar Building. Library construction, HCV enrichment and sequencing was performed by Amy Trebes, Palo Piazza and Rory Bowden at the Wellcome Centre for Human Genetics. The bioinformatics pipeline used to generate consensus HCV sequences and variant data was designed and built by David Bonsall, Camilla Ip and Azim Ansari. Andrea Magri taught me all the techniques used in the S52 replicon work and assisted with some RNA transfection and measurement of luciferase activity for the S52 In-vitro replicon work. The data from section 2.5.9 was generated by Dung Nguyen with help from myself. Clinical data for retreatment cohort from chapter 4 was collected by PHE and HCV Research UK. Sequencing for the retreatment cohort was performed at PHE, Glasgow and Oxford. Variant calling was performed by Josh Singer using HCV-GLUE.

Funding: The funding for my DPhil came from Oxford National Institute for Health Research Biomedical Research Centre research capability funding and the National Institute for Health Research Health Informatics Collaborative. This work was also supported by the Medical Research Council (MR/K01532X/1– STOP-HCV Consortium) and by core funding to the Wellcome Centre for Human Genetics provided by the Wellcome Trust (090532/Z/09/Z).

Acknowledgements

Throughout the process of completing the work for and the writing of this thesis, I have received a great deal of support and assistance from many people, without which I would not have been able to complete it. First I would like to thank my supervisors Professor Eleanor Barnes and Dr Azim Ansari, without who this work would not have been started let alone completed.

I must thank Tim Bradford, who first suggested I undertake this DPhil and arranged for the funding that allowed me to undertake it. I would also like to thank the rest of the Oxford Clinical Informatics Team and the National Institute for Health Research: Health Informatics Collaborative who continued to support me during my studies.

The in-vitro aspect of the thesis would not have been completed without support form Professor Peter Simmonds and his group in particular Dr Andrea Magri who taught me all the required lab techniques and Dr Dung Nguyen who's help on particular busy experiment days was invaluable. Thank you for all your guidance and help.

I would like to thank Professor Graham Foster and the BOSON Study team for collecting all the samples and data used in the first two chapters of this work. The BOSON data set is an incredible resource which has enabled a lot of great work.

The work from my last chapter would not have been possible without the cohorts collected by help Dr Daniel Bradshaw and Dr Tamyo Mbisa from the Public Health England Infection Service and Professors Will Irving and John McLauchlan from HCVRUK. Thank you for providing me with the opportunity to analyse such a great dataset. I must also mention Dr Josh

Singer who created HCV-GLUE which was used in this analysis and is an excellent resource for the community.

Finally, I would not have been able to complete this work, without the constant love and support from my family and in particular wife Kate who's constant support made over the years has allowed me to not just complete with thesis but enjoy the process.

Publications arising

Papers published from this work

- Pedergnana V., Smith D., STOP-HCV Consortium, Klenerman P., Barnes E., Spencer C. C. A., & Ansari M. A. (2016). Interferon lambda 4 variant rs12979860 is not associated with RAV NS5A Y93H in hepatitis C virus genotype 3a. Hepatology (Baltimore, Md.), 64(4), 1377–1378. https://doi.org/10.1002/hep.28533
- Smith D., Magri A., Bonsall D., Ip C. L., Trebes A., Brown A., ... Barnes E. (2018). Resistance analysis of genotype 3 HCV indicates subtypes inherently resistant to NS5A inhibitors. Hepatology (Baltimore, Md.), 69(5):1861-1872. https://doi.org/10.1002/hep.29837
- Nguyen D., Smith D., et al. (2019) Efficacy of NS5A inhibitors against under studied and potentially difficult-to-treat HCV subtypes commonly found in sub Saharan Africa and South East Asia. Journal of Hepatology, https://doi.org/10.1016/j.jhep.2020.05.029

Papers under review

- Smith D., et al. (2019). Hepatitis C Virus genome wide association study identifies novel sites associated with the outcome of sofosbuvir therapy.
- Smith D., et al. (2019). Effect of Resistance Associated Substitutions on Retreatment of HCV Infected Patients with Prior Failure to Direct Acting Antiviral Therapy.

Other related papers published

- Ansari M. A., Pedergnana V., Ip C., Magri A., Von Delft A., Bonsall D., Spencer C. C. A,

et al. (2017). Genome-to-genome analysis highlights the effect of the human innate and adaptive immune systems on the hepatitis C virus. Nat Genet. https://doi.org/10.1038/ng.3835

- Bonsall D., Black S., Howe A. Y. M., Chase R., Ingravallo P., Pak I., et al. (2018). Characterization of hepatitis C virus resistance to grazoprevir reveals complex patterns of mutations following on-treatment breakthrough that are not observed at relapse. Infection and Drug Resistance, (11), 1119–1135. https://doi.org/10.2147/IDR.S156581

- Downs L. O., Smith D. A., Lumley S. F., Patel M., McNaughton A. L., Mokaya J., et al. (2019). Electronic Health Informatics Data To Describe Clearance Dynamics of Hepatitis B Surface Antigen (HBsAg) and e Antigen (HBeAg) in Chronic Hepatitis B Virus Infection. MBio, 10(3). https://doi.org/10.1128/mBio.00699-19

- Wing P. A. C., Jones M., Cheung M., DaSilva S., Bamford C., Jason Lee W.-Y., et al. (2019). Amino Acid Substitutions in Genotype 3a Hepatitis C Virus Polymerase Protein Affect Responses to Sofosbuvir. Gastroenterology. https://doi.org/10.1053/j.gastro.2019.05.007

- Singer J., Thomson E., Hughes J., Aranday-Cortes E., McLauchlan J., da Silva Filipe A., et al. (2019). Interpreting Viral Deep Sequencing Data with GLUE. Viruses, 11(4), 323. https://doi.org/10.3390/v11040323

Contents

List of Figures

List of Tables

Abbreviations

HCV	Hepatitis C Virus
HCC	Hepatocellular Carcinoma
DAA	Direct Acting Antiviral
HIV	Human Immunodeficiency Virus
UTR	Untranslated Region
NS	Non Structural
gt	Genotype
INF	Interferon
SVR	Sustained Viral Response
RBV	Ribavirin
EASL	European Association for the Study of the Liver
AASLD	American Association for the Study of Liver Diseases
TVR	Telaprevir
BOC	Boceprevir
SOF	Sofosbuvir
SIM	Simeprevir
DCV	Daclatasvir
LDV	Ledipasvir
PTV	Paritaprevir
OBV	Ombitasvir
r	Ritonavir
DAS	Dasabuvir
GRZ	Grazoprevir
EBR	Elbasvir

VEL . . . . . Velpatasvir

GLE . . . . . Glecaprevir

PIB . . . . . Pibrentasvir

VOX . . . . . Voxilaprevir

RAS . . . . . Resistance Associated Substitution

RAV . . . . . Resistance Associated Variant

WT . . . . . Wild Type

EC50 . . . . . Half maximal effective concentration

PCR . . . . . Polymerase Chain Reaction

NGS . . . . . Next Generation Sequencing

WGS . . . . . Whole Genome Sequencing

IRES . . . . . Internal ribosome entry site

EMCV . . . . . Encephalomyocarditis virus

HCVcc . . . . . HCV cell cultured

IFNL4 . . . . . Interfron Lamda 4

SNP . . . . . Single nucleotide polymorphism

IL28B . . . . . Interleukin 28B

APASL . . . . . Asian Pacific Association for the Study of the Liver

RLU . . . . . Relative Light Unit

V-GWAS . . . . . Viral Genome Wide Association Study

FDR . . . . . False Discovery Rate

PHE . . . . . Public Health England

Chapter 1

Introduction and background

1.1 Hepatitis C Virus as a global health problem

By current World Health Organisation estimates, approximately 71 million people worldwide are chronically infected with the Hepatitis C Virus (HCV) which is equivalent to a global prevalence of 1% in 2015 [1]. The highest prevalence is in the Eastern Mediterranean and European regions with an estimated 15 and 14 million people with chronic HCV infection. This is equivalent to 2.3% and 1.5% of the population respectively [1]. HCV is responsible for approximately 399,000 deaths a year, through end stage liver failure due to cirrhosis or hepatocellular carcinoma (HCC) [1]. These numbers are thought to be an underestimate due to HCV infection not being recorded on death certificates. HCV infection was the leading cause of end-stage liver disease related liver transplantation [2]. However, the rate of liver transplantations due to HCV infection is now reducing [3] due to the high efficacy of Direct Acting Antiviral (DAA) therapy.

HCV is a blood born pathogen and there is currently no vaccine that can prevent HCV transmission. Before its discovery in 1989 [4] many people were infected with HCV via blood or blood product transfusions. The introduction of the heat treatment of blood products to inactivate viruses in the mid 1980s and the development of blood screening tests in the early 1990s have made this route of transmission rare in the developed world. However, HCV is still primarily spread via unsafe healthcare procedures in the developing world and in the developed world people who have injected drugs are at biggest risk of HCV infection. These routes of infection account for the majority of the estimated 1.75 million new infections in 2015 [1]. Other routes of infection include, vertical transmission from mother to child, piercing and tattooing. Additionally, Human immunodeficiency virus (HIV) positive men who have sex with men are at particularly high risk of sexual transmission [5]. The key risk groups in the UK for HCV infection are; people who have ever injected drugs, people who received a blood transfusion before 1991, people born or brought up in a country with a high prevalence of HCV, babies born to HCV infected mothers, prisoners, homeless people and HIV positive men who have sex with men [6].

Acute HCV infection is normally asymptomatic, therefore estimating the number of patients whom spontaneously resolve the infection is difficult. The rate of spontaneous resolution is

estimated to be between 20%-50%, with factors such as host genetics, race, age and sex being associated with spontaneous clearance [7, 8]. Individuals who do not spontaneously resolve HCV infection become chronically infected with HCV. Chronic HCV infection can remain asymptomatic for decades, but causes liver fibrosis and inflammation and as a consequence can cause cirrhosis and HCC [9]. High alcohol use, male sex, age and HIV co-infection all increase the risk of fibrosis, cirrhosis and HCC. Risk of cirrhosis after twenty years of infection has been estimated at 7-18% rising to 41% after 30 years [10]. Cirrhosis greatly increases the risk of developing HCC, but patients can develop HCC without having cirrhosis [9]. However, this is not a linear progression as fibrosis may accelerate with length of infection and some patients progress rapidly to HCC or end stage liver failure and some experience very mild symptoms.

1.2 Structure and life cycle of the Hepatitis C Virus

HCV is a $\sim$9.6kb long positive sense RNA virus. HCV is a member of hepacivirus genus which is one of four genera in the flaviviridae family the others being flavivirus, pegivirus and pestivirus [11]. The hepacivirus genus has 13 other species of virus which infect other non-primate mammals such as horses, dogs, rats and other rodents such as bank voles [11]. Some of the other hepaciviruses are used in animal models for vaccine challenge experiments [12, 13]. The flavivirus genus includes other human pathogens such as dengue, yellow fever, west nile and zika viruses [11]. The pegivirus genus includes two pegiviruses viruses that infect humans but do not appear to cause any disease [14–17], the first human pegivirus discovered was formally known as Hepatitis G Virus [14] but has since been renamed as it does not appear to cause disease [15]. The pestivirus genus includes viruses which infect and can cause severe disease in mammals such as cattle, sheep, goats and pigs [18–20].

The HCV genome codes for a 5' and 3' untranslated region (UTR) and one open reading frame and it is transcribed as one long polyprotein, $\sim$3000 amino acids long which is then cleaved into 10 proteins (figure 1.1). The core protein forms the main structure of the virion and the E1 and E2 proteins form the viral envelope. The p7 protein is a small viroporin which together with the nonstructural (NS) 2 protein is involved in viral release. The NS3, NS4A, NS4B, NS5A and NS5B protein create the viral replication complex. The NS3 and NS4A

proteins form the HCV protease complex responsible for cleaving the HCV polyprotein post transcription. The NS5A protein is known to have functions in the replication and assembly phases of the HCV lifecycle, but it's exact functions remain unclear [21]. The NS5B protein is the viral polymerase which creates new viral RNA.

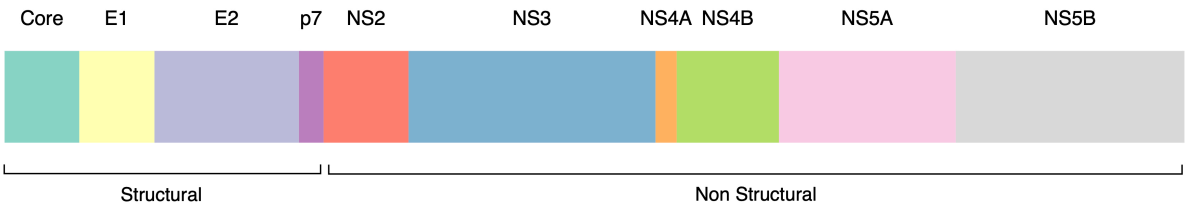

Figure 1.1: **Schematic representation of HCV genome.** The coloured rectangle represents the ∼3000 amino acid long HCV polyprotein. Each block is a scale representation of the HCV proteins. Core, E1, E2 and p7 are the structural proteins. NS2, NS3, NS4A, NS4B, NS5A and NS5B are the non-structural proteins involved in viral replication, assembly and release.

HCV replicates in hepatocytes and circulates in the blood. The cellular life cycle of HCV is broadly split into seven steps, attachment, entry, uncoating, translation, replication, assembly/maturation and release [22]. HCV virions circulate in the blood co-located with lipoproteins and together form a structure called the lipoviral particle [23, 24]. HCV enters the cell via cell membrane junctions and using cell surface markers such as CD81, claudin-1 and SR-B1 [25–27]. After the virion has entered the cytoplasm, the endosomes become acidified and the envelope protein conformation is changed which releases the viral RNA [28]. The RNA is then translated to the HCV polyprotein, via host proteases and the NS2, NS3 and NS4A protease complexes, the ten viral proteins are cleaved. At the endoplasmic reticulum the NS3, NS4A, NS4B, NS5A and NS5B proteins form membranous webs which act as an anchor for the HCV replication complex. The NS3 protein acts as a viral protease cleaving the viral polyprotein with the NS4A acting as a co-factor. The NS4B protein is an essential component in forming the membranous webs and in the assembly of new virus particles. The NS5B is a viral RNA dependent polymerase which creates new viral RNA. The NS5B polymerase does not have an error checking ability and this means a large portion of new viral RNA is attenuated and is incapable of further replication [29]. The newly created viral RNA is then packaged into viral capsids formed from the core, E1 and E2 proteins, these capsids are then released into the blood as lipoviral particles [30]. HCV can also infect neighbouring cells directly, this is mediated by the intercellular junction proteins claudin-1 and SR-B1.

1.3 Hepatitis C Virus diversity

HCV has a higher amount of genetic diversity compared to other persistent human pathogens such as HIV-1 [31–33]. HCV genotypes and subtypes are classified by phylogenetic clustering and a new genotype or subtype must form a new cluster on the tree of reference sequences maintained by the International Committee on Taxonomy of Viruses (ICTV) (Figure 1.2). In most cases the nucleotide sequence of a separate genotype diverges from other genotypes by $\sim$ 30% and a subtype differs by $\sim$15% from other subtypes [34]. The recent proliferation of high throughput methods for full genome sequencing of HCV [35, 36], has led to an increase in the amount of reported HCV genotypes and subtypes. Before 2014, HCV was classified into 6 genotypes and 67 subtypes [37]. In 2014 this was expanded to 7 reported genotypes and 75 subtypes [34]. In 2018 an additional genotype 8 (gt8) was isolated and described in Punjab, India [38] and currently HCV is split into 8 genotypes and 87 subtypes (Figure 1.2). Currently the vast majority of HCV sequences deposited in genbank are from patients in Europe and North America, so as sequencing technologies get cheaper and more viruses from Asia, Africa, South America and Australasia are sequenced it is likely that further genotype and subtypes will be discovered.

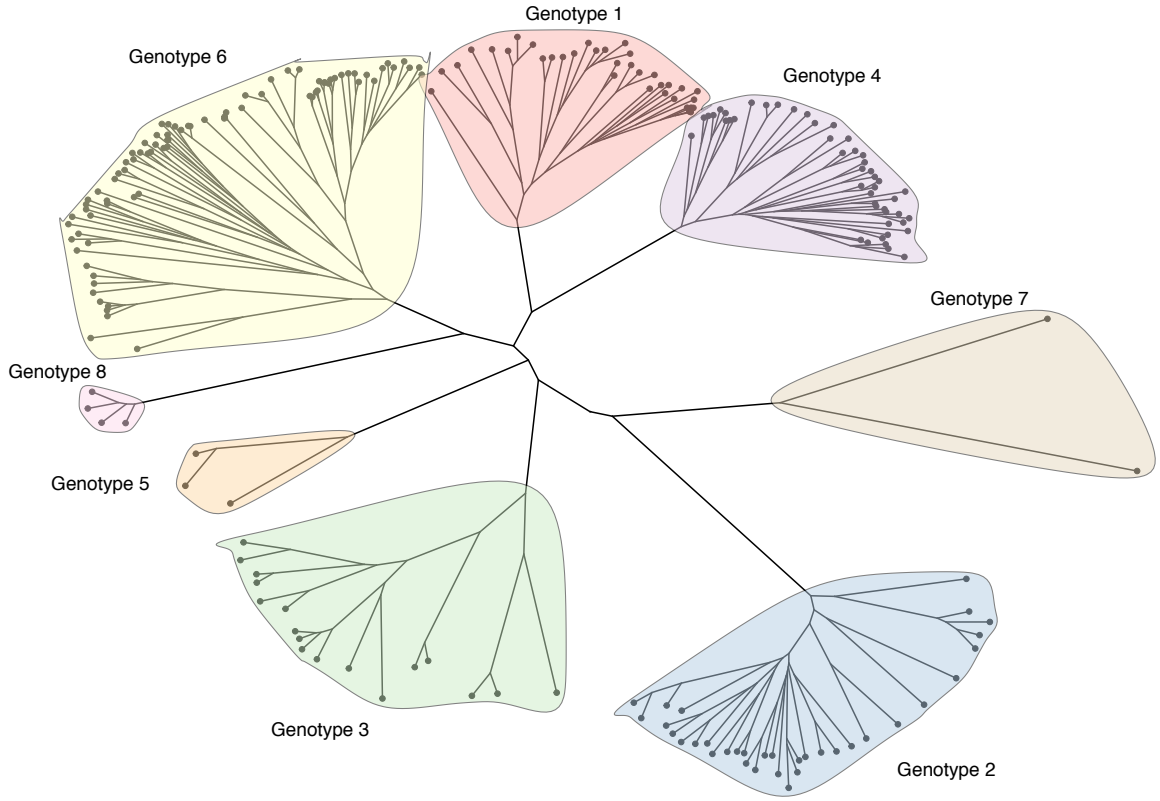

Figure 1.2: **Unrooted Maximum likelihood phylogenetic tree of HCV reference sequences.** Sequences downloaded from International Committee on Taxonomy of Viruses (ICTV) [39] on 27/06/19, tree estimated using RAxML [40].

gt1 is the most prevalent genotype worldwide with subtypes 1a and 1b being the most common worldwide [41]. Subtypes 2a and 2b are the most prevalent gt2 subtypes, they are distributed worldwide but highly prevalent in Western Africa. The second most prevalent gt is 3 with gt3a being the most prevalent subtype with a high prevalence in South Asia and Europe. In the Middle East, Northern and Eastern Africa, gt4 is highly prevalent. gt5 is relatively rare but is circulating in South Africa. The most genetically diverse gt is gt6 with 30 distinct subtype; gt6 has a high prevalence in South East Asia. Recently, gt7 was identified in Africa [34, 42]. The most recently discovered genotype was gt8 found in India in 2018 [38]. Although recombination is rare the inter-genotypic recombinant virus 1b/2k has infected multiple patients and is primarily found in Russia and Eastern Europe [43]. The 1a, 1b, 2a, 2b, and 3a subtypes are considered ”epidemic” subtypes as they are found globally with limited genetic diversity [41]. This suggests that the epidemic subtypes probably spread relativity recently due to modern efficient methods of transmission such as blood transfusions and injecting drug use in the 20th century [44, 45]. The other ”endemic” subtypes are found in

localised areas with more viral diversity, examples are the 67 gt6 subtypes found primarily in in south east asia and non gt-3a gt3 subtypes found in Asia [44, 46]. The increased diversity found in Asia suggests that the virus may have originated from there [45].

1.4 A history of Hepatitis C Virus treatment

Before the development of direct acting antivirals (DAAs), HCV infection treatment was based around interferon (IFNα). If HCV RNA remains undetectable after treatment cessation this is referred to as a sustained viral response (SVR). A patients HCV infection is considered cured if an SVR is maintained until 12 weeks after treatment has ended. If a patient's viral load never becomes undetectable then this is considered a non-response. If the viral load increases after treatment cessation this is considered a relapse. If the viral load becomes undetectable then becomes detectable on treatment, this is considered a breakthrough.

12 weeks is the normal period of time after which if HCV RNA remains undetectable after the end of treatment that an SVR can be said to have taken place. Treatment for Non-A Non-B hepatitis using IFNα mono-therapy was used even before HCV was isolated [47]; this treatment was not very effective with SVR rates of 10-25%. The addition of Ribavirin (RBV) to IFNα therapy increased SVR rates to approximately 40% [48]. RBV is a nucleotide analogue prodrug, with an antiviral effect against a broad spectrum of RNA viruses. There are several mechanisms of action proposed for ribavirin, some direct acting against the virus such as preventing RNA capping, polymerase inhibition and lethal mutagenesis, others indirect such as immune modulation [49]. Pegylated interferon (PEG-INF) was introduced in 2001 and the increased half-life and better pharmacokinetics increased SVR rates to approximately 60% when used with RBV [50]. However, SVR rates were varied with viral load, viral genotype and host genetics all playing a role in treatment outcomes. IFN based therapies have multiple effects on the host immune response, but also target the virus directly and is associated with significant side effects such as asthenia, neutropenia, cytopenias, depression and flu like symptoms. PEG-INFα-2b + RBV was the standard of care for HCV infection until 2011 when the first DAAs became available. Even after the introduction of DAAs, PEG-INFα-2b + RBV remained a key component of HCV treatment until the development of all-oral combinations of DAA's such as daclatasvir + sofosbuvir (DCV+SOF) and ledipasvir / sofosbuvir (LDV/SOF)

in 2014.

For clarity combinations of drugs separated by a + sign are administered as separate drugs in combination (for example DCV + SOF). However, when separated by a / they are administered as a fixed dose combination and are not available for use as individual compounds (for example LDV/SOF).

1.4.1 The development of direct acting antivirals against the Hepatitis C Virus

The development of DAAs, since 2011 has lead to a revolution in the treatment of HCV infection. Treatment durations have been dramatically reduced from 44 weeks or longer with PEG-INFα-2b and RBV to only 8 weeks for treatment naive patients without cirrhosis [51]. Current DAA therapies are all IFN free and are commonly RBV free, dramatically reducing the number of reported side effects since DAA based therapies are very well tolerated. The main benefit of DAA therapy is the dramatic increase in SVR rates to $>$95% in most groups of patients.

There have been a large number of compounds developed with many reaching phase 2 and 3 trials and a relatively small number now approved for clinical use in America and Europe (table 1.1). All DAAs fall into one of four classes each targeting a different viral protein or having a distinct mechanism of action. Protease inhibitors target the NS3 protein, binding to its active site and preventing it from cleaving the viral polyprotein. NS5A inhibitors bind to domain 1 of the NS5A protein and inhibits its action. NS5B inhibitors target the HCV polymerase and are split into two categories based on the mechanisms of action: Nucleotide analogues of which sofosbuvir is the primary example work by mimicking a nucleotide and being integrated into the viral RNA causing termination of the RNA chain; non-nucleotide analogues work by by binding to one of the palm domains of the NS5B protein and inhibiting its action.

The approved DAA treatments can be broadly categorised into three generations. The first generation of DAAs were used in combination with PEG-INF and RBV rather than in combination with other DAAs and were usually effective against only one or two HCV genotypes. First generation DAAs are no longer recommended for use by either the European Association for

the Study of the Liver (EASL) [51] or the American Association for the Study of Liver Diseases (ASSLD) , [52]. The second generation of DAAs are used in fixed dose combinations with each compound targeting a different HCV protein. They are INF free, but are still often used with RBV. Second generation compounds are also genotype specific, but generally cover a wider spectrum of genotypes than first generation DAAs. Most of the second generation DAAs are still in use and are highly effective [51, 52]. The third generation of DAA treatments are pangenotypic meaning this generation of DAAs are active against all HCV genotypes and are often used without RBV.

Class	Name	Code	Manufacture	Development Phase	Approval Year
NS3 Protease Inhibitors	Paritaprevir	ABT-450	Abbvie	Approved	2015
	Grazoprevir	MK-5172	Merck	Approved	2016
	Simeprevir	TMC-435	Janssen	Approved	2013
	Voxilaprevir	GS-9857	Gilead Sciences	Approved	2017
	Glecaprevir	ABT-493	Abbvie	Approved	2017
	Vaniprevir	MK-7009	Merck	Approved (Japan Only)	2014
	Asunaprevir	BMS-650032	Bristol-Myers Squibb (BMS)	Approved (Asia and Middle East Only)	2015
	Faldaprevir	BI-201335	Boehringer-Ingelheim	Development Stopped	NA
	Boceprivir	SCH-503034	Merck	Discontinued	2011
	Telaprivir	VX-950	Vertex	Discontinued	2011
	Sovaprevir	ACH-1625	Janssen and Achillion	Development Stopped	NA
NS5A Inhibitors	Ledipasvir	GS-5885	Gilead Sciences	Approved	2014
	Ombitasvir	ABT-267	Abbvie	Approved	2014
	Elbasvir	MK-8742	Merck	Approved	2016
	Daclatasvir	BMS-790052	Bristol-Myers Squibb (BMS)	Approved	2015
	Velpatasvir	GS-5816	Gilead Sciences	Approved	2016
	Pibrentasvir	ABT-530	Abbvie	Approved	2017
	Ruszasvir	MK-8408	Merck	Phase 3	NA
	Odalasvir	ACH-3102	Janssen and Achillion	Phase 2	NA
	Ravidasvir	PPI-668	Presidio	Phase 2	NA
Nucleotide Analogues	Sofosbuvir	GS-7977	Gilead Sciences	Approved	2013
	ACH-3422	ACH-3422	Janssen and Achillion	Phase 2	NA
	AL-335	AL-335	Janssen and Achillion	Phase 2	NA
	Uprifosbuvir	MK-3682	Merck	Phase 3	NA
Non-Nucleotide Analogues	Dasabuvir	ABT-333	Abbvie	Approved	2014
	Delobuvir	BI-207127	Boehringer-Ingelheim	Development Stopped	NA
	Beclabuvir	BMS-791325	Bristol-Myers Squibb (BMS)	Phase 3	NA

Table 1.1: **Summary of direct acting antivirals which have reached phase 2 trials since 2011.** NA:Non Applicable

1.4.2 First generation: single drugs combined with pegylated interferon

The first DAAs to be developed and approved for clinical use were the protease inhibitors telaprivir (TVR) and boceprevir (BOC). These were all used in combination with PEG-INF and RBV. The phase III SPRINT-2 trial [53] tested PEG-INF+RBV + BOC. This study had 3 arms, all started with 4 weeks of PEG-INF+RBV. The control arm continued with PEG-INF+RBV + placebo until 48 weeks. The second arm used response guided therapy if a patient had undetectable vial RNA at 8 and 24 weeks therapy was stopped at 28 weeks, if not they continued with PEG-INF+RBV until 48 weeks. The third arm used PEG-INF+RBV + BOC for 44 weeks. The control arm had an SVR rate of 38%. The treatment arms had rates of 63% and 66% respectively.

The ADVANCE trial [54] was a phase III trial for TVR. PEG-INF+RBV for 48 weeks was used for the control arm. Patients received 8 or 12 week of PEG-INF+RBV + TVR followed by PEG-INF+RBV until 24 weeks if HCV RNA was undetectable or PEG-INF+RBV until 48 weeks if it was. The SVR rates were 44% for the control arm, 69% for the 8 week arm and 75% for the 12 week arm.

In both SPRINT-2 and ADVANCE patients who had an undetectable viral load at 4 weeks which remained undetectable until the end of the protease inhibitor therapy had the best SVR rates. This lead to the recommendation of response guided therapy for first generation protease inhibitors. Whilst first generation protease inhibitors significantly increased SVR rates and reduced the length of treatment, they were limited to gt1 infected non-decompensated patients [55], for non-gt1 infected patients PEG-INF+RBV remained the standard of care.

Simeprevir (SMV) was the next protease inhibitor developed again for the treatment of gt1 infection and in combination with PEG-INF and RBV. The QUEST-2 study [56] showed that simeprevir had a SVR rate of 81% when given for 24 weeks compared to 50% for PEG-INF plus RBV for 48 weeks, with no difference in side effects between the two groups.

Sofosbuvir (SOF) is a uridine analogue pro-drug which competitively binds with the HCV NS5B polymerase. It is the only NS5B inhibitor in the nucleotide analogue class which has been approved for clinical use. It is active against all HCV genotypes and has a high barrier

to resistance. These attributes have seen it used in all three generations of DAA treatments combination with PEG-INF and or RBV, or in interferon and or RBV free combinations. The phase III NEUTRINO and FISSION studies [57] used SOF + PEG-INF + RBV for 12 weeks. NEUTRINO was primarily comprised of gt1 and gt4 infected patients, with limited numbers of gt5 and gt6 patients. FISSON was a non-inferiority study between SOF + RBV and PEG-INF + RBV for 24 weeks, in gt2 and 3 infected patients. The NEUTRINO study showed SVR Rates of 90%. However, the FISSON study reported no difference in SVR Rates between the two arms. The BOSON study [58] which tested SOF + PEG-INF + RBV for 12 weeks against, SOF + RBV for 16 or 24 weeks, showed that SOF + PEG-INF + RBV to be superior to either of the non-PEG-INF arms for gt2 and gt3 infected patients.

SOF + SIM was only recommended for use in gt1a, gt1b and gt4 infected patients. The COSMOS study [59] showed SVR rates $>$93% after 12 and 24 weeks of treatment with SOF + SIM in gt1 infected patients. The PLUTO study [60] showed an SVR rate of 100% in gt4 infected patients.

1.4.3 Second generation: interferon free combinations

Daclatasvir (DCV) was the first NS5A inhibitor approved for clinical use in combination with SOF. The ALLY-1 trial [61] showed SVR rates of 98% for DCV + SOF for 12 weeks for patients with gt1 infection and no prior treatment and 92% for patients with prior exposure to first generation protease inhibitors. gt3 infected patients with no prior exposure to first generation protease inhibitors faired slightly worse in the ALLY-1 trial with an SVR rate of 89%. The ALLY-3 [62] trial focused on patients with gt3 infection. The SVR rates for gt3 infected patients were lower than for gt1 infected patients with treatment naive 90% and treatment experienced 89%. Despite the reduced efficacy against gt3 infection SOF + DCV remained the only interferon free combination recommend for gt3 infection until the approval of sofsobuvir/velpatasvir and glecaprevir/pibrentasvir in 2016.

Ledipasvir (LDV) was the next NS5A inhibitor approved for the treatment of HCV infection, in a fixed dose combination with sofosbuvir. The use of LDV/SOF for the treatment of gt1 infected patients has been extensively studied in the ION-1,2,3 and 4 trials, as well as large

real life cohorts, where it shows SVR rates >95% [63–66]. The SYNERGY trial [67] showed an SVR rate of 95% in 21 patients with gt4 infection. Small numbers of gt2, gt3, gt5 and gt6 patients have been studied [68]. Therefore, SOF/LDV is recommended for gt5 and gt6 infection if the patient has not received prior therapy. The combination has not been shown to be effective in gt2 and gt3 based in-vitro assays. Furthermore, in cases were gt3 patients have been treated such as the Early Access Program [69] in the UK it has been shown not to be effective, therefore it is not recommend for the treatment of gt2 or gt3 infection.

The combination of the NS3 inhibitor paritaprevir (PTV), NS5A inhibitor ombitasvir (OMB) and ritonavir (r) has the trade name Viekirax. The same combination with the addition of the NS5B inhibitor dasabuvir (DAS) has the trade name Viekira Pak. For treatment naive, non cirrhotic patients, the SHAPHIRE-I [70] and PEARL-IV [71] trials both showed SVR rates of 95% and 97% against gt1a infection when used with RBV. The PEARL-III [71] trial showed SVR rates of 100% against gt1b infection with and without RBV. The PEARL-II [72] and the SHAPHIRE-II [73] studies also showed similar SVR rates in patients who had previously had HCV treatment with SVR Rates of 96-100% for both gt1a and gt1b infected patients. TURQUOISE-II [74] and TURQUOISE-III [75] showed an SVR rate of 92-100% for patients with compensated cirrhosis, with gt1b patients generally responding better. The PEARL-I [76] trial achieved SVR rates of 100% when using OMB/PTVr + RBV for gt4 infected patients without cirrhosis with and without treatment experience [77]. OMB/PTV/r + DAS was approved for the treatment of gt1a and gt1b infection in 2014 and OMB/PTV/r for gt4 infection in 2015.

The combination of the NS3 inhibitor grazoprevir (GZR) and the NS5A inhibitor elbasvir (ELB) is known as Zepatier. ELB/GZR currently only recommended by EASL for the treatment of gt1a and gt1b infections with the caveats of viral load being ≤800,000 IU/ml for gt1a and therapy being extended for gt1b patients with a fibrosis greater than F2 [51]. This recommendation is based on the data from the C-EDGE-TN [78], C-EDGE-COINFECTION [79], C-EDGE-TE [80], STREAGER [81] and C-CORAL [82] studies. Which showed SVR rates >90% with patients who are treatment experienced, treatment naive, HIV co-infected and/or have compensated cirrhosis. However, NS5A RAS were shown to have an effect in patients with a viral load >800,000 IU/ml (data un-published but referenced in EASL guidelines [51]). This led to the recommendation from EASL that ELB/GZR is only to be used in patients with a viral load lower than >800,000 IU/ml. AASLD recommends resistance testing before treat-

ment with ELB/GZR and an extension of therapy from 12 to 16 weeks if the NS5A resistance substitution Y93H is found. C-EDGE-TN [78] and C-CORAL [82] trials had small numbers of gt4 and gt6 infected patients and GZR/ELB was effective against these genotypes.

1.4.4 Third generation: pangenotypic, interferon and RBV free

The first pan-genotypic regimen to be approved for use was the once daily fixed dose combination of sofosbuvir and velpatasvir (SOF/VEL) with the trade name Epclusa. The ASTRAL-1 study [83] showed a SVR rate of 99% against gt1,2,4,5,6 after 12 weeks of therapy, with 19% of the enrolled patients having compensated cirrhosis. The numbers for gt5 and 6 were comparatively low in the study but both groups achieved 100% SVR rates. ASTRAL-3 [58] a study using SOF/VEL for 12 weeks to treat gt2 and gt3 patients with or without compensated cirrhosis. 99% of the gt2 patients achieved a SVR. 95% of gt3 patients achieved a SVR, but previous treatment and the presence of cirrhosis were associated with reduced SVR rates. Treatment naive patients without cirrhosis had an SVR rate of 98%, but treatment naive patients with cirrhosis had an SVR rate of 93%. Treatment experience patients without cirrhosis had a SVR rate of 91% and those with cirrhosis had an SVR rate of 89%. As a result EASL and AASLD do not recommend SOF/VEL for the treatment of gt3 infected cirrhotic patients [51].

The NS3 protease inhibitor glecaprevir (GLE) and the NS5A inhibitor pibrentasvir (PIB) make up the fixed dose combination (GLE/PIB) and have the commercial name MAVYRET. GLE/PIB is pan-genotypic and both the EASL [51] and AASLD [52] guidelines recommend 8 weeks of treatment for patients without cirrhosis with the exception of gt3 infected patients, where 12 weeks is recommenced by EASL and 8 weeks is recommend by AASLD. ENDURANCE-1 [84] showed SVR rates of 99% in non-cirrhotic patients with gt1 infection when treated for 8 weeks with GLE/PIB and 99.7% for patients treated for 12 weeks. EXPEDITION-2 [85] also showed SVR rates of 99% in patients with compensated cirrhosis and gt 1,2,4,5 or 6 infection after 12 weeks of therapy. ENDURANCE-3 [84] showed a SVR rate of 95% for patients with gt3 infection without cirrhosis and SURVERYOR-2 [86] showed that 98% of gt3 infected cirrhotic patients achieved SVR with 12 weeks of GLE/PIB treatment.

The addition of the NS3 protease inhibitor voxilaprevir to SOF/VEL increases its potency

and makes it an excellent pan-genotypic retreatment regimen. The treatment guidelines from EASL [51] and ASSLD [52] reflect this where it is only recommended for first line treatment of gt3 infected cirrhotic patients. Otherwise, it is kept in reserve as a retreatment option for patients who do not clear HCV after a first-line therapy. The POLARIS-1 [87] study included patients who had previously received an NS5A inhibitor containing regimen, one arm was treated with SOF/VEL/VOX, the other was given a placebo. The overall SVR rate was 96% in the SOF/VEL/VOX arm. The SVR rate for gt1a was 96%, 100% for gt1b, 100% for gt2, 95% for gt3 patients, 91% for gt4 and 100% for gt6. The POLARIS-4 [87] trial compared SOF/VEL/VOX against SOF/VEL for patients who had received a DAA regimen not including a NS5A inhibitor. The SVR rate was 98% and 94% respectively for patients without cirrhosis and 98% to 86% for patients with cirrhosis.

1.4.5 Current guidelines for the treatment of Hepatitis C Virus infection

A summary of the current guidelines for the treatment of HCV from EASL [51] and the AASLD. [52] can be found in Tables 1.3, 1.4. In these guidelines, treatment-naïve patients are defined as patients who have never been treated for their HCV infection and treatment-experienced patients defined as patients who were previously treated with pegylated IFN-a and RBV; pegylated IFN-a, RBV and sofosbuvir; or sofosbuvir and RBV. The EASL and AASLD recommendations are broadly similar, recommending the use of pangenotypic regimens such as GLE/PIB and SOF/VEL as first line treatment for patients with and without cirrhosis. The AASLD guidelines include more alternative therapies and recommends pre RAS testing for GZR/EBR based therapies, where as the EASL guidelines choose to implement a viral load limit of $<$800,000 IU/ml for a reduced treatment length.

Genotype	Treatment Experience	SOF/VEL	GLE/PIB	SOF/VEL/VOX	SOF/LDV	GZR/EBR	OBV/PTV/r + DSV
	Patients without Cirrhosis						
1a	Treatment-Naïve	12wk	8wk	NR	8-12wk	12wk*	NR
	Treatment-Experienced	12wk	8wk	NR	NR	12wk*	NR
1b	Treatment-Naïve	12wk	8wk	NR	8-12wk	8wk (F0-F2) 12wk (F3)	8wk (F0-F2) 12wk (F3)
	Treatment-Experienced	12wk	8wk	NR	12wk	12wk	12wk
2	Treatment-Naïve	12wk	8wk	NR	NR	NR	NR
	Treatment-Experienced	12wk	8wk	NR	NR	NR	NR
3	Treatment-Naïve	12wk	8wk	NR	NR	NR	NR
	Treatment-Experienced	12wk	12wk	NR	NR	NR	NR
4	Treatment-Naïve	12wk	8wk	NR	12wk	12wk*	NR
	Treatment-Experienced	12wk	8wk	NR	NR	NR	NR
5	Treatment-Naïve	12wk	8wk	NR	12wk	NR	NR
	Treatment-Experienced	12wk	8wk	NR	NR	NR	NR
6	Treatment-Naïve	12wk	8wk	NR	12wk	NR	NR
	Treatment-Experienced	12wk	8wk	NR	NR	NR	NR
	Patients with Cirrhosis						
1a	Treatment-Naïve	12wk	12wk	NR	8-12wk	12wk*	NR
	Treatment-Experienced	12wk	12wk	NR	NR	12wk*	NR
1b	Treatment-Naïve	12wk	12wk	NR	8-12wk	12wk	12wk
	Treatment-Experienced	12wk	12wk	NR	12wk	12wk	12wk
2	Treatment-Naïve	12wk	12wk	NR	NR	NR	NR
	Treatment-Experienced	12wk	12wk	NR	NR	NR	NR
3	Treatment-Naïve	NR	12wk	12wk	NR	NR	NR
	Treatment-Experienced	NR	16wk	12wk	NR	NR	NR
4	Treatment-Naïve	12wk	12wk	NR	12wk	12wk*	NR
	Treatment-Experienced	12wk	12wk	NR	NR	NR	NR
5	Treatment-Naïve	12wk	12wk	NR	12wk	NR	NR
	Treatment-Experienced	12wk	12wk	NR	NR	NR	NR
6	Treatment-Naïve	12wk	12wk	NR	12wk	NR	NR
	Treatment-Experienced	12wk	12wk	NR	NR	NR	NR

Table 1.3: **European Association for Study of the Liver (EASL) 2018 guidelines for the treatment of HCV infection.** Treatment-naïve patients defined as patients who have never been treated for their HCV infection and treatment-experienced patients defined as patients who were previously treated with pegylated IFN-a and RBV; pegylated IFN-a, RBV and sofosbuvir; or sofosbuvir and RBV. DSV, dasabuvir; EBR, elbasvir; GLE, glecaprevir; GZR, grazoprevir; HCV, hepatitis C virus; HIV, human immunodeficiency virus; LDV, ledipasvir; NR, Not Recommended ;OBV, ombitasvir; PIB, pibrentasvir; PTV, paritaprevir; r, ritonavir; SOF, sofosbuvir; VEL, velpatasvir; VOX: voxilaprevir. * Only recommended if HCV RNA $\leq$800,000 IU/ml

Genotype	Treatment Experience	SOF/VEL	GLE/PIB	SOF/VEL/VOX	SOF/LDV	GZR/EBR	OBV/PTV/r + DSV	DAC + SOF	GZR/EBR + SOF	SIM + SOF
Patients without Cirrhosis										
1a	Treatment-Naïve	12wk	8wk	NR	8wks** 12wks	12wk 16wk****	12wk + RBV*	12wk*	NR	12wk*
	Treatment-Experienced	12wk	12wk	NR	12wk	12wk 16wk****	12wk + RBV*	12wk*	NR	12wk*
1b	Treatment-Naïve	12wk	8wk	NR	8wks** 12wks	12wk	12wk + RBV*	12wk*	NR	12wk*
	Treatment-Experienced	12wk	8wk	NR	8wks** 12wks	12wk	12wk + RBV*	12wk*	NR	12wk*
2	Treatment-Naïve	12wk	8wk	NR	NR	NR	NR	12wk*	NR	NR
	Treatment-Experienced	12wk	8wk	NR	NR	NR	NR	12wk*	NR	NR
3	Treatment-Naïve	12wk	8wk	NR	NR	NR	NR	12wk*	NR	NR
	Treatment-Experienced	12wk	16wk*	12wk* (if Y93H Present)	NR	NR	NR	12wk*	NR	NR
4	Treatment-Naïve	12wk	8wk	NR	12wk	12wk	12wk*	NR	NR	NR
	Treatment-Experienced	12wk	8wk	NR	12wk	12wk 16wk +RBV***	12wk*	NR	NR	NR
5	Treatment-Naïve	12wk	8wk	NR	12wk	NR	NR	NR	NR	NR
	Treatment-Experienced	12wk	8wk	NR	12wk	NR	NR	NR	NR	NR
6	Treatment-Naïve	12wk	8wk	NR	12wk	NR	NR	NR	NR	NR
	Treatment-Experienced	12wk	8wk	NR	12wk	NR	NR	NR	NR	NR
Patients with Cirrhosis										
1a	Treatment-Naïve	12wk	12wk	NR	12wks	12wk 16wk****	NR	12wk*	NR	NR
	Treatment-Experienced	12wk	12wk	NR	12wk + RBV*	12wk 16wk****	NR	12wk*	NR	NR
1b	Treatment-Naïve	12wk	12wk	NR	12wks	12wk	12wk + RBV*	12wk*	NR	NR
	Treatment-Experienced	12wk	12wk	NR	12wk + RBV*	12wk	12wk + RBV*	12wk*	NR	NR
2	Treatment-Naïve	12wk	12wk	NR	NR	NR	NR	16-24wk*	NR	NR
	Treatment-Experienced	12wk	12wk	NR	NR	NR	NR	16-24wk*	NR	NR
3	Treatment-Naïve	12wk	12wk	12wk* (if Y93H Present)	NR	NR	NR	24 +/- RBV *	NR	NR
	Treatment-Experienced	12wk + RBV*	16wk*	12wk	NR	NR	NR	NR	12wk	NR
4	Treatment-Naïve	12wk	12wk	NR	12wk	12wk	12wk*	NR	NR	NR
	Treatment-Experienced	12wk	12wk	NR	12wk + RBV*	12wk 16wk +RBV***	12wk + RBV*	NR	NR	NR
5	Treatment-Naïve	12wk	12wk	NR	12wk	NR	NR	NR	NR	NR
	Treatment-Experienced	12wk	12wk	NR	12wk	NR	NR	NR	NR	NR
6	Treatment-Naïve	12wk	12wk	NR	12wk	NR	NR	NR	NR	NR
	Treatment-Experienced	12wk	12wk	NR	12wk	NR	NR	NR	NR	NR

Table 1.4: **American Association for the Study of Liver Diseases 2018 guidelines for the treatment of HCV infection.** Treatment-naïve patients defined as patients who have never been treated for their HCV infection and treatment-experienced patients defined as patients who were previously treated with pegylated IFN-a and RBV; pegylated IFN-a, RBV and sofosbuvir; or sofosbuvir and RBV. DSV, dasabuvir; EBR, elbasvir; GLE, glecaprevir; GZR, grazoprevir; HCV, hepatitis C virus; HIV, human immunodeficiency virus; LDV, ledipasvir; NR, Not Recommended ;OBV, ombitasvir; PIB, pibrentasvir; PTV, paritaprevir; r, ritonavir; SOF, sofosbuvir; SIM, simeprivir; RBV, ribavrin; VEL, velpatasvir; VOX: voxilaprevir. * Alternative therapy, **8Wks for patients who are non-black, HIV-uninfected, and whose HCV RNA level is $\leq$6 million IU/mL, **16wk +RBV for those with breakthrough or non-responder. ****16Wks if NS5A RAS present.

