Mixed Genotype Hepatitis C Virus Infections: Incidence in Scotland and Methods for Detection

Anna Lukats McNaughton

Submitted in fulfilment of the requirements for the Degree of
Doctor of Philosophy in Virology

Institute of Infection Immunity & Inflammation
College of Medical, Veterinary and Life Sciences
University of Glasgow

September, 2016

Abstract

The diagnosis of mixed genotype hepatitis C virus (HCV) infection is rare and information on incidence in the UK, where genotypes 1a and 3 are the most prevalent, is sparse. Considerable variations in the efficacies of direct-acting antivirals (DAAs) for the HCV genotypes have been documented and the ability of DAAs to treat mixed genotype HCV infections remains unclear, with the possibility that genotype switching may occur.

In order to estimate the prevalence of mixed genotype 1a/3 infections in Scotland, a cohort of 512 samples was compiled and then screened using a genotype-specific nested PCR assay. Mixed genotype 1a/3 infections were found in 3.8% of samples tested, with a significantly higher prevalence rate of 6.7% ($p<0.05$) observed in individuals diagnosed with genotype 3 infections than genotype 1a (0.8%). An analysis of the samples using genotypic-specific qPCR assays found that in two-thirds of samples tested, the minor strain contributed <1% of the total viral load. The potential of deep sequencing methods for the diagnosis of mixed genotype infections was assessed using two pan-genotypic PCR assays compatible with the Illumina MiSeq platform that were developed targeting the E1-E2 and NS5B regions of the virus. The E1-E2 assay detected 75% of the mixed genotype infections, proving to be more sensitive than the NS5B assay which identified only 25% of the mixed infections. Studies of sequence data and linked patient records also identified significantly more neurological disorders in genotype 3 patients. Evidence of distinctive dinucleotide expression within the genotypes was also uncovered. Taken together these findings raise interesting questions about the evolutionary history of the virus and indicate that there is still more to understand about the different genotypes.

In an era where clinical medicine is frequently more personalised, the development of diagnostic methods for HCV providing increased patient stratification is increasingly important. This project has shown that sequence-based genotyping methods can be highly discriminatory and informative, and their use should be encouraged in diagnostic laboratories. Mixed genotype infections were challenging to identify and current deep sequencing methods were not as sensitive or cost-effective as Sanger-based approaches in this study. More research is needed to evaluate the clinical prognosis of patients with mixed genotype infection and to develop clinical guidelines on their treatment.

Table of Contents

List of Tables

List of Figures

Acknowledgement

I would like to express my gratitude to the many colleagues, friends and family members who have provided assistance and support throughout this project. Those of you who know me well are aware that parts of this project were challenging for me and your continued encouragement has been greatly appreciated during my PhD.

Particular thanks are extended to –

Carol Leitch
John McLauchlan
Gavin Wilkie
Sreenu Vattipally
Ana Filipe
Rory Gunson
Members of the Thomson lab group
Members of the McLauchlan lab group
Residents of the Church St PhD office
Natalie Hutchinson
Gail Anderson
My flatmates
The Bush family
Gus McNaughton
Linda Lukats
Archie
Sadie McNaughton

Declaration

I, Anna Lukats McNaughton, declare that, except where explicit reference is made to the contribution of others, this thesis is the result of my own work and has not been submitted for any other degree at the University of Glasgow or any other institution.

Signature _____

Printed name _____

Abbreviations

3’UTR	3 prime untranslated region of HCV genome
5’UTR	5 prime untranslated region of HCV genome
ALT	Alanine transaminase
AST	Aspartate transaminase
bp	Base pair (of nucleotides)
cDNA	Complementary deoxyribonucleic acid
CpG	C-G dinucleotide
CVR	Centre for Virus Research
DAA	Direct-acting antiviral
dH_2O	Distilled water
DNA	Deoxyribonucleic acid
dsDNA	Double stranded deoxyribonucleic acid
dsRNA	Double stranded ribonucleic acid
E1	Envelope protein 1
E1-E2	Region spanning end of E1 and part of E2
E2	Envelope protein 2
EBNA-1	Epstein–Barr nuclear antigen 1
EIA	Enzyme immuno-assay
gt	Genotype
H77	Hepatitis c virus strain H77
HAART	Highly active anti-retroviral therapy
HLA	Human leukocyte antigen
IDU	Injecting drug use
IFN- β	Interferon- β
IFN-α	Interferon-α
IFN-λ3	Interferon- λ3
Il-1β	Interleukin-1β
IRES	Internal ribosome entry site
ISGs	Interferon stimulated genes

JAK-STAT	Janus kinase/signal transducers and activators of transcription pathway
LiPA	Line probe assay
MAVS	Mitochondrial antiviral signalling protein
min	Minute
MRC	Medical Research Council
MSM	Men who have sex with men
NS2	Non-structural protein 2
NS3	Non-structural protein 3
NS4A	Non-structural protein 4A
NS4B	Non-structural protein 4B
NS5A	Non-structural protein 5A
NS5B	Non-structural protein 5B
nt	Nucleotide
O/E	Observed/Expected ratio
PC	Postcode
PCR	Polymerase chain reaction
PEG-IFN-α	Pegylated interferon-α
PKR	Protein kinase R
PWID	People who inject drugs
qPCR	Quantitative polymerase chain reaction
RdRp	RNA-dependent RNA polymerase
RFLP	Restriction fragment length polymorphism
RIG-I	Retinoic acid-inducible gene 1
RNA	Ribonucleic acid
RSCU	Relative Synonymous Codon Use
RT-PCR	Reverse transcriptase polymerase chain reaction
TLR-3	Toll-like receptor
Tm	Melting temperature
TNF- α	Tumour-necrosis factor-α
TRIF	TIR-domain-containing adapter-inducing interferon-β
UpA	U-A dinucleotide
WSSVC	West of Scotland Specialist Virology Centre

Virus Abbreviations

CMV	Cytolomegavirus
DENV	Dengue virus
EBV	Epstein-Barr virus
HAV	Hepatitis A virus
HBV	Hepatitis B virus
HCV	Hepatitis C virus
HEV	Hepatitis E virus
HIV	Human immunodeficiency virus
TBEV	Tick-Borne Encephalitis virus
YFV	Yellow Fever virus

Chapter 1: Introduction

1.1 Hepatitis C virus

1.1.1 The discovery of the hepatitis C virus

Awareness of a transmissible hepatitis, distinct from both hepatitis A virus (HAV) and hepatitis B virus (HBV) dates back to the mid-1970s (Prince *et al.*, 1974). Later studies, including a prospective cohort of multiply-infused heart transplant patients in 1975 (Alter *et al.*, 1975) and a back-dated study re-analysing blood donated in the 1950s (Hoofnagle *et al.*, 1977) identified a number of transfused samples that were serologically negative for both HAV and HBV but were icterogenic when transfused into patients. Among individuals developing post-transfusion hepatitis, approximately 10% were shown to be serologically negative for HAV and HBV, implying there was an alternative cause (Aach *et al.*, 1981). The causative agent was designated non-A non-B hepatitis (NANBH) and a natural history of NANBH infection was defined (Purcell, Alter and Dienstag, 1976). Distinct histological findings were described for the liver and transmission from infected sera was demonstrated in a chimpanzee animal model (Hollinger *et al.*, 1978; Tabor *et al.*, 1978). In addition to well-described iatrogenic transmissions (Alter *et al.*, 1975)(Hoofnagle *et al.*, 1977), outbreaks of NANBH were also identified occurring intermittently within the community (Holland and Alter, 1981).