Retreatment of patients with prior exposure to NS3 Protease inhibitors and/or NS5A inhibitors

The retreatment of patients who have been failed by a second or third generation DAA regimen containing an NS5A inhibitor is made difficult by the fact that, NS5A inhibitors have a large amount of cross resistance between different drugs in the class (Table 1.6). NS5A inhibitor RAS have also been shown to persist for years after treatment failure [88]. This has lead to the recommendation of SOF/VEL/VOX as a rescue therapy by both EASL and AASLD, based on the results of the POLARIS-1 and POLARIS-4 trials previously discussed. However, the use of GLE/PIB + SOF represents an interesting possibility as pibrentasvir has a high barrier to resistance compared to other NS5A inhibitors. Furthermore, data from the small MAGELLAN-3 trial [89] suggests that this combination would be highly effective. However, larger studies of this combination have not taken place as GLE/PIB and SOF are manufactured and sold by different companies.

1.5 Resistance to direct acting antivirals

The high replication rate of HCV, estimated at 10^{12} virions per day [29], combined with the low fidelity of the HCV polymerase, (which does not have a proof reading capability and has an estimated error rate of 10^{-4} substitutions per site per replication [90]), means that HCV genetic mutation is happening at a very high frequency. As a consequence HCV exist more as a cloud of mutations than as a static genome[91]. These distributions of mutations are known as viral quasispeices [92]. This gives the HCV the ability to quickly develop mutations which provide resistance to DAAs. The development of these mutations are dependent on 3 factors: (i)The strength of the selection pressure, (ii)the genetic barrier to resistance and (iii)the viral fitness of the mutated variant. (i) The strength of the selection pressure or drug exposure, sub optimal treatment regimens, doses or treatment duration will all increase the likelihood of resistance developing. (ii) The genetic barrier to resistance, if only a single nucleotide substitution is required to develop resistance this is more likely than one which requires several substitutions. (iii) This is interconnected with the third factor which is the fitness of the resistant variant. If the mutation dramatically attenuates the viruses ability to replicate then it is less likely to become

a dominant variant. If the mutations which cause resistance to a drug have a large fitness cost then the drug can be described as having a “high barrier to resistance”. However, if the mutations are easily acquired and don’t dramatically attenuate the replicative fitness of the virus then they have a “low barrier to resistance”

HCV resistance to DAAs is normally the result of an amino acid substitution which is referred to as a Resistance Associated Substitution (RAS). Amino acid deletions have also been reported but are far less common. The viral populations carrying RAS are referred to as resistance associated variants (RAV). A RAS is reported with using the single letter amino acid code of the wild type (WT) amino acid, followed by the position in the protein, followed by the resistant amino acid, for example the NS5A RAS Y93H shows the amino acid tyrosine at position 93 in the NS5A protein has been replaced by histidine. In this work where it is not clear which amino acid is the WT amino acid the first letter will not be included to avoid confusion.

RAS have traditionally been identified by observing the viral amino acid substitutions which take place after a treatment failure which is known as treatment emergent. Treatment emergent RAS can be persistent or transient in nature. Pre-existing RAS are present before treatment begins and could be persistent treatment emergent RAS from a previous treatment or naturally occurring RAV.

1.5.1 NS3 protease inhibitor resistance associated substitutions

As previously described in section 1.4.2 protease inhibitors work by binding to and inhibiting the action of the NS3 protease which is essential for HCV replication. As a consequence there is a relatively low barrier for resistance as changes to the structure of the catalytic domain of the protease can dramatically limit the ability of the drug molecules to bind whilst still remaining functional. The NS3 inhibitor RAS are all found between positions 36 and 175 of the NS3 protein and there tends to be cross resistance between compounds (Table 1.5). The first generation of protease inhibitors such as BOC, TVR, SIM were shown to be sensitive to substitutions at positions 155, 156, and 168 [93]. However, the second generation protease inhibitors such as GLE and VOX have greater in-vitro activity against RAS than the preceding generation [94, 95].

Following treatment failure with protease inhibitor based therapies, it has been shown that treatment emergent RAS revert back to the WT relatively quickly compared to NS5A RAS. One study in Japan of gt1b patients showed that all NS3 D168 variants had reverted by 36 weeks post relapse where as none of the corresponding NS5A RAS had reverted [88]. Despite the evidence of the non-persistence of NS3 RAS, the question of whether clinicians should delay re-treatment with an NS3 inhibitor is still unanswered. Currently both EASL and AASLD do not recommend NS3 RAS testing [52, 96].

	Paritaprevir	**Grazoprevir**	**Voxilaprevir**	**Glecaprevir**
1a	36A/L/M	36A/L/M	36A/G/L	36M
	43L	54S	41R	56H
	80K/L/R	55A/I	43S	156G/T/V
	155G/K/Q/S/T/W	56F/H	55A	168A/V
	156G/S/T	80K/R	80K/L	
	168A/E/H/N/V/Y	107I	155G/W	
		109K	156L/T	
		122G/N/R/T	168A/F/I/K/L/R/T/V	
		155D/I/K/T		
		156G/L/M/S/T/V		
		158A		
		168A/C/E/G/N/V/Y		
		170T/V		
1b	56H	36A/L/M	36A/M	156T/V
	80R	43S	54A/S	168V
	122R	54S	122D	
	155K	55A	155W	
	156T	56F/H	156S/T/V	
	168A/E/H/T/V/Y	80R	168V/Y	
		107I	170A	
		122G		
		155G/K		
		156G/M/T/V		
		168A/E/G/K		
		170T		
2	168V/Y			
2a	123Q		43V	156T/V
	132V		156L/T/V	168E/V
	155K			
	156G/L/S/V			
	168A/E/G/H/N/V			
	177V			
2b				156T/V
				168E/V
3		80K		56H/N
		156G		80K/R

	Paritaprevir	Grazoprevir	Voxilaprevir	Glecaprevir
				156G
				168L/R
3a	80K		41K/R	156G
	156G/L/S/T/V		80K/R	166S/T
	166S		156T/V	168L/R
	168H/R/V		175M	
4	56H	156G/M/T/V		
	168H/V	168A/C/E/G/N/V/Y		
		170I		
4a	155K		41R	155C
	156G/S		156L/T/V	156T/V
	168V		168E/T/V	168H/V
	80R			
4d	56H			
	168V			
5a	155K		168A/E/H/K/R/Y	168E/H
	156S			
	168A/E/H/N/V			
6a	36A		41K/R	168A/H/V
	43C		56H	
	155K/T		168A/H	
	158I			
	168A/E/G/N/V			

Table 1.5: **Summary of resistance associated substitutions to NS3 protease inhibitors broken down by HCV genotype.**

1.5.2 NS5A inhibitor resistance associated substitutions

NS5A inhibitors all bind to domain 1 of the NS5A protein. The NS5A protein is involved in multiple processes in the HCV replication cycle as previously discussed, this means the exact mechanism of action of NS5A inhibitors is not understood. As a result of this shared target region NS5A inhibitors have similar RAS profile. The Y93H substitution has been shown to be associated with treatment failure for all NS5A inhibitors [97]. Y93H has also been shown to increase the half maximal effective concentration (EC50) of all NS5A inhibitors in in-vitro assays [98]. Pibrentasvir has been shown to be more effective against RAVs carrying the Y93H substitution but has been shown to be susceptible to substitutions at positions 30 and

31 in the NS5A [99]. Unlike treatment emergent NS3 inhibitor RAS treatment emergent NS5A inhibitor RAS have been shown to persist for at least 48 weeks [88]. This combined with the cross-resistance between NS5A inhibitors makes retreatment of patients who have been previously treated with NS5A inhibitors difficult.

	Daclatasvir	Elbasvir	Ledipasvir	Ombitasvir	Pibrentasvir	Velpatasvir
1a	28A/T/V 30D/E/G/H/K/L/R 31M/V 32del/L 58D 93C/H/N	28A/G/S/T/V 30D/E/G/H/K/L/R 31F/M/V 32del 58D 93C/H/N/S	24G/N/R 28A/G/I/T 30E/G/H/K/L/R/T/Y 31F/I/M/P/V 32L 38F 58D 92K/T 93C/F/H/N/S	28T/V 30E/G/H/R 31M 58D 93C/H/L/N/S	28G/T 29R 30D/del/E/G/K/R 58D 62A 93H/N	24R 28T 30E/H/K/R/T 31I/M/V 32del/L 58D 93C/F/H/N/R/S/W
1b	28I/M/T/V 29del/S 30E/G/H/L/Q/S/W 31F/I/M/V 32del/L 37I/L 54L 58S 92K 93H/N	28M 30Q 31F/I/M/V 32del 93H/N	31F/I/M/V 32del/L 58D 92E/K 93C/H/N/S	28M/T 30Q 31F/M/V 58S 93C/H/S		31M/V 32del 93H/N
1c						93H
1h						93H
2	31M			28F		
2a	24A/S 28C/S 31M/V 92R/S 93H/N	28S 31M/V 93H		24A/S 28S 31M		28S 31M/V 93H
2b	28L 31M 92R 93H	93H		28F 31V		31M 58S 93H 93H 93H
2j						93H
3	30S 93H	30K/L 31F 93H			28G 30G/K 31F/M 58T 93H	93H
3a	30K 31F/M/V 62L	30K 31F/V 93H		28T 31F 93H	28K 30K 31F	30K 31F/M/V 93H

	Daclatasvir	Elbasvir	Ledipasvir	Ombitasvir	Pibrentasvir	Velpatasvir
	93H				93H	
4	28M/S 30S 93C/H	30F/S 31V 58P 93H	28M 93C/H	28V 93C/H		
4a	30A/G/H/S 31V 58L 93H/R/S/W	28T 30H 58D		28V 58L		30R 93H/N
4d		28S 58T		28V 31I 58P/S		93H
4g		93C				
4r	30S					
5a	31F/V 56R	31F/V		28I 31F/V	31F	
6a	24H 28F 30H 31M/V 32L/S 58A/N/S	31M/V		28A/F/M/T/V 31I/M/V 58A/N/S 93I/S		28S 31M/V
7a	30G 31F	31F				

Table 1.6: **Summary of resistance associated substitutions to NS5A Inhibitors broken down by HCV genotype.**

1.5.3 NS5B polymerase inhibitor resistance associated substitutions

The NS5B polymerase is an enticing target for antivirals as it is both essential for viral replication and highly conserved. NS5B inhibitors are split into two categories by their mechanism of action. Nucleotide analogues are molecules which have a structure similar to a nucleotide molecule, which the NS5B polymerase attempts to incorporate into the new viral RNA which causes chain termination of the new viral RNA. Non-nucleotide analogues bind to the NS5B protein and prevent its action, by either preventing it's interaction with the other proteins which form the HCV replication complex or by preventing the formation of new viral RNA. Nucleotide

and non-nucleotide inhibitors have different RAS and different barriers to resistance, because of the differing mechanisms of action and differing targets on the protein.

Non-nucleotide analogues

The only approved non-nucleotide inhibitor is DAS which is only approved and recommended for use in patients infected with gt1 HCV in combination with paritaprevir, ombitasvir and ritonavir (OMB/PTVr + DAS). DAS works by binding to the NS5B protein and inhibiting its action it is therefore susceptible to changes and differences to the structure of this protein. Hence why it is only active against gt1 infection, and a large number of RAS have been reported. DAS RAS are primarily found between position 316 and 585 in the NS5B protein (Table 1.7), this is the palm 1 region of the protein. The RAS C316Y, M414T, Y448C, Y448H, and S556G have all been shown in resistance selection studies and in-vitro EC50 assays to increase the EC50 of dasabuvir [100]. M414T and S556G were the most commonly identified RAS in gt1a and gt1b patients who were failed by dasabuvir based therapy [72, 74].

Nucleotide analogues

SOF is the only approved drug in the nucleotide analogue class. SOF's mechanism of action is thought to be the reason for its high barrier to resistance. Any mutation which reduces the polymerases affinity for SOF will also reduce its affinity for the nucleotide uridine. This means that RAVs are either incapable of replication or have a dramatically reduced replicative fitness, making SOF RAS rare. There are reported SOF RAS such as the S282T substitution [101], however these RAS are normally treatment emergent and have a high chance of reverting back to WT after treatment cessation. The L159F, L320F and V321A RAS are commonly reported as treatment emergent RAS in both clinical trials [102] and real life cohorts [103]. However, when tested in-vitro they have only shown small increases in the EC50 of SOF [104].

	Dasabuvir	Sofosbuvir
1		159F
	316Y	159F
	395G	237G

1a

	Dasabuvir	Sofosbuvir
	414I/T	282T
	444K	320F
	448C/H	321A
	553T	
	556G/N/R	
	558R	
	559G	
	565F	
1b	159F	159F
	316N/Y	179A
	368T	282T
	411S	289L
	414T	316H/N
	445F	321I
	448C/H	
	451S	
	495A	
	553V	
	556G	
	585V	
2		159F
2a		282T
		289L
2b		282T
2d		282T
2j		282T
3		159F
		282T
		289I
		321A
3a		100R
		150V
		159F
		188D
		206E
		244I
		282T
		321A
4		282T
4a		282T
4r		282T
		321I
6a		282T
6l		282T

Dasabuvir	Sofosbuvir

Table 1.7: **Summary of resistance associated substitutions to NS5B Inhibitors broken down by HCV genotype.**

1.6 Detection and characterisation of resistance associated substitutions

1.6.1 Methods for sequencing the Hepatitis C Virus

HCV genotype/subtype and the presence of RAS are both important factors when choosing the optimal treatment for patients. Accurate genotyping is essential for most treatments as most are genotype specific. However, this is becoming less important for third-generation pan-genotyptic treatments. RAS testing before first-line treatment is not recommend by EASL, but is recommend by AASLD for some treatments such as ELB/GZR where the presence of NS5A RAS has been shown to affect treatment outcomes. Testing for RAS before re-treatment has been shown in limited studies to be effective in guiding the choice of therapy [105].

The clinical assessment of viral genotype is primarily done using reverse hybridisation with genotype-specific oligonucleotide probes which target the highly conserved 5'UTR, Core and NS5B regions of the HCV genome. Polymerase Chain Reaction (PCR) based HCV genotyping kits such as TruGene 5'NC HCV Genotyping kit (Siemens Healthcare Diagnostics Division, Tarrytown, NY) and Versant™ HCV Genotype Assay LiPA (version I, Siemens Medical Solutions, Diagnostics Division, Fernwald, Germany) and automated systems such as the cobas® 4800 (Roche, USA) are easy and quick to use. Whilst these assays can accurately genotype viruses, they are not as accurate when calling less common HCV subtypes, mixed infections and recombinant viruses such as gt2k/1b. These genotyping methods are also unable to detect the presence of RAS.

Traditional Sanger sequencing uses PCR amplified 5'UTR, Core and NS5B regions to genotype the virus. If the relevant parts of the NS3, NS5A and NS5B are also sequenced it can identify RAS in majority variants. However, this method is still reliant on accurate PCR

primers design. This is especially important for resistance calling, as the genotype of the virus is required to correctly select primers. This is complicated by the high genetic diversity of HCV. The detection of low frequency variants is also not possible unless multiple clones are sequenced which can be time consuming.

Next generation sequencing (NGS) using platforms like Illumina allows the generation of whole genome sequences (WGS). It also allows the identification of RAS in minority viral variants and can be used to accurately call mixed infections. NGS methods such as metagenomic and probe based enrichment also allows a higher throughput of samples with multiplexing. Most clinical guidelines state that for a RAS detected by NGS to be clinically relevant it must be detected in $>$15% [106] of the sequencing reads. This guideline assumes that adequate sequence depth has been achieved at that site, where a depth of at least 100 reads is considered acceptable.

There are a variety of ways NGS can be used to sequence HCV. Metagenomic sequencing involves the use of NGS with no enrichment or amplification of the HCV RNA in the sample. Depending on the method used depletion of ribosomal RNA can take place. This results in a large amount of non-HCV or off target reads, that allows the identification of other pathogens, but results in less coverage of the HCV genome and lower sequencing depth. It is also more expensive due to the increased amount of sequencing required, to achieve a useful sequencing depth and coverage of the HCV genome.

Specific PCR primers can be used to amplify the RNA, this is a similar method to that used or Sanger sequencing but the use of NGS allows the detection of lower frequency variants and can also be used to generate whole genome sequences. However, it requires primers which cover the whole genome and are specific to the genotype being sequenced. This method also has the disadvantage of being very time consuming and hence is not high throughput. Unlike metagenomics, the number of HCV reads generated using this method and the amount of sample RNA does not follow linear relationship [107], meaning that this method may be more effective for low viral load samples.

Probe based enrichment techniques such as ve-seq [35] use a panel of oligonucleotide probes that compliment the HCV genome and have been designed to cover all HCV geno-

types, to capture HCV RNA from a metagenomic library. This method can be used to generate HCV WGS with a greater read depth than metagenomic sequencing, and so allows the identification of both low frequency RAS and mixed infections. It also is easy to multiplex samples which allows a high throughput and a lower cost per sample for sequencing. However, information about other circulating pathogens is reduced post HCV probe based enrichment.

1.6.2 In-vitro analysis of resistance associated substitutions

The development of DAAs was made possible because of the insights gained into the viral life-cycle of HCV from in-vitro studies of the virus [108] and RAS can be both identified and characterised using a variety of in-vitro methods. HCV isolated from patient samples does not replicate in standard cell culture models, with the exception of the JFH-1 strain. Therefore it was nearly 10 years after the discovery of HCV that an in-vitro model was developed (the subgenomic replicon in 1999) [109]. The subgenomic replicon contains the HCV NS3-NS5B genes and the HCV internal ribosomal entry site (IRES) is also used to initiate translation. The HCV structural genes are replaced with an encephalomyocarditis virus (EMCV) promoter gene and in some cases a neomycin resistance gene which can be used to create stable replicon producing cell lines via antibiotic selection. A luciferase reporter gene can also be added which allows for direct quantification of HCV replication (figure 1.3). HCV subgenomic replicons have been created for all major HCV genotypes and provide an excellent model for the study of the HCV replication cycle. However, several adaptive mutations are required for replication [109, 110] and cells are transfected with the viral RNA and no infectious particles are produced. The lack of the structural genes mean attachment, entry, uncoating, assembly/maturation and release aspects of the HCV life cycle can not be studied with subgenomic replicons. A transient replication assay, based on the transfection of HCV replicon RNA in to a hepatoma cell line, followed by a short term monitoring of viral replication through the reporter gene, is the preferred method for RAS testing. This is because of the reduced chance of adaptive mutations and a higher throughput than models which utilise stable replicon cell lines expressing viral RNA or models which utilise infectious systems.

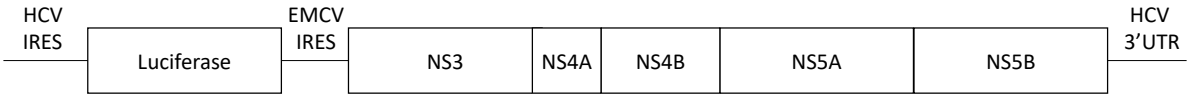

Figure 1.3: **Schematic of subgenomic replicon.** Not to Scale

The isolation of the JFH-1 gt2a virus in 2005, which can replicate in-vitro unlike any other naturally occurring strain of HCV [111, 112], allowed for the first time the study of the entire HCV lifecycle using one in-vitro system. The replacement of the JHF-1 Core-NS2 genes with the same genes for the J6 gt2a isolate, dramatically increased the replicative fitness of the model and allowed the creation of other cell culture derived (HCVcc) models. The isolation of the JFH-1 virus enabled the creation of chimeric viruses where the NS3-NS5A genes from different genotype are inserted into the JFH-1 viral backbone. This boosted the replicative fitness and also allowed the study of different HCV genotypes [113]. Full-length non-chimeric infectious clones have since been developed for gt1a [114], gt2b [115, 116] and gt3a [104], however these infectious clones have several adaptive mutations which allows replication in hepatoma cell lines such as Huh7 and Huh7.5.

Incubation of replicon expressing cells or cells infected with HCVcc particles with a DAA can be used to select RAS against a specific DAA. However, the adaptive mutations required to allow replication and the genetic difference of the cell lines used compared to normal human liver cells means that these mutations may not be representative of in-vivo mutations and may not be clinically relevant. Site directed mutagenesis can introduce RAS into a replicon or HCVcc which contains some sort of reporter gene such as luciferase. The replication of the virus can be measured with serial dilutions of the compound and a regression analysis used to calculate an EC50 value, which can be compared to the EC50 of the un-mutated virus. The EC50 fold change measurement is used to assess the impact of RAS on the activity of DAAs, an increase in EC50 of greater than a 5 fold increase is considered to be resistant, but RAS can increase the EC50 of a DAA by more than 10,000 fold.

1.7 Aims of work

As discussed, the development of DAAs has led to SVR rates routinely exceeding 95%. However, RAS have been described for all DAAs and the presence of these RAS have been shown by multiple studies to reduce SVR rates. The aim of this work is to better characterise viral resistance to HCV DAAs by:

1. Studying the natural prevalence of RAS in a DAA treatment naive population and characterise them using in-vitro assays to:
   (a) Establish the natural prevalence of RAS in a gt3 population.
   (b) Identity phylogenetic patterns to the prevalence of RAS.
   (c) Understand the fitness cost and increase of EC50 of common naturally occurring RAS.
2. Investigate the viral determinate of failure in a cohort of DAA treatment naive patients treated with sofosbuvir by:
   (a) Establish the prevalence of known SOF RAS.
   (b) Identify novel RAS using genome wide association studies.
   (c) Asses the effect of RAS on treatment response.
   (d) Study the effect of viral diversity on treatment outcome.
3. Study the effect of RAS on retreatment of HCV infected patients with prior failure to DAA therapy using SOF/VEL/VOX.
   (a) Establish the prevalence of RAS is a population of patients failed by DAA therapy.
   (b) Identify the factors effect SVR rates for SOF/VEL/VOX retreatment of patients failed by first-line therapy.

Chapter 2

Analysis of naturally occurring resistance associated substitutions in a population of genotype 3 Hepatitis C Virus infected direct acting antiviral treatment naïve patients

A slightly modified version of this work was published in Hepatology in 2018. Smith D, Magri A, Bonsall D, et al. Resistance analysis of genotype 3 HCV indicates subtypes inherently resistant to NS5A inhibitors. Hepatology 2018;00:1–12. doi:10.1002/hep.29837

2.1 Abstract

Background and Aims: HCV gt3a is highly prevalent globally, with non-gt3a subtypes common in Southeast Asia. RAS have been shown to play a role in treatment failure. However, the role of RAS in gt3 is not well understood. This chapter reports the prevalence of RAS in a cohort of DAA treatment-naïve gt3 infected patients, including those with rarer subtypes and then evaluates the effect of these RAS on DAAs in-vitro.

Methods: Baseline samples from 496 gt3 infected patients enrolled in the BOSON clinical trial were analysed by next-generation sequencing after probe-based enrichment for HCV. Whole viral genomes were analysed for the presence of RAS to approved DAAs. The resistance phenotype of RAS in combination to daclatasvir, velpatasvir, pibrentasvir, elbasvir and sofosbuvir was measured using the S52-ΔN gt3a replicon model.

Results: The NS5A A30K and Y93H substitutions were the most common at 8.8% (n=44) and 12.2% (n=61) respectively and showed a 10-fold and 11-fold increase in EC50 for daclatasvir compared to the unmodified replicon. Paired RAS (A30K + L31M and A30K + Y93H) were identified in 18 patients (9 of each pair); these combinations were shown to be highly resistant to daclatasvir, velpatasvir, elbasvir and pibrentasvir. The A30K + L31M combination was found in all gt3b and gt3g samples.

Conclusions: This chapter demonstrates the high frequencies of RAS to NS5A inhibitors in gt3 HCV. The paired A30K + L31M substitutions occur in all patients with gt3b and gt3g virus. In-vitro analysis suggests that these subtypes may be inherently resistant to all approved NS5A Inhibitors for gt3 HCV.

2.2 Background

The recent development of DAAs for the treatment of HCV infection has led to a dramatic increase in SVR rates, with many studies reporting >90% SVR rates [61, 83, 117, 118]. Despite this dramatic increase in effectiveness of DAA treatment for chronic HCV infection, the treatment of gt3 infection has shown lower SVR rates compared to other genotypes, especially in patients with cirrhosis [62, 119]. This contrasts with the treatment of gt3-infected individuals with IFN-based therapy in whom SVR rates were consistently higher than in gt1 [120, 121]. More recently, pan-genotypic regimens such as GLE/PIB, SOF/VEL and SOF/VEL/VOX have been developed that very effectively target HCV gt3 [58, 87, 122–124].

The reasons for the reduced efficacy of some IFN-free DAA therapy against gt3 infection remain unclear. Host genetics are known to play a factor in HCV treatment outcomes. The C/C genotype of the IFNL4 SNP rs12979860 (formally referred to as the IL28B) is associated with better outcomes in HCV treatment [8, 125]. Infection with gt3 HCV has also been associated with clinical phenotypes that may affect response to DAA therapy, including hepatic steatosis, increased rates of liver fibrosis [126] and increased chance of progression to hepatocellular carcinoma [127, 128], each of which has been linked to poor outcomes to DAA therapy [129], and this may help to explain the reduced efficacy of DAAs in gt3. Presence of RAS in viral sequences could be another factor contributing to lower SVR rates in gt3. For instance the Y93H Substitution has a high prevalence in gt3 sequences and has been shown to be associated with lower SVR rates especially in patients with cirrhosis [62, 97].

At the time of publication of this work, the treatment recommendations for the treatment of gt3 infection from EASL 2016 was a combination of one of the NS5A inhibitors, DCV or VEL, with SOF [96]. AASLD 2017 recommended one of the following combinations depending on previous treatment experience and the presence or absence of cirrhosis and hepatic decompensation; GLE/PIB, SOF/VEL, SOF/VEL/VOX or EBR/GRZ + SOF [130]. The Asian-Pacific Association for the Study of the Liver (APASL) recommendations advocated the use of either SOF + RBV or the combination of DCV+ SOF ± RBV depending on treatment experience and liver disease state [131].

Viral variants carrying RAS have been reported in clinical trials for all current DAAs [43, 132, 133], many of which have been characterised in vitro using replicon- or virus-based resistance assays. For non-gt2 HCV variants, the role of RAS has been typically evaluated using sub-genomic replicons where the structural protein region has been replaced with a luciferase reporter that allows direct quantitation of replication [109]. The transient-replication assay, based on viral RNA transfection followed by a short term monitoring of viral replication through the reporter gene, is the preferred method for RAS testing, because of the reduced chance of adaptive mutations and a higher throughput than models which utilise stable replicon cell lines expressing viral RNA [110]. The gt3a replicon S52/SG-Feo, used in a transient replication assay, was recently improved by removing the neomycin resistance gene (ΔN) [110]. For this work the gt3a S52ΔN replicon with a modified Huh 7.5 cell line expressing a stable, high level of the SEC14-L2 gene [110] was used, to assess the phenotype of RAS in a gt3a background in a transient-replication model. The increased expression of the SEC14-L2 has been shown to enhance HCV replication [134], and is essential for the efficient replication of genotype 3 replicons [110].

2.3 Aims

To investigate the frequencies of potential RAS in a large (n=496) gt3 cohort (prior to SOF based treatment regimens in the BOSON clinical study) using a probe-based sequence capture approach for next-generation sequencing to generate full-length HCV genomes [35] and bioinformatics tools to detect viral variants at frequencies of $<$1% [135, 136]. The phenotypic effect of RAS was evaluated in vitro both individually and in combination using the gt3a replicon system, and their potential roles in treatment failure were evaluated.

2.4 Methods

2.4.1 Subjects and samples

The samples used in this work were from the BOSON clinical trial [137]. The BOSON trial was an open-label, randomised, phase 3 trial to determine the efficacy and safety of SOF and ribavirin, with and without peginterferon-alfa, in DAA treatment naive gt2 and gt3 patients. 592 patients were enrolled in the BOSON trial, of which the majority (n=544) were infected with gt3 HCV, the remaining 48 patients were infected with gt2 HCV. All the gt2 infected patients had compensated cirrhosis and had prior treatment experience with an interferon based therapy. Of the gt3 infected patients 265 were previously treated with interferon based therapies and 158 had compensated cirrhosis.

Plasma samples were obtained from patients before treatment commenced, during treatment and at 12 and 24 weeks post treatment. Theses samples were then sent to Oxford and the RNA was extracted and sequenced from the viremic samples using the method outlined below in section 2.4.2. Plasma samples were obtained from patients before treatment commencement. All patients have not previously received DAA based therapy.

2.4.2 HCV sequencing

RNA was extracted from 500 μl of plasma using the NucliSENS® easyMAG system (bioMérieux) into 30 μl of water, of which 5 μl was processed with the NEBNext® Ultra™Directional RNA Library Prep Kit for Illumina® (New England Biolabs) with previously published modifications to the manufacturers protocol [138]. A 500 ng aliquot of the pooled library was enriched using the xGen® Lockdown® protocol from IDT (Rapid Protocol for DNA Probe Hybridisation and Target Capture Using an Illumina TruSeq® Library (v1.0), Integrated DNA Technologies) with a comprehensive panel of HCV-specific 120 nt DNA oligonucleotide probes (IDT), designed using a previously published algorithm [35]. The enriched library was sequenced using Illumina MiSeq v2 chemistry to produce paired 150 b reads. Reads were demultiplexed and 150 b paired-end reads were trimmed using QUASR v7.01 [139] then adaptor sequences re-

moved using Cutadapt v1.7.1 [140] reads were discarded if less than 50p long. Reads were mapped to human reference sequence using Bowtie 2.2.4 [141] and host derived sequences were removed. The remaining reads were then mapped against a database containing all 165 ICTV (International Committee of the Taxonomy of Viruses) [39] HCV genomes using BWA mem v0.7.10 [142] and STAMPY v1.0.23 [143]. An appropriate reference sequence is selected from this database and reads which conform to the majority population are then used fro de novo assembly using Vicuna v1.3 [144] and V-FAT v1.0 [145]. All viral reads are then mapped back to this assembly using Mosaik v2.2.28 [146] to generate a BAM file containing viral reads. Further detail of the techniques used can be found in Bonsal et al, 2015 [35].

2.4.3 Resistance associated substitution definitions

At the time of this work DCV + SOF was the only approved INF free treatment for gt3 HCV. A search through the available literature for gt3 RAS to DCV and SOF revealed five sites in the NS5A region associated with DCV resistance and three sites in the NS5B region associated with SOF resistance. Sites 32, 54 and 92 were also included as they have been shown to be associated to resistance in other genotypes however their effect in gt3 is unclear. The RAS at these sites are summarised in figure 2.1 [62, 97, 132, 147]

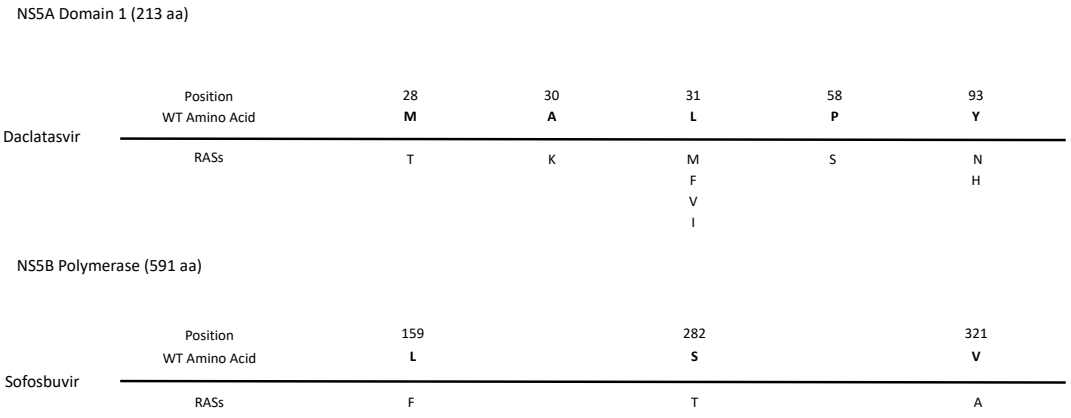

Figure 2.1: **Summary of daclatasvir and sofosbuvir RAS identified in the literature.** Key genotype 3 RAS positions for daclatasvir and sofosbuvir are shown (not to scale) along the NS5A Domain 1 and NS5B (Polymerase) protein respectively, labelled according to the H77 gt1 HCV reference protein sequence. The wild-type (WT) amino acid for genotype 3 is shown above and the relevant RAS are indicated below. Data taken from references [62, 97, 132, 147]

2.4.4 S52-$\triangle$N replicon system

The HCV Gt3a replicon S52-$\triangle$N plasmid contains the HCV IRES driving the translation of the firefly luciferase gene and the EMCV IRES responsible for the synthesis of the HCV polyprotein (NS3-NS5B) (Supplementary Fig 1). This replicon has been recently improved by removing the neomycin resistance gene [110]. The plasmid DNA was linearised using XbaI (NEB) and in vitro transcribed (IVT) as previously described [148, 149].

Generation of S52-$\triangle$N mutant replicons

RAS were introduced into HCV gt3a replicon using QuikChange II XL Site-Directed Mutagenesis Kit (Agilent Technologies). Positions in this paper refer to H77 consensus; for a full list of positions in S52-$\triangle$N and H77 see supplementary table 1. All the reactions were transformed into Top10 bacteria (Life Technologies) and for each mutation 3 clones were isolated and sequenced to confirm the viral sequence.

Cell culture and replicon transfection

Huh7.5 cells over expressing SEC14L2 protein (Huh7.5-SEC14L2), were generated as previously reported by Witteveldt, et al [110]. In brief, Huh7.5 cells were transduced with a lentiviral vector to overexpress SEC142 protein. After puromycin selection, single cell clones were isolated and characterised as described [110]. Cells were maintained in DMEM supplemented with 10% foetal calf serum (FCS), 10 mM HEPES buffer, 2 mM L-glutamine, 100 μg/ml penicillin, 100 μg/ml streptomycin and 0.1 M nonessential amino acids. RNA transfection was performed by electroporating 4-5x106 cells for each reaction with 2 μg of RNA as previously described [110, 148].

Evaluation of replicative fitness

To infer the replicative fitness of RAS-containing viruses, cells transfected with each construct were resuspended in complete medium at the approximate concentration of $3x10^5 cells/ml$ and seeded into 96-well plates, then luciferase activity (RLU) was measured after 4, 24, 48 and 72hr using Bright Glo (Promega) on the GloMax® 96 Microplate Luminometer (Promega). The replication capacity of the variant was calculated using the equation:

$$\text{Replication Capacity} = \frac{\text{Variant 72hr RLU} \ / \ \text{Variant 4hr RLU}}{\text{WT 72hr RLU} \ / \ \text{WT 4hr RLU}}$$

The replicative-defective pSGR-JFH1/GND (GND) replicon, containing a self-inactivating mutation in the NS5B gene, was used as a negative control.

Direct acting antiviral drugs

DCV, VEL and EBV were purchased from MedchemExpress, SOF was purchased Selleckchem and PIB from MedKoo.

Evaluation of resistance phenotype

Huh7.5-Sec14L2 cells transfected as described above were seeded in the presence of DCV or SOF over the appropriate concentration range. For the single RAS 3-fold dilutions were prepared starting at 3 μM for SOF and 300 pM for DCV. The susceptibility of multiple NS5A RAS were evaluated using 10-fold dilutions of DCV, velpatasvir, PIB and EBV starting at 100 μM. Each experiment included at least 2 replicates and was performed at least twice. EC50 values were calculated using non-linear regression on Graphpad Prism 7.00 for Mac(GraphPad Software, La Jolla California USA).

2.5 Results

2.5.1 Whole-genome sequencing of Hepatitis C Virus to identify resistance associated substitutions

There was insufficient coverage of the HCV genome to create consensus sequences for 12 of the 530 gt3 infected patients enrolled in the study, these samples were excluded from further analysis. The assembly lengths of the remaining 518 sequences ranged from 7519-9861 (mean = 9443) and the average sequence depth across the genomes ranged from 26 to 130236 (mean = 5673). These sequences were analysed for the presence of RAS at the sites listed in section 2.4.3 and figure 2.1. If there was insufficient sequence depth for variant calling at any of the listed RAS codons the sequence was discarded. Complete sets of viral variant data for the 11 RAS sites in the NS3, NS5A and NS5B proteins were obtained for 496 gt3 patients.

2.5.2 Prevalence resistance associated substitutions to daclatasvir and sofosbuvir

The HCV sequences were analysed for the presence of RAS to DCV and SOF. Five DCV RAS were detected; M28T, A30K, L31M, Q54H, P58S and Y93H in NS5A gene and three SOF RAS were observed in the NS5B gene; L159F, S282T and V321A. The most common DCV RAS were A30K and Y93H, these were detected in 8.9% (n=44) and 12.3% (n=61) of samples respectively (figure 2.2A). The L31M RAS was present in 1.8% (n=9), P58S in 2.4% (n=12), M28T in 0.6% (n=3) and Q54H in 0.2% (n=1) of the samples. However, only the A30K, L31M, P58S, Y93H substitutions were detected as majority variants present in $>$50% of the reads, in 28, 9, 7 and 22 samples respectively. The A30K, P58S and Y93H substitutions were detected as both low frequency and majority variants (figure 2.2B). It is interesting that the polymorphisms were largely found as either low frequency variants or majority variants. The M28T RAS was detected in $<$1% of the sequencing reads, in all the samples within which those substitutions were detected. The DCV RAS Q54H was only detected as a low frequency variant present in $<$15% of viral reads in one sample. RAS to SOF were found far more

infrequently, with L159F being detected as a majority variant in only one patient. The RAS S282T and V321A were only found in <1% of reads in two and three patients respectively. Amino acids substiturions not associated with resistance were detected and their prevalence can be seen in figure A.2. Of note were the A30S, A30T, Q54T substitutions in the NS5A gene as these were detected in 13, 23 and 61 patients respectively.

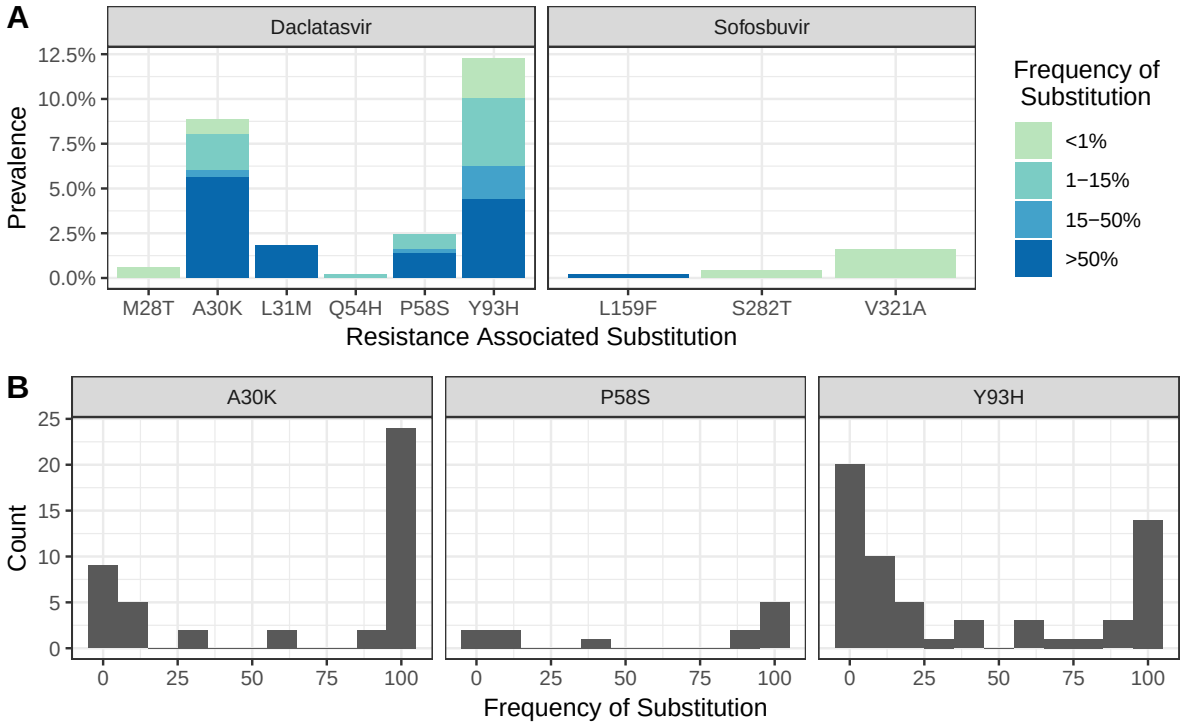

Figure 2.2: **Prevalence of individual resistance associated substitutions to daclatasvir and sofosbuvir. A)** The percentage of the cohort with an individual RAS is shown, with the bars further broken down by the frequency of the substitution in an individual patients sequencing reads(n=428). **B)** Histograms showing the distribution of substitution frequency in individual patient HCV sequencing reads. Daclatasvir resistance associated substitutions A30K, P58S and Y93H shown with a bin size of 10%.

2.5.3 Effect of individual substitutions on replicative fitness

All NS5A RAS detected in more than 1% of the samples (A30K, L31M, P58S, Y93H) and all three detected NS5B RAS (L159F, S282T and V321A) were selected for phenotypic characterisation using the gt3 S52-ΔN replicon system. The RAS were introduced into the S52-ΔN replicon. Initially, individual RAS affect on viral fitness in the absence of DAA was investigated. The unmodified gt3 S52-ΔN replicon was used as a positive control and the replicative-defective pSGR-JFH1/GND (GND) replicon was used as a negative control (figure 2.3). All of

the RAS containing replicons were capable of replicating. However, all the NS5B RAS had a substantial negative effect on replication capacity of the S52-ΔN replicon, with L159F showing a ∼7-fold , S282T a ∼16-fold and V321A a ∼5-fold reduction. Interestingly, compared to WT, A30K exhibited a ∼0.5-fold increase in replication capacity, while P58S and L31M showed a ∼3-fold reduction in replication, Y93H showed the largest reduction in replication capacity when compared to wild type of all the NS5A substitutions with a ∼13-fold reduction.

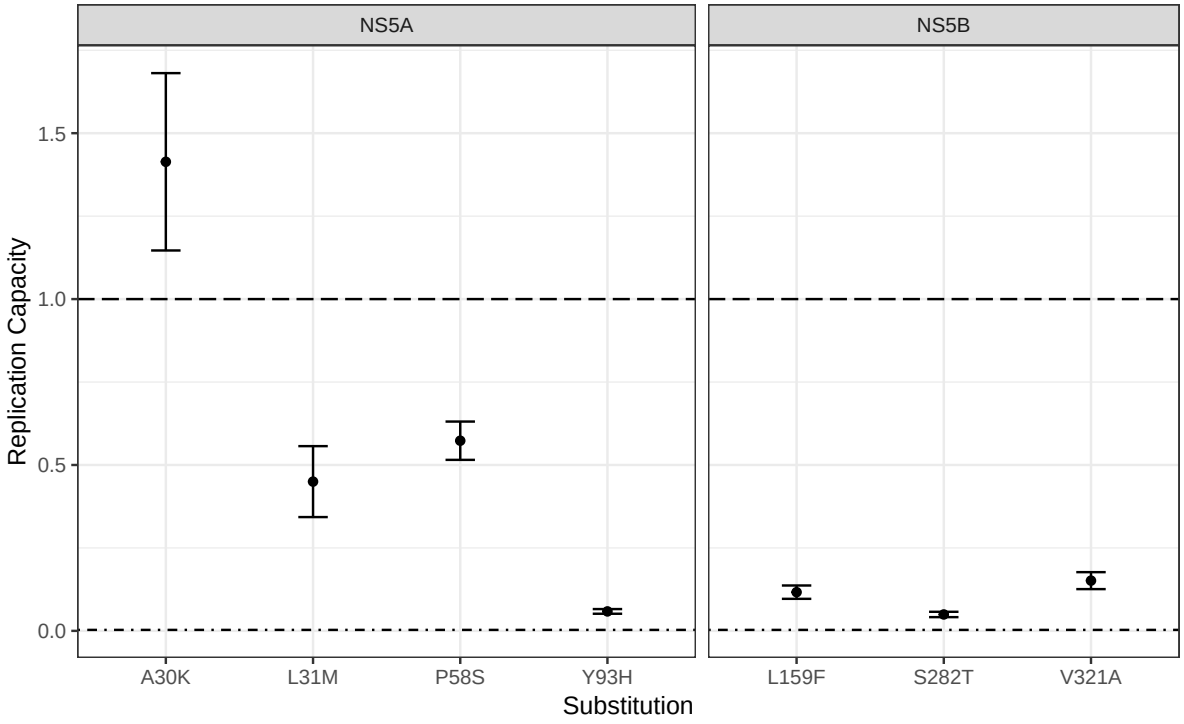

Figure 2.3: **Replication capacity of S52-ΔN replicon carrying candidate RAS.** Huh7.5 SEC14L2 cells were electroplated the luciferase expressing S52-ΔN replicon containing the A30K, L31M, P58S and Y93H NS5A resistance associated substitutions (RAS) and L159F, S282T and V321A NS5B RAS, the luciferase activity was measured at 4 and 72hrs. The replication capacity was then calculated using the equation in section 2.4.4 the mean replication capacity is plotted on the x axis with 95% CI from 3 replicates technical replicates from 3 experiments. The top dashed and bottom dotted lines indicates the replication capacity of the positive control an unmodified WT S52-ΔN replicon and negative control replicative-defective pSGR-JFH1/GND replicon respectively.

2.5.4 Effect of individual substitutions on resistance phenotype

To assess the effect of the RAS on the S52-ΔN replicons susceptibility to DCV or SOF, the EC50 of each RAS containing replicon to the relevant drug was measured using a dose-response scale, the does response curve for each RAS carrying replicon can be seen in

figure A.3 and figure A.4 in the appendices. The mean fold change in EC50 from the wild type S52-ΔN replicon was then calculated. (figure 2.4). For DCV, the NS5A substitutions A30K and Y93H, showed a $\sim$9-fold and $\sim$11-fold increase in EC50, while the L31M substitution only showed a $\sim$2-fold increase in EC50 and the P58S substitution showed no significant change in EC50 when compared to WT. For SOF, the NS5B substitutions S282T and V321A showed a $>$5-fold increase in EC50. However the heavily reduced replicative fitness of the replicons carrying these substitutions meant there was large variations in EC50 between experiments. The L159F substitution showed a $\sim$2-fold increase of EC50 compared to WT.

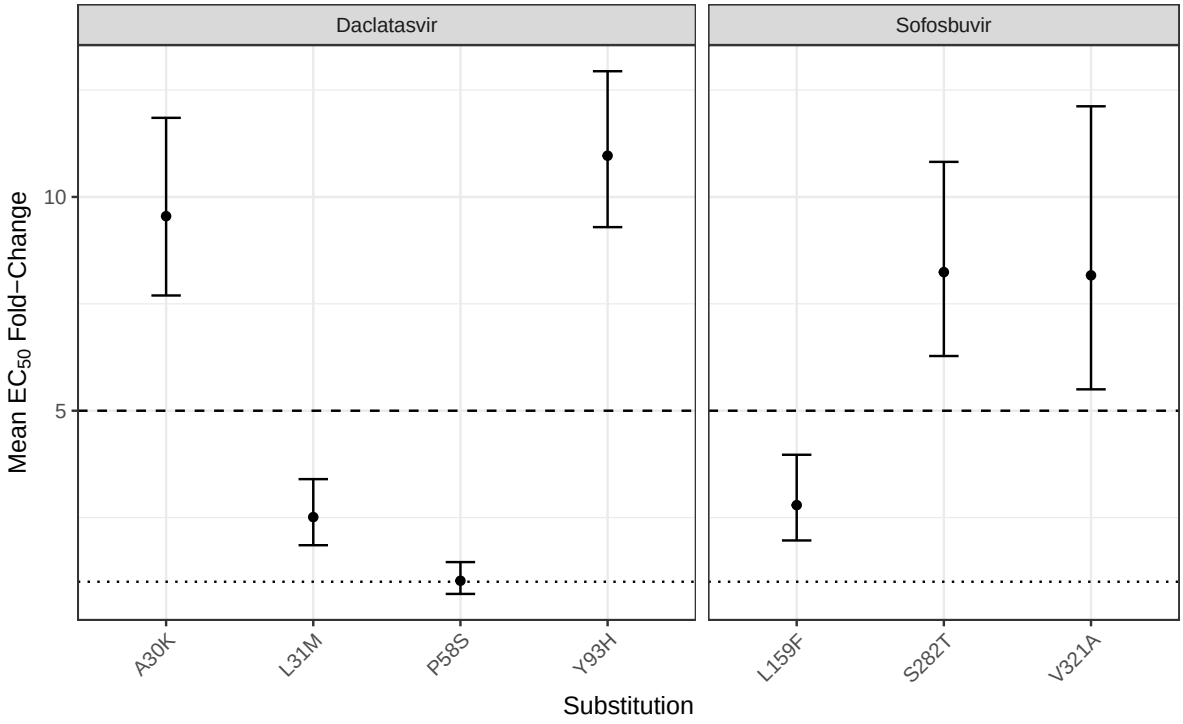

Figure 2.4: **Fold-changes in resistance (based on EC50 measurements) for daclatasvir or sofosbuvir against S52-ΔN replicons carrying candidate RAS.** Huh7.5 SEC14L2 cells were electroplated the luciferase expressing S52-ΔN replicon containing the A30K, L31M, P58S and Y93H NS5A resistance associated substitutions (RAS) and L159F, S282T and V321A NS5B RAS. The cells were then treated with a 3-fold dilutions of daclatasvir or sofosbuvir starting at 300 pM and 3 μM respectively, the EC50 was then calculated using regression analysis for each mutant and the fold change from the WT S52-ΔN EC50 calculated. The mean EC50 fold change is plotted for each RAS carrying replicon with 95% CI calculated from at least 2 experiments. The dotted line indicates WT susceptibility, the dashed line indicates a 5-fold increase in EC50 when compared to the WT S52-ΔN replicon.

2.5.5 NS5A resistance associated substitution combinations

The sequences were scanned for the presence of combinations of RAS which had a EC50 change of >2-fold to DCV (A30K, L31M and Y93H) and SOF (L159F, S282T and V321A). A total of 9 patients had the combination of A30K + Y93H variant and 9 patients had the combination of A30K + L31M in the NS5A protein figure 2.5. No NS5B RAS were detected in combination. In samples carrying the A30K + L31M combination, both RAS were present as the majority variant. In all the samples with the A30K + Y93H combinations at least one of the RAS was detected at a frequency of < 50%.

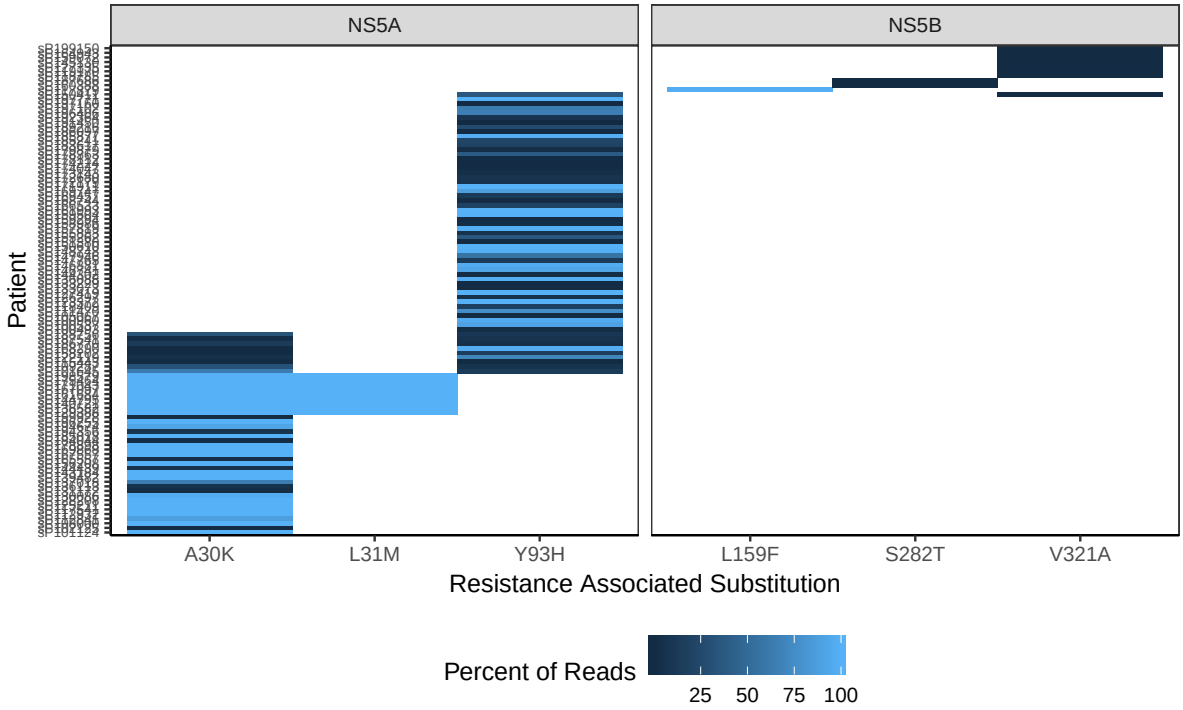

Figure 2.5: **Prevalence of RAS combinations to daclatasvir.** Each column on the x axis represents a RAS site and each row on the y axis is a patient. The frequency of the substitution within the HCV sequencing reads is shown in each cell, expressed as a percentage of reads and coloured from black to blue. Blue meaning the RAS is present in 100% of HCV reads and Black being 1%. White cells mean the RAS are not present.

2.5.6 Effect of resistance associated substitution combinations on replicative fitness

The effect of paired RAS on replication capacity and NS5A inhibitor susceptibility was first evaluated using DCV in an in-vitro system. A30K + L31M had a replication capacity 3-fold greater than WT, while the combination A30K + Y93H had 1-fold lower replication capacity than WT (figure 2.6). The triple-RAS variant of A30K + L31M + Y93H combination of RAS was also generated. This was not found among the pre-treatment sequences from the cohort, but may represent possible additional synergistic interactions. A30K + L31M + Y93H showed a 3-fold increase in replication capacity when compared to WT figure 2.6.

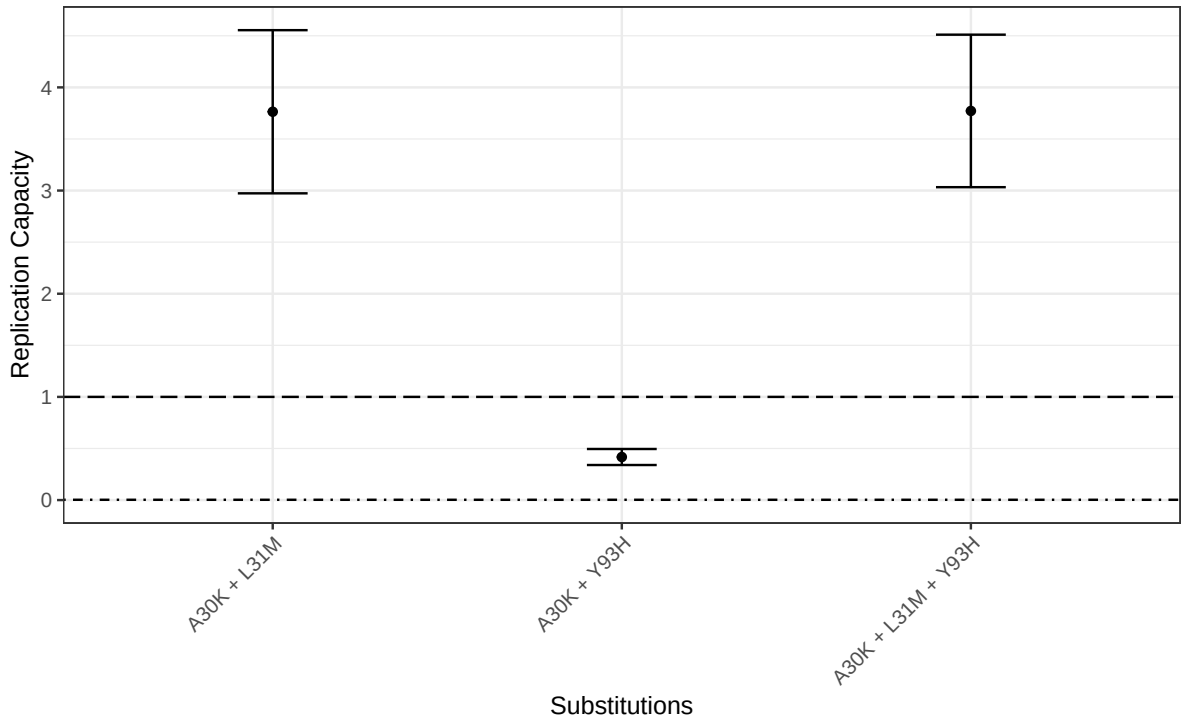

Figure 2.6: **Replication capacity of S52-ΔN replicon carrying candidate NS5A RAS combinations.**Huh7.5 SEC14L2 cells were electroplated the luciferase expressing S52-ΔN replicon containing the A30K + L31M, A30K + Y93H and A30K + L31M + Y93H resistance associated substitutions, the luciferase activity was measured at 4 and 72hrs. The replication capacity was then calculated using the equation in section 2.4.4 the mean replication capacity is plotted on the x axis with 95% CI from 3 replicates technical replicates from 3 experiments. The top dashed and bottom dotted lines indicates the replication capacity of the positive control an unmodified WT S52-ΔN replicon and negative control replicative-defective pSGR-JFH1/GND replicon respectively.

2.5.7 Effect of resistance associated substitution combinations on resistance phenotype

Both A30K + L31M and A30K + Y93H combinations showed a >10,000-fold increase in EC50 figure 2.7. A30K + L31M + Y93H showed a >10,000-fold increase in the EC50 value for DCV when compared to WT figure 2.7. Due to the large increase in DCV EC50 values for the combined RAS, these RAS combinations were tested against the more modern NS5A inhibitors velpatasvir, PIB and EBV which were approved for the treatment of gt3 HCV infection [130], whilst this work was taking place. The double combinations A30K + L31M and A30K + Y93H and the triple combination A30K + L31M + Y93H, also showed a highly resistant phenotype to both velpatasvir (>10,000 fold increase in EC50) and EBV (>100,000,000 fold increase in EC50). Resistance to PIB was also observed, but at lower levels (>20 fold increase in EC50). The relatively higher fold change in EC50 with EBV may be explained by the increased susceptibility of the WT replicon to this drug figure A.9.

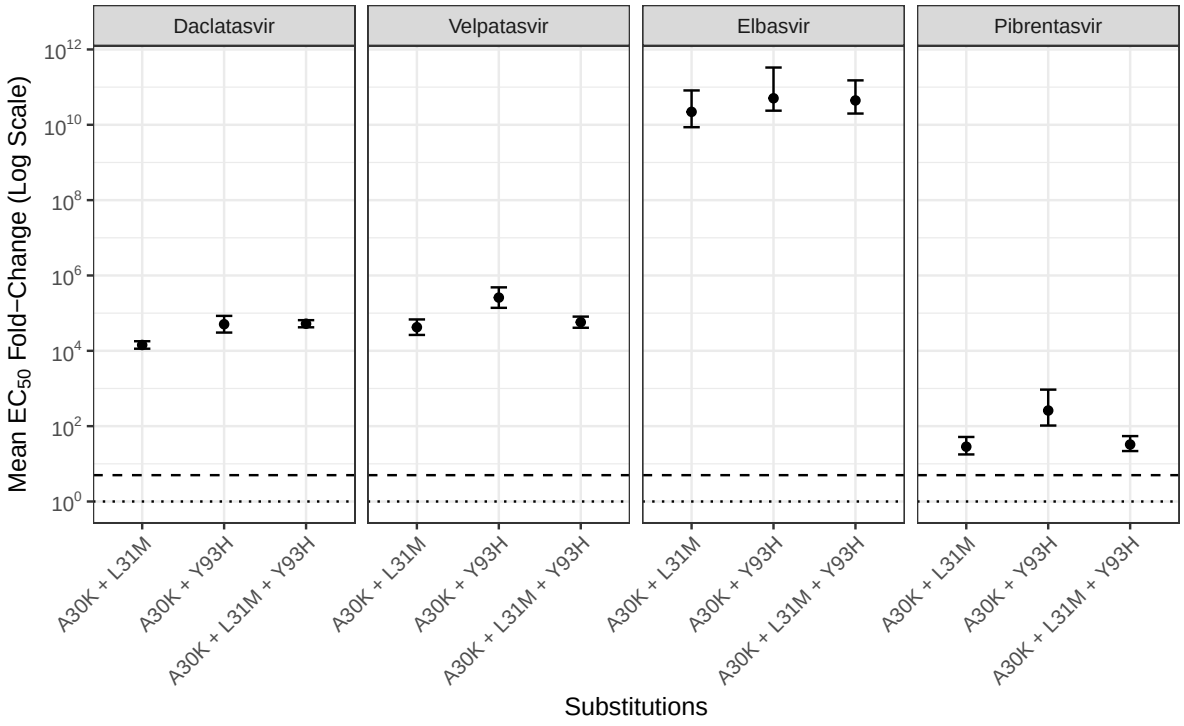

Figure 2.7: **EC50 fold change for daclatasvir, velpatasvir, elbasvir or pibrentasvir against S52-ΔN replicon carrying candidate RAS combinations.** Huh7.5 SEC14L2 cells were electroplated the luciferase expressing S52-ΔN replicon containing the A30K + L31M, A30K + Y93H and A30K + L31M + Y93H resistance associated substitutions. The cells were then treated with 10-fold dilutions of daclatasvir, velpatasvir, elbasvir or pibrentasvir starting at 100 μM, the EC50 was then calculated using regression analysis for each mutant and the fold change from the WT S52-ΔN EC50 calculated. The mean EC50 fold change is plotted for each RAS carrying replicon with 95% CI calculated from at least 2 experiments. The dotted line indicates WT susceptibility, the dashed line indicates a 5-fold increase in EC50 when compared to the WT S52-ΔN replicon.

2.5.8 Distribution of resistance associated substitutions within genotype 3 subtypes

To better understand the population structure of RAS carrying genotype 3 HCV viruses a maximum likelihood phylogenetic tree of all 496 consensus HCV sequences was constructed using RAxML [40] (figure 2.8). The tree breaker software was used to look for phylogenetic clustering of RAS and identify evolutionary change points. The distribution of RAS conferring high level resistance to NS5A inhibitors was strongly associated with the genetic background of the virus figure 2.8. The L31M RAS was only observed in non-gt3a subtypes (gt3b, 5 samples and gt3g, 2 samples). Additionally, all non-gt3a samples carried the A30K RAS

(gt3b, 5 samples; gt3g, 2 samples; gt3i, 2 samples). Of the seven patients carrying gt3b and gt3g viruses, six were of Asian origin and one was of unknown origin. No other clusters of RAS where identified and the Y93H RAS was distributed throughout the phylogeny. To further evaluate the global distribution of the A30K + L31M combination in the gt3 subtypes, 27 gt3 full genome consensus sequences from the NCBI together with the 9 non-gt3a sequences from this cohort, were analysed for the presence of the A30K + L31M substitutions figure 2.9. Interestingly the A30K + L31M substitutions were found in all but two gt3b and gt3g sequences as well as the one gt3d and the two gt3k sequences, suggesting that the A30K + L31M substitutions are the wildtype amino acids for these subtypes.

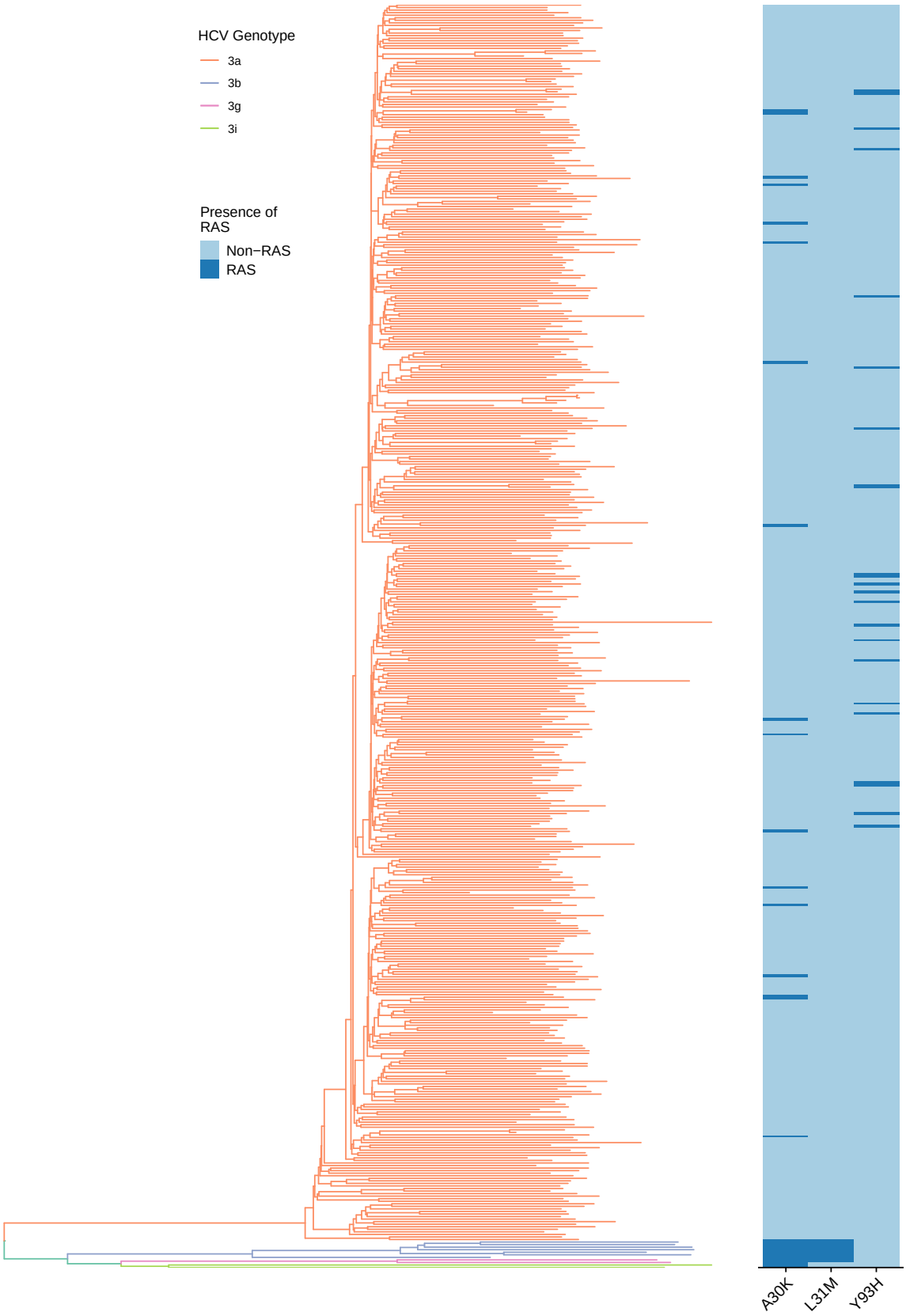

Figure 2.8: **Whole-genome phylogeny of 496 gt3 HCV sequences and heat map showing the presence of NS5A RAS and the HCV genotype of each taxa.** A neighbour-joining, mid-point rooted phylogenetic tree was estimated from the whole genome sequence of all 496 sequences in cohort, using the R package ape [150]. The clades are coloured by the HCV subtype and the tree is annotated with a heat map showing the presence of the A30K, L31M, Y93H RAS in each sequence. The A30K + L31M RAS combination is only found in the gt3b and gt3g sequences.

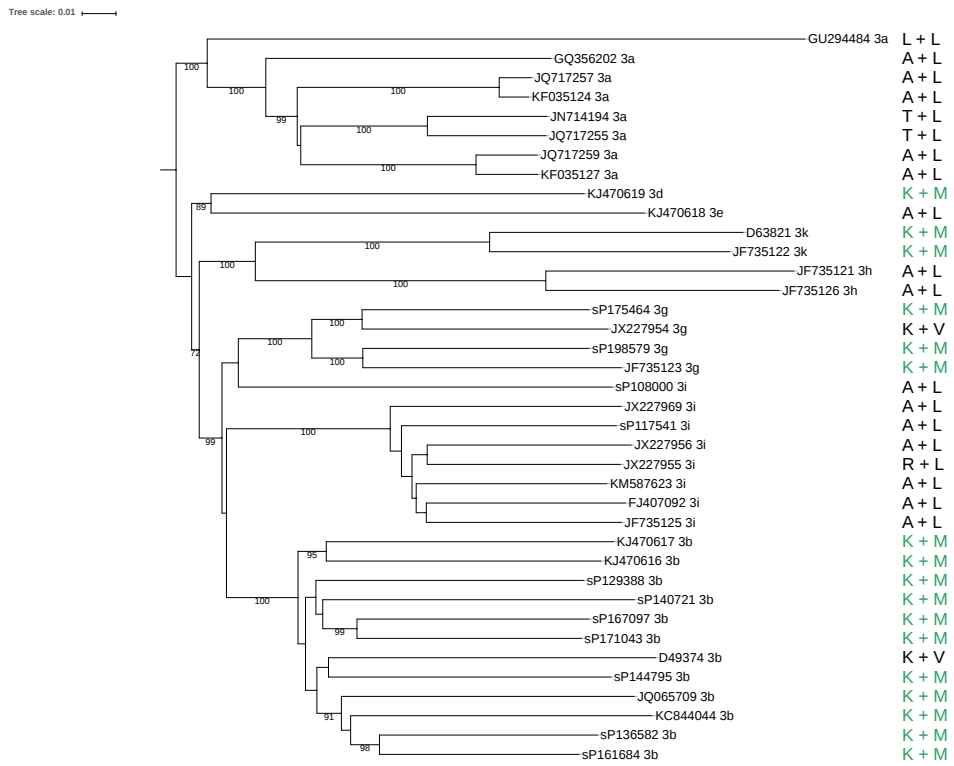

Figure 2.9: **Whole-genome phylogeny of gt3 subtypes labelled with NS5A position 30 and 31 amino acids.** A neighbour-joining, mid-point rooted phylogenetic tree was estimated using the whole genome sequence of the gt3b, gt3g and gt3i subtypes from this cohort and gt3a, gt3d, gt3e, gt3k, gt3h, gt3i, gt3g and gt3b sequences downloaded from the NCBI database, and annotated with the NS5A position 30 and 31 amino acids. The A30K + L31M RAS are coloured green.

2.5.9 Efficacy of NS5A inhibitor against replicon containing gt3b or gt3g NS5A gene

To test if the NS5A gene from gt3b and gt3g HCV provides inherent resistance to NS5A inhibitors. The NS5A gene from a gt3b and gt3g HCV sequence was inserted into the pSGR-JFH-1 replicon and the efficacy of DCV, velpatasvir, EBV and PIB was measured via a dose response curve figure 2.10. The 3b-JFH-1 and 3g-JFH-1 chimeric replicons showed similar resistance to DCV, velpatasvir and EBV to that demonstrated using the S52-ΔN replicons carrying the A30K + L31M substitutions, however the EC50 fold change was lower. 3b-JFH-1 showed a modest increase in EC50 when compared to the WT JFH-1 replicon and 3g-JFH-1 showed a similar EC50 compared to JFH-1. However, PIB showed a decrease in EC50 when the 3b-JFH-1 and 3g-JFH-1 are compared to the S52-ΔN replicon. This suggests that gt3b

and gt3g are more susceptible to PIB than other NS5A inhibitors.

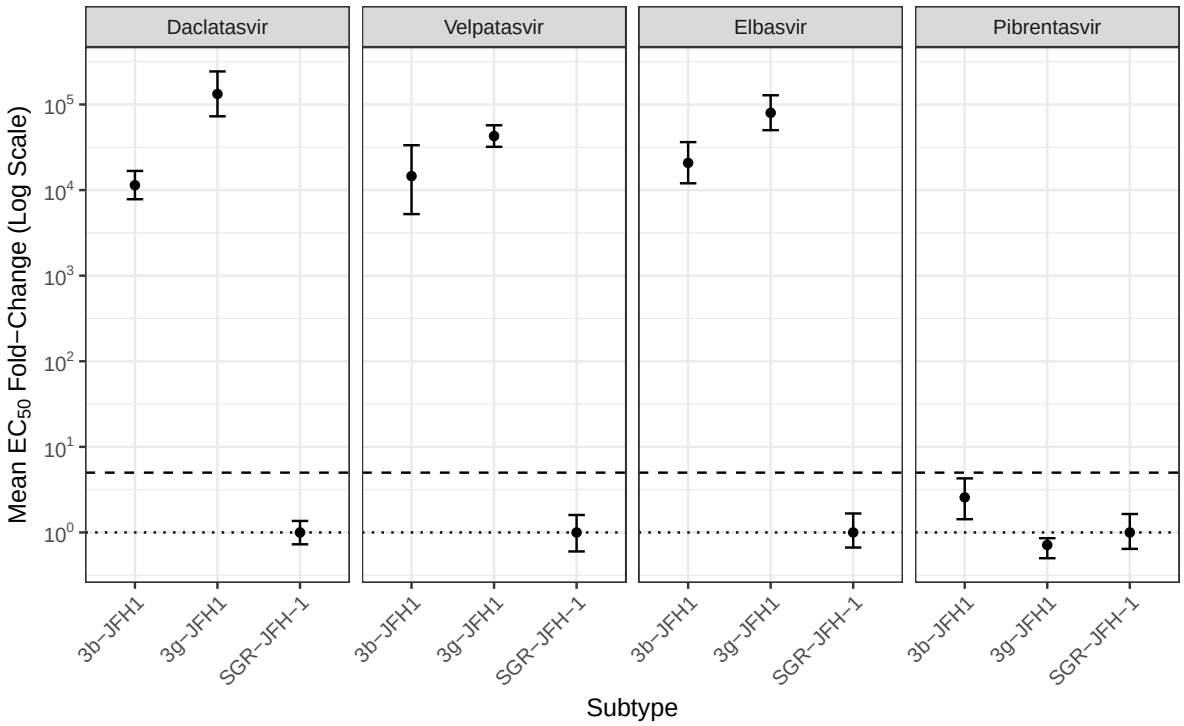

Figure 2.10: **Mean EC50 fold change for daclatasvir, velpatasvir, elbasvir or pibrentasvir against pSGR-JFH-1 replicon carrying NS5A gene from gt3b and gt3g HCV.** Huh7.5 SEC14L2 cells were electroplated the luciferase expressing pSGR-JFH-1 replicon containing the NS5A gene from a gt3b and gt3g seqeunce. The luciferase activity was measured at 4 and 72hrs. The replication capacity was then calculated using the equation in section 2.4.4 the mean replication capacity is plotted on the x axis with 95% CI from 3 replicates technical replicates from 3 experiments. The top dashed and bottom dotted lines indicates the replication capacity of the positive control an unmodified WT pSGR-JFH-1 replicon and negative control replicative-defective pSGR-JFH1/GND replicon respectively.

2.6 Discussion

This work reports the prevalence of RAS in a large cohort of DAA treatment-naive gt3 infected patients using next-generation virus whole-genome sequencing. From these data, a list of RAS was compiled for further characterisation using the S52 gt3a replicon.

This data show that the frequency of RAS to DCV or SOF in gt3 was 22% and 2% respectively, comparable to those described recently by Chen et al [151] (27.6% and 3.9%) and Patino-Galindo et al [152], although both studies used consensus sequences obtained from

publicly available databases. In these studies substitutions representing deviations from the consensus sequence were considered to be potential RAS and were based on smaller sample sizes for gt3 (Chen et al had a sample size of 48 and Patino-Galindo et al used sequences of individual genes with samples sizes of 117 NS3, 226 NS5A and 81 NS5B respectively). These data highlight a need for more study into gt3 RAS to allow clearer definitions.

Next generation sequencing techniques such as the ve-seq [35] method, used in this work, allow the detection of very low frequency viral variants 0.5 – 1% [135]; this sensitivity is much better than Sanger capillary based population sequencing used in the majority of studies of RAS in which minor variants at a lower frequency than $\sim$20% cannot be resolved unless individual clones are sequenced which is time consuming and labour intensive [153]. For example, 40% of the occurrences of gt3 NS5A RAS had variant frequencies of less than 15% that would have been undetectable by older methodologies.

This work shows that in gt3, the individual RAS A30K and Y93H provide modest resistance to DCV and that the combinations of A30K + L31M, A30K + Y93H, A30K + L31M + Y93H provide a dramatic increase in resistance to velpatasvir, DCV, EBV and PIB. This data agree with Hernandez et al [154] showing a similar phenotype for both A30K + L31M and A30K + Y93H using a JFH1-gt3a NS5A chimera replicon when treated with DCV. Additionally, this data show, that L31M and P58S show little or no evidence of resistance to DCV in the gt3a replicon system. There is also, evidence that S282T and V321A substitutions in NS5B provide small levels of resistance to SOF in this assay.

The relationship between the occurrence of RAS and clinical resistance to treatment is additionally influenced by the impact on viral fitness of substitutions that depart from the consensus sequence. Clearly, RAS, however effective they may be at reducing viral susceptibility, may be of no clinical relevance if the mutant viruses are unable to replicate in-vivo. On this basis, the A30K + L31M combination of substitutions should be viewed with substantial concern clinically as they have been observed in patients, confer both a large increase in EC50 when compared to WT and substantially enhance viral replication competence, a surrogate measure of basic fitness, albeit in a cell culture-based replicon format.

Unexpectedly, this combination was only detected in gt3b and gt3g samples and suggests

that its occurrence in gt3a may therefore be associated with an undocumented fitness cost or incompatibility not replicated in in vitro assays. The preferential distribution of this RAS combination in non-gt3a subtypes, typically distributed in South and South East Asia [155] suggests these gt3 subtypes may be naturally more resistant to NS5A Inhibitors. A clinical trial of SOF and Velpatasvir based in China, Thailand, Vietnam, Singapore and Malaysia has shown that the SVR rate of patients infected with gt3b and with cirrhosis was only 50% compared to 89% in gt3b infected patients without cirrhosis and 97% for the entire cohort [156]. There is also growing evidence that subtypes which have been under represented in clinical trials, due to their rarity in western Europe and north America also show reduced susceptibility to DAA combinations. Da Silva Filpe, et al show in a cohort from the UK that gt1l and gt4r infected patients failed multiple rounds of therapy and that these subtypes contain RAS motifs which could cause resistance[69]. The combination of LBV and SOF (Harvoni) has been shown by Fourati, et al to have dramatically reduced efficacy against gt4r than gt4a. It has also been shown to develop the S282T substitution, which is associated with resistance to SOF more readily [157]. This is potentially worrying as these subtypes appear to be most commonly found in countries with limited resources and where the epidemiology of HCV infection is not well understood [158]. In the BOSON cohort, all patients were treated by SOF + ribavirin +/- PEG-interferon and all the non-gt3a-infected patients achieved sustained viral clearance, except one gt3i infected patient, precluding direct testing of NS5A inhibitor resistance on re-treatment, in this cohort. The high resistance phenotype was observed when these RAS combination A30K + L31M were introduced onto the gt3a replicon and may display a different phenotype when present in their native subtype. The data from the 3b-JFH-1 and 3g-JFH-1 replicons, shows similar levels of resistance to DCV, LBV and EBV when the entire NS5A gene from gt3b is inserted into a JFH-1 replicon. However, PIB seems to be more effective against both 3b-JFH-1 and 3g-JFH-1 and the A30K + L31M S52-ΔN, this suggests that this drug in combination with the protease inhibitor glecaprevir may be more effective than other combinations against non-gt3a gt3 subtypes.

The question of at what frequency a RAS becomes clinically relevant remains challenging. EASL guidelines recommend the use of resistance testing for the NS5A region only for patients who are to be retreated having failed previous all-oral DAA treatments [51]. When NGS techniques are used, it is recommended that variants with a frequency of less than 15% not be considered clinically relevant [51]. Interestingly in all samples in which the A30K +

Y93H substitutions were detected, at least one of the two substitutions were detected with a frequency of less than 15%. Because of the distance between the two sites and the short fragments used in Illumina sequencing, the linkage of these two RAS cannot be confirmed using this data. This data also suggests that the A30K + Y93H combination may not be a fit variant in-vivo as well as in vitro.

Overall, this work shows relatively high frequencies of RAS to the approved NS5A inhibitor DCV in gt3 HCV. It also shows that the A30K + L31M RAS combination is the WT for gt3b and gt3g HCV and this combination provides resistance to all NS5A inhibitors approved for the treatment of gt3 HCV. These findings support the recommendation from AASLD and EASL [51, 159] and EASL that support the use of NS3 inhibitors in combination with NS5A inhibitors and especially the use of the GLE/PIB combination as PIB was shown to be the most effective against these subtypes. Additionally, these findings highlight the need for more study into the efficacy of DAA therapy in less common gt3 subtypes.

2.7 Summary of key findings

- High frequency of RAS to NS5A inhibitors in treatment naive gt3 infected patients.
- A30K + L31M RAS combination is WT for gt3b and gt3g.
- A30K + L31M and A30K + Y93H RAS combinations dramatically increase the EC50 of all NS5A inhibitors approved for the treatment of gt3 infection.

Chapter 3

Viral determinates of outcome in sofosbuvir based treatment of genotype 3 Hepatitis C Virus infection.

3.1 Abstract

Background and Aims: Sofosbuvir (SOF) is a key component in the treatment of HCV infection, with a higher barrier to resistance. Novel amino acid substitutions associated with treatment failure identified from a large cohort (n=507) of gt3a infected patients treated with SOF + RBV +/- PEG-INF are reported in this chapter.

Methods: Next generation whole genome sequencing (WGS) of HCV was performed on all baseline (BL) and 12 week post-treatment (12WPT) viraemic samples. A viral genome wide association study (V-GWAS) was undertaken to study the effect of BL consensus polymorphisms across the viral polyprotein on treatment outcome. The effect of BL polymorphisms identified in the V-GWAS on the initial reduction of viral load was investigated. BL and 12WPT sequences were compared and treatment emergent substitutions investigated.