NANBH was demonstrated to be infectious when transfused from both acute and chronically infected individuals, implying that the causative pathogen was able to establish a chronic carrier state and remained infectious throughout the duration of the infection (Alter *et al.*, 1978). Treatment of infected sera with chloroform prior to inoculation was sufficient to prevent NANBH transmission and implicated a lipid-soluble pathogen (Feinstone *et al.*, 1983). Filtration of infected sera from chimpanzees through polycarbonate membranes of various sizes indicated that the causative agent was likely to be 30-60nm in diameter, suggesting that a small enveloped virus was the most likely causative agent of NANBH (He *et al.*, 1987). Despite accumulating evidence implicating a

viral causative agent for NANBH, researchers struggled to identify any antigen, antibody or nucleic acid linked to the disease or to propagate an NANBH virus via cell culture for more than a decade after initial identification.

A cDNA clone of the virus was eventually isolated from chimpanzees infected with human NANBH positive sera by Michael Houghton (Chiron Corporation) and Daniel Bradley (CDC) and their pioneering work was published in 1989 (Choo *et al.*, 1989). Immunological assays based on the detection of anti-NANBH using a small yeast-expressed recombinant part of the NANBH genome were developed in 1989 and were used demonstrate conclusively that the virus detected was the causative agent associated with NANBH (Kuo *et al.*, 1989). The virus was named hepatitis C virus shortly afterwards and was established to be distantly related to other known *flaviviruses* and *pestiviruses* (Miller and Purcell, 1990). In the years following these discoveries, HCV was demonstrated to be the cause of most community acquired NAHBH in the developed world (Choo *et al.*, 1990) and was linked to large numbers of hepatocellular carcinoma cases worldwide (Colombo *et al.*, 1989).

1.1.2 Classification of the hepatitis C virus

Within the Baltimore Classification System (Baltimore, 1971), HCV is classified as group IV, alongside other viruses with positive sense single-stranded RNA genomes. Group IV is the largest group of RNA viruses, including 30 distinct families. All viruses within the group are able to replicate in the cytoplasm and are able to interact directly with ribosomes in the cell to trigger protein synthesis.

Taxonomically, on the basis of its genomic structure, HCV is classified as a *hepacivirus*, within the *Flaviviridae* family (Lindenbach and Rice, 2007). The family is split into 4 genera, *flaviviruses*, *pestiviruses*, *pegiviruses* and *hepaciviruses* (figure 1-1). This family of viruses are all small, enveloped, positive-sense, single-stranded RNA viruses that share similar replication strategies and common morphological and genomic features. All the *Flaviviridae* replicate in the cytoplasm and their genomes range in size from 9.6-12kb (Lindenbach and Rice, 2007). The viral proteins are translated directly from their RNA genomes as long polyproteins which are cleaved by both host and viral proteases (Lindenbach and Rice, 2007). The viral capsids are made from a single core protein (C) and the viruses encode 2-3 membrane-bound proteins that constitute the envelope

(Lindenbach and Rice, 2007). The viruses exhibit a diverse range of biological properties and are able to infect a wide range of both vertebrate and invertebrate hosts via a range of different transmission routes (Lindenbach and Rice, 2007). There are a number of important medically and veterinary pathogens within the family.

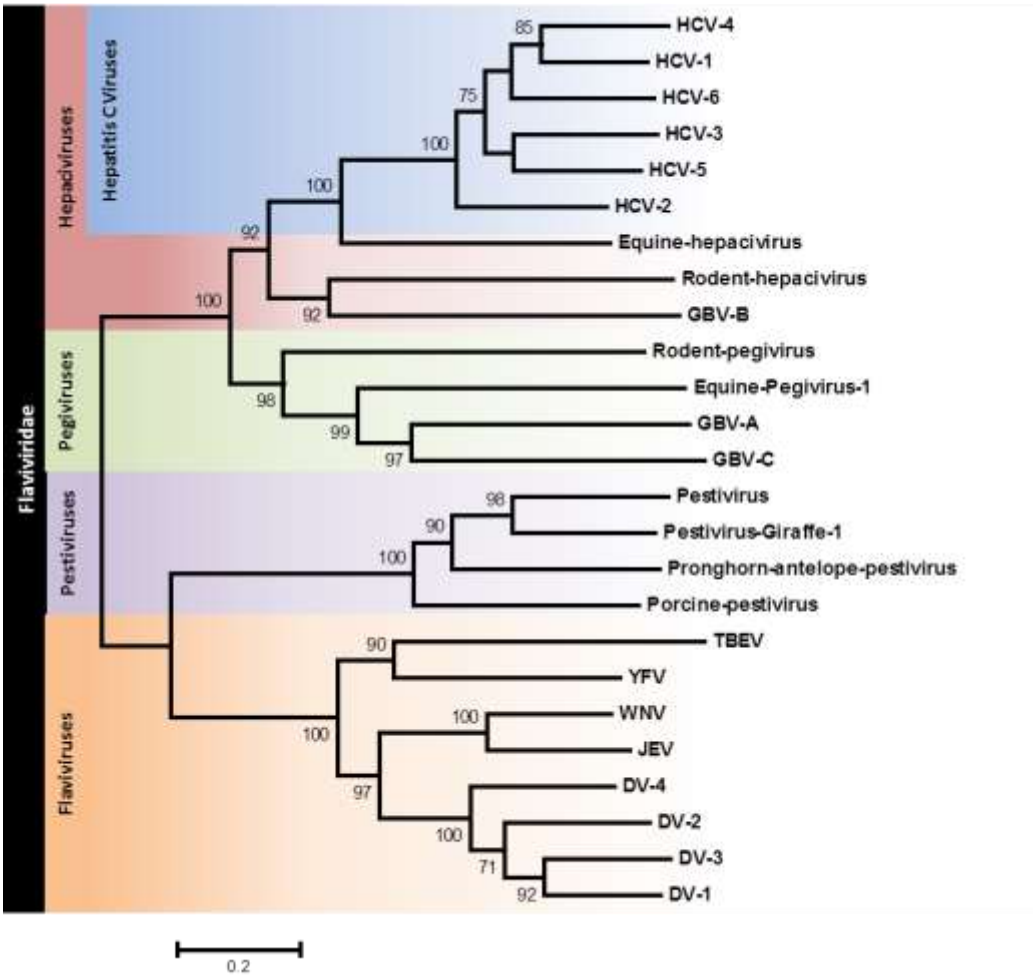


Figure 1-1 The *Flaviviridae*

A maximum likelihood phylogenetic tree of full-length reference *Flaviviridae* sequences showing the major genera within the group, *flavivirus*, *pestivirus*, *pegivirus* and *hepacivirus*. Viruses in the tree include Tick-Borne Encephalitis Virus (TBEV), Yellow Fever Virus (YFV), West Nile Virus (WNV), Japanese Encephalitis Virus (JEV) and Dengue Viruses (DV-). Bootstrap values ≥70% after 1000 replicates are shown.

Our understanding of the evolutionary origins of HCV is still incomplete and our knowledge of viruses closely related to HCV is still expanding with the recent identification of numerous rodent, equine, bovine and canine *hepacivirus* homologs of HCV (Kapoor *et al.*, 2011, 2013; Lyons *et al.*, 2012; Corman *et al.*, 2015). Many of these viruses do not appear to cause significant disease in their hosts and may prove to be useful

animal models for HCV work. The recent discovery of a novel human virus, human hepegivirus 1 (HHpgV-1), detected in the blood serum of four blood product transfusion recipients, suggests that there may yet be more to discover within this group of viruses (Kapoor *et al.*, 2015).

1.1.3 Genomic structure

The HCV genome encodes 10 proteins in a large, single open reading frame (ORF). During viral replication, the virus is translated as a polyprotein of around 3000 amino acids which is post-translationally cleaved into 10 distinct viral proteins (Okamoto *et al.*, 1991). The ORF is flanked at both the 5' and 3' ends by highly conserved untranslated regions (UTR). The viral genes can be broadly grouped into structural and non-structural genes and are shown in figure 1-2 (Selby *et al.*, 1993). The structural genes are core (C), envelope proteins 1 and 2 (E1 and E2) and the p7 ion channel (P7). The non-structural genes outnumber the structural genes and they encode a transmembrane protein (NS2), a viral serine protease (NS3), an NS3 co-factor (NS4A), a hydrophobic protein (NS4B), a phosphoprotein (NS5A) and an RNA-dependant RNA-polymerase (RdRp) (NS5B). The non-structural proteins are largely involved in facilitating viral replication within infected cells.

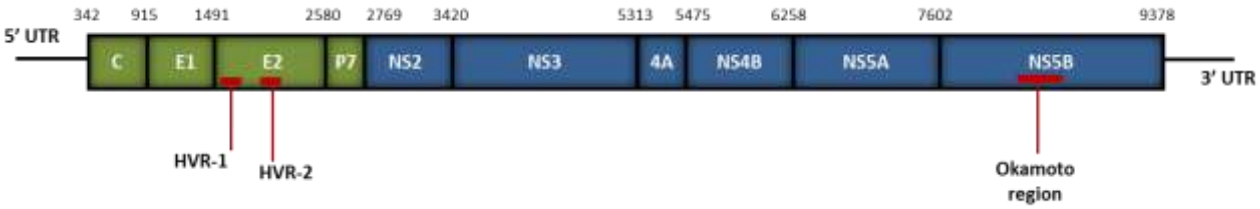


Figure 1-2; Organisation of the HCV genome

Schematic diagram of the HCV genomic organisation with nucleotide coordinates of gene products based on H77 (accession reference NC_004102)

The HCV 5' UTR is a highly conserved region of the virus and has been shown to have similarity to the 5'UTRs of GBV-B and *pestiviruses* (Choo *et al.*, 1991). Along with the first 12-30 nucleotides of the core region, folding of the RNA in the 5'UTR gives rise to a complex secondary structure, forming an internal ribosome entry site (IRES). The resulting IRES is capable of initiating cap-independent translation within cells, playing a key role in the early stages of HCV polyprotein translation (Honda *et al.*, 1999). The 3'UTR at the opposite end of the HCV genome has several features, notably a long poly U/UC tract and

a highly conserved 98 nucleotide sequence at its 3' terminal end. Stem-loop structures formed by the region have been shown to interact with the encoded RdRp and play a role in the initiation of RNA replication (Friebe and Bartenschlager, 2002).

Post-translational cleavage of the structural proteins from the polyproteins is thought to be mediated by host proteases (Lohmann, Koch and Bartenschlager, 1996). The HCV core protein is an RNA-binding protein that forms the viral capsid. Cleavage from the polyprotein and assembly into viral capsid are both thought to occur at the cytoplasmic face of the endoplasmic reticulum (ER). The protein has been shown to have additional roles influencing transcriptional events, apoptotic pathways, cellular transformation, immune presentation and lipid metabolism (McLauchlan, 2000). Core, along with the viral genome forms the nucleocapsid which is surrounded by a host-derived lipid envelope (Lindenbach and Rice, 2007). Embedded within the envelope are two envelope glycoproteins, E1 and E2, both of which are type I transmembrane glycoproteins (Lindenbach and Rice, 2007). They have been shown to assemble as noncovalent heterodimers and are crucial mediators of viral entry, facilitating numerous steps in the attachment, entry and fusion processes (Cocquerel *et al.*, 2003). Until recently, E2 was thought to be a class II fusion protein but recently published structural studies of its core region suggest a novel globular structure (Kong *et al.*, 2013; Khan *et al.*, 2014). The E2 protein contains two hypervariable regions within it, HVR-1 and HVR-2, and HVR-1 is thought to be the major target for anti-HCV neutralising antibodies during infection (Kato *et al.*, 1992).

P7 is a small polypeptide, belonging to the viroporin family of proteins. The protein is an integral membrane protein and it is thought to form hexameric and heptameric ring structures, conferring cation channel functions (Steinmann and Pietschmann, 2010). These channels are most likely to function as calcium ion channels and, whilst not essential for RNA replication, mutations in the protein have been shown to be detrimental to the production of infectious viruses implicating a probable role for the protein in the assembly and egress of viral particles (Steinmann and Pietschmann, 2010; Chandler *et al.*, 2012).