Results: The V-GWAS identified three novel BL polymorphisms significantly associated with a negative treatment outcome (at a 25% false discovery rate) 119A and 132V in the NS2 protein and 67V in the NS3 protein, the presence of these polymorphisms at baseline was associated with a lower reduction in viral load after 1 week of therapy. Within the NS5B protein the substitution A150V had strongest association with outcome and was significantly enriched in the 12WPT samples (p=9x10-4).

Conclusions: This chapter for the first time reports the use of HCV WGS and V-GWAS to identify novel amino acid substitutions associated with SOF treatment failure in HCV gt3a infection.

3.2 Background

SOF was the first non-protease inhibitor DAA approved, and is a nucleotide analogue that competitively binds to and blocks the HCV polymerase (coded by non-structural 5B (NS5B) gene) and inhibits viral replication [160]. SOF is a key component in many HCV treatment regimens and is now only recommended for use in combination with other DAAs such as NS5A inhibitors (ledipasvir, daclatasvir or velpatasvir) and NS3 protease inhibitors such as voxilaprevir [51, 52]. Additionally, SOF in combination with velpatasvir and voxilaprevir has become the backbone of most of the recommended regimens for patients who have been failed by previous NS5A inhibitor containing therapy because of its pan-genotypic activity and high barrier to development of resistant virus. Presence of RAS in viral sequences has been shown to be a factor in treatment failure (10) and viral genetics is still one of the main determinants of treatment responsiveness, particularly virus geno/subtype. Baseline RAS for both NS3 and NS5A inhibitors are reported to be common in many genotypes [43] and certain subtypes such as gt3b and gt4r have been shown to be inherently resistant to current NS5A inhibitors and SOF containing therapies [69, 98, 157]. NS5A RAS cause a large reductions in the activity of NS5A inhibitors which would make SOF the only active compound in NS5A + SOF treatment.

SOF has a high barrier to resistance and very few RAS have been reported [102, 161]. In multiple clinical trials using SOF, the treatment emergent NS5B RAS L159F, S282T, C316H/N, L320F and V321A have been detected in patients whom do not achieve SVR [102]. With the exception of S282T these substitutions have all showed negligible increases in the EC50 of SOF using in-vitro models [98]. The treatment emergent SOF RAS have been shown to have a large fitness costs. However, S282T has been shown to persist during viral selection studies if compensatory mutations are also acquired [103] and certain HCV subtypes such as genotype 4r have been shown to acquire the S282T substitution more readily [157]. However, many patients fail treatment without any detectable RAS, and no comprehensive study into pre existing polymorphisms within a HCV subtype which may impact of the efficacy of SOF has been performed. A comprehensive understanding of the viral determinants of response to SOF is crucial to ensure that resistant HCV will not become a significant problem in the coming years and will help us to stratify patients for more effective treatments, especially in hard to treat patient groups or patients who have already been failed by DAA treatment.

3.3 Aims

This analysis uses full length HCV genomes along side host genotyping and clinical data from patients infected with gt3a HCV and treated with SOF to identify viral determinants of resistance to SOF in gt3a HCV infection.

3.4 Methods

3.4.1 Subject and samples

The patients from this work were enrolled in the BOSON Study further details of which can be found in chapter 2 section 2.4.1. Briefly, 592 gt2 and gt3 infected patients were randomised into three treatment arms of (i)12 weeks SOF + RBV + IFN, (ii) 16 weeks SOF + RBV and (iii) 24 weeks SOF + RBV. All patients were DAA treatment-naïve. Each patient's HCV viral load was measured prior to treatment commencing (BL), weeks 1,2,4 and 8. Samples were obtained from patients at BL and 12 weeks after the end of treatment (12WPT).

3.4.2 Viral sequencing

A detailed description of the how the patient samples were sequenced and the bioinformatic pipeline used to process the sequencing data can be found in chapter 2. The outputs from the bioinformatics pipeline were analysed using custom code and the R statistical package.

3.4.3 Statistical analysis

Identification of viral amino acid substitutions associated with outcome using a viral genome wide association study

To investigate the effect of individual viral amino acid substitutions on SVR a genome wide association study (V-GWAS) was performed using a logistic regression model. To ensure possible confounders were accounted for in the model, a multivariate analysis was performed using logistic regression to identify non-viral factors, which are associated with SVR. Akaike Information Criterion was used to choose the set of covariates to include in the model. The resulting model included, patient's cirrhosis status, IFNL4 genotype, prior treatment status, gender and log baseline viral load. To ensure that host-virus population co-structuring had minimal impact on the analysis, the cohort was limited to patients infected with HCV gt3a. The first three viral principal components calculated from viral sequences and the first three host principal components calculated from the host genome-wide genotyping data were also included as covariates in the model, to further control for both host and virus population structures.

Phylogenetics and association of a phenotype with tree structure

RAxML [40] was used to build a maximum likelihood phylogenetic tree using the full length HCV genome sequences. A general time reversible model of nucleotide substitution under the gamma model of rate heterogeneity was used. The treeBreaker [162] software was then used to investigate if a phenotype of interest or any novel RAS are associated with specific clades on the tree.

Covariation and interaction analysis

Logistic regression was used to investigate covariation of any novel RAS identified and other sites in the viral genome and to test for interaction between covarying sites and its impact on

SVR.

Enrichment of an amino acid in the post-treatment Hepatitis C Virus sequences

To test for enrichment of an amino acid in the post-treatment samples of patients who did not achieve SVR, the binomial test was used to detect if the frequency of an amino acid has increased in the post-treatment samples relative to the baseline samples.

Viral diversity analysis

Using NGS data, within patient diversity of each site in the genome was calculated using nucleotide pairwise diversity. The relevant values for the gene or region of interest were averaged to get a measure of viral quasispecies per patient per region of interest. As an alternative measure of within patient diversity, Shannon entropy of each codon in the coding region of the HCV genome as well as each amino acid site in the polyprotein was calculated and averaged across the genes and regions of interest.

Response to therapy analysis

Each patient's response to therapy was estimated by calculating the log10 change in viral load from BL (prior to therapy) to Wk1 of therapy. The Wilcoxan test was used to test for differences between patients carrying novel RAS and those not.

Replication cohort

The results were replicated in an independent cohort. Using whole HCV genome sequences from 144 patients enrolled in the HCV Research UK cohort who were treated with SOF containing regimens. One-sided logistic regression was used to replicate the results from the BOSON cohort. All the patients were cirrhotic and the host genotyping data was not available

so no co-variables were included.

3.5 Results

3.5.1 Patient demographics and non-viral factors associated with outcome

Baseline whole genome HCV sequences were obtained for 568 patients, 518 of which were gt3, 49 were gt2 and one was a 1a/2b recombinant virus. HCV is a highly diverse pathogen and to limit the impact of virus population structure on the analysis, only samples from patients infected with subtype 3a were included (N=507). In this cohort 82% (416/507) patients achieved an SVR. To account for non-viral factors associated with SVR, logistic regression was used to test for association between treatment outcome and patient's liver cirrhosis status, gender, baseline viral load, prior IFN-based treatment and IFNL4 SNP rs12979860 genotypes (CC vs non-CC) in a multivariate model. Liver cirrhosis status was significantly associated with SVR status ($P=1.5x10^{-4}$), having the largest impact on treatment response (failure rate in cirrhotic patients 29%=47/163 and in non-cirrhotic patients 13%=44/344). Male gender and the non-CC genotype of the IFNL4 SNP rs12979860 were also significantly associated with increase in treatment failure rate ($P_{gender}=0.016$, $P_{IFNL4}=6.4x10^{-3}$). Previous IFN-based treatment and higher viral load were associated with increase in treatment failure although the effect was not statistically significant (Figure 3.1). In all subsequent regression-based analysis these factors were added as covariates to account for possible confounders.

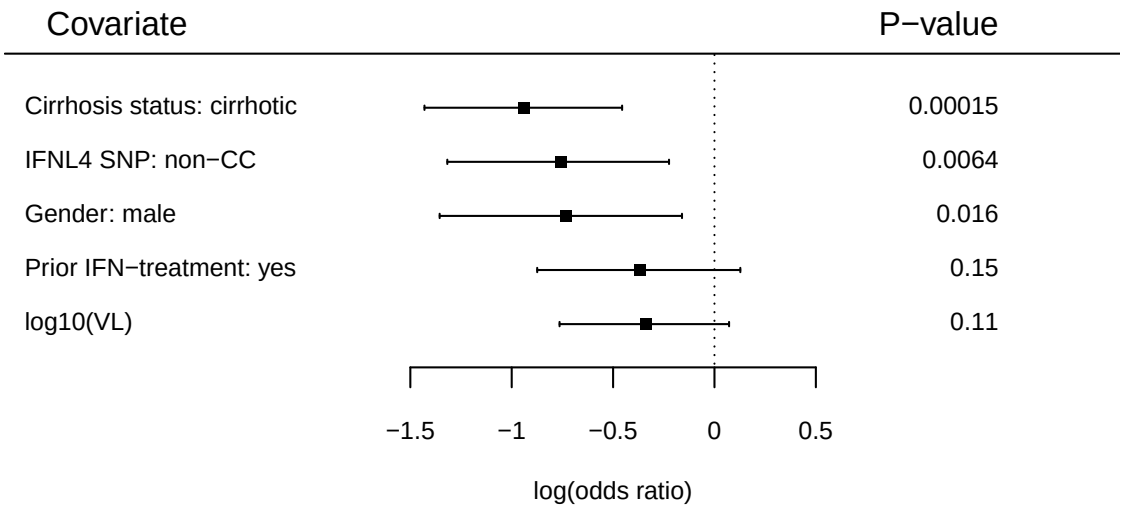

Figure 3.1: **Non-viral factors impact on SVR using a multivariate logistic regression.** The squares show the estimated effect size (log(OR)) for each covariate and the lines show its 95% confidence interval, the p value is shown on the right.

3.5.2 Previously reported resistant associated substitutions

Sequences were scanned for the presence of previously reported SOF RAS (NS5B L159F, S282T, C316H/N L320F and V321A) in the baseline sequences. L159F was detected in one patient's sample at the consensus sequence level. No other RAS were detected in patients viral consensus sequences. However, S282T (n=2), C316N (n=1), L320F(n=2) and V321A (n=8) were detected in $<$1% of reads from each individual patient.

3.5.3 Viral population structure and outcome

The viral population structure was analysed for HCV lineages associated with changes in SVR rate. To do this a maximum likelihood phylogenetic tree from baseline consensus sequences was constructed using RAxML [40]. The treeBreaker software [162] was used to estimate the evidence for the association between the HCV phylogenetic tree and the SVR status. No evidence for any clades having a different SVR rate from the rest of the tree was found (Bayes factor= 0.969) (figure 3.2).

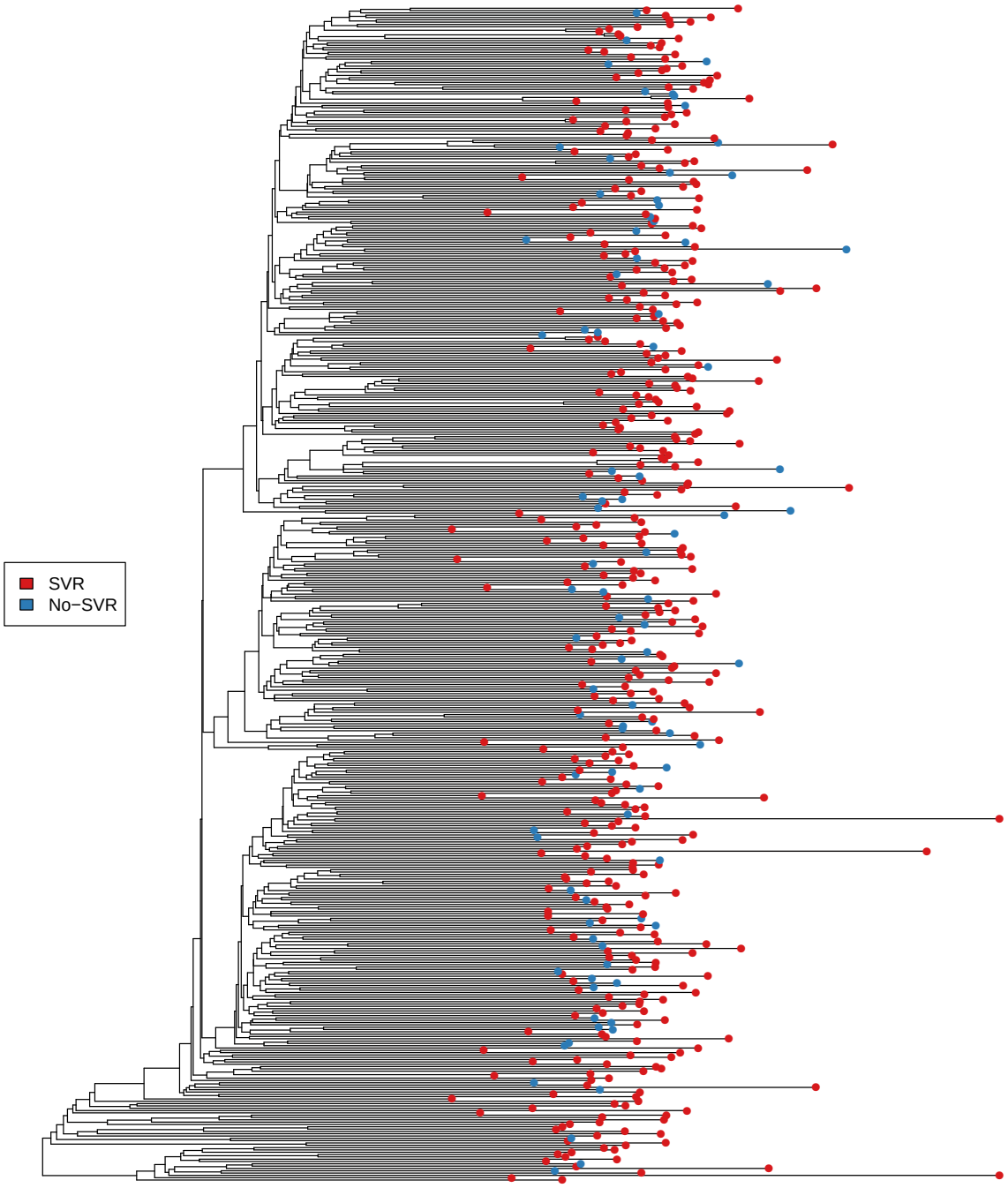

Figure 3.2: **Association between virus population structure and SVR.** A maximum likelihood phylogenetic tree for the hepatitis C virus (HCV) of genotype 3a HCV infected patients was constructed using. The colour of the tips shows if the patient achieved a sustained viral response (red) or not (blue). Red lines show change points from which the samples within that clade are different from the rest of the tree. The phylogenetic tree was constructed using RAxML [40] and the change point analysis conducted using treeBreaker [162].

3.5.4 Identification of novel viral polymorphisms associated with sustained viral response using viral genome wide association study

A major advantage of the whole genome sequencing of HCV is the possibility of testing viral genetic variants for association with SVR at a genome-wide scale to identify potentially novel associations between HCV genetic variants and the response to SOF treatment. Logistic regression was used to test for association between response to SOF treatment and HCV amino acids, including 3 viral principal components (PCs) and 3 host PCs as covariates to account for host-virus population co-structuring as well as the possible confounders previously mentioned. SVR status was used as the response variable and the presence and absence of each amino acid as the explanatory variable. This resulted in 1010 tests at 484 sites. At a 15% false discovery rate (FDR), three viral sites were significantly associated with SVR (Figure 3.3, Table 3.1 and Supplementary Figure B.6). The most significant association was at site 132 in NS2 protein ($P=1.05x10^{-5}$). At this site the amino acids isoleucine (80%=405/506) and valine (20%=100/506) were observed and one isolate carried leucine. Valine was associated with a reduction in treatment response (SVR 66%=66/100) relative to the isoleucine residue (86%=348/405). The other two sites significantly associated with SVR were in NS3 (site 67, $P=4.71x10^{-4}$) and NS2 (site 119, $P=5.18x10^{-4}$) proteins.

In the NS5B protein, site 150 was the most associated site with SVR ($P=4.92x10^{-3}$) and the fifth most associated site genome-wide. The most common amino acid residues at this site were valine (41%=204/501), alanine (38%=191/501), threonine (13%=63/501), isoleucine (5%=27/501) and a small number of other residues. At site 150, Valine was associated with the lowest rate of SVR (75%=153/204) vs all other amino acids (86%=257/297). This site has been previously reported to be associated with host IFNL4 SNP rs12979860 [163] and SVR in a subset of this cohort [164]. Among CC patients the impact of valine on SVR rate was minimal (SVRvaline 85%=35/41, SVRnon-valine 87%=129/148) while among non-CC patients valine reduced SVR rate substantially (SVRvaline 72% = 117/162, SVRnon-valine 85%=127/148, Fisher's exact test P=0.0036). However, when tested using an interaction term in the model this was not statistically significant (P=0.36). The amino acid that accounted for the largest reduction in SVR rate at the three sites significantly associated with SVR (NS2 119A, NS2 132V, NS3 67V) and the top most associated site in the NS5B protein (NS5B 150V) were taken forward as putative RAS for further investigation (Table 3.2) .

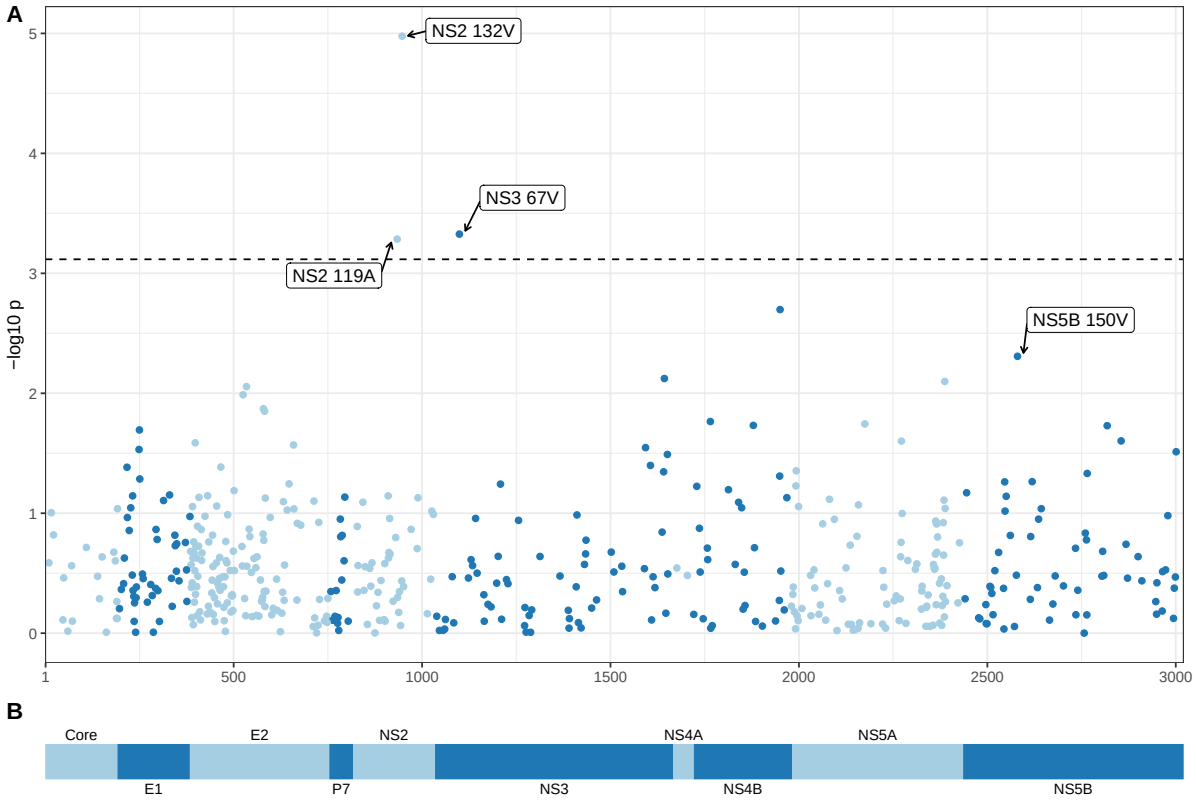

Figure 3.3: **Association between HCV amino acids and sustained virologic response after sofosbuvir based treatment.** (a) Manhattan plot of association tests between HCV amino acids and SVR. For each site only the most associated p-value is plotted. The dashed line indicates 15% FDR. For the three sites significantly associated with SVR and the most associated site in the NS5B protein, the putative resistance associated amino acid and its position within its respective protein is indicated. (b) Schematic of HCV polyprotein.

Position on polyprotein relative to H77	Protein	Position in protein and amino acid	SVR rate (%)	Log odds ratio	Standard error	P value	Q value
941	NS2	132V	66%	-1.27	0.29	1.05E-05	9.65E-03
1093	NS3	67V	77%	-1.17	0.33	4.71E-04	1.31E-01
928	NS2	119A	59%	-1.88	0.54	5.18E-04	1.31E-01
1944	NS4B	233V	58%	1.54	0.50	2.01E-03	2.95E-01
2570	NS5B	150V	75%	-0.76	0.27	4.92E-03	6.21E-01

Table 3.1: **Top 5 most associated amino acids with SVR in V-GWAS.** Only the most associated amino acid is reported at each site. At 15% FDR, the first three sites are significantly associated with SVR.

Protein	Position in Protein	Amino Acid	Frequency at Baseline %	SVR Rate % (n)
NS2	132	I	80.0%	86% (348/405)
		V	19.8%	66% (66/100)
NS3	67	A	53.6%	86% (232/271)
		V	43.9%	77% (171/222)
NS2	119	A	4.4%	59% (13/22)
		M	4.2%	81% (17/21)
		V	90.5%	83% (381/457)
NS5B	150	A	38.1%	88% (168/191)
		T	12.6%	83% (52/63)
		V	40.7%	75% (153/204)

Table 3.2: **Baseline frequency and SVR Rates for amino acids at novel RAS sites.** Only amino acids detected in more than 20 patients are included.

3.5.5 Phylogenetic analysis of Novel RAS

To understand if any particular HCV lineage was associated with any of the sites identified in the V-GWAS and to further rule out viral population structure confounding, a maximum likelihood phylogenetic tree from baseline consensus sequences was used with the treeBreaker software [162] to detect if the distribution of putative RAS was distinct on any branches of the tree (Figure 3.4). One lineage was associated with the presence of NS2 132V (Bayes factor = 13.12, Supplementary Figure B.2) and two lineages were associated with NS3 67V (Bayes factor = 18180.81, Supplementary Figure B.3), NS2 119V (Bayes factor = 0.96, Supplementary Figure B.1) and NS5B A150V (Bayes factor = 0.99, Supplementary Figure B.4) were not associated with any specific lineages. In the case of NS2 132V and NS3 67V the association with treatment outcome was retested using the previous model with the addition of variable indicating if the sequence was from a lineage associated with the putative RAS. Both putative RAS remained significantly associated with treatment outcome (NS2 132V: $p=1.2x10^{-5}$, NS3 67V: p=0.001).

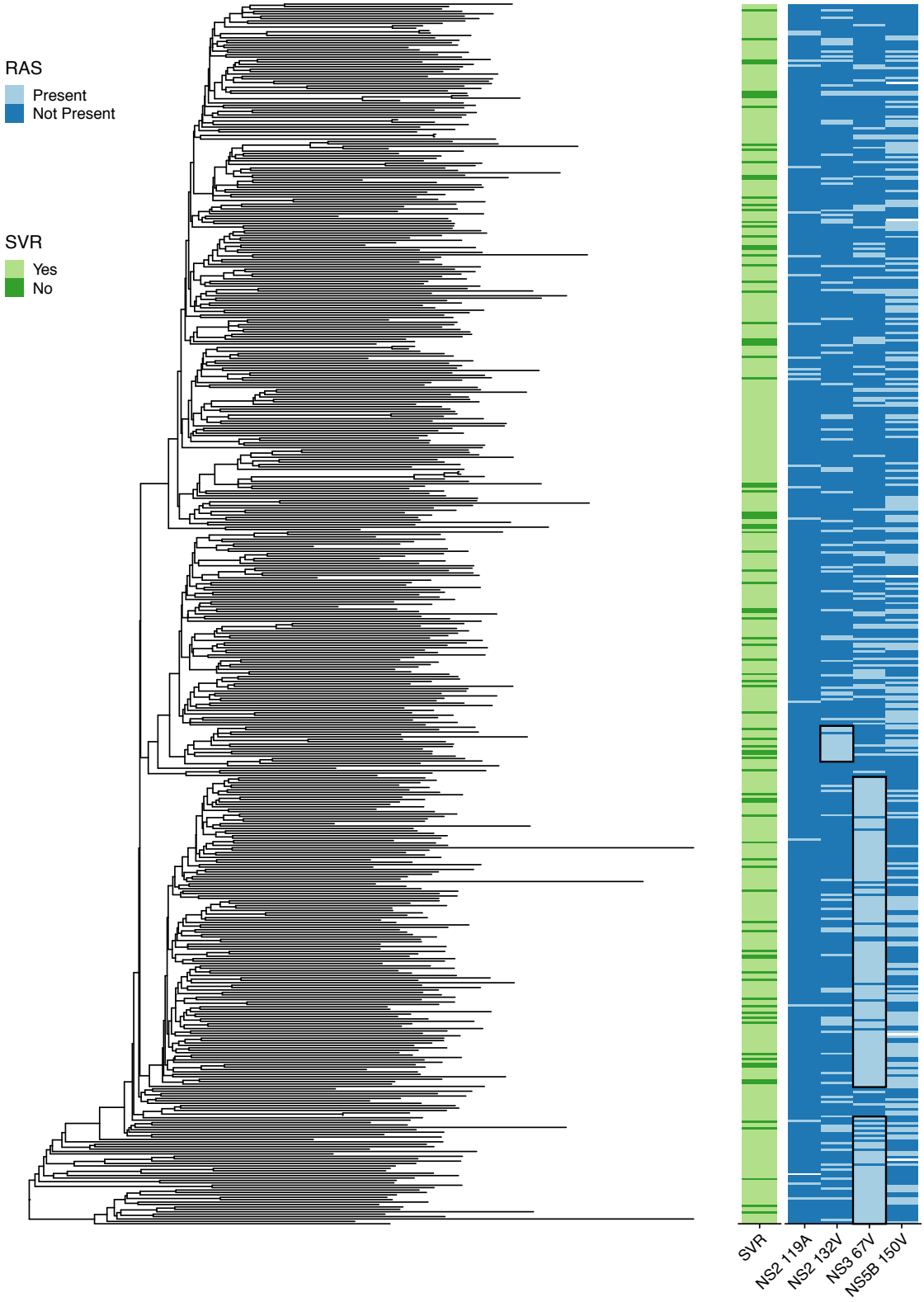

Figure 3.4: **Distribution of treatment outcome (SVR) and putative RAS within HCV phylogeny.** The phylogenetic tree was constructed using RAxML [40] using whole HCV genome sequences. Each row in the heat maps is aligned with the corresponding taxa on the phylogenetic tree and shows if the patient infected with that virus achieved an SVR and the presence or absence of the putative RAS in the consensus viral sequence. The clusters of novel RAS are highlighted in black.

3.5.6 Interaction and co-variation of novel resistance associated substitutions

To understand the impact of presence of multiple putative RAS in each sequence on SVR, patients were stratified based on the number of RAS present in their baseline consensus sequences (Figure 3.5). The SVR rate for patients whose virus carried no RAS was 95% (133/140) which reduced to 50% (13/26) for patients with 3 RAS. Only one patient carried all four putative RAS and this patient achieved SVR. The step-wise reduction in the SVR rate associated with the increase in the number of RAS present at baseline was highly significant (logistic regression $P=6.6x10^{-10}$).

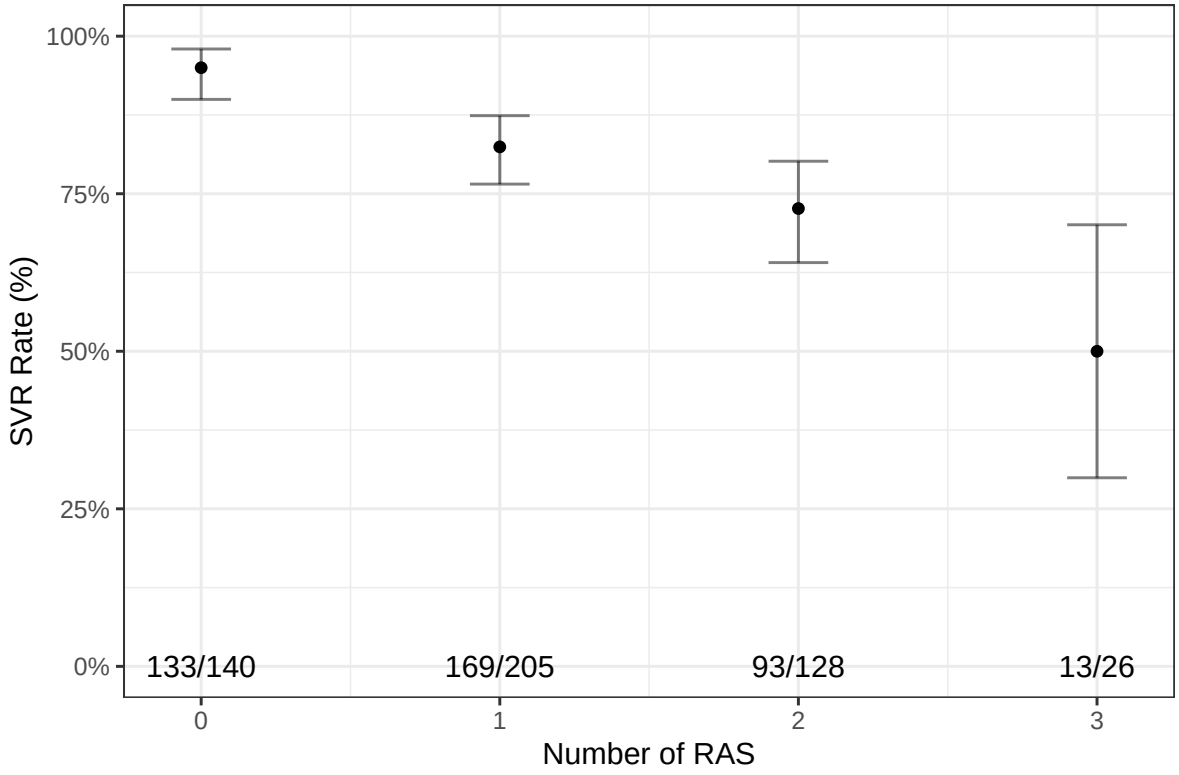

Figure 3.5: **SVR rate for patients with different numbers of RAS at baseline.**The dots indicate the SVR rate in each group and the lines indicate its 95% confidence intervals calculated using Clopper-Pearson interval. The numbers in each group are shown at the bottom of the figure.

To investigate if any specific combination of these putative RAS were driving the drop in SVR rate, the co-variation between these sites was investigated using logistic regression. The presence and absence of the putative RAS at each site was tested against the other RAS, the first 3 viral PCs and host IFNL4 SNP rs12979860 genotypes (CC vs non-CC) were included as

covariates to account for virus population structure and the impact of IFNL4 on the HCV amino acids. NS2 119A was excluded from this analysis because of its low frequency (4%=22/505). Among the six co-variation tests performed, NS5B 150V was significantly associated with NS2 132V (p =0.004) and nominally with NS3 67V (P=0.02). Stratifying the patients into four groups based on the presence and absence of valine at sites NS2 132 and NS5B 150 (presence and absence of the putative RAS), patients carrying non-valine residues at both sites had the highest rate of SVR (91%=228/251). Carrying valine only at NS2 132 was associated with the lowest SVR rate (63% = 29/46, P = $2.3x10^{-6}$). Carrying a valine only at NS5B 150 was also associated with a significant reduction in SVR rate but to a smaller degree (77%=117/51, P = $1.0x10^{-3}$). Carrying valine at both sites results in significant reduction in SVR rate (67%=36/53, P=$1.3x10^{-4}$). However, the effects are not additive (p-value of interaction term is significant $8.4x10^{-3}$) and presence of both moderately recovers SVR rate.

Stratifying the patients by presence and absence of valine at the three sites of NS2 132, NS3 67 and NS5B 150, patients carrying non-valine residues at all three sites had the highest SVR rate (94%=140/149). Carrying any one of these putative RAS individually reduced SVR rate significantly (Figure 3.6) as well as any combinations of them. Patients carrying valine only at NS2 132 and NS3 67 had the lowest SVR rate (50%=7/14).

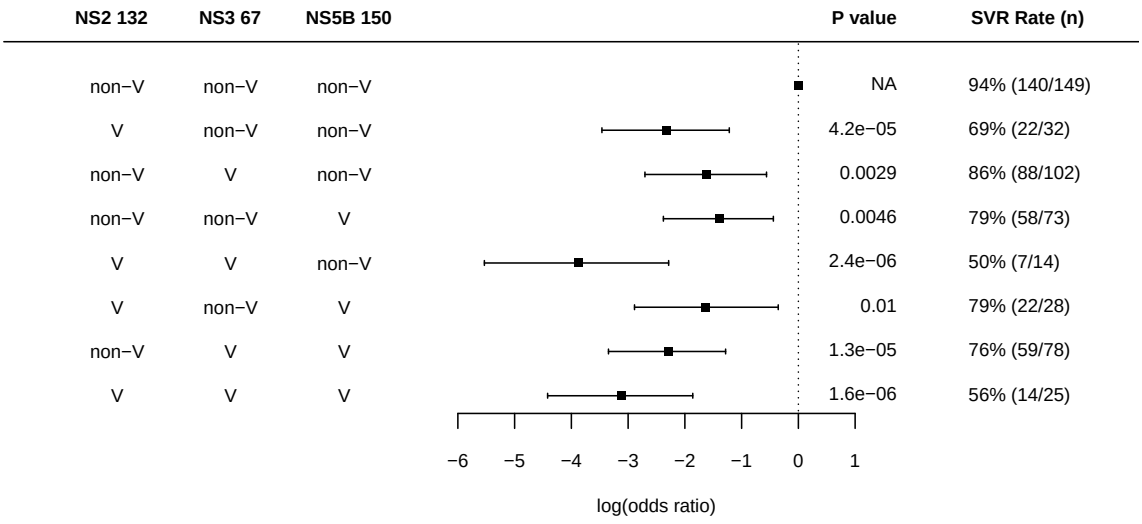

Figure 3.6: **Covariation between the 3 putative RAS and their association with outcome.** Combinations tested are listed in the table on the left. The squares show the estimated effect size (log(OR)) for each variable and the lines show its 95% confidence interval and the P value of the association is shown on the right alongside the SVR Rate for patients infected with virus harbouring these putative RAS combinations.

3.5.7 Replication cohort

The impact of the putative RAS was tested in a replication cohort of 144 patients infected with gt3a HCV. All patients had cirrhosis and 75% (107=/144) were treated with SOF and an additional NS5A inhibitor such as Daclatasvir, ledipasvir or velpatasvir, the remaining 37 patients were either treated with SOF + RBV or SOF + RBV + PEG-IFN. This cohort has a much smaller sample size and many of the patients had received other DAAs in combination with SOF (unlike BOSON in which patients received DAA mono-therapy of SOF) which will reduce the likelihood of replication. Since the direction of the effect tested was known, a one-sided test was used for this analysis. NS2 119A (P=0.045) and the combination of NS2 119A and NS5B 150V were nominally associated with SVR (P=0.034) in the replication cohort. None of the other individual novel RAS or combinations was significantly associated with SVR in this cohort, however the effect size magnitudes and directions were consistent for three out of the four sites (Figure 3.7 and Figure 3.8).

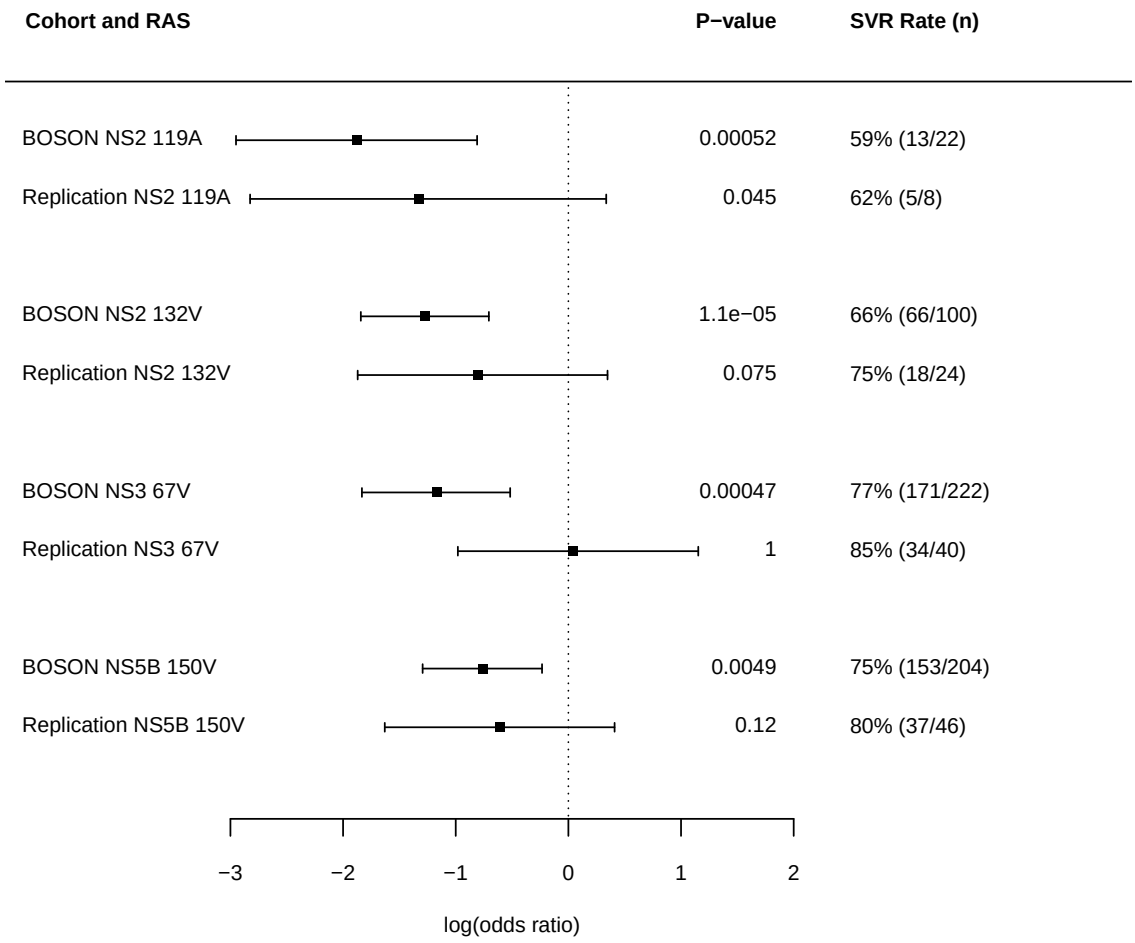

Figure 3.7: **Association between the presence of putative RAS and outcome in BOSON and replication Cohort.** The Cohort and putative RAS is shown on the right. The squares show the estimated effect size (log(OR)) for each variable and the lines show its 95% confidence interval and the P value of the association is shown on the right alongside the SVR rate of patients infected with virus harbouring these mutation combinations.

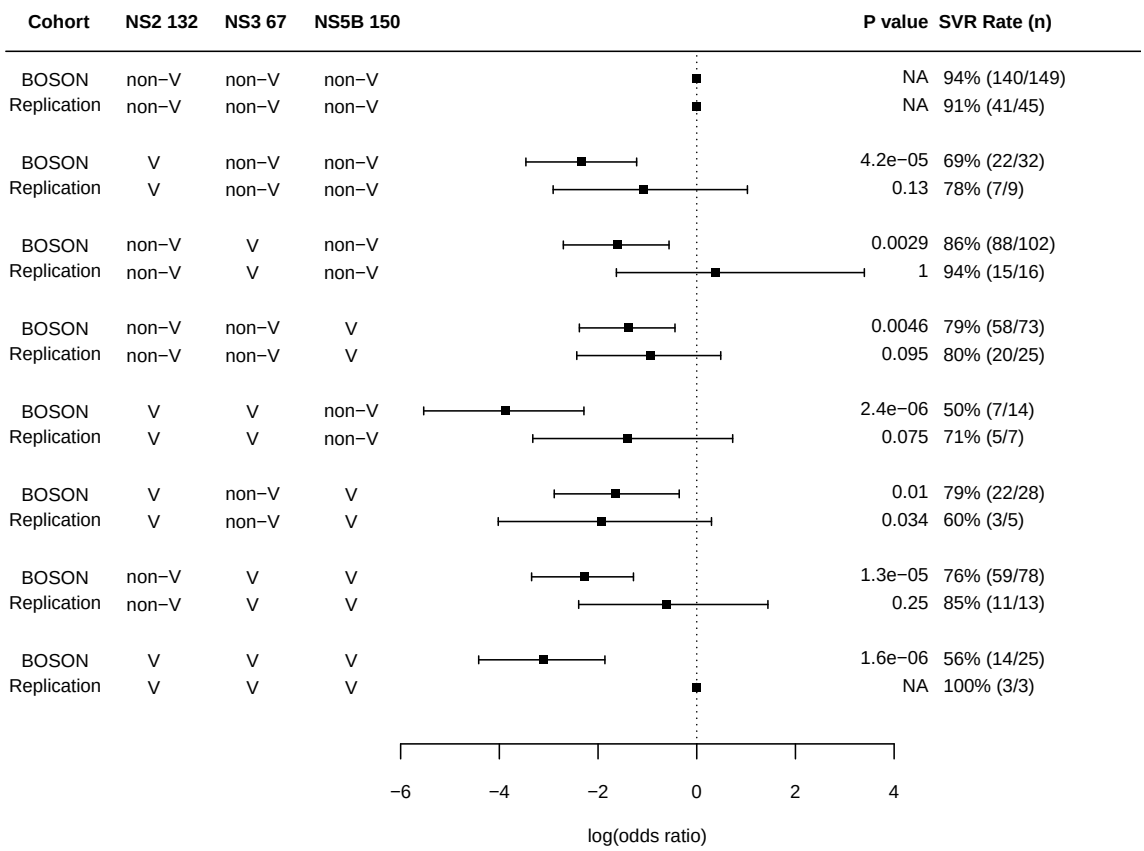

Figure 3.8: **Covariation between the 3 putative RAS combinations and their association with outcome in both the BOSON and Cirrhosis cohort.** Combinations tested are listed in the table on the left. The squares show the estimated effect size (log(OR)) for each variable and the lines show its 95% confidence interval and the P value of the association is shown on the right alongside the SVR rate of patients infected with virus harbouring these mutation combinations. The combination of NS2 132 V, NS367V and NS5B 150V is not shown for the cirrhosis cohort as it was not present in any sequences.