NS2, along with the terminal region of NS3, forms a zinc-dependant metalloprotease (NS2-3) which is required for the cleavage of NS2 and NS3 (Grakoui *et al.*, 1993). The protein is also thought to mediate processes associated with virus assembly, through a complex series of interactions with the other HCV proteins (Jirasko *et al.*, 2010). The NS3

protein has multiple functions and is comprised of an N-terminal serine proteinase region and a C-terminal helicase/NTPase domain. NS4A is an important co-factor for NS3, and NS3-4A functions as a serine proteinase, post –translationally cleaving all non-structural protein products, apart from NS2/NS3 (Bartenschlager, 1999). The proteinase is a popular target for therapeutics and numerous protease inhibitors have been developed and licenced for the treatment of HCV (Clark, Peter and Nelson, 2013). The NS3 helicase-NTPase is thought to be involved in RNA binding and the unwinding of RNA secondary structure. A role facilitating translocation along the RNA strand during RNA replication has been hypothesized (Levin, Gurjar and Patel, 2005).

The role of NS4B remains to be fully characterised although it is thought to be involved in the formation of replication complexes. The protein is usually closely associated with membranes and it is believed to initiate the formation of the membranous web, a conformational change to local membrane vesicles that generates a scaffold for the formation of replication complexes (Gouttenoire *et al.*, 2009). NS5A is a phosphoprotein that is thought to play a central role in the regulation of HCV replication. The protein appears to be closely associated with lipid-rich regions of intracellular membranes and it has been hypothesized that changes to the level of NS5A phosphorylation provide a regulatory switch between replication and assembly (Tellinghuisen, Foss and Treadaway, 2008). The protein is also thought to be a determinant of interferon sensitivity, possibly via protein kinase R inhibition (Gale *et al.*, 1997). NS5B functions as an RNA-dependant RNA polymerase and it plays a key role in viral replication, synthesizing negative-stranded templates from the positively-stranded genome and vice versa. The protein has been shown to lack proof-reading activity and consequently, HCV replication is highly error-prone (Steinhauer, Domingo and Holland, 1992). The protein is essential for HCV replication and is a major target for the design of anti-viral therapeutics.

1.1.4 The hepatitis C virus lifecycle

The HCV lifecycle is generally divided into 3 stages comprising of attachment and entry, translation and replication and budding and release. Initial transmission of the virus is usually via parenteral routes, resulting in the release of infectious virions into the bloodstream. Whilst circulating within the bloodstream, HCV virions are thought to associate with low-density lipoprotein (LDL) and very low-density lipoprotein (VLDL)

(André et al., 2002). Viruses are trafficked in the bloodstream towards the liver where infection is established within hepatocytes. Initial exposure to the hepatocytes occurs after crossing the sinusoidal epithelia and viral interaction at the basolateral surface is mediated by HCV envelope proteins E1 and E2. Initial interactions are believed to occur with the dendritic cell-specific intercellular adhesion molecule-3-grabbing non-integrin (DC-SIGN) and the liver/lymph node-specific intracellular adhesion molecule-3 grabbing integrin (L-SIGN), both of which are understood to play a central role in the tissue tropism of HCV (Gardner et al., 2003; Lozach et al., 2003). Evidence suggesting binding to low-density lipoprotein receptors (LDL-R) and heparan sulfate proteoglycans has also been published (Wünschmann et al., 2000; Barth et al., 2003). These initial binding steps are then thought to trigger a coordinated series of interactions with specific viral entry factors, eventually resulting in the release of the HCV genome into the cytosol (figure 1-3). E2 has been shown to mediate attachment to the tetraspanin CD81 and the scavenger receptor B1 (SR-B1) (Pileri et al., 1998; Scarselli et al., 2002). Studies have indicated that CD81 may be involved in priming the viral envelope for pH-dependent membrane fusion (Sharma et al., 2011). Interactions with these proteins then facilitate the transfer of the HCV virons onto the tight junction proteins claudin-1 (CLN-1) and occludin (OCN). CD81 and OCN have been shown to be key determinants of HCV's human-specific tropism (Ploss et al., 2009; Bitzegeio et al., 2010). Eventual uptake of the virus into the cytosol is thought to occur via clathrin-mediated endocytosis and release of the viral genome is dependent on a low pH. Once within hepatocytes, HCV has also been shown to be capable of cell to cell spread in *in vitro* systems, suggesting that cell to cell transmission may also occur in vivo, contributing to the establishment of infectious foci throughout the liver during infection (Timpe et al., 2008).

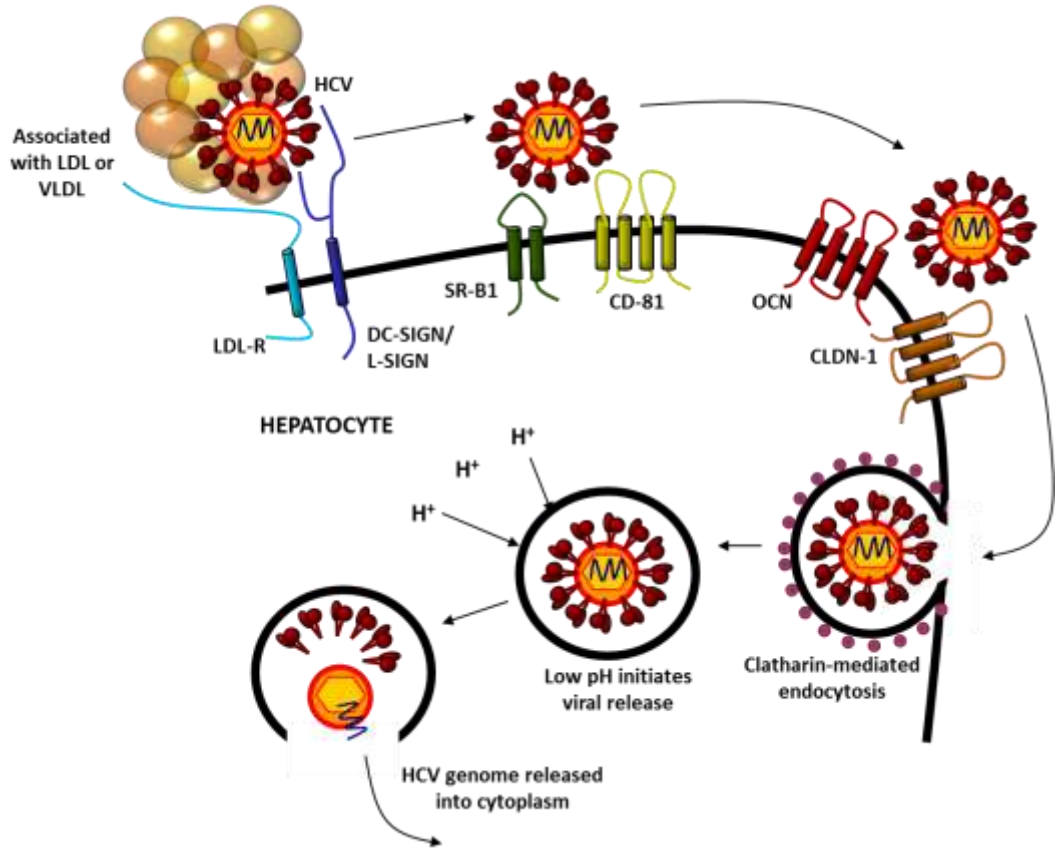


Figure 1-3; Entry of the HCV virus into hepatocytes

Figure adapted from a figure in an MRes thesis submitted by AL McNaughton (2012) (McNaughton, 2012).

Once the viral genome has been uncoated and released into the cytoplasm, the next phase of the lifecycle is translation and replication. As a positive-sense single-stranded RNA virus, the HCV genome is able to interact directly with host proteins and can be processed in a similar way to mRNA, resulting in the rapid production of viral products. The IRES contained within the 5'UTR facilitates the interaction of the viral genome with host cell translation complexes (Friebe *et al.*, 2001). The genome is then translated as a single ORF into a long polyprotein which is then cleaved by both host and viral-encoded proteases. The cleaved protein products are associated with membranes originating from the endoplasmic reticulum. In order to replicate the viral genome, HCV induces the formation of the membranous web, a complex rearrangement of the intracellular membranes enabling the formation of HCV replication complexes (Egger *et al.*, 2002). Comprised of NS3-4A, NS4B, NS5A and NS5B, these replication complexes produce positive-stranded HCV genomes from negative-sense intermediates. The complexes also function to generate negative-sense anti-genomes to function as templates for the production of positive-sense genomes although these are usually produced in relatively low quantities relative to the positive-sense genomes (Craggs *et al.*, 2001). These positive-sense genomes are used to produce increased quantities of viral proteins, allowing a dramatic increase in the replicative rate of the virus.

The final phase of the HCV lifecycle is the assembly and release of new HCV virions, enabling a continuation of the infection process. The process of HCV assembly and release has proven difficult to study however and our understanding of the process remains limited. During the process, core, E1, E2, p7 and HCV genomes must be brought together and packaged in a regulated manner to produce infectious virions. Recent work has identified domains in NS5A as key regulatory mediators of this process (Zayas *et al.*, 2016). Roles for p7 and NS2 contributing to the mediation of the assembly process have also been described (Popescu *et al.*, 2011). Cytosolic lipid droplets are thought to have a major function in the process and are associated with trafficking the viral core proteins from the endoplasmic reticulum to the site of assembly. Depletion of apolipoproteins, in particular, apoE, has been shown to restrict the production of viral progeny, implying that they are key viral infectivity factors (Jiang and Luo, 2009).

Determination of the HCV lifecycle has proved challenging as progress in the field has been restricted by the lack of permissive cell culture systems or easily accessible animal models. Much of our current understanding has therefore developed from the use of

surrogate systems allowing the modelling of aspects of the viral lifecycle. Early characterisation of the replicative process within hepatocytes relied heavily on the use of sub-genomic replicons, fragments of the HCV genome that are able to self-replicate within hepatoma cell lines (Lohmann *et al.*, 1999; Blight, Kolykhalov and Rice, 2000). Infectious culture systems (HCVcc) have also been developed from the genotype 2 JFH-1 strain, isolated from a patient in Japan and have been widely used to study all aspects of the viral lifecycle (Wakita *et al.*, 2005). Chimeric variants of the JFH-1 HCVcc system have been refined to enable the study of other genotypes but considerable variation in the replicative capacity of the chimeras is frequently observed, meaning they often cannot be directly compared (Pietschmann *et al.*, 2006). Pseudoparticle systems (HCVpp), using the gag and pol retroviral genes to form particles expressing the HCV glycoproteins E1 and E2 on their surface, have also been used to examine the viral entry process (Bartosch, Dubuisson and Cosset, 2003). Recent work by Saeed *et al*., (2015) has demonstrated that the addition of SEC14L2 makes hepatoma cell lines permissive for the replication of multiple HCV genotypes, enabling the propagation of clinical isolates within *in vitro* systems (Saeed *et al.*, 2015). It is anticipated that this novel system will provide powerful new tools to study the infection process within and facilitate the *in vitro* drug resistance testing of patient-derived isolates for those responding poorly to treatment.

1.2 The genotypes & epidemiology of the hepatitis C virus

1.2.1 The hepatitis C virus genotypes

Currently, the hepatitis C viruses are classified into 7 distinct genotypes (1 to 7) on the basis of their nucleotide sequence (Smith *et al.*, 2014). The HCV genotypes show considerable diversity and may exhibit as much as 25-30% difference between each other at the nucleotide level (P. Simmonds *et al.*, 1993). After genotypic classification, HCV viruses are further classified into 67 distinct subtypes which may show up to 15% nucleotide sequence heterogeneity (figure 1-4). A standardised system for the classification of HCV genotypes was proposed by Simmonds *et al*., in 2005 requiring phylogenetic analysis based on the core/E1, NS5B and complete genome sequences (Simmonds *et al.*, 2005). The proposed classification system has since been adopted by the

International Committee on Taxonomy of Viruses and is used to this day (Smith *et al.*, 2014).

There is a large range of subtype diversity among the different genotypes with genotypes 4 and 6 being divided into 17 and 24 viral subtypes respectively (Smith *et al.*, 2014). Conversely, genotypes 5 and 7 are restricted to a single subtype each, possibly a result of a lack of sequences and/or geographical restriction of these genotypes (Smith *et al.*, 2014). A particularly high diversity within genotypes 1, 2 and 4 is observed in Africa and a similar, high divergence within genotypes 3 and 6 is found throughout Asia, suggesting that these particular viral genotypes have a long history of transmission within these regions (Tokita *et al.*, 1994; Jeannel *et al.*, 1998). Between the different genotypes, there are also documented differences in clinical associations (Zein, 2000), immune responses (M. W. Robinson *et al.*, 2015), treatment outcomes (Zein, 2000) and geographical prevalence (Gower *et al.*, 2014).