3.5.8 Within Host Viral Diversity and its Effect on Outcome

The NGS data for each sample was used to calculate nucleotide diversity for each site across the whole genome which was then averaged to estimate a single value per genome or gene of interest. A logistic regression was performed to investigate the effect of diversity across the genome and of individual genes on treatment outcome, the possible confounders identified earlier (Figure 3.1) and the host and viral PCs were all included as cofactors. No significant associations were found between either genome-wide diversity, individual gene diversity or diversity at putative or previously reported (NS2 119, 132, NS3 67, NS5B 150, 159, 184, 282, 316, 320 and 321) RAS and treatment outcome (Supplementary Table B.1 and Supplementary

Table B.4).

Shannon entropy was also used to investigate amino acid diversity across the entire poly protein (Figure 3.9), each protein and at each putative and previously reported RAS site and their association with SVR using the same logistic regression model. No significant difference in Shannon entropy was detected at these sites in association with SVR (Supplementary Table B.2 and Supplementary Table).

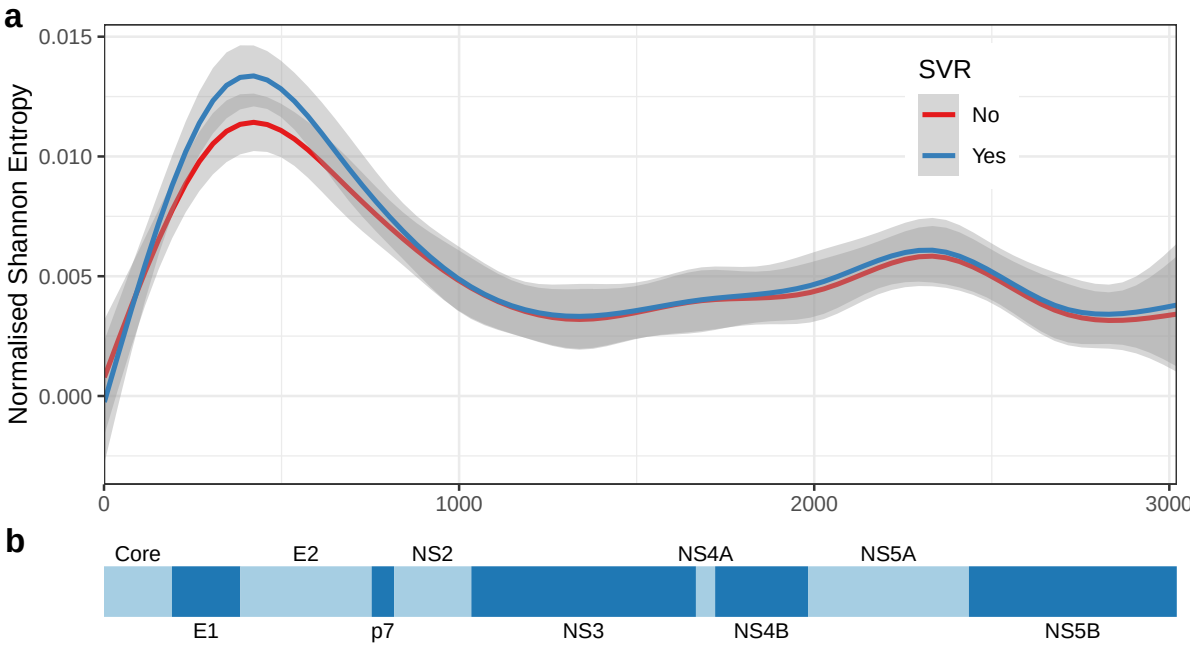

Figure 3.9: **Amino acid Shannon Entropy across HCV polyprotein for patients who achieved and did not achieve SVR.** Shannon entropy for each site in the HCV polyprotein was calculated for each patient and then averaged across for the whole cohort. A generalised additive model was used to create the smoothed line for each group.

3.5.9 Novel resistance associated substitutions effect on response to therapy

To understand the relevance of these putative RAS on patients response to therapy the impact of these novel RAS on the viral load reduction during therapy was investigated. The reduction in viral load was calculated by subtracting the log10 Wk1 VL from the Log10 BL Viral load and the difference in mean reduction between patients with and without RAS at each site and in combination was calculated. Differences in means were then tested using a Wilcoxan test (Figure 3.10A). At all of the sites, the mean reduction in viral load was smaller for RAS relative to non-RAS residues. The reduction in viral load was nominally significant for patients

whose virus carried RAS at sites NS2 119 (P=0.013), NS2 132 (P=0.018) or NS3 67 (P=0.029) relative to those whose virus did not carry the RAS. The NS3 67V also resulted in a smaller reduction of viral load, however the p value is not significant (P=0.057). NS5B 150V had no impact on reduction of log10 viral load from BL to Wk1 of therapy (P=0.317). The impact of the total number of RAS carried by each patient on the viral load reduction was also significant (linear regression, P=0.006, Figure 3.10B). The presence of valine at the two sites of NS2 132 and NS3 67 and non-valine at NS5B150 resulted in the smallest reduction in viral load during therapy (P=0.00091, Supplementary Figure 3.10C).

A V-GWAS was performed using a linear regression model with Log10 reduction in VL as the response variable and the presences or absence of every variable amino acid in the HCV genome was performed. The same co-variables identified in figure 3.1 were used. There were no significant associations using a 15% FDR (Table B.6). However, NS2 119A had the smallest P value and NS2 132V had the fifth smallest. This shows that the RAS associated with viral outcome also have a effect on the initial response to therapy.

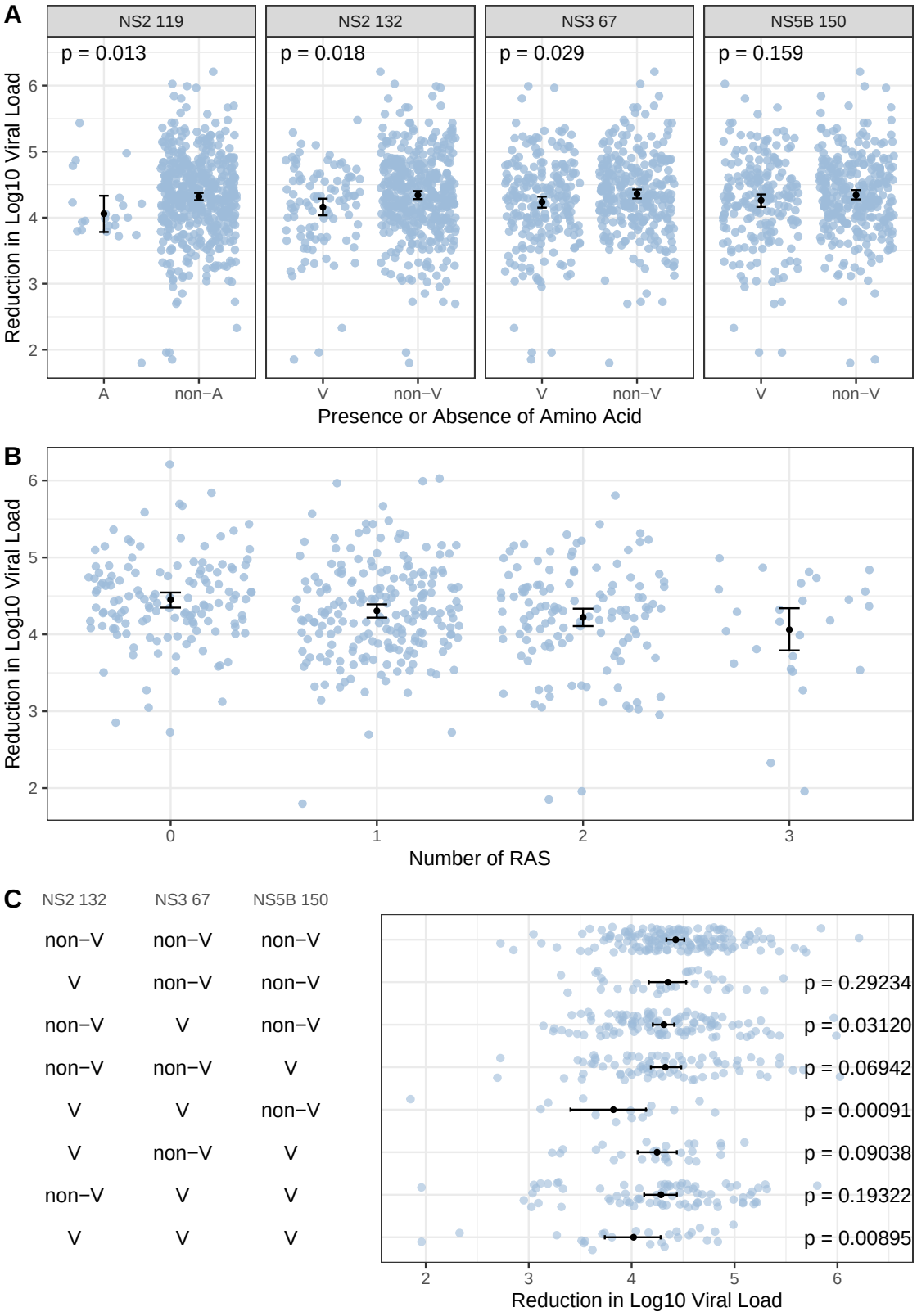

Figure 3.10: **Reduction in log10 viral loads between baseline and week 1 of therapy. A) Viruses carrying individual novel RAS. B) Viruses carrying increasing numbers of novel RAS. C) Viruses carrying specific combinations of novel RAS** The mean change in viral load is shown in black for each group of patients with confidence intervals calculated using bootstrapping. P values for difference in mean calculated using Wilcoxan tests.

3.5.10 Prevalence and enrichment of resistance associated substitutions in post treatment sequences

91 patients in this cohort trial did not achieve SVR, from these 73 whole genome consensus sequences from plasma samples taken at 12WPT were assembled. These consensus sequences where scanned for the presence of previously reported SOF RAS. NS5B 159F was detected in four sequences. This represents a significant change in the distribution of this RAS in the post treatment sequences relative to the baseline sequences (binomial test, $P=1.65x10^{-5}$). The NS5B RAS 282T, 316H/N 320F and 321A were not detected in any of the post treatment consensus sequences. On analysis of the deep sequencing data the 159F RAS was detected in two additional 12WPT samples as minor variant ($<$15% of reads). The 321A RAS was detected at very low frequencies ($<$1% of reads) in two samples.

The distribution of amino acids at the sites identified as putative RAS in this study were investigated in the 12WPT sequences. A binomial test was performed to identify changes in the distribution of RAS at these sites between the baseline population and the 12WPT sequences (Table 3.3). NS2 132V ($P=1.94x10^{-2}$), NS3 67V ($P=2.33x10^{-2}$) and NS5B 150V ($P=8.77x10^{-4}$), were significantly enriched in the 12WPT sequences. The majority of these putative RAS were present at baseline and were carried forward to the post-treatment isolates suggesting that the increase in prevalence is due to enrichment of pre-existing polymorphisms rather than substitutions during therapy. However, this was not the case for the previously reported NS5B 159F, where three of the four RAS had developed during therapy.

To identify any additional treatment emergent RAS, the same binomial test was performed for every site across the viral genome. Using a 15% FDR no amino acids were significantly enriched in the 12WPT samples (Table B.5). However, the L159F and the 150V RAS in the NS5B were the first and third most enriched amino acids across the whole genome.

Protein	RAS	Prevalence at BL % (n)	Prevalence at 12WPT % (n)	p value	Proportion of RAS at 12WPT not present at BL % (n)
NS5B	159F	0% (0/422)	4% (3/67)	5.69x10-4	100% (3/3)
NS5B	150V	42% (176/424)	61% (41/67)	9.08x10-4	7% (3/41)
NS2	132V	19% (81/427)	31% (21/67)	1.05x10-2	0% (0/21)
NS2	119A	4% (18/428)	9% (6/67)	6.31x10-2	16% (1/6)
NS2	119I	1% (4/428)	1% (1/67)	4.68x10-1	0% (0/1)
NS2	119M	4% (18/428)	4% (3/67)	5.41x10-1	6% (1/18)
NS3	67V	39% (170/428)	55% (37/68)	1.04x10-2	0% (0/37)
NS4B	233I	5% (21/428)	10% (7/67)	4.62x10-2	14% (1/7)

Table 3.3: **Prevalence of most enriched amino acids between baseline (BL) and 12 weeks post treatment (12WPT) at existing RAS sites or putative RAS sites.**

3.5.11 Patients who experienced an on treatment viral breakthrough

For the majority of patients who did not achieve SVR, the virus reached undetectable levels during treatment and rebounded post-treatment. However, three patients experienced a viral breakthrough i.e. viral load did not reach undetectable levels or rebounded to detectable levels whilst on treatment (Figure 3.11). Two patients withdrew from the trial after viral breakthrough, therefore post breakthrough samples were not available for these patients. Sequencing was attempted on all viraemic samples and reads were mapped to the baseline sequence from the respective patients. Whole genome HCV sequences were retrieved for patient sP126297 at week one of therapy and for patient sP114734 at Week 12 of therapy, 4, 12 and 24 weeks post treatment time points and no whole genome sequences were assembled for patient sP145389. No previously characterised RAS were detected for either patients at any time points. However, patient sP114734 did have NS2 132V, NS3 67V and NS5B 150V and patient sP126297 had NS2 119A at all successfully sequenced timepoints.

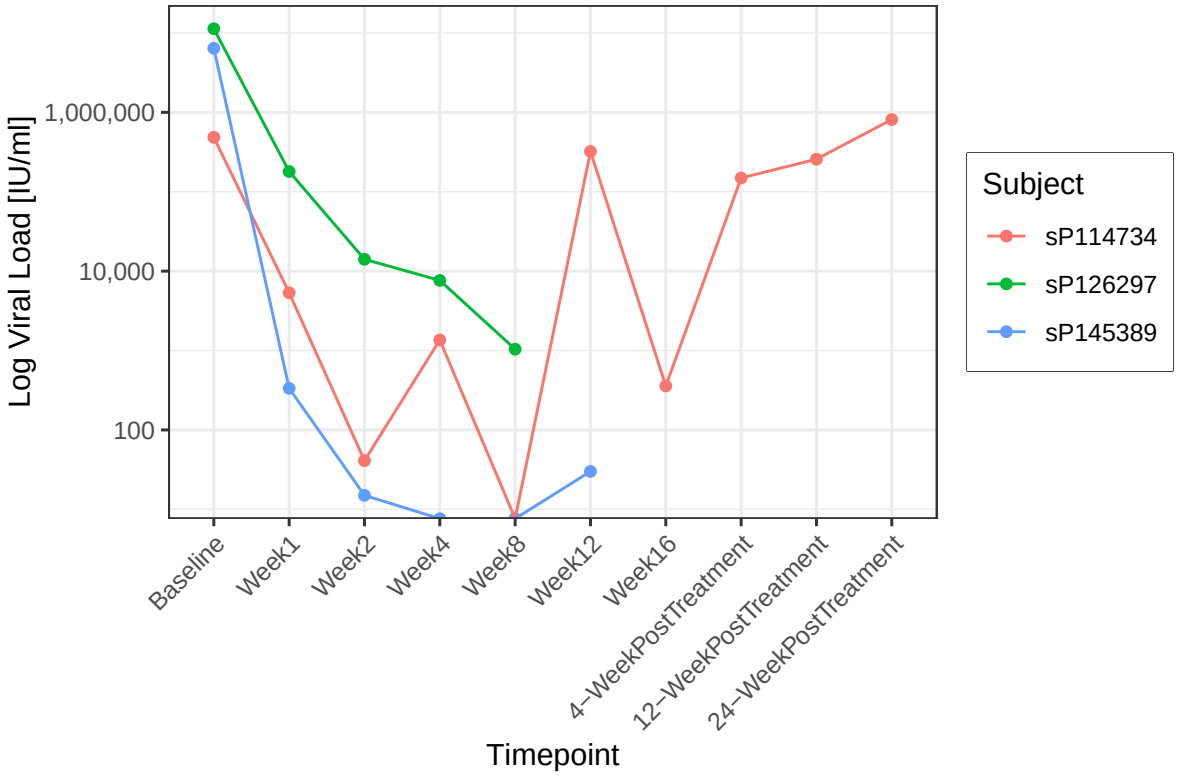

Figure 3.11: **Viral dynamics of patients who experienced an on treatment viral breakthrough.** 3 of the 91 patients who where failed by treatment experienced an on-treatment breakthrough. The Viral load for these patients at each timepoint is shown in IU/ml is shown on a logarithmic scale

3.6 Discussion

This is the first systematic, hypothesis free and comprehensive investigation of the viral determinants of treatment outcome in a large cohort of HCV gt3a infected patients treated with SOF DAA mono-therapy. NGS technologies were used to generate whole genome HCV consensus sequences as well as within-patient viral diversity data before and after treatment. The impact of previously reported SOF RAS on treatment outcome was investigated and genome wide association studies were used to test for association between all HCV amino acid sites and treatment outcome correcting for human and virus population structures by using PCs of the host and virus genome-wide data. V-GWAS was also used to test for associations between HCV amino acids and Treatment response and for the enrichment of HCV amino acids after treatment. This approach identified potential SOF RAS outside the NS5B protein for the first time and could be applied to other cohorts for other antiviral drugs.

Traditionally, resistance sites have been identified by comparing baseline sequences to post treatment failure sequences to identify changes in a specific protein which is believed to be the target of the drug [101]. These RAS are then tested using in-vitro assays to determine their impact on viral replication in cell culture. This method has identified SOF RAS such as NS5B 282T which are treatment emergent and tend to be transient due to their high fitness cost to the virus [165]. This methodology has two drawbacks. Firstly, it cannot detect wild type polymorphisms that potentially interact with the drug and contribute to the treatment response. Secondly, virus replication and production happens due to complex interactions between viral proteins and polymorphism in other viral protein could impact on the treatment outcome, which will be ignored if only the target protein is investigated. This genome-wide, hypothesis free approach has successfully identified common polymorphisms in NS2 and NS3 proteins which are present in wild type HCV and are associated with treatment outcome that would not normally be identified. It has also successfully identified treatment emergent RAS such as L159F.

Strong evidence that common polymorphisms in HCV NS2 and NS3 proteins modify SOF treatment response has been identified for the first time. Given the complex interactions among the HCV proteins as well as with host proteins, it should not be surprising to see polymorphisms in proteins other than NS5B (which is the target of SOF) to impact on treatment outcome. The identified polymorphisms in NS2 and NS3 proteins could impact SOF treatment through modifying their interactions with NS5B protein in presence of SOF. The function of the NS2 protein in the viral lifecycle is not as well understood as other proteins. It's exclusion from HCV replicons demonstrate it is not required for replication [134] but there is evidence to suggest it interacts with other viral proteins and has a role in virus assembly [166]. NS2 119, NS2 132 and NS3 67 are all found in the NS2-3 protease which is responsible for the cleavage of NS2-3, but they are not found near the active site [167]. At all of the four investigated sites, valine residue is either the resistant residue or the most common residue and all the residues found at these sites are small hydrophobic molecules. This combined with the fact they are all commonly found in gt3a virus suggest the substitutions do not have a large effect on the structure of these proteins. However, further investigation is required to deduce the mechanism of action of these sites and their impact on SOF treatment. All the patients received ribavirin as well as SOF and another explanation for association os sites in the NS2-3 protase could be ribavirin resistance. However, none of these sites have been previously identified

when ribavirin resistant viruses have been selected [168, 169], and ribavirin resistance isn't commonly observed in patients.

The impact of the novel RAS was studied in an independent cohort made up of primarily of patients who are treated with other DAAs such as NS5A inhibitors such as Daclatasvir or Ledipasvir as well as SOF. The association between treatment outcome and three of the four putative sites were not statistically significant in the replication cohort, however the direction and estimated effect sizes were consistent with the BOSON study. It is important to note that this is not an ideal replication cohort as the BOSON study was a SOF DAA mono-therapy where as in the replication cohort used other DAAs as part of the treatment regime. These other DAAs will interfere with other aspects of HCV life cycle and reduce the observed effect of any SOF RAS. Additionally, the reduced sample size (144 patients) will also reduce the power for detecting associations.

In studying the impact of the novel RAS on reduction of viral load from baseline to week one of therapy, it was identified that all four on average resulted in smaller reductions in viral load. As viral load is a surrogate for viral replication and production, its reduction during therapy is an independent measure that can inform whether these putative RAS have an impact on SOF treatment.

In conclusion, this data show that common HCV gt3a polymorphisms are associated with SOF treatment outcome, which until now has been thought of having high barrier to resistance. Assessing these polymorphisms may be useful in directing therapy duration and type for difficult to treat patients such as those with advanced cirrhosis or those who have failed DAA therapy previously. SOF is usually used in combination with NS5A inhibitors and it is known that RAS to NS5A inhibitors are common in baseline HCV sequences. The addition of a polymorphism that reduces the impact of SOF especially in difficult to treat groups maybe important in deciding the duration and the combination of drugs required to achieve a SVR.

3.7 Summary of key findings

- First comprehensive use of Next Generation Sequencing and Viral Genome Wide Association Studies to identify viral determinants of outcome in HCV infection.
- NS5B L159F only treatment emergent RAS identified.
- Identification of novel viral substitutions outside the NS5B protein associated with sofosbuvir treatment outcome and treatment response.
- Viral diversity was not associated with treatment outcome.

Chapter 4

Effect of resistance associated substitutions on the retreatment of Hepatitis C Virus infected patients with prior failure to direct acting antiviral therapy using sofsobuvir, velpatasvir and voxilaprevir

4.1 Abstract

Background and Aims: SVR rates to first-line DAA therapy routinely exceed 95%. However, a small number of patients require retreatment. Sofosbuvir/Velpatasvir/Voxilaprevir (SOF/VEL/VOX) is a potent DAA combination primarily used for the retreatment of patients failed by first line DAA therapies. Previous studies have shown that SOF/VEL/VOX has reduced effectiveness in gt3 infected individuals. This chapter evaluates virologic outcomes and the effects of RAS in the largest real-world cohort of gt3 infected patients, retreated with SOF/VEL/VOX to date.

Methods: Whole genome, next-generation sequencing of HCV using target enrichment was performed on 348 patients following virologic failure with DAA treatment regimens in the UK. Patients were retreated with 8-24 weeks of SOF/VEL/VOX, GLE/PIB or SOF/VEL + RBV. HCV subtypes were assigned and RAS relevant to each genotype were identified (15% read cut-off) using a bespoke pipeline based on HCV-GLUE.

Results: Of the 348 patients prior to retreatment, HCV subtypes 1a (39%, n=136) and 3a (39%, n=136) were the most common. RAS against NS5A inhibitors were the most frequently detected RAS (68%, n=147). Treatment outcomes are currently known for 162 patients. 144 of these patients were retreated with SOF/VEL/VOX and for these patients the overall re-treatment SVR12 rate was 90% (129/144). The SVR rate for non-gt3 infected patients was 96% (79/82) whereas the SVR for gt3 infected patients was significantly lower 81% (50/62) ($p = 0.002$). On univariate analysis cirrhosis ($p = 0.01$) was also significantly associated with retreatment failure. Patients infected with gt3, cirrhosis and who have been previously treated with a pangenotypic regimen have a retreatment SVR rate of 67%.

Conclusions: This work shows that retreatment outcomes for patients failed by first-line DAA therapy is good for non-gt3 patients. For gt3 infected patients particularly ones failed by pan-genotypic regimens SVR rates were significantly lower and alternative treatment regimens such as GLE/PIB + SOF should be considered for gt3 infected patients.

4.2 Background

SVR rates for patients treated with second or third generation generation DAA therapies in clinical trials and real world cohorts are often >95% [156, 170, 171]. Failure of treatment is associated with multiple factors including; advanced fibrosis or cirrhosis and the presence of RAS [62]. The subtype of HCV with which the patient is infected is also known to have an effect on treatment outcomes; gt4r, gt3b and gt1l all have reduced susceptibility to NS5A inhibitor containing regimens such as SOF/LED or SOF/VEL [69, 97]. This leaves a small subsection of HCV-infected patients who have been failed by first-line therapies and are therefore by definition ”difficult to treat”. Additionally retreatment of the patient population may be more difficult, as treatment failure commonly leads to the development of treatment emergent RAS which have been shown to persist for years after treatment failure [88]. For patients who have been failed by second generation DAA therapies such as SOF/LED, DCV + SOF or OMB/PTV/r+/-DAS, the option of third generation treatments such as SOF/VEL or GLE/PIB is available, with the additional options of adding in ribavirin and extending therapy. For those who have been failed by third generation therapies the options are limited, to SOF/VEL/VOX which is the currently the only recommended retreatment option for these patients [51, 130]. Retreatment options for patients with decompensated cirrhosis are further limited regardless of the first-line therapy, as these patients do not tolerate treatment with NS3 protease inhibitors [51], for these patients SOF/VEL + RBV is the only pangenotypic regimen available and in this setting treatment extension to 24 weeks should be considered.

Retreatment of patients previously failed by DAA therapy with SOF/VEL/VOX has been evaluated in both the POLARIS-1 and POLARIS-4 phase-II and III studies [87], with SVR rates in excess of 95%. In these trials 46% of the patients had cirrhosis and 49-83% had detectable RAS before retreatment. SOF/VEL/VOX showed a very high SVR rate in gt1 infected patients (97% SVR); however a higher virologic failure rate in gt3 patients was observed (95% SVR). These trials lead to the recommendation of SOF/VEL/VOX as a retreatment regime by both EASL and AASLD for patients with compensated cirrhosis. There have been some studies looking into the observed real-world outcomes of SOF/VEL/VOX retreatment, in both the US and Europe. The largest cohort (n=573) was from the US and reported by Belperio, et al, [172]. This showed SVR rates of 90.7% (429/473), 90.0% (18/20), 91.3% (42/46) and 100.0%

(12/12) for genotypes 1-4, respectively. Degasperi, et al, reported a SVR rate of 96% in a cohort of 179 Italian patients, with cirrhosis and HCC being associated with treatment failure [173]. However, Llaneras, et al, reported a Spanish cohort of 137 patients retreated with SOF/VEL/VOX. The overall SVR rate was 95%, although SVR12 was lower in patients with cirrhosis (89%, p = 0.05) and those with gt3 infection (80%, p $<$0.001) and patients with gt3 infection and cirrhosis had the lowest SVR12 rate (69%) [174].

Despite these studies there is a lack of data from the real world setting on the prevalence of RAS amongst patients not achieving SVR with first line DAA treatment, in particular following unsuccessful therapy with a third generation DAA containing treatment. Additionally the potential impact of RAS on SOF/VEL/VOX outcomes, has not been explored. These data are urgently required to inform optimal retreatment strategies.

4.3 Aims

The aim of this work is to investigate the prevalence of RAS in 348 patients who have not been cured by first line DAA therapy in the UK. Then to investigate the effect of RAS and other clinical features on the retreatment of 162 patients with GLE/PIB, SOF/VEL + RBV and SOF/VELVOX.

4.4 Methods

4.4.1 Subjects and samples

Blood samples and clinical data were collected by HCV Research UK on 111 patients who were treated with first line therapies unsuccessfully and about to undergo retreatment for HCV infection. 237 samples were submitted to Public Health England (PHE) for resistance testing, prior to retreatment, this data was collected as part of PHE's HCV antiviral resistance testing service.

Inclusion criteria

- Patient previously failed interferon free DAA treatment and retreated with GLE/PIB, VEL/SOF + RBV or SOF/VEL/VOX.
- Plasma sample available from before retreatment.
- Clinical dataset available for patient.

Exclusion criteria

- Patients included in both HCVRUK and PHE Cohort and this duplication is unresolvable.

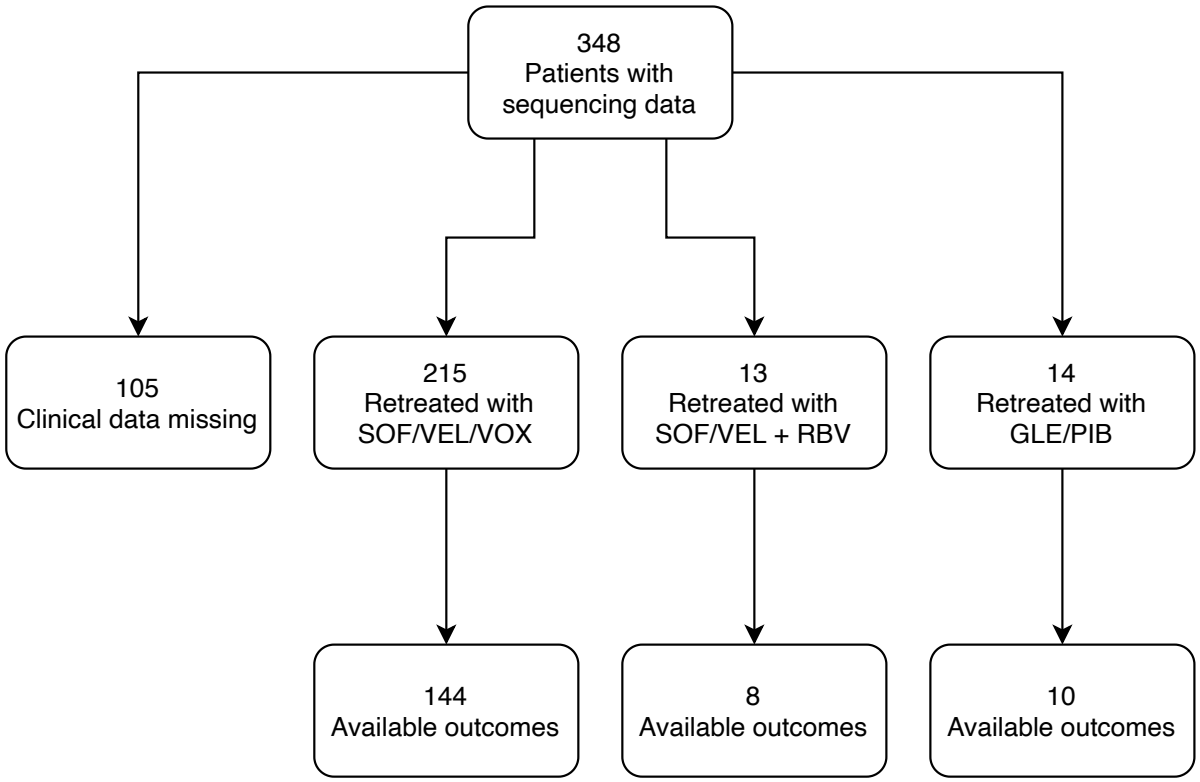

Figure 4.1: **Data availability for each retreatment regimen used in a cohort of 348 patients failed by first-line DAA therapy.**

4.4.2 HCV sequencing

The viral sequencing of these samples was completed at the Peter Medawar Building and Wellcome Trust Centre for Human Genetics in Oxford, the Centre for Virus Research in Glasgow and P in Colindale, London. Details of how the HCV sequencing was completed in Oxford can be found in chapter 2 section 2.4.2.

Viral sequencing at Centre for Virus Research in Glasgow

RNA was isolated from 200 μL of plasma using the RNAdvance Blood extraction kit (Beckman Coulter, Brea, CA, United States) and collected in 27 μL of water. Following conversion of RNA to double-stranded DNA, libraries were prepared for Illumina sequencing using the KAPA DNA LTP Library Preparation Kit (Roche, Basel, Switzerland), and NEBNext Multiplex Oligos for Illumina (New England Biolabs, Ipswich, MA, United States). Libraries were quantified using Qubit dsDNA HS Assay Kit (Invitrogen, Carlsbad, CA, United States) and size distribution assessed using Agilent TapeStation with D1K High Sensitivity Kit (Agilent, Santa Clara, CA, United States); libraries were normalised according to viral load and mass. A 500 ng aliquot of the pooled library was enriched using SeqCap EZ Developer Probes (Roche), following the manufacturer's protocol. Following a 14 cycle post-enrichment PCR, the cleaned pool was sequenced with 151-base paired-end reads on a NextSeq cartridge (Illumina, San Diego, CA, United States). Reads were trimmed and filtered using TrimGalore [175] with quality threshold 30 and minimum read length 75. The most appropriate HCV reference sequence was identified via a k-mer-based approach, using k-mers unique to each genotype. BAM files were generated by mapping against the best-matching HCV reference using Tanoti [176].

Viral sequencing at Public Health England in Colindale

The PHE library preparation protocol is the laboratory component of a pipeline aimed at clinical use; a manuscript describing the full pipeline is in preparation. RNA was extracted from 350 μL of plasma using the NUCLISENS easyMAG system (bioMerieux). Total eluates were subjected to Turbo DNAse treatment (Thermo Fisher, Waltham, MA, United States) followed by

library preparation using KAPA RNA HyperPrep kit (Roche). Libraries were pooled based upon DNA concentration and HCV quantity, assessed using the Quant-iT kit on the Glomax platform (Promega, Madison, WI, United States) and the Qiagen QuantiTect kit with primers and probes from Davalieva et al. [177]. Pools were enriched by hybridisation to a biotinylated probe set (Integrated DNA Technologies, described by Bonsall et al. [35]) followed by further PCR cycles depending upon HCV quantity. The two pools were pooled, again by concentration and HCV quantity. The final pool was quantified using the KAPA SYBR FASTA qPCR kit (Roche) on a PRISM 7500 (Applied BioSystems, Foster City, CA, United States) before being sequenced on a MiSeq using Reagent kit v2 (Illumina). Human reads were filtered out from trimmed FASTQ files using SMALT [178], remaining reads were then assembled using VICUNA de novo assembly [144]. Contigs were matched to HCV reference genomes using BLAST [179] and gaps filled using LASTZ [180] to generate a draft assembly. Reads were then mapped to the draft assembly with BWA [142].

4.4.3 HCV sub-typing and resistance associated substitutions calling with HCV-GLUE

HCV subtypes were assigned and NS3, NS5A and NS5B RAS relevant to each genotype were identified (15% of reads cut off) using HCV-GLUE and RAS definitions provided by PHE (Tables 1.5, 1.6 and 1.7). HCV GLUE is a resource created as part of the STOP-HCV consortium based on the GLUE (Genes Linked by Underlying Evolution) platform [181]. HCV-GLUE uses the ICCT HCV reference sequences and a database of publicly available HCV sequences, to construct phylogenies using RAxML to assign sequences to HCV clades, it also contains a database of HCV RAS created by PHE and this is used to identify RAS relevant to the HCV genotype and subtype. The web version of HCV-GLUE only accepts consensus HCV sequences as an input, however the offline version is capable of accepting BAM files and using the data to identify low frequency viral variants [176].

4.5 Results

4.5.1 HCV genotypes and prevalence of resistance associated substitutions in patients prior to retreatment

Whole genome HCV sequences were obtained from 348 patients prior to retreatment. HCV subtypes 1a (39%, n=136) and 3a (39%, n=136) were the most common; the remaining sequences included 26 subtypes within gt1, gt2, gt3, gt4 and gt6 (Figure 4.2). Five sequences gt1, one gt2, one gt3 and one gt4 could not be assigned to subtypes, and potentially represent novel subtypes.

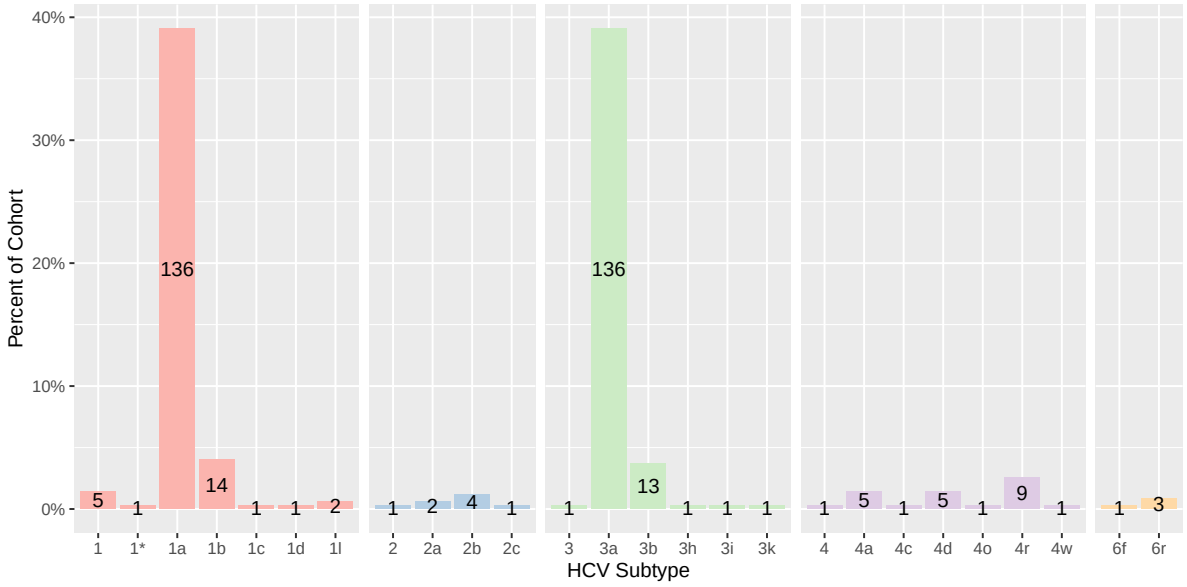

Figure 4.2: **HCV Subtypes of the 348 patients failed by first-line DAA therapy.** The percentage prevalence of each subtype within the cohort (n=348) of patients failed by first-line HCV DAA therapy is shown broken down by the HCV genotype. Sequences which could not be assigned to subtypes have been labelled with the respective genotype. *This sequence clusters with the unassigned reference sequence 1KJ439779.

RAS relevant to each sequence's respective genotype were called using HCV-GLUE. 88% (n = 315) of patients had a detectable (15% of reads cut off) RAS to a DAA. 59% (n = 206), 70% (n = 242) and 38% (n = 132) of patients had NS3, NS5A and NS5B RAS respectively. The prevalence of individual RAS is shown in Figure 4.3. A total of 822 RAS were detected in the 348 samples. Low frequency RAS were rare, 84% of RAS are found in >95% of sequencing

reads from the respective samples. Across the whole cohort the most prevalent RAS was the NS5A RAS Y93H being detected in 23% (n=81) of patients, interestingly it was most prevalent in the gt3a 41% and gt1b 57% viruses. However, in gt1a infected patients Q30R was the most prevalent. The HCV subtypes gt3b and gt4r are known to commonly carry the RAS A30K+L31M and M28V+R30R+L31M respectively, and this is reflected in the high RAS prevalence numbers in these subtypes. gt3b and gt4r are also the only two subtypes where the NS5B S282T RAS is detected.

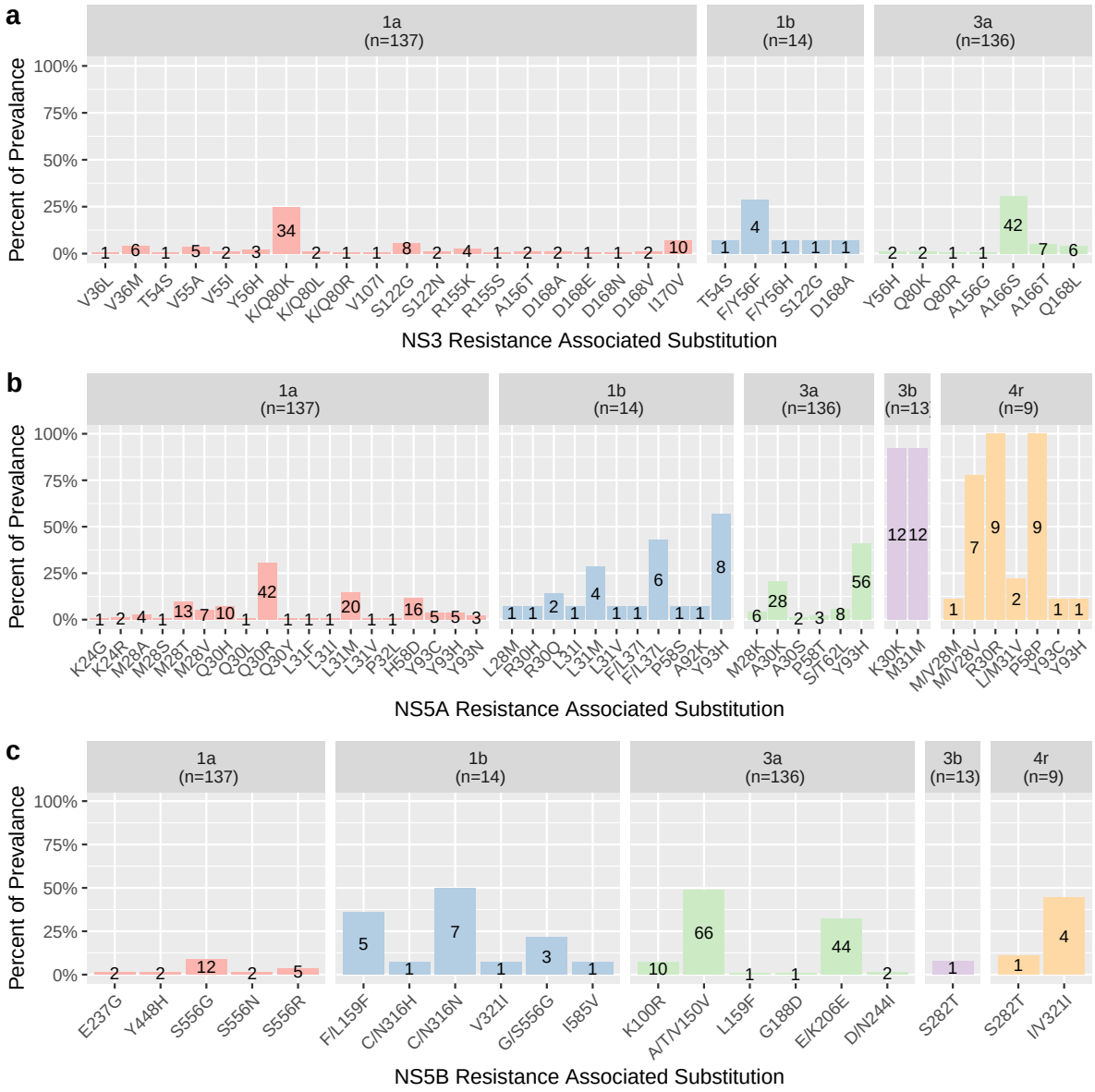

Figure 4.3: **The prevalence of individual resistance associated substitutions in a cohort (n=348) of patients failed by first-line HCV DAA therapy.** The percentage prevalence of RAS (15% of reads cut off) in each subtype with more than 5 samples available is shown.