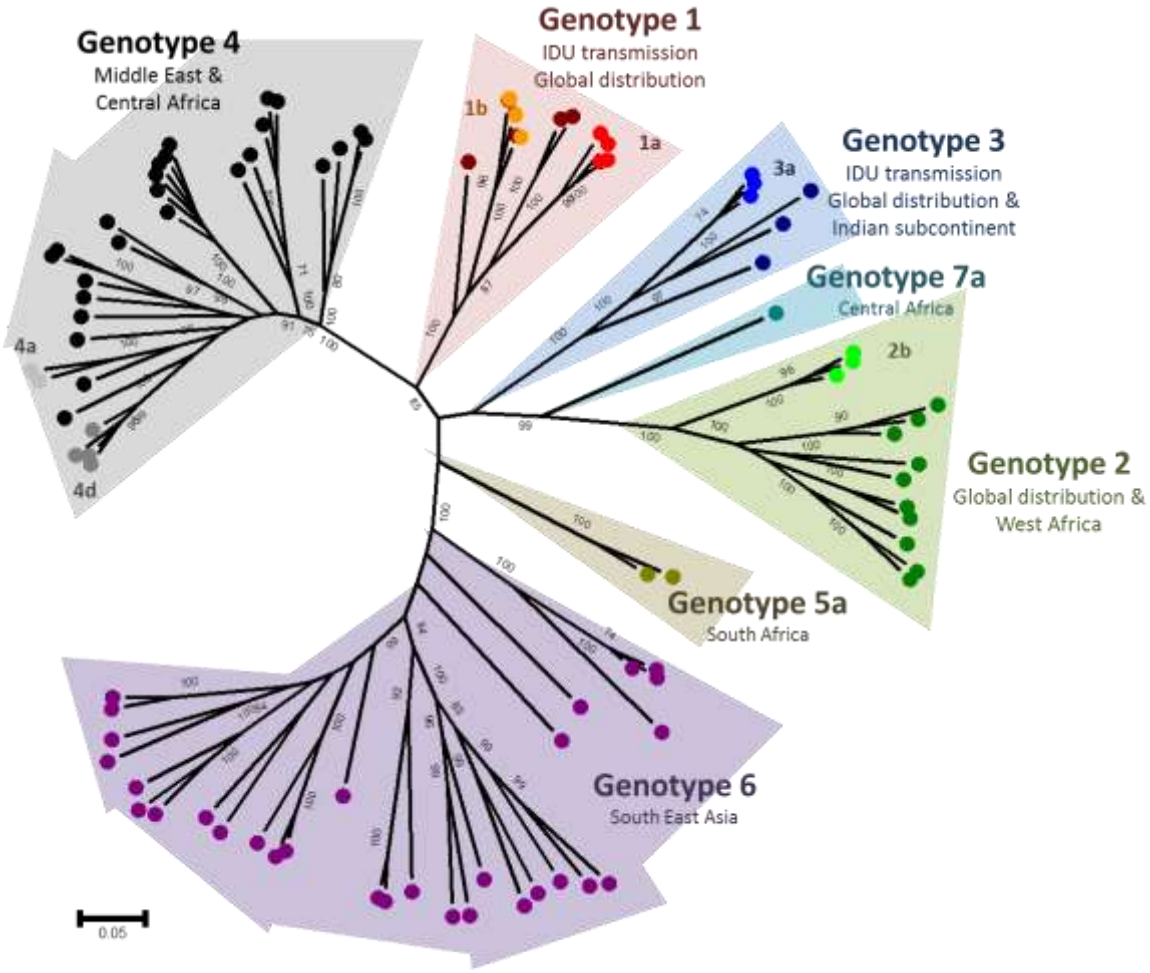


Figure 1-4; The phylogeny of the HCV genotypes

A maximum-likelihood phylogenetic tree of full-length HCV reference sequences as available from the Los Alamos HCV database (Kuiken *et al.*, 2004). Some of the most common subtypes, 1a, 1b, 2b, 3a, 4a and 4d have been highlighted. Well-defined geographical and transmission associations are indicated for the genotypes. Bootstrap values ≥70% after 1000 replicates are shown.

The highest prevalence rates are observed in Africa, particularly in Egypt and the Cameroon (both >10%) (Hajarizadeh, Grebely and Dore, 2013). Approximately 50% of individuals infected with HCV are estimated to reside in Asia where large populations in China, India, Pakistan and Indonesia mean that despite low-to-mid level prevalence rates, high numbers of HCV-infected individuals (>9 million) are present in these countries (Hajarizadeh, Grebely and Dore, 2013). Genotype 1 is generally assumed to be the most prevalent genotype globally, infecting an estimated 46.2% of the total infected population(Gower *et al.*, 2014; Messina *et al.*, 2015). It should however be noted that genotype 1 viruses are most frequently diagnosed in the Americas and Europe where a disproportionately high number of studies also originate from (Messina *et al.*, 2015) and it is therefore probable that prevalence rates for other genotypes are underestimated.

1.2.2 Transmission of the hepatitis C virus

Humans are the single host species for HCV and the virus is thought to have a long evolutionary history with humans, likely dating back several thousand years (Simmonds, 2004). The majority of contemporary HCV transmissions occur via parenteral exposure with the most common transmission routes linked to the use of unclean needles during recreational drug use, medical procedures and tattooing (World Health Organisation, 2015). Prior to the introduction of blood-borne viral (BBV) testing, blood and blood product transfusions were significant risk factors for HCV infection. Other, less frequently reported transmission routes include vertical transmission from mother to child, sexual transmission and needle stick injuries.

There is a significant disparity between the most common HCV transmission routes in high and low resource nations. In high resource nations, sterilization of medical equipment is highly standardised and donated blood samples are screened according to rigorous guidelines. The majority of new HCV infections occurring are therefore in individuals with a history of injecting drug use (Razali *et al.*, 2007; Cornberg *et al.*, 2011; Williams *et al.*, 2011). In middle and low-income nations, external quality control schemes for the testing of donated blood are in place in 82% and 53% of nations respectively and as of 2011, there were still an estimated 39 countries that are unable to screen donated blood for transfusion-transmissible infections (World Health Organization, 2011). Whilst transmissions will still occur via recreational drug use in these nations, the prevalence of HCV acquired through unsafe medical procedures is likely to be significantly higher.

Changes in global migration and the development of parenteral medical procedures are thought to have contributed to the increased spread of HCV from endemic regions to industrialised nations during the mid-1900s. Several studies, including work by McNaughton *et al.*, have inferred an exponential growth in genotype 1a and 3a infected populations occurring in the 1940s (Pybus *et al.*, 2005; Mcnaughton *et al.*, 2015). It is likely that the population mixing and increased use of blood transfusions during WWII played a pivotal role in the global dissemination of HCV. Since WWII the increasing availability of hypodermic syringes has led to a significant HCV epidemic among injecting drug users (Trepo and Pradat, 1999; Drucker, Alcabes and Marx, 2001; Aceijas and Rhodes, 2007). Widespread re-use of hypodermic syringes during vaccination initiatives has also contributed significantly to the spread of HCV, particularly in Egypt where an anti-*schistosomiasis* campaign in the 1960s led to millions becoming infected with HCV (Frank *et al.*, 2000).