4.5.2 Clinical characteristics of patients prior to retreatment

Retreatment outcomes were available for 162 patients. The majority (n=144) of patients where retreated with SOF/VEL/VOX, 10 patients where retreated with GLE/PIB and 8 where retreated with SOF/VEL + RBV. A breakdown of their clinical characteristics by retreatment regimen can be found in table 4.1. 45% of patients had liver cirrhosis with 5% having decompensated cirrhosis. 13% had a prior liver transplant and 10% had HCC. A higher proportion of patients retreated with SOF/VEL + RBV had decompensated cirrhosis (62% vs 5%); this is to be expected as protease inhibitors are not recommend for patients with decompensated cirrhosis.

A further breakdown of the clinical characteristics of the patients treated with SOF/VEL/VOX by genotype, revealed significant differences between the pre retreatment characteristics of different patients (Table 4.2). The main differences were the previous treatment regimens and the presence of RAS, this is to be expected though as treatments are guided by the genotype of HCV the patient is infected with and RAS are often genotype specific.

	GLE/PIB	SOF/VEL + RBV	SOF/VEL/VOX
Demograhics			
Age	58 (23-67)	56 (40-69)	56 (26-81)
Male	9/10 (90)	6/8 (75)	121/144 (84)
Female	1/10 (10)	2/8 (25)	22/144 (15)
Disease State			
Cirrhosis	8/10 (80)	8/8 (100)	58/144 (40)
Decompensated cirrhosis	0/8 (0)	5/8 (62)	4/58 (7)
Prior HCC	1/10 (10)	2/8 (25)	14/144 (10)
Prior Liver Transplant	0/10 (0)	1/8 (12)	12/144 (8)
Pretreatment Viral Load	$1.72x10^6$ ($2.95x10^5$-$2.6x10^6$)	$2.07x10^7$ ($9.48x10^2$-$1.6x10^8$)	$4.55x10^5$ ($2.21x10^3$-$8.5x10^7$)
Previous Treatment			
SOF/LDV	4/10 (40)	6/8 (75)	35/144 (24)
SOF/VEL	-	-	17/144 (12)
GLE/PIB	-	-	19/144 (13)
DCV + SOF	-	-	19/144 (13)
EBR/GZR	-	-	12/144 (8)
OBT/PTVr +/- DAS	1/10 (10)	-	25/144 (17)
SOF	3/10 (30)	2/8 (25)	8/144 (6)
Unknown	-	-	9/144 (6)
Presence of RAS			
None	4/10 (40)	2/8 (25)	43/144 (30)
NS3	1/10 (10)	-	10/144 (7)
NS5A	1/10 (10)	4/8 (50)	30/144 (21)
NS5B	-	2/8 (25)	17/144 (12)
NS3 + NS5A	3/10 (30)	-	12/144 (8)
NS5A + NS5B	1/10 (10)	-	31/144 (22)
NS3 + NS5A + NS5B	-	-	1/144 (1)
Genotype			
gt 1	6/10 (60)	3/8 (38)	66/144 (46)
gt 2	-	2/8 (25)	3/144 (2)
gt 3	3/10 (30)	3/8 (38)	62/144 (43)
gt 4	1/10 (10)	-	10/144 (7)
gt 6	-	-	3/144 (2)

Table 4.1: **Clinical characteristics of 162 patients with clinical outcomes prior to re-treatment with GLE/PIB, SOF/VEL + RBV or SOF/VEL/VOX.** *Subtype unknown, Daclatasvir (DCV), Sofosbuvir (SOF), Glecaprevir (GLE), Pibrentasvir (PIB), Grazoprevir (GZR), Elbasvir(EBR), Ledipasvir(LDV), Paritaprevir (PTV), Ombitasvir (OBV) Ritonavir (r), Dasabuvir (DAS), Velpatasvir (VEL).

	gt1	gt2	gt3	gt4	gt6
Demographics					
Age	54 (26-80)	64 (49-71)	57 (39-80)	62 (44-81)	27 (27-62)
Male	86% (56/65)	67% (2/3)	85% (53/62)	70% (7/10)	100% (3/3)
Female	14% (9/65)	33% (1/3)	15% (9/62)	30% (3/10)	-
Treatment Length					
$<$8 Weeks	2% (1/66)	-	2% (1/62)	-	-
8 Weeks	15% (10/66)	-	13% (8/62)	10% (1/10)	-
12 Weeks	76% (50/66)	67% (2/3)	77% (48/62)	90% (9/10)	67% (2/3)
16 Weeks	2% (1/66)	-	3% (2/62)	-	-
24 Weeks	6% (4/66)	33% (1/3)	5% (3/62)	-	33% (1/3)
Disease State					
Cirrhosis	24% (16/66)	67% (2/3)	60% (37/62)	30% (3/10)	0% (0/3)
Decomp	0% (0/16)	0% (0/2)	11% (4/37)	0% (0/3)	0% (0/0)
HCC	9% (6/66)	0% (0/3)	13% (8/62)	0% (0/10)	0% (0/3)
Liver Transplant	3% (2/66)	33% (1/3)	15% (9/62)	0% (0/10)	0% (0/3)
Viral Load	5.0e+06 (3.6e+03-2.7e+07)	1.3e+06 (4.2e+05-2.1e+06)	3.1e+06 (2.2e+03-2.9e+07)	1.4e+07 (2.1e+04-8.5e+07)	1.4e+06 (2.3e+04-2.1e+06)
Previous Treatment					
SOF/LDV	38% (25/66)	-	10% (6/62)	40% (4/10)	-
SOF/VEL	2% (1/66)	-	26% (16/62)	-	-
GLE/PIB	2% (1/66)	-	24% (15/62)	-	100% (3/3)
DCV + SOF	-	-	31% (19/62)	-	-
EBR/GZR	15% (10/66)	-	-	20% (2/10)	-
OBT/PTVr +/- DAS	32% (21/66)	-	2% (1/62)	30% (3/10)	-
SOF	2% (1/66)	100% (3/3)	6% (4/62)	-	-
Unkown	11% (7/66)	-	2% (1/62)	10% (1/10)	-
Presence of RAS					
NS3	15% (10/66)	-	-	-	-
No RAS	29% (19/66)	33% (1/3)	23% (14/62)	60% (6/10)	100% (3/3)
NS3 + NS5A	17% (11/66)	-	2% (1/62)	-	-
NS5A + NS5B	9% (6/66)	-	40% (25/62)	-	-
NS5A	29% (19/66)	67% (2/3)	11% (7/62)	20% (2/10)	-
NS5B	2% (1/66)	-	23% (14/62)	20% (2/10)	-
NS3 + NS5A + NS5B	-	-	2% (1/62)	-	-

Table 4.2: **Clinical characteristics of 144 patients prior to retreatment with SOF/VEL/VOX split by genotype.** Daclatasvir (DCV), Sofosbuvir (SOF), Glecaprevir (GLE), Pibrentasvir (PIB), Grazoprevir (GZR), Elbasvir(EBR), Ledipasvir(LDV), Paritaprevir (PTV), Ombitasvir (OBV) Ritonavir (r), Dasabuvir (DAS), Velpatasvir (VEL).

4.5.3 Retreatment outcomes

The SVR rates for patients treated with GLE/PIB and SOF/VEL + RBV were 90% and 88% respectively with only one failure in each group. Full details of each patient can be found in Supplementary Table C.1. One patient was infected with gt2b and one gt1a, both patients had cirrhosis and both experienced a post treatment relapse.

The overall SVR rate for patients retreated with SOF/VEL/VOX was 91%. For gt3 infected patients the SVR Rate was only 81% compared to 97% for non-gt3 infected patients; this represents a statistically significant difference (P=0.002). The majority of patients who did not achieve an SVR experienced a post treatment relapse (n=10). Interestingly, one patient experienced an on-treatment breakthrough, two more were non-responders and this information was not recorded for the two other patients.

A breakdown of the clinical factors associated with outcome can be found in Figure 4.4a. The presence of Cirrhosis (p=0.01) and gt3 (p=0.009) infection were both significantly associated with outcome in the whole cohort. When a multivariable analysis is performed gt3 infection remains significantly associated with outcome (p=0.03), Cirrhosis is only nominally associated (p=0.08). Both associations retain their negative direction. Of the 15 patients who did not achieve a SVR, 12 were infected with gt3. When only gt3 infected patients are analysed treatment for 24 weeks was the only variable significantly associated with outcome (p=0.04) but only 3 patients received treatment for this long (Figure 4.4b). However, cirrhosis still had a negative effect on outcome, 75% SVR rate for patients with cirrhosis vs 88% without. When limited to patients who previously received an approved treatment for gt3 infection (SOF/VEL, GLE/PIB and SOF +DCV, n=50) this number is even lower 67% vs 86% however this is not significant(p=0.13). Interestingly, when patients are grouped by their previous regimen SOF/VEL, GLE/PIB or SOF + DCV cirrhosis seems to have the largest effect on patients previously treated with SOF/VEL with only 42% of patients with cirrhosis achieving an SVR vs 100% for patients without cirrhosis (Figure 4.5).

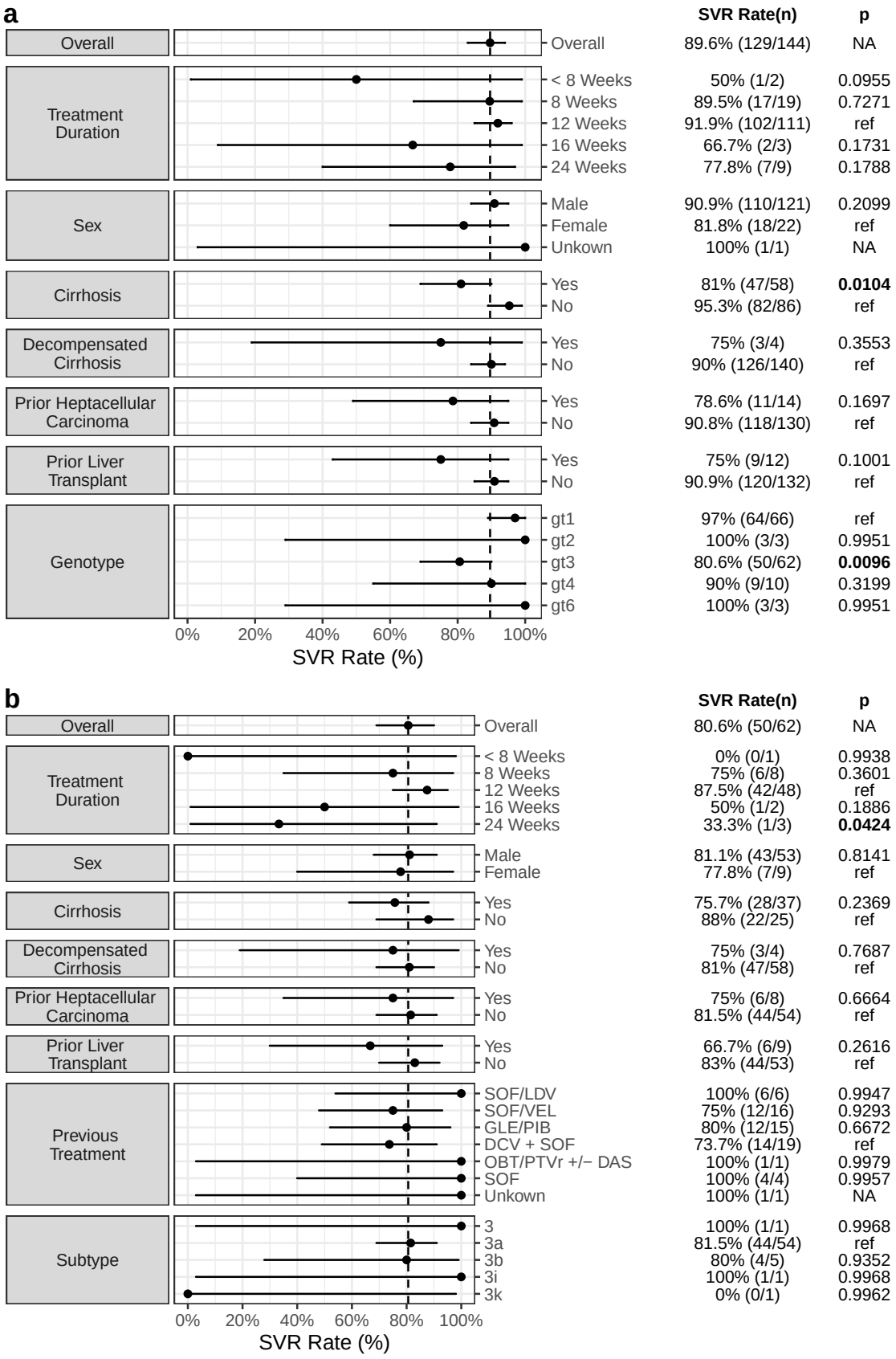

Figure 4.4: **Univariate analysis of clinical factors associated with SVR rates for patients retreated with SOF/VEL/VOX according to baseline and treatment characteristics. a) Whole cohort (n=144). b) Genotype 3 infected patients (n=62).** The SVR rate for each variable is shown with the black dot with 95% confidence intervals shown, the SVR rate of the whole cohort is shown with a dashed line. P values calculated using logistic regression, with reference value shown for each variable.

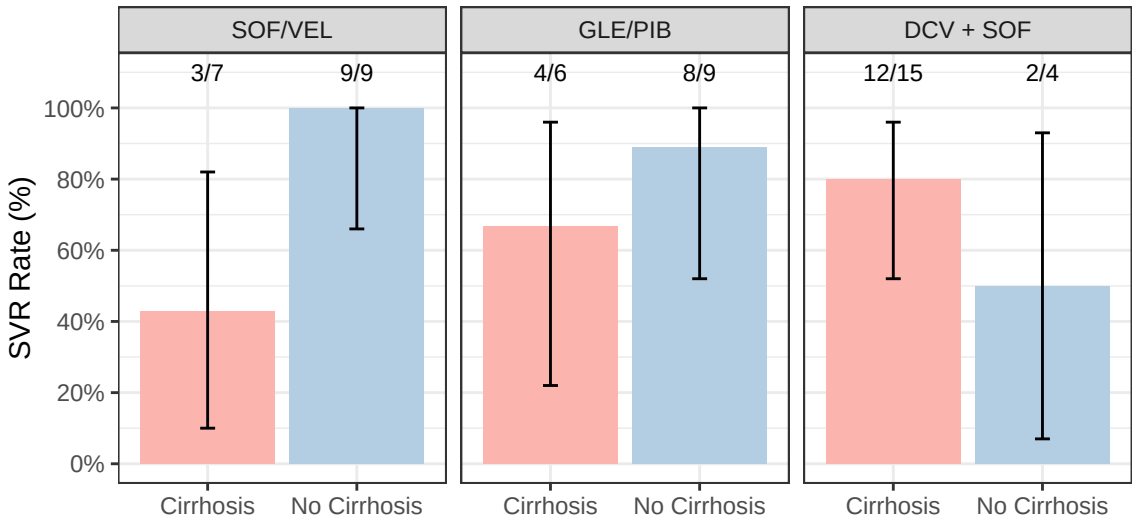

Figure 4.5: **SVR rates of gt3 infected patients for patients previously treated with different treatment regimes with and without cirrhosis.** SVR rate for each group is shown with 95% confidence intervals shown, numbers in each group are shown on the top. Sofosbuvir (SOF), Glecaprevir (GLE), Pibrentasvir (PIB), Velpatasvir (VEL).

4.5.4 Effect of resistance associated substitutions on SOF/VEL/VOX retreatment

Among the 144 patients retreated with SOF/VEL/VOX, 70% (101/144) had detectable RAS prior to retreatment. 16% had NS3 inhibitor RAS, 51% had NS5A inhibitor RAS and 34% had SOF RAS. Of the 15 patients failed by SOF/VEL/VOX retreatment 12/15 had detectable RAS prior to retreatment commencing (Supplementary Table C.1). Despite the high prevalence of RAS, the presence of a RAS in a particular protein or combination of proteins was not significantly associated with outcome. When each RAS is looked at independently using logistic regression the NS5A RAS A30K and Y93H were associated with retreatment failure in the whole cohort (A30Kp = 0.02, Y93Hp = 0.01) (Table C.2). However, the A30K RAS is unique to gt3 and the Y93H substitution was much more common in the gt3 patients. When the analysis is limited to just the gt3 patients, the SVR rates for patients with these RAS were still reduced A30K 71% and Y93H 78% but no longer significant (A30Kp = 0.2, Y93Hp =0.3). Whilst individually the A30K and Y93H RAS is common, the combination of A30K + Y93H is found rarely and is not associated with SVR in this cohort. No other RAS was significantly associated with outcome in the whole cohort or within a specific genotype of patients.

4.6 Discussion

This cohort represents both the largest cohort of patients retreated with SOF/VEL/VOX in the UK and the largest cohort of gt3 infected patients retreated with SOF/VEL/VOX reported. This work shows that retreatment of patients failed by first-line therapy with SOF/VEL/VOX is very effective. However, patients with gt3 infection and cirrhosis do not respond as well to SOF/VEL/VOX retreatment.

At present SOF/VEL/VOX is the only recommended retreatment regimen for patients previously failed by a NS5A inhibitor containing regimen. This data suggests that whilst this recommendation is correct for non-gt3 infected patients, there is a need for an alternative retreatment regimen for gt3 infected patients with cirrhosis and previous exposure to a third generation DAA regimen. One current alternative to SOF/VEL/VOX for these patients is the off label use of GLE/PIB + SOF, this combination has been shown to be effective in limited trials [89] and could represent an effective retreatment option for gt3 patients with prior exposure to SOF/VEL. Other options, include increasing the SOF/VEL/VOX treatment duration, adding ribavirin and a controversial option could be the use of INF+RBV+SOF/VEL/VOX for these patients as gt3 infected patients are known to have good response rates to INF.

The strength of this work is that it has a good representation of the real world patient population that requires retreatment compared to that of clinical trials. However, this also leads to a clinically heterogeneous cohort with a varied treatment history. For example, patients with gt3 infection did not respond as well in this cohort. However, this observation is confounded by the higher prevalence of cirrhosis, NS5A RAS and an increased likelihood of previous treatment with a third-generation DAA regimen for patients infected with gt3. Multivariable analysis using logistic regression, suggests that prior treatment with SOF/VEL or SOF + DCV is the main explanatory variable, however most of the patients failed by these regimens were infected with gt3 HCV. Our data showed that gt3 infected patients did not respond well to SOF/VEL/VOX retreatment, and that if they have cirrhosis and had been previously treated with SOF/VEL they did dismally with an SVR rate of only 42%. This is in line with the data from Llaneras et al [174], which also showed a SVR rate of 81% for gt3 infected patients and 69% for gt3 infected patients with cirrhosis. Other real world trials Belperio, et al, [172]

and Degasperi, et al [173], showed SVR rates for gt3 infected patients comparable to that of non-gt3. However, this cohort has a higher proportion of gt3 infected patients previously treated with SOF/VEL, GLE/PIB or SOF + DCV, which could explain the discrepancies. These findings are of particular interest in areas where gt3 infection is common, such as the UK and South Asia [41].

This work shows that whilst retreatment outcomes for patients failed by first-line DAA therapy is good for non-gt3 patients. For gt3 infected patients particularly ones failed by pangenotypic regimens SVR rates were significantly lower and alternative treatment regimens such as GLE/PIB + SOF should be considered for gt3 infected patients.

4.7 Summary of key findings

- Patients who have been failed by first line therapies, have a very high prevalence of RAS.
- The presence of the NS5A RAS A30K or Y93H both result in reduced SVR rates.
- Despite the high prevalence of RAS, retreatment SVR rates are high for non-gt3 with or without cirrhosis.
- Patients infected with gt3, cirrhosis and who have been previously treated with a pangenotypic regimen have a retreatment SVR rate of 67%.
- gt3 infected patients previously treated with SOF/VEL and with cirrhosis had an SVR rate of 43% compared to 100% for those who didn't have cirrhosis.

Chapter 5

Concluding Remarks

5.1 Summary of main findings

Chapters 2 and 3 of this work analysed whole-genome HCV sequences generated from the BOSON clinical trial [137]. This represents one of the largest cohorts of gt3 whole genome sequences collected to date. Chapter 2 focuses on the use of these sequences to identify the prevalence of HCV RAS to approved HCV treatments in a DAA treatment naive population. This work revealed a high prevalence of pre-existing wild type RAS to NS5A inhibitors in the gt3a sequences and identified a pattern of the A30K+L31M RAS as being the wild-type in the gt3b and gt3g sequences. When this pattern of RAS was tested in a gt3 in-vitro assay, the A30K+l31M RAS combination was shown to infer a dramatic increase in the EC50 of all NS5A inhibitors approved and recommended for the treatment of gt3 HCV infection. This led to the hypothesis that gt3b and gt3g infected patients would not respond as well to DAA therapy. A recent clinical trial of SOF/VEL in China [156], supports this hypothesis. The study reports SVR rate of only 50% in patients with cirrhosis and gt3b infection vs 97% in patients with gt3a infection. This data is of particular concern for parts of the world where gt3b is endemic such as Thailand and Laos [155].

Next, a systematic analysis of the pre treatment sequences from gt3a infected patients was used to identify viral determinates of treatment outcome with SOF+RBV+/-INF. This analysis included the first use of a viral genome wide association studies to identify novel pre-existing RAS to sofosbuvir. Two of these novel RAS were identified in the NS2 protein 119A and 132V, one in the NS3 67V and one in the NS5B 150V. These RAS were then shown to be significantly enriched in the sequences assembled from post-treatment samples from patients who did not achieve a SVR. Three of these RAS where also shown to significantly reduce the initial response to treatment of these patients. The presence of three of these novel RAS reduced the SVR rate dramatically from 94% to 51%. The NS5B 150V substitution identified using the V-GWAS had previously been identified as a sofosbuvir RAS using in-vitro assays [164]. Position 150 in the NS5B protein has also been linked to the host INFL4 genotype [163].

Finally, whole-genome HCV sequences were generated from a cohort of 348 patients who have been failed by first-line DAA therapy in the UK, prior to retreatment using GLE/PIB, SOF/VEL + RBV or SOF/VEL/VOX. A very high prevalence of RAS was identified in this cohort

of patients. Retreatment outcomes were available for 162 of these patients. In the largest real-world cohort of gt3 infected patients retreated with SOF/VEL/VOX, gt3 infection was found to reduce the SVR rate of SOF/VEL/VOX treatment dramatically from 97% in non-gt3 infected patients to 81% in gt3 infected patients. This work along with other data from real-world cohorts in Europe [174] shows that gt3 infected patients with cirrhosis previously treated with a pan genotypic regimen, do not respond well to retreatment and additional retreatment regimens such as GLE/PIB + SOF need to be made available for the retreatment of these patients.

5.2 The identification of novel pre-existing DAA resistance

The work from chapters 2 and 3, identifies pre-existing RAS which have an effect on treatment outcomes in samples from DAA treatment naive patients. This pre-existing resistance falls into two categories; the first is a HCV subtype where the RAS is the part of the WT variant. The second is where the RAS is present within a subtype which does not normally have the RAS in the WT variant. Both of these categories of RAS have been shown to effect the treatment outcomes and their presence in treatment naive samples at high frequencies implies that they have a high viral fitness in-vivo. Traditionally, RAS have been identified by looking for treatment emergent substitutions and testing them in-vitro. The work from chapter 3, demonstrates the utility of whole genome sequencing and systematic bioinformatics led approach to identifying RAS which can both identify treatment emergent RAS and preexisting RAS. The primary limitation of this approach is that it requires both large sample sizes and a large amount of additional data. However, given the appropriate datasets, this approach also allows the identification of co-variation of substitutions and interaction. This could prove useful in the identification of compensatory mutations required to improve viral fitness after a treatment emergent RAS is acquired.

5.3 HCV Resistance in the context of HCV elimination

The WHO have set the ambitious goal of eliminating viral hepatitis as a public health problem by 2030 [1]. It is clear that the main barrier to achieving this goal for HCV is the identification

of new cases and increasing access to treatment [1]. However, as access to treatment is scaled up, there will be an increase in the absolute number of patients who have been failed by first-line DAA therapy. The work from chapter 4 of this thesis and other studies have shown that these patients virus will likely harbour multiple RAS, of which some are highly fit variants and have been shown to persist for years [88, 182]. Individual case studies have shown that HCV RAVs can be transmitted to new hosts [183–185]. More work is needed to confirm that HCV RAVs aren't be transmitted from host to host, but if this is the case it is possible that the SVR rates we see today may decrease. The second danger which DAA resistance may pose to the HCV elimination goal is that of understudied and undiscovered subtypes which are circulating in areas such as central Africa and Asia with less developed health care systems [157, 158]. The data from the chapter 2 of this thesis identified gt3b, gt3g as being inherently resistance to all NS5A inhibitors. Other work has shown the gt4r and gt1l have reduced SVR rates when compared to other subtypes within the respective genotypes. With more whole genome sequencing of HCV taking place and the high genetic diversity of HCV, it is likely that new HCV subtypes and potentially new genotypes will be identified that are inherently resistant to DAAs. Further understanding of HCV resistance as well as identifying circulating resistant subtypes and monitoring the transmission of resistance variants will all be required if the WHO goals are to be met.

Bibliography

1. World Health Organization. *Global hepatitis report, 2017* 62. ISBN: 9789241565455. doi:ISBN978-92-4-156545-5. http://www.who.int/hepatitis/publications/global-hepatitis-report2017/en/ (World Health Organization, 2017).
2. Kim, W. R. *et al.* OPTN/SRTR 2013 Annual Data Report: Liver. *American Journal of Transplantation* **15,** 1–28. ISSN: 16006135 (Jan. 2015).
3. Kim, W. R. *et al.* OPTN/SRTR 2017 Annual Data Report: Liver. *American journal of transplantation : official journal of the American Society of Transplantation and the American Society of Transplant Surgeons* **19 Suppl 2,** 184–283. ISSN: 1600-6143 (Feb. 2019).
4. Alter, H. J. *et al.* Detection of Antibody to Hepatitis C Virus in Prospectively Followed Transfusion Recipients with Acute and Chronic Non-A, Non-B Hepatitis. *New England Journal of Medicine* **321,** 1494–1500. ISSN: 0028-4793 (Nov. 1989).
5. Westbrook, R. H. & Dusheiko, G. Natural History of Hepatitis C. *Gastroenterology Clinics of North America* **61,** S58–S68. ISSN: 01688278 (2014).
6. *Who to screen/test — Diagnosis — Hepatitis C — CKS — NICE* https://cks.nice.org.uk/topics/hepatitis-c/diagnosis/who-to-screen-test/ (2020).
7. Grebely, J. *et al.* The effects of female sex, viral genotype, and IL28B genotype on spontaneous clearance of acute hepatitis C virus infection. *Hepatology.* ISSN: 15273350. doi:10.1002/hep.26639 (2014).
8. Thomas, D. L. *et al.* Genetic variation in IL28B and spontaneous clearance of hepatitis C virus. *Nature* **461,** 798–801. ISSN: 1476-4687 (Oct. 2009).
9. Hashem, E.-S. Epidemiology of Viral Hepatitis and Hepatocellular Carcinoma. *Gastroenterology.* ISSN: 0016-5085. doi:10.1053/j.gastro.2011.12.061 (2012).

10. Thein, H.-H., Yi, Q., Dore, G. J. & Krahn, M. D. Estimation of stage-specific fibrosis progression rates in chronic hepatitis C virus infection: a meta-analysis and meta-regression. *Hepatology (Baltimore, Md.)* **48,** 418–31. ISSN: 1527-3350 (Aug. 2008).

11. Walker, P. J. *et al.* Changes to virus taxonomy and the International Code of Virus Classification and Nomenclature ratified by the International Committee on Taxonomy of Viruses (2019). *Archives of Virology* **164,** 2417–2429. ISSN: 14328798 (Sept. 2019).

12. Trivedi, S. *et al.* Viral persistence, liver disease, and host response in a hepatitis C–like virus rat model. *Hepatology* **68,** 435–448. ISSN: 15273350 (Aug. 2018).

13. Atcheson, E. *et al.* Use of an Outbred Rat Hepacivirus Challenge Model for Design and Evaluation of Efficacy of Different Immunization Strategies for Hepatitis C Virus. *Hepatology* **71,** 794–807. ISSN: 0270-9139 (Mar. 2020).

14. Simons, J. N. *et al.* Isolation of novel vims-like sequences associated with human hepatitis. *Nature Medicine* **1,** 564–569. ISSN: 1546170X (1995).

15. Stapleton, J. T., Foung, S., Muerhoff, A. S., Bukh, J. & Simmonds, P. *The GB viruses: A review and proposed classification of GBV-A, GBV-C (HGV), and GBV-D in genus Pegivirus within the family Flaviviridae* Feb. 2011. doi:10.1099/vir.0.027490-0. https://pubmed.ncbi.nlm.nih.gov/21084497/%20https://pubmed.ncbi.nlm.nih.gov/21084497/?dopt=Abstract.

16. Kapoor, A. *et al.* Virome analysis of transfusion recipients reveals a novel human virus that shares genomic features with hepaciviruses and pegiviruses. *mBio* **6.** ISSN: 21507511. doi:10.1128/mBio.01466-15. /pmc/articles/PMC4600124/?report=abstract%20https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4600124/ (2015).

17. Berg, M. G. *et al.* Discovery of a Novel Human Pegivirus in Blood Associated with Hepatitis C Virus Co-Infection. *PLoS Pathogens* **11.** ISSN: 15537374. doi:10.1371/journal.ppat.1005325. /pmc/articles/PMC4676677/?report=abstract%20https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4676677/ (2015).

18. Lamp, B. *et al.* Novel pestivirus species in pigs, Austria, 2015. *Emerging Infectious Diseases* **23,** 1176–1179. ISSN: 10806059 (July 2017).

19. Smith, D. B. *et al.* Proposed revision to the taxonomy of the genus Pestivirus, family Flaviviridae. *Journal of General Virology* **98,** 2106–2112. ISSN: 14652099 (2017).

20. Hoppe, I. B. A. L., de Souza-Pollo, A., de Medeiros, A. S. R., Samara, S. I. & Carvalho, A. A. B. HoBi-like pestivirus infection in an outbreak of bovine respiratory disease. *Research in Veterinary Science* **126,** 184–191. ISSN: 15322661 (Oct. 2019).

21. Ross-Thriepland, D. & Harris, M. Hepatitis C virus NS5A: enigmatic but still promiscuous 10 years on! *Journal of General Virology* **96,** 727–738. ISSN: 0022-1317 (Apr. 2015).

22. Dustin, L. B., Bartolini, B., Capobianchi, M. R. & Pistello, M. *Hepatitis C virus: life cycle in cells, infection and host response, and analysis of molecular markers influencing the outcome of infection and response to therapy* 2016. doi:10.1016/j.cmi.2016.08.025.

23. Agnello, V., Ábel, G., Elfahal, M., Knight, G. B. & Zhang, Q. X. Hepatitis C virus and other flaviviridae viruses enter cells via low density lipoprotein receptor. *Proceedings of the National Academy of Sciences of the United States of America* **96,** 12766–12771. ISSN: 00278424 (Oct. 1999).

24. Andre, P. *et al.* Characterization of Low- and Very-Low-Density Hepatitis C Virus RNA-Containing Particles. *Journal of Virology* **76,** 6919–6928. ISSN: 0022-538X (July 2002).

25. Allander, T., Forns, X., Emerson, S. U., Purcell, R. H. & Bukh, J. Hepatitis C virus envelope protein E2 binds to CD81 of tamarins. *Virology* **277,** 358–367. ISSN: 00426822 (Nov. 2000).

26. Scarselli, E. *et al.* The human scavenger receptor class B type I is a novel candidate receptor for the hepatitis C virus. *EMBO Journal* **21,** 5017–5025. ISSN: 02614189 (Oct. 2002).

27. Evans, M. J. *et al.* Claudin-1 is a hepatitis C virus co-receptor required for a late step in entry. *Nature* **446,** 801–805. ISSN: 14764687 (Apr. 2007).

28. Zeisel, M. B., Felmlee, D. J. & Baumert, T. F. in, 87–112 (2013). doi:10.1007/978-3-642-27340-7_4. http://link.springer.com/10.1007/978-3-642-27340-7%7B%5C_%7D4.

29. Zeuzem, S. Clinical implications of hepatitis C viral kinetics. *Journal of hepatology* **31 Suppl 1,** 61–4. ISSN: 0168-8278 (Jan. 1999).

30. Lindenbach, B. D. in, 199–218 (2013). doi:10.1007/978-3-642-27340-7_8. http://link.springer.com/10.1007/978-3-642-27340-7%7B%5C_%7D8.

31. Ribeiro, R. M. *et al.* Quantifying the Diversification of Hepatitis C Virus (HCV) during Primary Infection: Estimates of the In Vivo Mutation Rate. *PLoS Pathogens* **8.** ISSN: 15537366. doi:10.1371/journal.ppat.1002881 (Aug. 2012).

32. Martinez, M. A., Nevot, M., Jordan-Paiz, A. & Franco, S. Similarities between Human Immunodeficiency Virus Type 1 and Hepatitis C Virus Genetic and Phenotypic Protease Quasispecies Diversity. *Journal of Virology* **89,** 9758–9764. ISSN: 0022-538X (Oct. 2015).

33. Powdrill, M. H. *et al.* Contribution of a mutational bias in hepatitis C virus replication to the genetic barrier in the development of drug resistance. *Proceedings of the National Academy of Sciences of the United States of America* **108,** 20509–20513. ISSN: 00278424 (Dec. 2011).

34. Smith, D. B. *et al.* Expanded classification of hepatitis C virus into 7 genotypes and 67 subtypes: Updated criteria and genotype assignment web resource. *Hepatology* **59,** 318–327. ISSN: 15273350 (2014).

35. Bonsall, D. *et al.* ve-SEQ: Robust, unbiased enrichment for streamlined detection and whole-genome sequencing of HCV and other highly diverse pathogens. *F1000Research* **4,** 1062 (2015).

36. Cuypers, L. *et al.* Next-generation sequencing for the clinical management of hepatitis C virus infections: Does one test fits all purposes? *Critical Reviews in Clinical Laboratory Sciences,* 1–15. ISSN: 1040-8363 (July 2019).

37. Simmonds, P. *et al.* Consensus proposals for a unified system of nomenclature of hepatitis C virus genotypes. *Hepatology (Baltimore, Md.)* **42,** 962–73. ISSN: 0270-9139 (Oct. 2005).

38. Hedskog, C. *et al.* Identification of 19 Novel Hepatitis C Virus Subtypes-Further Expanding HCV Classification. *Open Forum Infectious Diseases.* ISSN: 2328-8957. doi:10.1093/ofid/ofz076. https://academic.oup.com/ofid/advance-article/doi/10.1093/ofid/ofz076/5362553 (Feb. 2019).

39. *International Committee on Taxonomy of Viruses (ICTV)* https://talk.ictvonline.org/ictv%7B%5C_%7Dwikis/flaviviridae/w/sg%7B%5C_%7Dflavi/56/hcv-classification (2019).

40. Stamatakis, A. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30,** 1312–1313. ISSN: 14602059 (2014).

41. Messina, J. P. *et al.* Global distribution and prevalence of hepatitis C virus genotypes. *Hepatology* **61,** 77–87. ISSN: 15273350 (2015).

42. Salmona, M. *et al.* First next-generation sequencing full-genome characterization of a hepatitis C virus genotype 7 divergent subtype. *Clinical Microbiology and Infection* **22,** 947.e1–947.e8. ISSN: 1198743X (Nov. 2016).

43. Welzel, T. M. *et al.* Global epidemiology of HCV subtypes and resistance-associated substitutions evaluated by sequencing-based subtype analyses. *Journal of Hepatology* **67,** 224–236. ISSN: 16000641 (2017).

44. Pybus, O. G. *et al.* The epidemic behavior of the hepatitis C virus. *Science* **292,** 2323–2325. ISSN: 00368075 (June 2001).

45. Pybus, O. G. *et al.* Genetic history of hepatitis C virus in East Asia. *Journal of virology* **83,** 1071–82. ISSN: 1098-5514 (Jan. 2009).

46. Smith, D. B. *et al.* The origin of hepatitis C virus genotypes. *Journal of General Virology* **78,** 321–328. ISSN: 00221317 (1997).

47. Davis, G. L. *et al.* Treatment of chronic hepatitis C with recombinant interferon alfa. A multicenter randomized, controlled trial. *The New England journal of medicine* **321,** 1501–6. ISSN: 0028-4793 (1989).

48. McHutchison, J. G. *et al.* Interferon alfa-2b alone or in combination with ribavirin as initial treatment for chronic hepatitis C. Hepatitis Interventional Therapy Group. *The New England journal of medicine* **339,** 1485–92. ISSN: 0028-4793 (Nov. 1998).

49. Graci, J. D. & Cameron, C. E. Mechanisms of action of ribavirin against distinct viruses. *Reviews in Medical Virology* **16,** 37–48. ISSN: 1052-9276 (Jan. 2006).

50. Manns, M. P. *et al.* Peginterferon alfa-2b plus ribavirin compared with interferon alfa-2b plus ribavirin for initial treatment of chronic hepatitis C: a randomised trial. *Lancet (London, England)* **358,** 958–65. ISSN: 0140-6736 (Sept. 2001).

51. Pawlotsky, J. M. *et al.* EASL Recommendations on Treatment of Hepatitis C 2018. *Journal of Hepatology* **69,** 461–511. ISSN: 16000641 (2018).

52. American Association for the Study of Liver Diseases. *Recommendations for Testing, Managing, and Treating Hepatitis C — HCV Guidance* `http://www.hcvguidelines.org/` (2017).

53. Poordad, F. *et al.* Boceprevir for untreated chronic HCV genotype 1 infection. *The New England journal of medicine* **364,** 1195–206. ISSN: 1533-4406 (Mar. 2011).

54. Jacobson, I. M. *et al.* Telaprevir for previously untreated chronic hepatitis C virus infection. *New England Journal of Medicine* **364,** 2405–2416. ISSN: 1533-4406 (June 2011).

55. EASL Clinical Practice Guidelines: management of hepatitis C virus infection. English. *Journal of hepatology* **60,** 392–420. ISSN: 1600-0641 (Feb. 2014).

56. Manns, M. *et al.* 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 (London, England)* **384,** 414–26. ISSN: 1474-547X (Aug. 2014).

57. Sheikh, A. M. *et al.* Sofosbuvir for Previously Untreated Chronic Hepatitis C Infection. *New England Journal of Medicine* **368,** 1878–1887. ISSN: 0028-4793 (2013).

58. Foster, G. R. *et al.* Sofosbuvir and Velpatasvir for HCV Genotype 2 and 3 Infection. *New England Journal of Medicine* **373,** 2608–2617. ISSN: 1533-4406 (Nov. 2015).

59. Lawitz, E. *et al.* 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. *The Lancet* **384,** 1756–1765. ISSN: 1474547X (Nov. 2014).

60. Buti, M. *et al.* Simeprevir in combination with sofosbuvir in treatment-naïve and -experienced patients with hepatitis C virus genotype 4 infection: a Phase III, open-label, single-arm study (PLUTO). *Alimentary Pharmacology & Therapeutics* **45,** 468–475. ISSN: 02692813 (Feb. 2017).

61. Sulkowski, M. S. *et al.* Daclatasvir plus Sofosbuvir for Previously Treated or Untreated Chronic HCV Infection. *New England Journal of Medicine* **370,** 211–221. ISSN: 15334406 (2014).

62. Nelson, D. R. *et al.* All-oral 12-week treatment with daclatasvir plus sofosbuvir in patients with hepatitis C virus genotype 3 infection: ALLY-3 phase III study. *Hepatology (Baltimore, Md.)* **61,** 1127–35. ISSN: 1527-3350 (Apr. 2015).

63. Afdhal, N. *et al.* Ledipasvir and sofosbuvir for untreated HCV genotype 1 infection. *New England Journal of Medicine* **370,** 1889–1898. ISSN: 15334406 (2014).

64. Afdhal, N. *et al.* Ledipasvir and sofosbuvir for previously treated HCV genotype 1 infection. *New England Journal of Medicine* **370,** 1483–1493. ISSN: 15334406 (2014).

65. Buggisch, P. *et al.* Real-world effectiveness of 8-week treatment with ledipasvir/sofosbuvir in chronic hepatitis C. *Journal of Hepatology* **68.** ISSN: 16000641. doi:10.1016/j.jhep.2017.11.009 (2018).

66. Naggie, S. *et al.* Ledipasvir and sofosbuvir for HCV in patients coinfected with HIV-1. *New England Journal of Medicine* **373,** 705–713. ISSN: 15334406 (Aug. 2015).

67. Kohli, A. *et al.* Ledipasvir and sofosbuvir for hepatitis C genotype 4: A proof-of-concept, single-centre, open-label phase 2a cohort study. *The Lancet Infectious Diseases* **15,** 1049–1054. ISSN: 14744457 (Sept. 2015).

68. Abergel, A. *et al.* Ledipasvir-sofosbuvir in patients with hepatitis C virus genotype 5 infection: An open-label, multicentre, single-arm, phase 2 study. *The Lancet Infectious Diseases* **16,** 459–464. ISSN: 14744457 (Apr. 2016).

69. da Silva Filipe, A. *et al.* Response to DAA therapy in the NHS England Early Access Programme for rare HCV subtypes from low and middle income countries. *Journal of Hepatology* **67,** 1348–1350. ISSN: 16000641 (Aug. 2017).

70. Feld, J. J. *et al.* Treatment of HCV with ABT-450/r-ombitasvir and dasabuvir with ribavirin. *New England Journal of Medicine* **370,** 1594–1603. ISSN: 15334406 (2014).

71. Ferenci, P. *et al.* ABT-450/r-ombitasvir and dasabuvir with or without ribavirin for HCV. *New England Journal of Medicine* **370,** 1983–1992. ISSN: 15334406 (2014).

72. Zeuzem, S. *et al.* Retreatment of HCV with ABT-450/r-ombitasvir and dasabuvir with ribavirin. *New England Journal of Medicine* **370,** 1604–1614. ISSN: 15334406 (2014).

73. Andreone, P. *et al.* 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.** ISSN: 15280012. doi:10.1053/j.gastro.2014.04.045 (2014).

74. Poordad, F. *et al.* ABT-450/r-ombitasvir and dasabuvir with ribavirin for hepatitis C with cirrhosis. *New England Journal of Medicine* **370,** 1973–1982. ISSN: 15334406 (2014).

75. Feld, J. J. *et al.* Sustained virologic response of 100% in HCV genotype 1b patients with cirrhosis receiving ombitasvir/paritaprevir/r and dasabuvir for 12 weeks. *Journal of Hepatology* **64,** 301–307. ISSN: 16000641 (Feb. 2016).

76. Hézode, C. *et al.* Ombitasvir plus paritaprevir plus ritonavir with or without ribavirin in treatment-naive and treatment-experienced patients with genotype 4 chronic hepatitis C virus infection (PEARL-I): A randomised, open-label trial. *The Lancet* **385,** 2502–2509. ISSN: 1474547X (June 2015).

77. Keating, G. M. Ombitasvir/Paritaprevir/Ritonavir: A Review in Chronic HCV Genotype 4 Infection. *Drugs* **76,** 1203–1211. ISSN: 11791950 (Aug. 2016).

78. Zeuzem, S. *et al.* Grazoprevir-Elbasvir combination therapy for treatment-naive cirrhotic and noncirrhotic patients with chronic hepatitis C virus genotype 1, 4, or 6 infection: A randomized trial. *Annals of Internal Medicine* **163,** 1–13. ISSN: 15393704 (July 2015).

79. Rockstroh, J. K. *et al.* Efficacy and safety of grazoprevir (MK-5172) and elbasvir (MK-8742) in patients with hepatitis C virus and HIV co-infection (C-EDGE CO-INFECTION): A non-randomised, open-label trial. *The Lancet HIV* **2,** e319–e327. ISSN: 23523018 (Aug. 2015).

80. Kwo, P. *et al.* Effectiveness of Elbasvir and Grazoprevir Combination, With or Without Ribavirin, for Treatment-Experienced Patients With Chronic Hepatitis C Infection. *Gastroenterology* **152,** 164–175.e4. ISSN: 15280012 (Jan. 2017).

81. Zeuzem, S. *et al.* The safety and efficacy of elbasvir and grazoprevir in participants with hepatitis C virus genotype 1b infection. *Journal of Gastroenterology* **53,** 679–688. ISSN: 14355922 (May 2018).

82. Wei, L. *et al.* Efficacy and safety of elbasvir/grazoprevir in participants with hepatitis C virus genotype 1, 4, or 6 infection from the Asia–Pacific region and Russia: Final results from the randomized C-CORAL study. *Journal of Gastroenterology and Hepatology (Australia)* **34,** 12–21. ISSN: 14401746 (Jan. 2019).

83. Feld, J. J. *et al.* Sofosbuvir and Velpatasvir for HCV Genotype 1, 2, 4, 5, and 6 Infection. *The New England journal of medicine* **373,** 2599–607. ISSN: 1533-4406 (Nov. 2015).

84. Zeuzem, S. *et al.* Glecaprevir-pibrentasvir for 8 or 12 weeks in HCV genotype 1 or 3 infection. *New England Journal of Medicine* **378,** 354–369. ISSN: 15334406 (Jan. 2018).

85. Forns, X. *et al.* Glecaprevir plus pibrentasvir for chronic hepatitis C virus genotype 1, 2, 4, 5, or 6 infection in adults with compensated cirrhosis (EXPEDITION-1): a single-arm, open-label, multicentre phase 3 trial. *The Lancet Infectious Diseases* **17,** 1062–1068. ISSN: 14744457 (Oct. 2017).

86. Wyles, D. *et al.* Glecaprevir/pibrentasvir for hepatitis C virus genotype 3 patients with cirrhosis and/or prior treatment experience: A partially randomized phase 3 clinical trial. *Hepatology* **67,** 514–523. ISSN: 15273350 (Feb. 2018).

87. Bourlière, M. *et al.* Sofosbuvir, Velpatasvir, and Voxilaprevir for Previously Treated HCV Infection. *New England Journal of Medicine* **376,** 2134–2146. ISSN: 0028-4793 (June 2017).

88. Yoshimi, S. Long term persistence of NS5A inhibitor-resistant hepatitis C virus in patients who failed daclatasvir and asunaprevir therapy. *Journal Of Medical Virology* (2015).

89. Wyles, D. *et al.* Retreatment of patients who failed glecaprevir/pibrentasvir treatment for hepatitis C virus infection. *Journal of Hepatology* **70,** 1019–1023. ISSN: 16000641 (May 2019).

90. Bartenschlager, R. & Lohmann, V. Replication of hepatitis C virus. *Journal of General Virology* **81,** 1631–1648. ISSN: 0022-1317 (July 2000).

91. Perales, C., Quer, J., Gregori, J., Esteban, J. I. & Domingo, E. *Resistance of hepatitis C virus to inhibitors: Complexity and clinical implications* Nov. 2015. doi:10.3390/v7112902.

92. Andino, R. & Domingo, E. *Viral quasispecies* May 2015. doi:10.1016/j.virol.2015.03.022.

93. Lenz, O. *et al.* In Vitro Resistance Profile of the Hepatitis C Virus NS3/4A Protease Inhibitor TMC435. *Antimicrobial Agents And Chemotherapy* **54,** 1878–1887. ISSN: 00664804 (May 2010).

94. Summa, V. *et al.* MK-5172, a selective inhibitor of hepatitis C virus NS3/4a protease with broad activity across genotypes and resistant variants. *Antimicrobial Agents and Chemotherapy* **56,** 4161–4167. ISSN: 00664804 (Aug. 2012).

95. Sarrazin, C. *et al.* No impact of resistance-associated substitutions on the efficacy of sofosbuvir, velpatasvir, and voxilaprevir for 12 weeks in HCV DAA-experienced patients. *Journal of Hepatology* **69,** 1221–1230. ISSN: 16000641 (Dec. 2018).

96. European Association for the Study of the Liver. EASL Recommendations on Treatment of Hepatitis C 2016. *Journal of Hepatology* **66,** 153–194 (2017).

97. Leroy, V. *et al.* Daclatasvir, sofosbuvir, and ribavirin for hepatitis C virus genotype 3 and advanced liver disease: A randomized phase III study (ALLY-3+). *Hepatology (Baltimore, Md.)* **63,** 1430–41. ISSN: 1527-3350 (May 2016).

98. Smith, D. *et al.* Resistance analysis of genotype 3 HCV indicates subtypes inherently resistant to NS5A inhibitors. *Hepatology* **00,** 1–12. ISSN: 02709139 (Feb. 2018).

99. Poordad, F. *et al.* Glecaprevir and pibrentasvir for 12 weeks for hepatitis C virus genotype 1 infection and prior direct-acting antiviral treatment. *Hepatology* **66,** 389–397. ISSN: 15273350 (2017).

100. Kati, W. In vitro activity and resistance profile of dasabuvir, a nonnucleoside hepatitis C virus polymerase inhibitor. *Antimicrob Agents Chemother* **59,** 1505–1511 (2015).

101. Donaldson, E. F., Harrington, P. R., O'Rear, J. J., Naeger, L. K. & O'Rear, J. J. Clinical evidence and bioinformatics characterization of potential hepatitis C virus resistance pathways for Sofosbuvir. *Hepatology (Baltimore, Md.)* n/a–n/a. ISSN: 1527-3350 (2014).

102. Svarovskaia, E. S. *et al.* L159F and V321A Sofosbuvir-Associated Hepatitis C Virus NS5B Substitutions. *Journal of Infectious Diseases.* ISSN: 15376613. doi:10.1093/infdis/jiv564 (2016).

103. Dietz, J. *et al.* Patterns of Resistance-Associated Substitutions in Patients With Chronic HCV Infection Following Treatment With Direct-Acting Antivirals. *Gastroenterology* **154,** 976–988.e4. ISSN: 15280012 (Mar. 2018).

104. Ramirez, S., Mikkelsen, L. S., Gottwein, J. M. & Bukh, J. Robust HCV Genotype 3a Infectious Cell Culture System Permits Identification of Escape Variants With Resistance to Sofosbuvir. *Gastroenterology* **151,** 973–985.e2. ISSN: 15280012 (2016).

105. Vermehren, J. *et al.* Retreatment of Patients who Failed Daa-Combination Therapies: Real-World Experience from a Large Hepatitis C Resistance Database. *Journal of Hepatology* **64,** S188. ISSN: 01688278 (2016).

106. Pawlotsky, J.-M. Hepatitis C Virus Resistance to Direct-Acting Antiviral Drugs in Interferon-Free Regimens. *Gastroenterology.* ISSN: 00165085. doi:10.1053/j.gastro.2016.04.003. http://linkinghub.elsevier.com/retrieve/pii/S0016508516300555 (Apr. 2016).

107. Thomson, E. *et al.* Comparison of Next-Generation Sequencing Technologies for Comprehensive Assessment of Full-Length Hepatitis C Viral Genomes. *Journal of Clinical Microbiology* **54** (ed Loeffelholz, M. J.) 2470–2484. ISSN: 0095-1137 (Oct. 2016).

108. Catanese, M. T. & Dorner, M. Advances in experimental systems to study hepatitis C virus in vitro and in vivo. *Virology* **479-480,** 221–233. ISSN: 10960341 (2015).

109. Lohmann, V. *et al.* Replication of subgenomic hepatitis C virus RNAs in a hepatoma cell line. *Science (New York, N.Y.)* **285,** 110–3. ISSN: 0036-8075 (July 1999).

110. Witteveldt, J., Martin-Gans, M. & Simmonds, P. Enhancement of the replication of HCV replicons of genotypes 1-4 by manipulation of CpG and UpA dinucleotide frequencies and use of cell lines expressing SECL14L2 – application for antiviral resistance testing. *Antimicrobial Agents and Chemotherapy* **60,** AAC.02932–15. ISSN: 0066-4804 (2016).

111. Lindenbach, B. D. *et al.* Complete replication of hepatitis C virus in cell culture. *Science (New York, N.Y.)* **309,** 623–6. ISSN: 1095-9203 (July 2005).

112. Wakita, T. *et al.* Production of infectious hepatitis C virus in tissue culture from a cloned viral genome. *NIH Public Access* **11,** 791–796. ISSN: 10788956 (July 2010).

113. Gottwein, J. M. *et al.* Efficacy of NS5A Inhibitors Against Hepatitis C Virus Genotypes 1–7 and Escape Variants. *Gastroenterology* **154,** 1435–1448. ISSN: 00165085 (2017).

114. Li, Y. P. *et al.* Highly efficient full-length hepatitis C virus genotype 1 (strain TN) infectious culture system. *Proceedings of the National Academy of Sciences of the United States of America* **109,** 19757–19762. ISSN: 00278424 (Nov. 2012).

115. Li, Y. P. *et al.* Robust full-length hepatitis C virus genotype 2a and 2b infectious cultures using mutations identified by a systematic approach applicable to patient strains. *Proceedings of the National Academy of Sciences of the United States of America* **109.** ISSN: 00278424. doi:10.1073/pnas.1203829109 (May 2012).

116. Ramirez, S. *et al.* Highly efficient infectious cell culture of three hepatitis C virus genotype 2b strains and sensitivity to lead protease, nonstructural protein 5A, and polymerase inhibitors. *Hepatology* **59,** 395–407. ISSN: 02709139 (Feb. 2014).

117. Kowdley, K. V. *et al.* Ledipasvir and Sofosbuvir for 8 or 12 Weeks for Chronic HCV without Cirrhosis. *New England Journal of Medicine* **370,** 1879–1888. ISSN: 0028-4793 (May 2014).

118. Wyles, D. L. *et al.* Daclatasvir plus Sofosbuvir for HCV in Patients Coinfected with HIV-1. *New England Journal of Medicine* **373,** 714–725. ISSN: 0028-4793 (Aug. 2015).

119. Hézode, C. *et al.* Daclatasvir plus sofosbuvir, with or without ribavirin, for hepatitis C virus genotype 3 in a French early access programme. *Liver International* **37,** 1314–1324. ISSN: 14783223 (Sept. 2017).

120. Ampuero, J., Romero-Gómez, M. & Reddy, K. R. Review article: HCV genotype 3 - the new treatment challenge. *Alimentary Pharmacology & Therapeutics* **39,** 686–698. ISSN: 02692813 (2014).

121. Sarwar, S., A.Khan, A. & Tarique, S. Response Guided Interferon Therapy for Genotype 3 of Chronic Hepatitis C : Compliance and Outcome. **31,** 843–847 (2015).

122. Jacobson, I. M. *et al.* Efficacy of 8 Weeks of Sofosbuvir, Velpatasvir, and Voxilaprevir In Patients With Chronic HCV Infection: 2 Phase 3 Randomized Trials. *Gastroenterology* **153,** 113–122. ISSN: 00165085 (July 2017).

123. Pianko, S. *et al.* Sofosbuvir Plus Velpatasvir Combination Therapy for Treatment-Experienced Patients With Genotype 1 or 3 Hepatitis C Virus Infection. *Annals of Internal Medicine* **163,** 809. ISSN: 0003-4819 (Dec. 2015).

124. Foster, G. R. *et al. C-ISLE: Grazoprevir/Elbasvir plus Sofosbuvir in Treatment-naive and Treatment-experienced HCV GT3 Cirrhotic Patients Treated for 8, 12 or 16 weeks.[Abstract 74]* in *67th Meeting of the American Association for the Study of Liver Diseases: Liver Meeting* (2016). http://www.natap.org/2016/AASLD/AASLD%7B%5C_%7D58.html.

125. Ge, D. *et al.* Genetic variation in IL28B predicts hepatitis C treatment-induced viral clearance. *Nature* **461,** 399–401. ISSN: 0028-0836 (Sept. 2009).

126. Bochud, P.-Y. *et al.* Genotype 3 is associated with accelerated fibrosis progression in chronic hepatitis C. *Journal of Hepatology* **51,** 655–666. ISSN: 01688278 (Oct. 2009).

127. Nkontchou, G. *et al.* HCV genotype 3 is associated with a higher hepatocellular carcinoma incidence in patients with ongoing viral C cirrhosis. *Journal of Viral Hepatitis* **18,** e516–e522. ISSN: 13520504 (Oct. 2011).

128. Larsen, C. *et al.* Hepatitis C virus genotype 3 and the risk of severe liver disease in a large population of drug users in France. *Journal of medical virology* **82,** 1647–1654. ISSN: 1096-9071 (Sept. 2010).

129. Poynard, T., Bedossa, P. & Opolon, P. Natural history of liver fibrosis progression in patients with chronic hepatitis C. The OBSVIRC, METAVIR, CLINIVIR, and DOSVIRC groups. *Lancet* **349,** 825–32. ISSN: 0140-6736 (Mar. 1997).

130. American Association for The Study of Liver Diseases. Recommendations for Testing, Managing, and Treating Hepatitis C. http://www.hcvguidelines.org.

131. Omata, M. *et al.* APASL consensus statements and recommendation on treatment of hepatitis C. *Hepatol Int* **10,** 702–726 (2016).

132. Lontok, E. *et al.* Hepatitis C virus drug resistance-associated substitutions: State of the art summary. *Hepatology* **62,** 1623–32. ISSN: 02709139 (Nov. 2015).

133. Sarrazin, C. *et al.* Prevalence of Resistance-Associated Substitutions in HCV NS5A, NS5B, or NS3 and Outcomes of Treatment with Ledipasvir and Sofosbuvir. *Gastroenterology* **151,** 501–512. ISSN: 00165085 (2016).

134. Saeed, M. *et al.* Efficient replication of genotype 3a and 4a hepatitis C virus replicons in human hepatoma cells. *Antimicrobial agents and chemotherapy* **56,** 5365–73. ISSN: 1098-6596 (Oct. 2012).

135. Macalalad, A. R. *et al.* Highly sensitive and specific detection of rare variants in mixed viral populations from massively parallel sequence data. *PLoS computational biology* **8,** e1002417. ISSN: 1553-7358 (Jan. 2012).

136. Henn, M. R. *et al.* Whole genome deep sequencing of HIV-1 reveals the impact of early minor variants upon immune recognition during acute infection. *PLoS pathogens* **8,** e1002529. ISSN: 1553-7374 (Jan. 2012).

137. Foster, G. R. *et al.* Efficacy of Sofosbuvir Plus Ribavirin With or Without Peginterferon-Alfa in Patients With Hepatitis C Virus Genotype 3 Infection and Treatment-Experienced Patients With Cirrhosis and Hepatitis C Virus Genotype 2 Infection. *Gastroenterology* **149,** 1462–1470. ISSN: 00165085 (2015).

138. Batty, E. M. *et al.* A Modified RNA-Seq Approach for Whole Genome Sequencing of RNA Viruses from Faecal and Blood Samples. *PLoS ONE* **8.** ISSN: 19326203. doi:10.1371/journal.pone.0066129 (2013).

139. Gaidatzis, D., Lerch, A., Hahne, F. & Stadler, M. B. QuasR: quantification and annotation of short reads in R. *Bioinformatics* **31,** 1130–1132. ISSN: 1460-2059 (Apr. 2015).

140. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17,** 10. ISSN: 2226-6089 (May 2011).

141. Langmead, B. Aligning short sequencing reads with Bowtie. *Current protocols in bioinformatics* **Chapter 11,** Unit 11.7. ISSN: 1934-340X (Dec. 2010).

142. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics (Oxford, England)* **25,** 1754–60. ISSN: 1367-4811 (July 2009).

143. Lunter, G. & Goodson, M. Stampy: a statistical algorithm for sensitive and fast mapping of Illumina sequence reads. *Genome research* **21,** 936–9. ISSN: 1549-5469 (June 2011).

144. Yang, X. *et al.* De novo assembly of highly diverse viral populations. *BMC Genomics* **13,** 475. ISSN: 1471-2164 (Sept. 2012).

145. Charlebois, P., Yang, X., Newman, R., Henn, M. & Zody, M. V-FAT: a post-assembly pipeline for the finishing and annotation of viral genomes. `https://www.broadinstitute.org/viral-genomics/v-fat`.

146. Lee, W.-P. *et al.* MOSAIK: A Hash-Based Algorithm for Accurate Next-Generation Sequencing Short-Read Mapping. *PLoS ONE* **9** (ed Hsiao, C. K.) e90581. ISSN: 1932-6203 (Mar. 2014).

147. Kurosaki, M. *et al.* Complex pattern of resistance-associated substitutions of hepatitis c virus after daclatasvir/asunaprevir treatment failure jun Itakura. *PLoS ONE* **11,** 1–13. ISSN: 19326203 (2016).

148. Magri, A. *et al.* Rethinking the old antiviral drug moroxydine: Discovery of novel analogues as anti-hepatitis C virus (HCV) agents. *Bioorganic and Medicinal Chemistry Letters* **25,** 5372–5376. ISSN: 14643405 (2015).

149. Magri, A. *et al.* Exploration of acetanilide derivatives of 1-(ω-phenoxyalkyl)uracils as novel inhibitors of Hepatitis C Virus replication. *Scientific Reports* **6,** 29487. ISSN: 2045-2322 (Sept. 2016).

150. Schliep, E. P. & K. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in {R}. *Bioinformatics* **25,** 526–528 (2018).

151. Chen, Z.-w., Li, H., Ren, H. & Hu, P. Global prevalence of pre-existing HCV variants resistant to direct-acting antiviral agents (DAAs): mining the GenBank HCV genome data. *Scientific Reports* **6,** 20310. ISSN: 2045-2322 (2016).

152. Patiño-Galindo, J. Á., Salvatierra, K., González-Candelas, F. & López-Labrador, F. X. Comprehensive Screening for Naturally Occurring Hepatitis C Virus Resistance to Direct-Acting Antivirals in the NS3, NS5A, and NS5B Genes in Worldwide Isolates of Viral Genotypes 1 to 6. *Antimicrobial agents and chemotherapy* **60,** 2402–16. ISSN: 1098-6596 (Apr. 2016).

153. Vermehren, J. & Sarrazin, C. The role of resistance in HCV treatment. *Best Practice and Research: Clinical Gastroenterology* **26,** 487–503. ISSN: 15216918 (2012).

154. Hernandez, D., Zhou, N., Ueland, J., Monikowski, A. & McPhee, F. Natural prevalence of NS5A polymorphisms in subjects infected with hepatitis C virus genotype 3 and their effects on the antiviral activity of NS5A inhibitors. *Journal of Clinical Virology* **57,** 13–18. ISSN: 13866532 (May 2013).

155. Wasitthankasem, R. *et al.* Genotypic Distribution of Hepatitis C Virus in Thailand and Southeast Asia. *PLOS ONE* **10** (ed Hotta, H.) e0126764. ISSN: 1932-6203 (May 2015).

156. Wei, L. *et al.* Sofosbuvir–velpatasvir for treatment of chronic hepatitis C virus infection in Asia: a single-arm, open-label, phase 3 trial. *The Lancet Gastroenterology and Hepatology.* ISSN: 24681253. doi:10.1016/S2468-1253(18)30343-1 (2019).

157. Fourati, S. *et al.* Frequent antiviral treatment failures in patients infected with hepatitis C virus genotype 4, subtype 4r. *Hepatology.* ISSN: 02709139. doi:10.1002/hep.30225. http://doi.wiley.com/10.1002/hep.30225 (2018).

158. Davis, C. *et al.* Highly Diverse Hepatitis C Strains Detected in Sub-Saharan Africa Have Unknown Susceptibility to Direct-Acting Antiviral Treatments. *Hepatology,* hep.30342. ISSN: 0270-9139 (Mar. 2019).

159. AASLD. *AASLD-IDSA. Recommendations for testing, managing, and treating hepatitis C* 2019. (2019).

160. Niu, C. *et al.* PSI-7851, a Pronucleotide of -D-2'-Deoxy-2'-Fluoro-2'-C-Methyluridine Monophosphate, Is a Potent and Pan-Genotype Inhibitor of Hepatitis C Virus Replication. *Antimicrobial Agents and Chemotherapy* **54,** 3187–3196. ISSN: 0066-4804 (2010).

161. Svarovskaia, E. S. *et al.* 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.** ISSN: 15376591. doi:10.1093/cid/ciu697 (2014).

162. Azim Ansari, M. & Didelot, X. Bayesian inference of the evolution of a phenotype distribution on a phylogenetic tree. *Genetics* **204,** 89–98. ISSN: 19432631 (2016).

163. Ansari, M. A. *et al.* Genome-to-genome analysis highlights the effect of the human innate and adaptive immune systems on the hepatitis C virus. *Nat Genet* **advance on.** ISSN: 1546-1718. doi:10.1038/ng.3835. http://dx.doi.org/10.1038/ng.3835%7B%5C%%7D0Ahttp://10.0.4.14/ng.3835%7B%5C%%7D0Ahttp://www.nature.

com/ng/journal/vaop/ncurrent/abs/ng.3835.html%7B%5C#%7Dsupplementary-information (2017).

164. Wing, P. A. *et al.* Amino Acid Substitutions in Genotype 3a Hepatitis C Virus Polymerase Protein Affect Responses to Sofosbuvir. *Gastroenterology.* ISSN: 00165085. doi:10.1053/j.gastro.2019.05.007. https://linkinghub.elsevier.com/retrieve/pii/S0016508519408652 (2019).

165. Wyles, D. *et al.* Glecaprevir/Pibrentasvir for HCV Genotype 3 Patients with Cirrhosis and/or Prior Treatment Experience: A Partially Randomized Phase III Clinical Trial. *Hepatology.* ISSN: 02709139. doi:10.1002/hep.29541. http://www.ncbi.nlm.nih.gov/pubmed/28926120%20http://doi.wiley.com/10.1002/hep.29541 (Sept. 2017).

166. Popescu, C. I. *et al.* NS2 protein of hepatitis C virus interacts with structural and non-structural proteins towards virus assembly. *PLoS Pathogens* **7.** ISSN: 15537374. doi:10.1371/journal.ppat.1001278 (2011).

167. Lorenz, I. C., Marcotrigiano, J., Dentzer, T. G. & Rice, C. M. Structure of the catalytic domain of the hepatitis C virus NS2-3 protease. *Nature* **442,** 831–835. ISSN: 14764687 (Aug. 2006).

168. Feigelstock, D. A., Mihalik, K. B. & Feinstone, S. M. Selection of hepatitis C virus resistant to ribavirin. *Virology Journal* **8,** 402. ISSN: 1743-422X (Dec. 2011).

169. Mihalik, K. B. & Feigelstock, D. A. Sensitivity of a Ribavirin Resistant Mutant of Hepatitis C Virus to Other Antiviral Drugs. *PLoS ONE* **8** (ed Hotta, H.) e74027. ISSN: 1932-6203 (Sept. 2013).

170. Everson, G. T. *et al.* Sofosbuvir with velpatasvir in treatment-naive noncirrhotic patients with genotype 1 to 6 hepatitis c virus infection. *Annals of Internal Medicine* **163.** ISSN: 15393704. doi:10.7326/M15-1000 (2015).

171. Hezode, C. *et al.* Resistance analysis in patients with genotype 1–6 HCV infection treated with sofosbuvir/velpatasvir in the phase III studies. *Journal of Hepatology* **68,** 895–903. ISSN: 16000641 (2018).

172. Belperio, P. S., Shahoumian, T. A., Loomis, T. P. & Backus, L. I. Real-world effectiveness of sofosbuvir/velpatasvir/voxilaprevir in 573 direct-acting antiviral experienced hepatitis C patients. *Journal of Viral Hepatitis.* ISSN: 13652893. doi:10.1111/jvh.13115 (2019).

173. Degasperi, E. *et al.* Real-life effectiveness and safety of sofosbuvir/velpatasvir/voxilaprevir in hepatitis C patients with previous daa failure. *Journal of Hepatology* **71,** 1106–1115. ISSN: 01688278 (Aug. 2019).

174. Llaneras, J. *et al.* Effectiveness and safety of sofosbuvir/velpatasvir/voxilaprevir in patients with chronic hepatitis C previously treated with DAAs. *Journal of Hepatology* **71,** 666–672. ISSN: 01688278 (June 2019).

175. Krueger, F. *TrimGalore: A Wrapper around Cutadapt and FastQC to Consistently Apply Adapter and Quality Trimming to FastQ Files, with Extra Functionality for RRBS Data* https://github.com/%20FelixKrueger.

176. Singer, J. *et al.* Interpreting Viral Deep Sequencing Data with GLUE. *Viruses* **11,** 323. ISSN: 1999-4915 (2019).

177. Davalieva, K., Kiprijanovska, S. & Plaseska-Karanfilska, D. Fast, reliable and low cost user-developed protocol for detection, quantification and genotyping of hepatitis C virus. *Journal of Virological Methods* **196,** 104–112. ISSN: 0166-0934 (Feb. 2014).

178. Ponstingl, H. *SMALT: A Mapper for DNA Sequencing Reads* https://www.sanger.ac.uk/science/tools/smalt-0.

179. Camacho, C. *et al.* BLAST+: architecture and applications. *BMC Bioinformatics* **10,** 421 (2009).

180. Harris, R. *Improved pairwise alignment of genomic DNA* PhD thesis (he Pennsylvania State University, 2007). http://www.bx.psu.edu/%7B~%7Drsharris/lastz/.

181. Singer, J. B., Thomson, E. C., McLauchlan, J., Hughes, J. & Gifford, R. J. GLUE: A flexible software system for virus sequence data. *bioRxiv,* 269274 (2018).

182. Jensen, S. B. *et al.* Evolutionary Pathways to Persistence of Highly Fit and Resistant Hepatitis C Virus Protease Inhibitor Escape Variants. *Hepatology,* hep.30647. ISSN: 0270-9139 (Apr. 2019).

183. Abravanel, F., Métivier, S., Chauveau, M., Péron, J. M. & Izopet, J. *Transmission of HCV NS5A Inhibitor-Resistant Variants among HIV-Infected Men Who Have Sex with Men* Nov. 2016. doi:10.1093/cid/ciw554.

184. Buuren, N. V., Tellinghuisen, T. L., Richardson, C. D., Kirkegaard, K. & Scotia, N. Transmission Genetics of Drug-Resistant Hepatitis C Virus. *eLife* **e32579.** doi:10.7554/eLife.32579 (2018).

185. Newsum, A. M. *et al. Persistence of NS5B-S282T, a sofosbuvir resistance-associated substitution, in a HIV/HCV-coinfected MSM with risk of onward transmission* 2018. doi:10.1016/j.jhep.2018.06.021.

Appendix A

Protein	Position	Position in H77 Polyprotein	Amino Acid in H77	Codon in H77	Position in S52 Polyprotein	Amino Acid in S52	Codon in S52	Substitution
NS5A	30	2002	Q	caa	2008	A	gca	K
NS5A	31	2003	L	ctg	2009	L	ctc	M
NS5A	58	2030	H	cac	2036	P	cct	S
NS5A	93	2065	Y	tac	2071	Y	tac	H
NS5B	159	2579	L	ctc	2589	L	ctc	F
NS5B	282	2702	S	agc	2712	S	agt	T
NS5B	321	2741	V	gtg	2751	V	gtg	A

Table A.1: **Positions, amino acids and codons in H77 reference and S52-△N replicon sequences.**

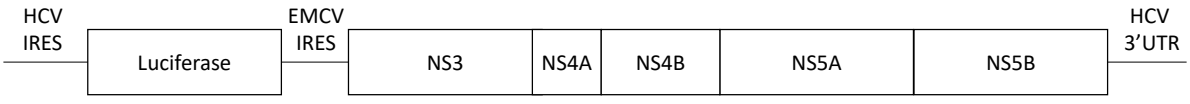

Figure A.1: **Schematic of S52-△N replicon.** Not to Scale

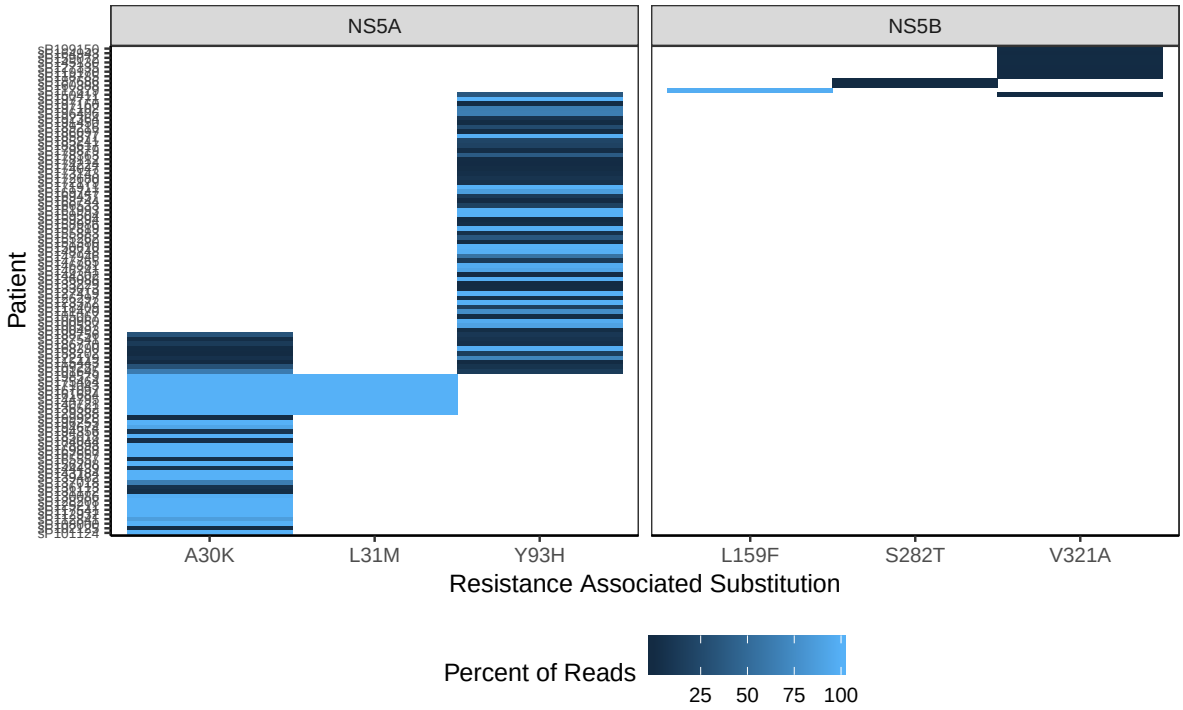

Figure A.2: **Prevalence of substitutions not associated with resistance at resistance associated substitution sites.**