Sexual transmission of HCV is thought to be relatively rare and studies have shown that rates of infection among long-term co-habiting heterosexual couples are low (Terrault *et al.*, 2013). A number of factors have however been found to increase the likelihood of sexual transmission of HCV including co-infection with HIV and engaging in risky sex practices. Studies have shown that changes to gut-associated lymphoid tissue in the early phases of HIV infection can lead to the depletion of CD4+ cells in the gut. It is thought probable that HIV-induced changes to mucosal immunity may also be facilitating the sexual transmission of HCV (Guadalupe *et al.*, 2006; Kim and Chung, 2009). Since 2000, it has become clear that HIV-positive men who have sex with men (MSM) are particularly at risk of contracting HCV via sexual routes and numerous large outbreaks within this population have been documented (Gotz *et al.*, 2005; Urbanus *et al.*, 2009; Laar *et al.*, 2013). Several mono-phyletic clusters of gt4d viruses transmitting across Europe have been linked to transmission networks in HIV-infected MSM, suggesting this genotype is relatively adept at sexual transmission (Serpaggi *et al.*, 2006; van de Laar *et al.*, 2007; Vogel *et al.*, 2010; Thomson, Smith and Klenerman, 2011).

1.2.3 Hepatitis C virus in the UK

Within the UK an estimated 214,000 individuals are thought to be infected with chronic HCV (Public Health England, 2015). The distribution of HCV infections across the UK

varies significantly with the estimated prevalence in Scotland (0.7%) being almost twice the estimated prevalence in England (0.4%) (Public Health England, 2015). Urban centres are also disproportionately affected, with high levels of infection reported in London, the North West and the Greater Glasgow region (Hutchinson *et al.*, 2002; Harris *et al.*, 2012). The prevalence of HCV is also known to vary significantly between different ethnic groups throughout the UK with relatively high numbers of Asian, British Asian and Eastern European individuals testing positive for HCV currently (Harris *et al.*, 2012; Public Health England, 2015). HCV genotypes 1a and 3 are the most frequently diagnosed throughout the UK, being responsible for an estimated 90% of infections (Public Health England, 2015). Both genotypes are strongly associated with transmission by IDU, the most commonly recorded risk factor for HCV in the UK. Evidence suggests that IDU is currently declining within the UK and that increasing numbers of injecting drug users are enrolling in drug treatment programs (National Treatment Agency for Substance Misuse, 2010). Awareness of the disease among primary healthcare providers has also improved considerably and HCV testing rates have dramatically improved in all regions of the UK over the past 5 years (Public Health England, 2015). Despite these developments, hospital admissions for HCV-associated end-stage liver disease and hepatocellular carcinomas have tripled within the UK over the past decade, and the number of deaths from HCV has doubled (Public Health England, 2015).

Since the introduction of screening for HCV in donor blood in 1990, transmission of HCV by infected blood or blood products has been virtually eliminated and is generally observed in older individuals or those who have received treatment abroad. In all regions of the UK, injecting drug use remains the major risk factor for HCV infection and is associated with an estimated 70-90.6% of infections (Public Health England, 2015). Numerous studies have indicated that there is a heterogeneous gender prevalence of HCV within the UK with males accounting for 60-70% of the infected population (Mohsen, *et al*., 2001; Public Health England, 2015). Asian and British Asian individuals contribute a significant proportion of the individuals (approaching 50%) who are HCV positive with no previous history of injecting drug use suggesting that treatment abroad may be contributing to the transmission of HCV in these groups (Harris *et al.*, 2012). Other genotypes frequently diagnosed within the UK include gt2 and gt1b although these are generally observed in older individuals with a history of a blood transfusion prior to 1980.

1.3 Infection with the hepatitis C virus

1.3.1 Disease progression

Today, HCV is recognised as a global problem, infecting an estimated 130-170 million individuals worldwide (Lavanchy, 2009, 2011; Hajarizadeh, Grebely and Dore, 2013). Infection with HCV can be broadly split into acute and chronic phases. In 15-40% individuals, infection with HCV is acute, with the infection lasting up to 6-12 months before being naturally cleared by the host (Westbrook and Dusheiko, 2014). Acute HCV infection is often asymptomatic and causes few issues for the host. Patients are rarely sufficiently ill (often describing relatively non-specific 'flu-like' symptoms) to consider seeking medical attention and as a result, cases of acute HCV are usually only detected in those attending clinics and treatment centres for alternative reasons (Cox *et al.*, 2005). In the 60-85% individuals who do not clear HCV, progression towards chronic HCV infection occurs after the acute infection phase, with the virus likely to persist for the lifetime of the host unless treated. The virus can cause significant damage to the liver and consequently is a major source of morbidity and mortality, killing more than 350,000 people annually (WHO and World Health Organization, 2012).

Acute HCV is characterised by detectable viral loads in the blood 2-14 days post exposure and the subsequent elevation of ALT and AST levels, indicating increased liver stress (Cox *et al.*, 2005; Glynn *et al.*, 2005). After initial infection, there is an increase in HCV viral load followed by a plateau phase, usually occurring 45-68 days post infection (Glynn *et al.*, 2005). After the plateau phase, the viral load decreases and either spontaneous clearance or progression to chronic disease ensues. Antibody to the virus is typically detected 20-150 days after infection (Cox *et al.*, 2005; Glynn *et al.*, 2005; Page-Shafer *et al.*, 2008). Numerous prognostic indicators for viral clearance during acute infection have been identified in the literature including, infection with a gt1 virus, female gender and favourable host genetics inducing broad and neutralising antibody responses (Takaki *et al.*, 2000; Harris *et al.*, 2007; Page *et al.*, 2009; Lemon, 2010). Polymorphisms in the IL28B gene, particularly for the CC genotype at rs12979860, have been strongly associated with viral clearance during acute HCV (Thomas *et al.*, 2009; Tillmann *et al.*, 2010).

In those who fail to clear the virus during the acute phase, a persistent chronic infection is established. Symptoms of chronic HCV are generally mild and non-specific, with anorexia,

nausea and right-upper quadrant pain commonly reported (Zaltron *et al.*, 2012). Increased ALT levels are frequently observed but there is considerable variation, and a poor correlation is observed between ALT levels and disease severity (Alberti *et al.*, 1992; Barrera *et al.*, 1995; Massard *et al.*, 2006). Fluctuations in the viral load to the magnitude of 0.5log are often observed throughout chronic infection and thought to have little clinical significance (Ito *et al.*, 2004; Barreiro *et al.*, 2013). Chronic HCV is usually a slowly progressing disease with persistent infection resulting in hepatic inflammation, often resulting in fibrotic scarring over time. In an estimated 10-20% of individuals with HCV, cirrhosis develops after 20-30 years (Thein *et al.*, 2008; Westbrook and Dusheiko, 2014). The development of cirrhosis has been shown to be highly variable however with estimates ranging considerably with study methodology and population sampling (Thein *et al.*, 2008). Progression to cirrhosis is a multi-factorial process and numerous risk factors have been associated with an increased risk and accelerated rate of fibrotic progression including male gender, age >40 years, excessive alcohol intake, immunosuppression, diabetes and co-infection with HIV or HBV (T Poynard, Bedossa and Opolon, 1997; Ramos Paesa *et al.*, 1998; Minola *et al.*, 2002; Gaeta *et al.*, 2003; Ratziu *et al.*, 2003; Hutchinson, Bird and Goldberg, 2005). Hepatocellular carcinoma (HCC) is estimated to occur in 2-3% of individuals infected with HCV after 30 years (Grebely and Dore, 2011). Individuals with cirrhosis have an increased risk of developing HCC with studies indicating that annually 1-3% of cirrhotic individuals will develop HCC (El-Serag and Rudolph, 2007). A higher prevalence of HCC has also been associated with age >55 years, male gender and an excessive alcohol intake (Fattovich *et al.*, 2004).