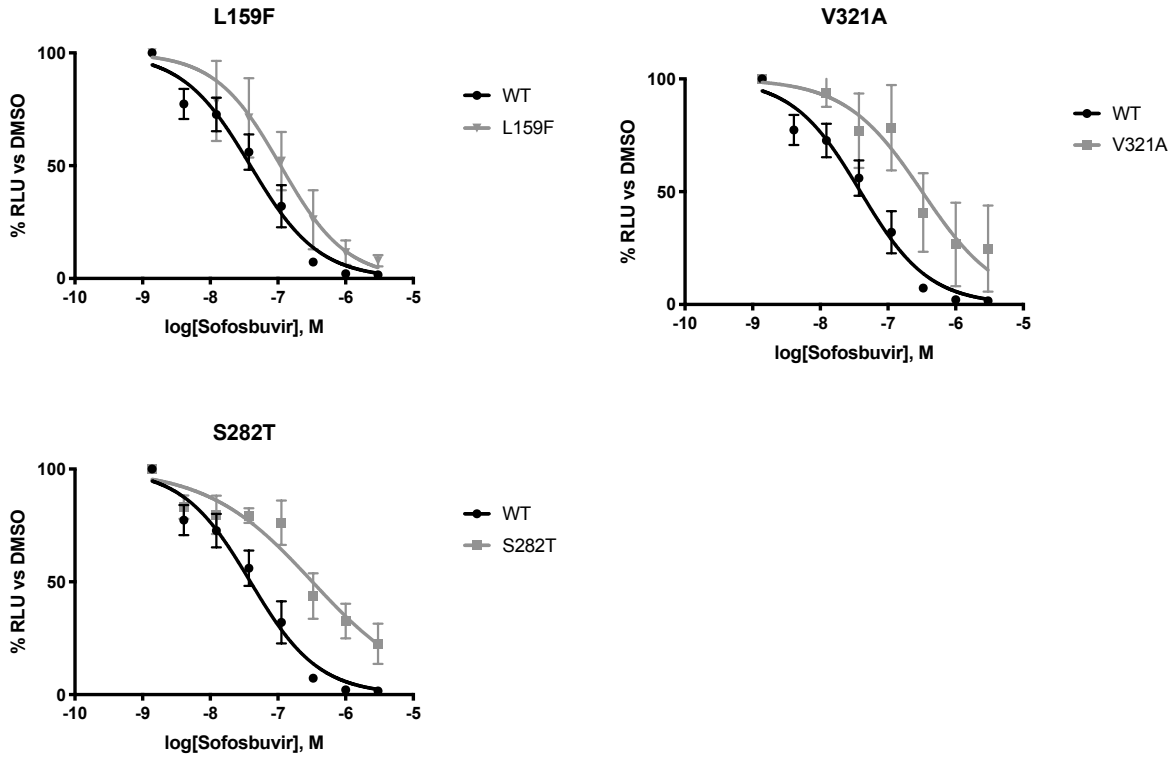

Figure A.3: **Dose-response profiles of individual RAS to sofosbuvir.**

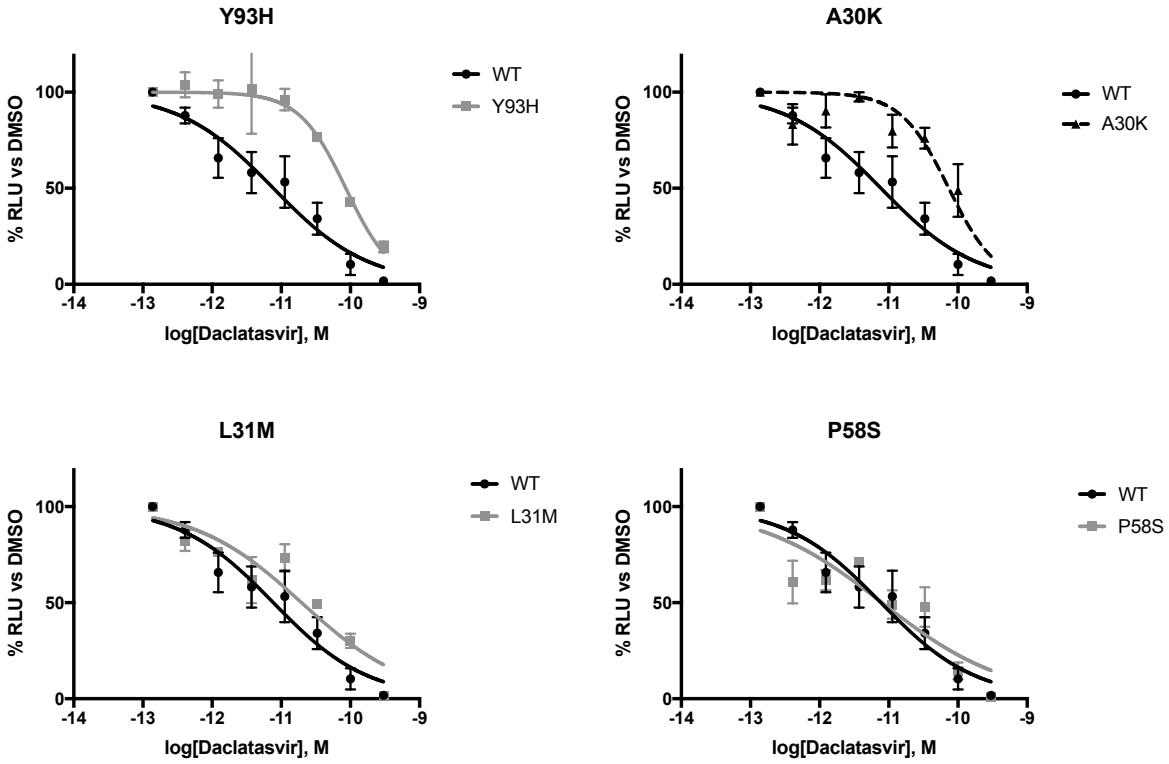

Figure A.4: **Dose-response profiles of individual RAS to daclatasvir.**

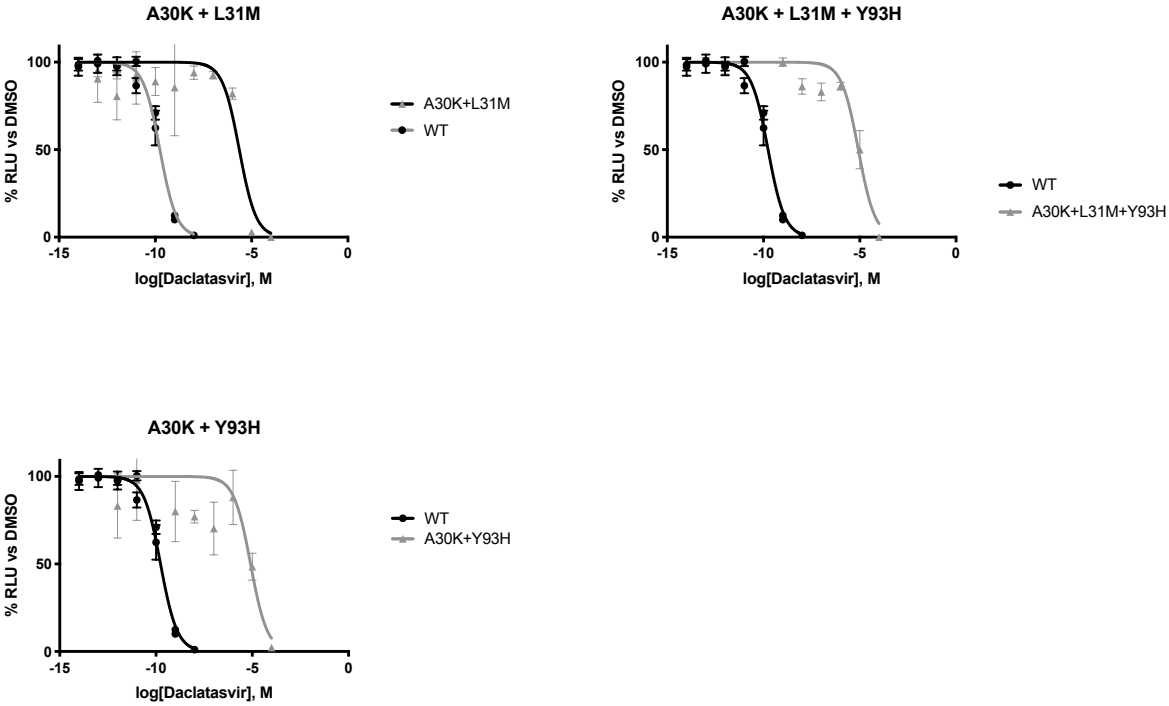

Figure A.5: **Dose-response profiles of RAS combinations to daclatasvir.**

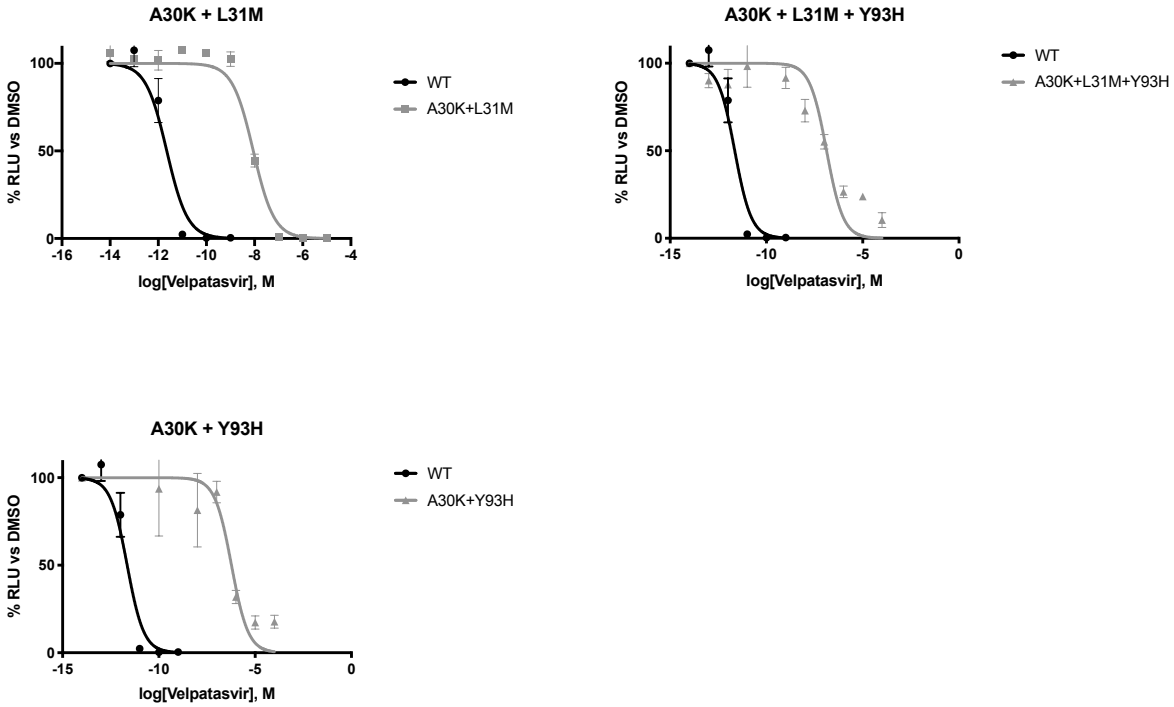

Figure A.6: **Dose-response profiles of RAS combinations to velpatasvir.**

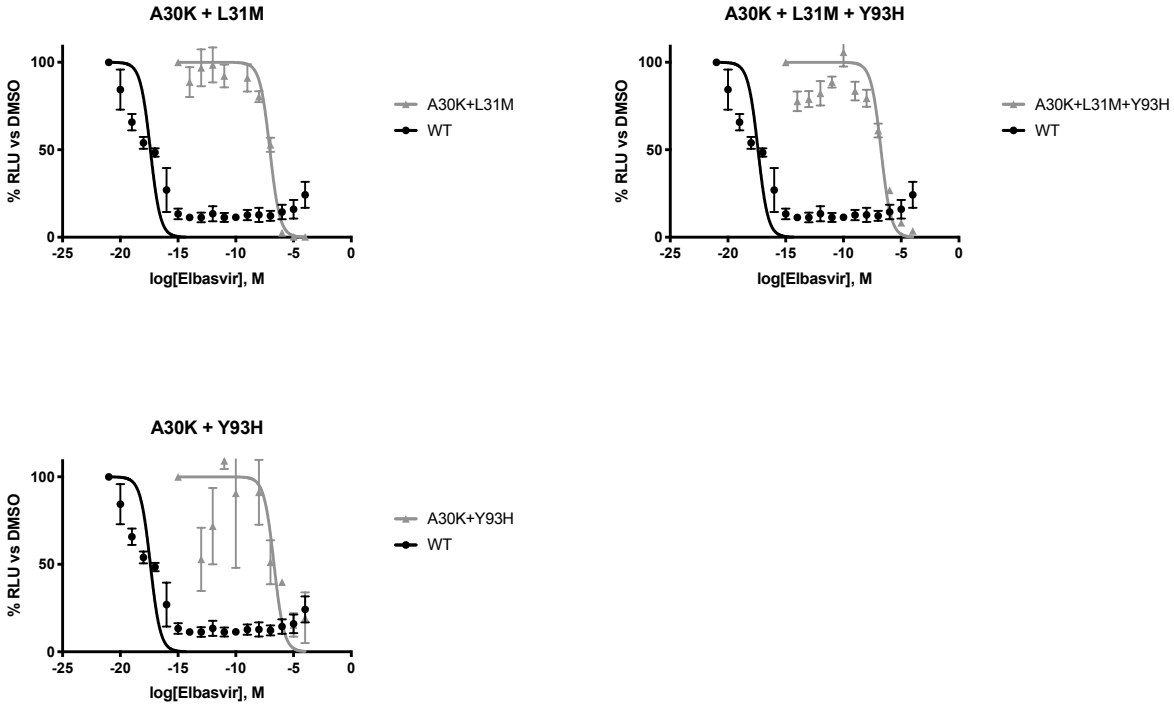

Figure A.7: **Dose-response profiles of RAS combinations to elbasvir.**

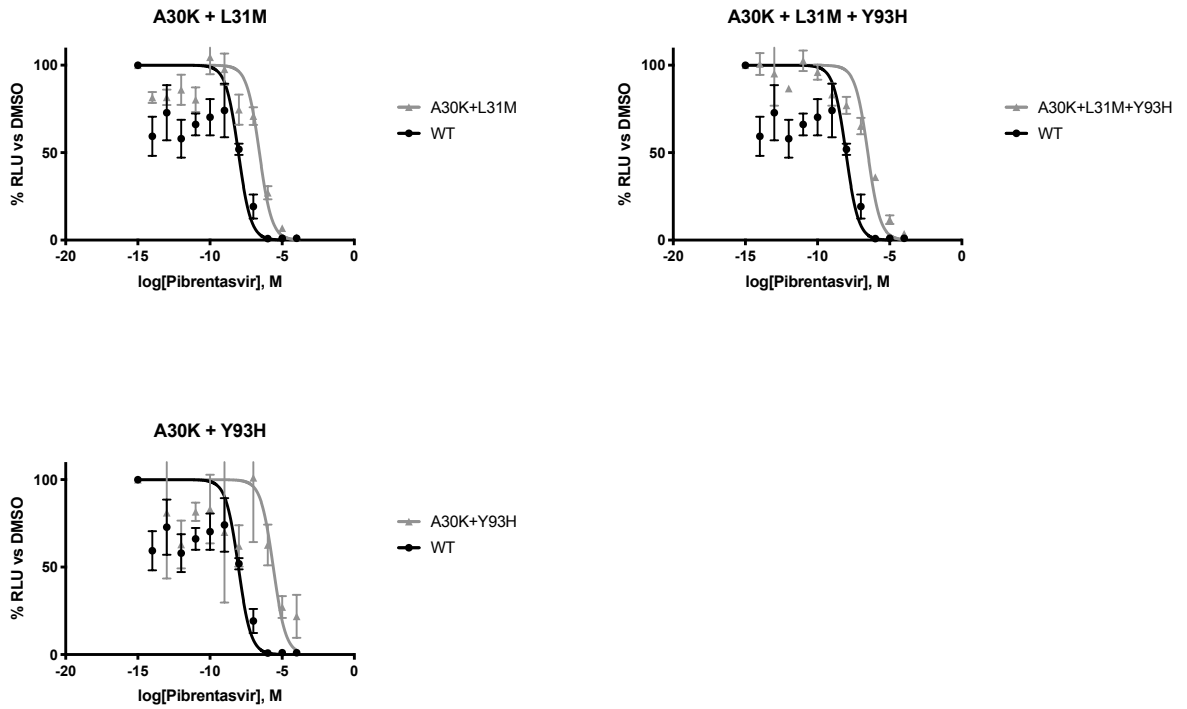

Figure A.8: **Dose-response profiles of RAS combinations to pibrentasvir.**

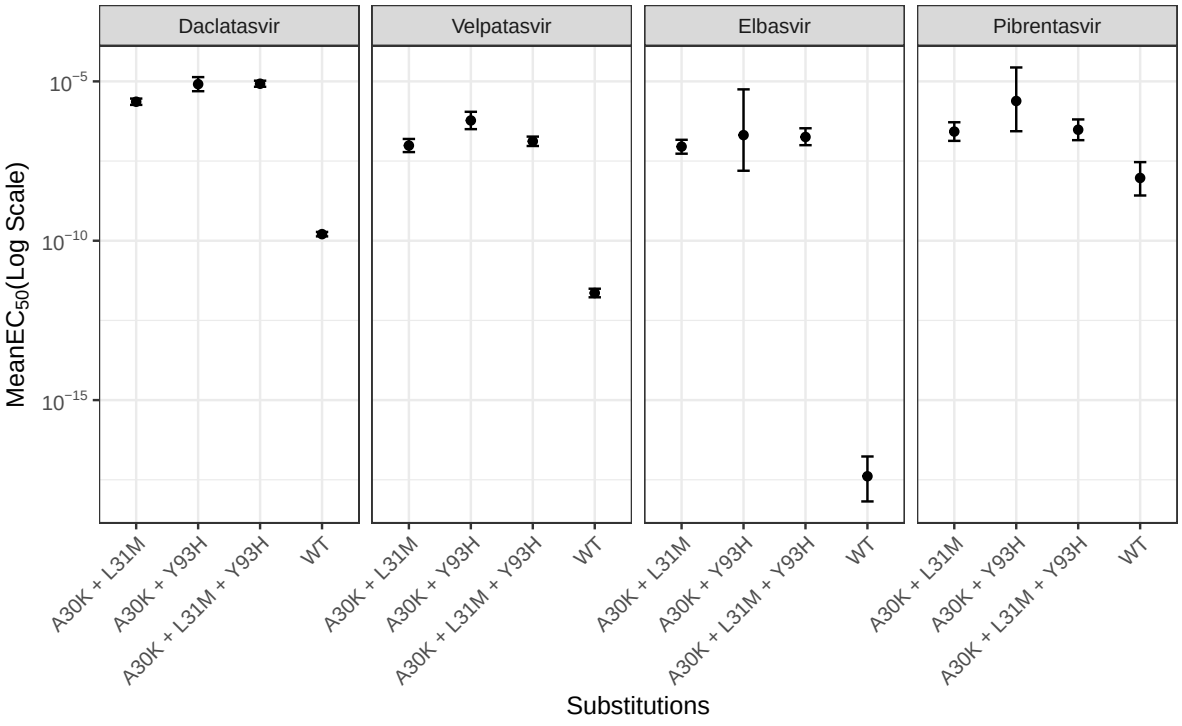

Figure A.9: **IC50 values for daclatasvir, velpatasvir, elbasvir or pibrentasvir against WT S52-∆N replicon and S52-∆N carrying candidate RAS combinations.**

Appendix B

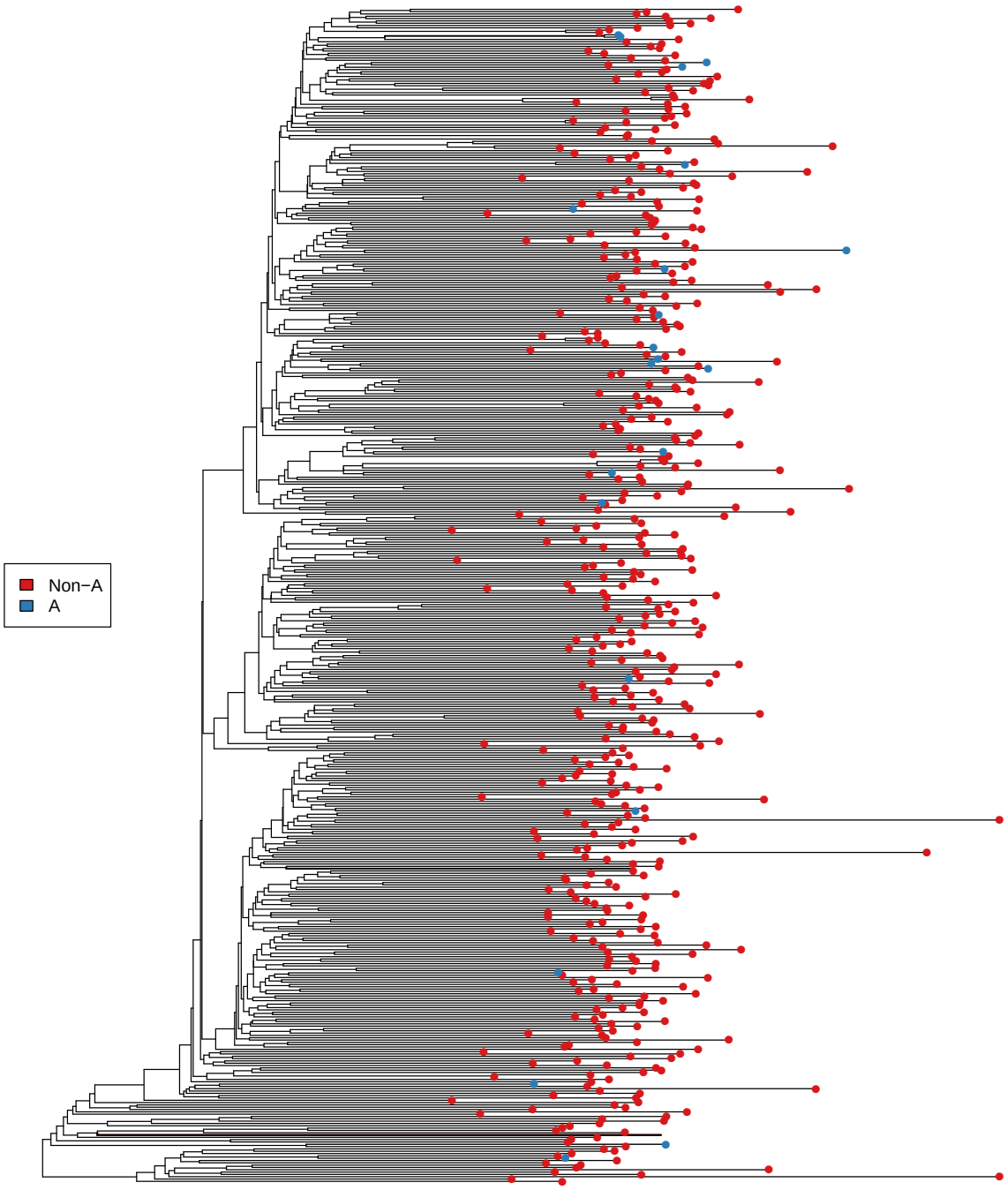

Figure B.1: **A maximum likelihood phylogenetic tree of 507 gt3a HCV sequence with the presence or absence of the putative RAS NS2 119A shown on the tree tips.** The change point for any clade associated with SVR is coloured red and the thickness pf the line shows the significance of that association. The phylogenetic tree was constructed using RAxML [40] and the change point analysis conducted using treeBreaker [162].

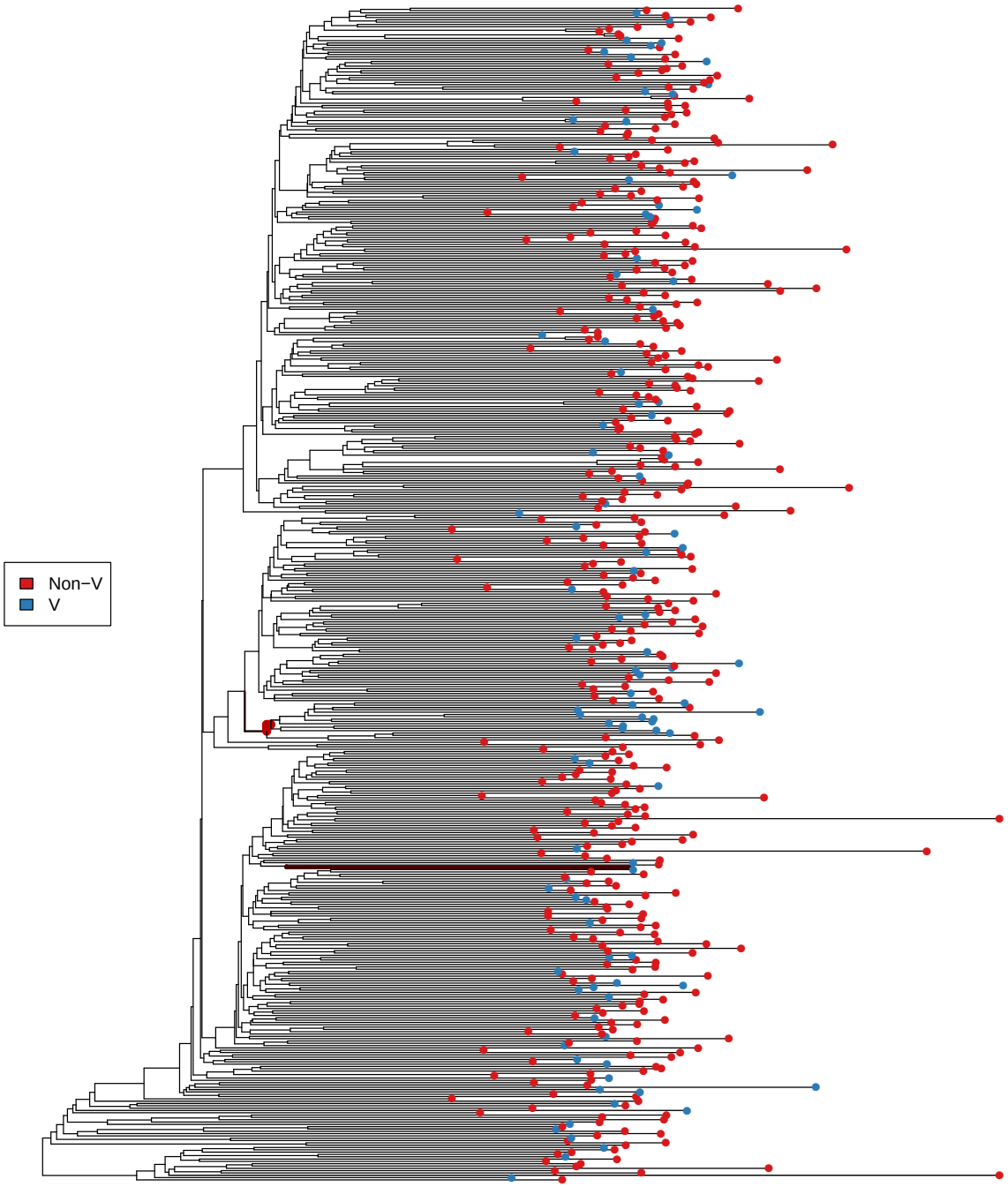

Figure B.2: **A maximum likelihood phylogenetic tree of 507 gt3a HCV sequence with the presence or absence of the putative RAS NS2 132V shown on the tree tips.** The change point for any clade associated with SVR is coloured red and the thickness pf the line shows the significance of that association. The phylogenetic tree was constructed using RAxML [40] and the change point analysis conducted using treeBreaker [162].

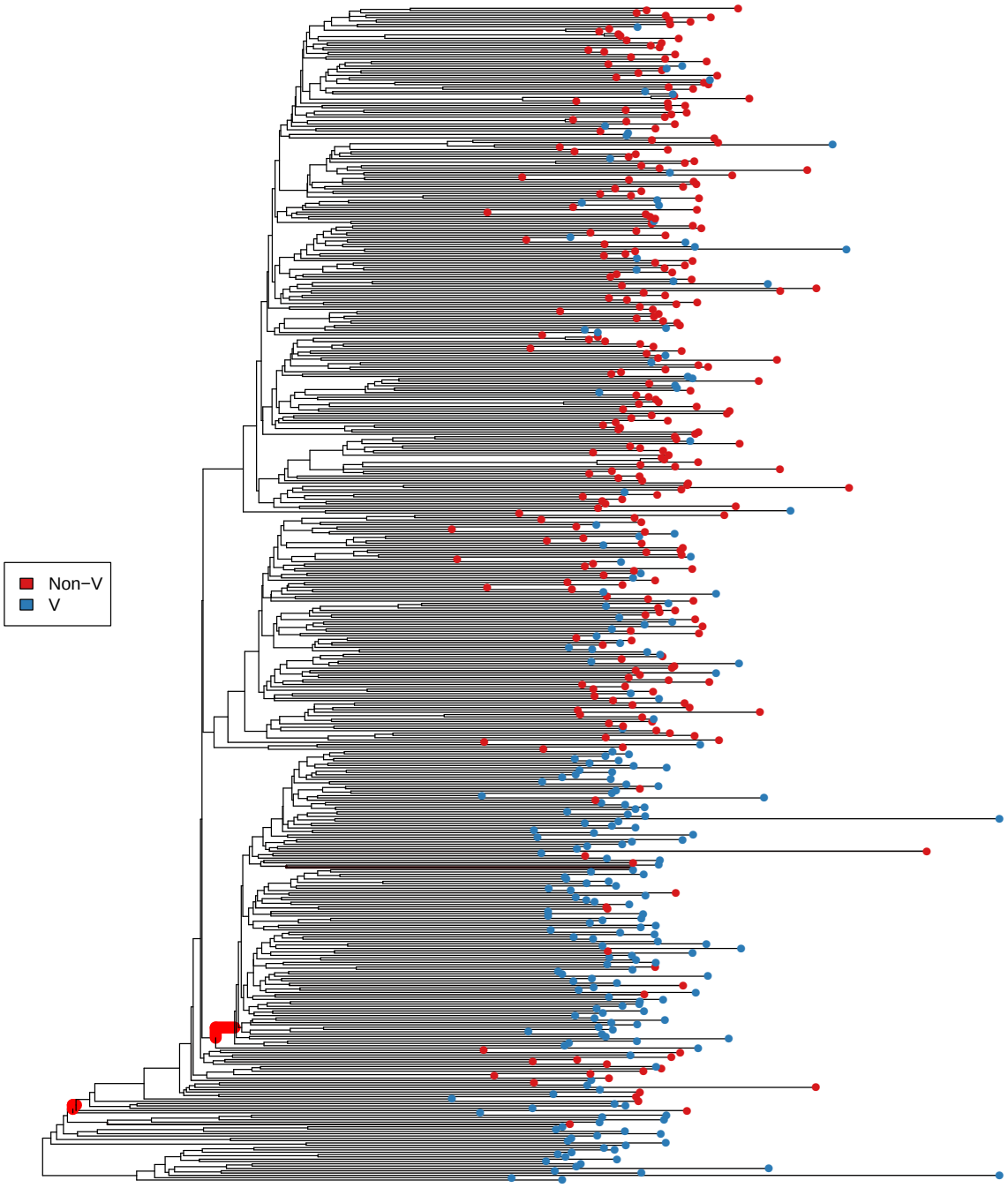

Figure B.3: **A maximum likelihood phylogenetic tree of 507 gt3a HCV sequence with the presence or absence of the putative RAS NS3 67V shown on the tree tips.** The change point for any clade associated with SVR is coloured red and the thickness pf the line shows the significance of that association. The phylogenetic tree was constructed using RAxML [40] and the change point analysis conducted using treeBreaker [162].

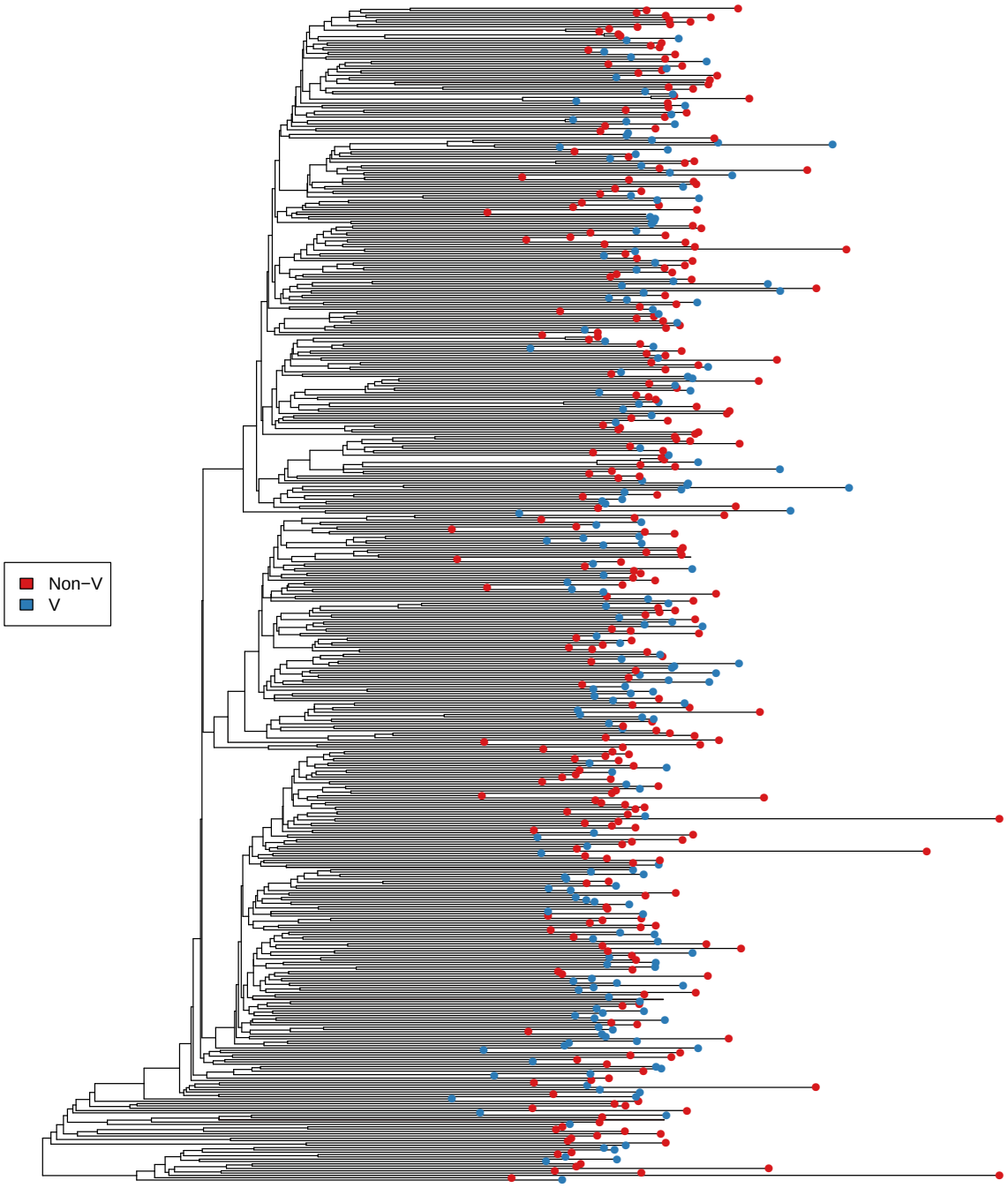

Figure B.4: **A maximum likelihood phylogenetic tree of 507 gt3a HCV sequence with the presence or absence of the putative RAS NS5B 150V shown on the tree tips.** The change point for any clade associated with SVR is coloured red and the thickness pf the line shows the significance of that association. The phylogenetic tree was constructed using RAxML [40] and the change point analysis conducted using treeBreaker [162].

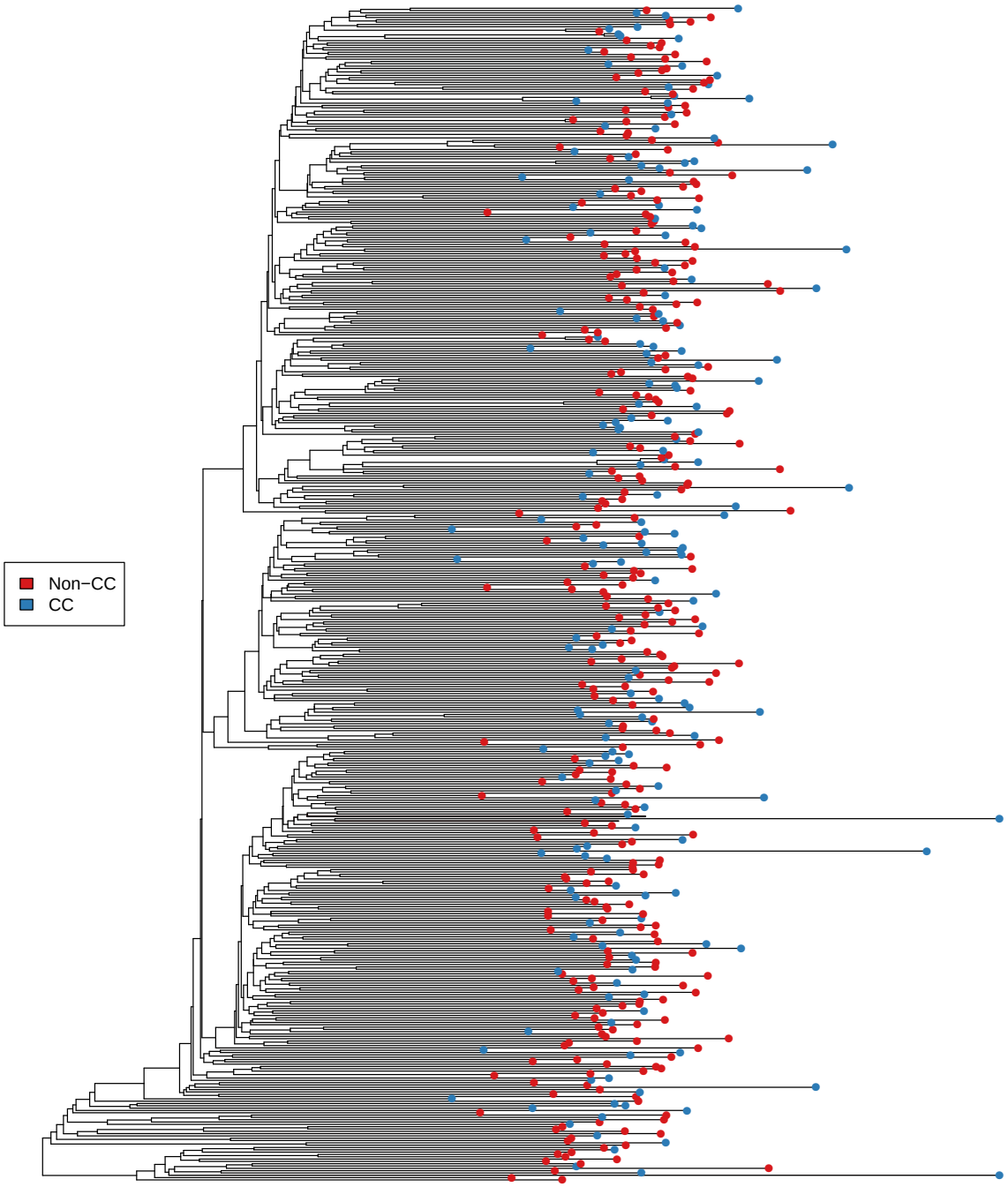

Figure B.5: **A maximum likelihood phylogenetic tree of 507 gt3a HCV sequence with the IFNL4 C/C or non-C/C status of each infected patient shown.** The change point for any clade associated with SVR is coloured red and the thickness pf the line shows the significance of that association. The phylogenetic tree was constructed using RAxML [40] and the change point analysis conducted using treeBreaker [162].

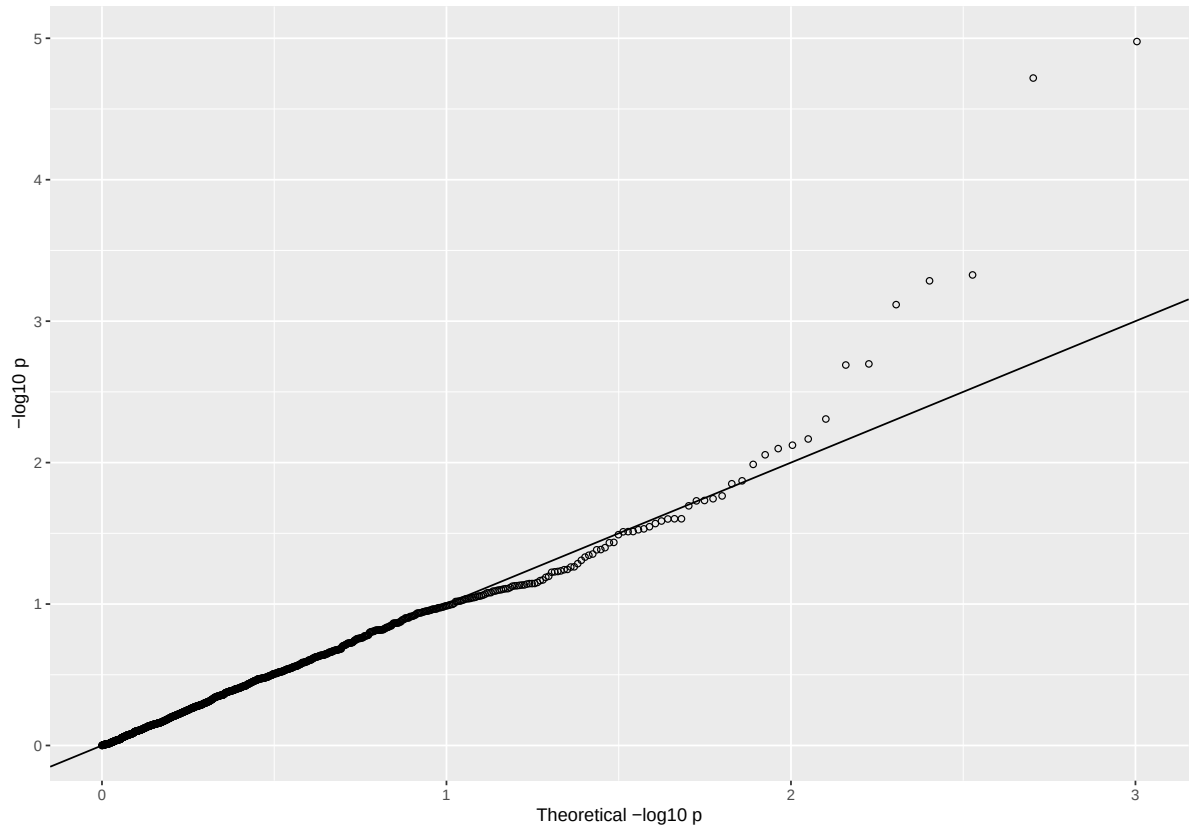

Figure B.6: **Q-Q plot for viral genome wide association study between viral amino acids and treatment outcome.**

Protein/Region	P value	Log Odds Ratio	Standard Error
Polyprotien	0.594	13.785	25.869
Core	0.922	-2.770	28.337
E1	0.448	11.959	15.756
E2	0.165	16.591	11.953
P7	0.947	0.963	14.416
NS2	0.913	1.897	17.457
NS3	0.940	2.601	34.556
NS4A	0.489	-19.246	27.815
NS4B	0.874	4.660	29.333
NS5A	0.863	3.675	21.344
NS5B	0.699	13.071	33.810
HVR1	0.277	2.950	2.711
HVR2	0.590	2.635	4.887
HVR3	0.862	1.169	6.735

Table B.1: **Associations between mean normalised amino acid shannon entropy of the HCV polyprotein, protein and regions of interest and treatment outcome (SVR12).** The mean normalised Shannon entropy of the HCV polyprotein, protein and regions of interest was calculated for each patient. Logistic regression was used to test with SVR as the dependent variable and mean normalised Shannon entropy as the explanatory variable and IFNL4 genotype as a cofactor.

Position in polyprotein relative to H77	Protein	Position in Protein	Log odds ratio	Standard Error	P value
Previously characterised RAS sites					
2579	NS5B	159	-43.7713	42.6427	0.3047
2702	NS5B	282	236.0770	616.1905	0.7016
2736	NS5B	316	95.3591	105.9016	0.3679
2740	NS5B	320	0.2431	16.3641	0.9881
2741	NS5B	321	21.8900	39.6323	0.5807
Newly Identified RAS Sites					
941	NS2	132	-1.2990	1.8853	0.4908
1093	NS3	67	-1.4040	1.0603	0.1854
928	NS2	119	-1.4817	1.3453	0.2707
2570	NS5B	150	-0.8093	1.0997	0.4618

Table B.2: **Associations between mean normalised Shannon entropy of the RAS sites and treatment outcome (SVR12).** The normalised Shannon entropy of each RAS site was calculated for each patient. Logistic regression was used to test with SVR as the dependent variable and normalised Shannon entropy as the explanatory variable and IFNL4 genotype as a cofactor to investigate the association.

Gene/Region	P value	Log Odds Ratio	Standard Error
Genome	0.94	0.87	12.31
UTR_5	0.17	53.87	39.11
C	0.83	3.32	15.10
E1	0.75	3.28	10.30
E2	0.83	1.81	8.66
P7	0.91	-1.03	8.96
NS2	0.76	-3.09	9.92
NS3	0.96	-0.56	12.11
NS4A	0.61	-5.66	11.24
NS4B	1.00	-0.04	11.50
NS5A	0.98	-0.29	10.91
NS5B	0.85	3.04	15.67
UTR_3	0.40	11.66	13.99
HVR1	0.59	1.90	3.53
HVR2	0.49	3.32	4.75
HVR3	0.76	2.21	7.23

Table B.3: **Associations between mean pairwise diversity of the HCV genome, genes and regions of interest and treatment outcome (SVR12).** The mean pairwise diversity of the HCV genome, genes and regions of interest was calculated for each patient. Logistic regression was used to test with SVR as the dependent variable and mean normalised pairwise diversity as the explanatory variable and IFNL4 genotype, cirrhosis sate, gender, prior treatment and viral and the first 3 principle components of the viral genome and host genotyping as a cofactors.

Position in HCV genome relative to H77	Gene	Position in gene	Log odds ratio	Standard Error	P value
Previously characterised RAS sites					
7737	NS5B	477	-3.55	3.49	0.31
8106	NS5B	846	1.30	9.58	0.89
8208	NS5B	948	1.21	2.03	0.55
8220	NS5B	960	12.68	7.70	0.10
8223	NS5B	963	0.79	5.42	0.88
Newly Identified RAS Sites					
2823	NS2	396	-0.45	2.69	0.87
3279	NS3	201	-0.87	3.64	0.81
2784	NS2	357	5.08	3.44	0.14
7710	NS5B	450	-1.62	8.24	0.84

Table B.4: **Associations between mean pairwise diversity of the RAS sites and treatment outcome (SVR12).** The mean pairwise diversity of each RAS site was calculated for each patient. Logistic regression was used to test with SVR as the dependent variable and mean normalised pairwise diversity as the explanatory variable and IFNL4 genotype, cirrhosis sate, gender, prior treatment and viral and the first 3 principle components of the viral genome and host genotyping as a cofactors.

Position in polyprotien relative to H77	Protien	Position and Amino Acid in Protien	P Value	Q Value
2579	NS5B	159F	1.47E-05	0.08
583	E2	206A	4.16E-04	1
2570	NS5B	150V	8.77E-04	1
1637	NS3	611L	1.05E-03	1
774	P7	28T	3.16E-03	1
1975	NS5A	3E	3.32E-03	1
2510	NS5B	90T	8.86E-03	1
978	NS2	169S	9.11E-03	1
2452	NS5B	32K	9.18E-03	1
2797	NS5B	377R	9.27E-03	1

Table B.5: **Top 10 results from BL to 12WPT Enrichment GWAS.**

Position in Polyprotein Relative to H77	Protein	Position and Amino Acid in Protein	Log OR	SE	P Value	Q Value
928	NS2	119A	-0.434	0.132	0.001	0.616
2487	NS5B	67A	0.284	0.089	0.001	0.616
1282	NS3	256N	-0.375	0.120	0.002	0.616
714	E2	337I	0.338	0.116	0.004	0.637
941	NS2	132V	-0.195	0.068	0.004	0.637
2487	NS5B	67V	-0.279	0.097	0.004	0.637
2034	NS5A	62S	0.169	0.059	0.004	0.637
941	NS2	132I	0.189	0.068	0.005	0.676
49	Core	49T	-0.351	0.129	0.007	0.717
182	Core	182L	-0.203	0.076	0.008	0.717

Table B.6: **Top 10 results from Wk1 Response to Therapy GWAS.**

Appendix C

Patient Number	Treatment	HCV Subtype	Treatment Length	Outcome	Presence of Cirrhosis	Decompensated Cirrhosis	Previous Hepatocellular Carcinoma	Previous Liver Transplant	Previous Treatment	Presence of RAS
1	SOF/VEL/VOX	3a	12 Weeks	Relapse	Yes	No	No	No	GLE/PIB	NS5A + NS5B
2	SOF/VEL/VOX	3b	12 Weeks	Relapse	Yes	No	No	Yes	SOF/VEL	None
3	SOF/VEL/VOX	1b	12 Weeks	Relapse	Yes	No	Yes	No	SOF/VEL	NS5A + NS5B
4	SOF/VEL/VOX	3a	8 Weeks	Non-Responder	No	No	No	No	SOF + DAC + RBV	NS5A
5	SOF/VEL/VOX	3a	12 Weeks	Relapse	Yes	No	No	No	DCV + SOF + RBV	NS5A + NS5B
6	SOF/VEL/VOX	3a	12 Weeks	Relapse	No	No	No	No	DCV + SOF + RBV	NS5A + NS5B
7	SOF/VEL/VOX	3a	$<$8 Weeks	Reason Not Recorded	Yes	No	No	No	SOF/VEL	NS5B
8	SOF/VEL/VOX	3a	24 Weeks	Breakthrough	Yes	Yes	Yes	No	SOF/VEL	None
9	SOF/VEL/VOX	3a	24 Weeks	Relapse	Yes	No	No	Yes	DCV + SOF	NS5A
10	SOF/VEL/VOX	3a	12 Weeks	Relapse	Yes	No	No	No	DCV + SOF + RBV	NS5A + NS5B
11	SOF/VEL/VOX	3a	8 Weeks	Non-Responder	No	No	No	No	GLE/PIB	NS5A
12	SOF/VEL/VOX	4r	12 Weeks	Relapse	Yes	No	No	No	SOF/LDV	NS5B
13	SOF/VEL/VOX	3a	12 Weeks	Relapse	Yes	No	No	Yes	SOF/VEL	NS5A + NS5B
14	SOF/VEL/VOX	3k	16 Weeks	Relapse	Yes	No	Yes	No	GLE/PIB	NS5A
15	SOF/VEL/VOX	1a	12 Weeks	Reason Not Recorded	No	No	No	No	OBT/PTVr + DAS	None
16	SOF/VEL + RBV	2b	24 Weeks	Relapse	Yes	No	No	No	SOF + RBV	NS5A
17	GLE/PIB	1a	12 Weeks	Relapse	Yes	No	No	No	SOF/LDV	NS3 + NS5A

Table C.1: **Clinical characteristics of the 17 individual patients who were failed by retreatment.** Daclatasvir (DCV), Sofosbuvir (SOF), Glecaprevir (GLE), Pibrentasvir (PIB), Grazoprevir (GZR), Elbasvir(EBR), Ledipasvir(LDV), Paritaprevir (PTV), Ombitasvir (OBV) Ritonavir (r), Dasabuvir (DAS), Velpatasvir (VEL).

Protein	RAS	SVR Rate	p	log Odds Ratio	Standard Error
NS3	V55A	100 (4/4)	0.990	14.446	1199.772
NS3	K/Q80K	100 (14/14)	0.992	16.331	1638.472
NS3	K/Q80L	100 (2/2)	0.993	14.430	1696.734
NS3	Q80R	100 (1/1)	0.993	13.422	1455.398
NS3	A156T	100 (1/1)	0.993	13.422	1455.398
NS3	L175M	100 (1/1)	0.993	13.422	1455.398
NS5A	M28T	100 (5/5)	0.993	15.454	1769.258
NS5A	A30K	71 (10/14)	0.029	-1.465	0.670
NS5A	K/R30K	100 (1/1)	0.993	13.422	1455.398
NS5A	L/M31M	100 (2/2)	0.993	14.430	1696.734
NS5A	L31M	100 (16/16)	0.992	16.373	1543.764
NS5A	L31V	100 (3/3)	0.992	14.438	1385.378
NS5A	H58D	100 (9/9)	0.990	15.183	1201.731
NS5A	Y93C	100 (4/4)	0.990	14.446	1199.772
NS5A	Y93H	78 (29/37)	0.014	-1.371	0.559
NS5A	Y93N	100 (2/2)	0.993	14.430	1696.734
NS5B	A/T/V150V	93 (26/28)	0.531	0.495	0.791
NS5B	F/L159F	67 (2/3)	0.229	-1.512	1.257
NS5B	L159F	100 (2/2)	0.993	14.430	1696.734
NS5B	E/K206E	80 (16/20)	0.142	-0.943	0.642
NS5B	E237G	100 (1/1)	0.993	13.422	1455.398
NS5B	D/N244I	100 (1/1)	0.993	13.422	1455.398
NS5B	C/N316H	100 (1/1)	0.993	13.422	1455.398
NS5B	C/N316N	75 (3/4)	0.355	-1.099	1.189
NS5B	I/V321I	50 (1/2)	0.125	-2.213	1.442
NS5B	V321I	100 (1/1)	0.993	13.422	1455.398

Table C.2: **SVR rates for 128 patients retreated with SOF/VEL/VOX with RAS.** Logistic regression was used to test RAS association with outcome