1.3.2 Immunity to hepatitis C

In up to 80% of individuals infected with HCV, the immune system is unable to effectively clear the virus and progression towards chronic infection occurs. Once chronic infection is established, patients are likely to carry the virus for the rest of their lifetime, unless treatment is successful. This makes HCV unusual as it is one of the few RNA viruses to establish chronic infection in immunocompetent hosts. The diversity observed between the HCV genotypes and their ability to rapidly mutate mean that heterologous protective immunity is difficult to generate and requires the specific targeting of conserved regions. Recent work has identified genetic components of the host immune system as being critical determinants in the outcome of acute HCV infection. Individuals carrying the human leukocyte antigen (HLA) B27 and several short nucleotide polymorphisms of the IFN-λ3

allele have been found to mount vigorous anti-HCV responses and have a significantly higher chance of clearing the virus during the acute phase of infection (Neumann-Haefelin *et al.*, 2006, 2011; Mancuso *et al.*, 2014; Moqueet *et al.*, 2015).

Initial infection with HCV is characterised by a phase of rapid viral replication which stimulates the innate immune system (figure 1-5). Specific viral features including the IRES, the 3'UTR poly U/UC tract and dsRNA replication intermediates are recognised by the pattern recognition receptors protein kinase R (PKR), toll-like receptor 3 (TLR-3) and retinoic acid-inducible gene I (RIG-I) (Horner and Gale, 2013). PKR and RIG-I activation initiate signal cascades mediated by mitochondrial anti-viral signalling protein (MAVS), triggering interferon-β (IFN- β) synthesis and the production of tumour necrosis factor-α (TNF- α) and CXCL10. TLR-3 signalling is cascaded by Toll/IL-1 receptor domain-containing adaptor (TRIF), also resulting in the induction of IFN- β. HCV has been shown however to abrogate IFN induction by NS3-4A mediated cleavage of MAVS and TRIF (Foy *et al.*, 2003; Li *et al.*, 2005). Phagocyctic uptake of HCV RNA by Kupffer cells and dendritic cells in the intrahepatic environment has been shown to stimulate the synthesis of IFN-α , IFN-β and interleukin-1β (Il-1β), providing additional pathways for immune activation (Takahashi *et al.*, 2010; Negash *et al.*, 2013). IFN signalling via the JAK-STAT pathway also results in the induction of the interferon stimulated genes (ISGs). Distinct ISG profiles have been attributed to different HCV genotypes, indicating significant disparity in the host-pathogen interactions which occur for each genotype (M. Robinson *et al.*, 2015; M. W. Robinson *et al.*, 2015).

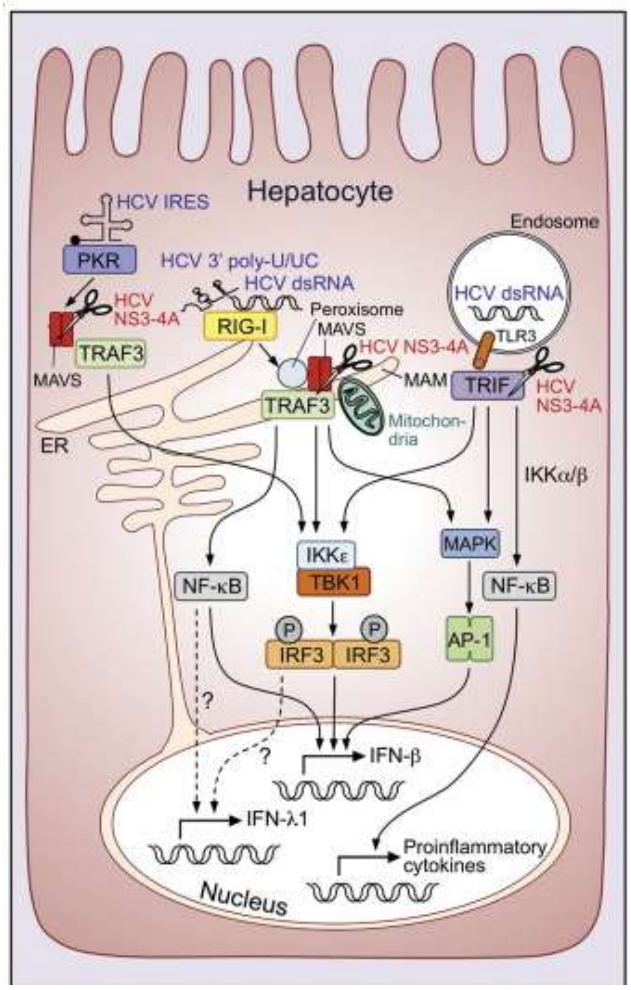


Figure 1-5; Intrinsic immune responses against HCV in hepatocytes

Specific HCV RNA structures recognized by pattern recognition receptors and the immune signalling cascades activated. Dotted lines indicate pathways that have been demonstrated to be activated in other hepatic infections but have not yet been confirmed for HCV. Abbreviations not discussed elsewhere include MAM, mitochondria-associated endoplasmic reticulum membrane; TRAF3, TNF receptor-associated factor; TBK1, TANK-binding kinase 1; AP-1, activator protein-1. Image published by Park and Rehermann, 2014 (Park and Rehermann, 2014).

Following HCV infection, activation of the adaptive immune response is delayed by around 8-12 weeks, regardless of whether or not chronic infection eventually establishes (Park and Rehermann, 2014). Individuals clearing their HCV infection have been shown to generate broad-acting CD4+ T cell responses early in the infection process. Responses are characterised by strong T cell proliferation and synthesis of IL-2, TNF-α and IFN-λ, and appear to be critical determinants of viral clearance (Missale *et al.*, 1996; Smyk-Pearson *et al.*, 2008). Among those developing chronic infection, the CD4+ T cell responses are typically poorly targeted and weaker with T cell exhaustion frequently observed (Gerlach *et al.*, 2005). Vigorous CD8+ T cell responses targeting a broad range of epitopes with strong IFN-ɣ production have also been shown to be crucial in the resolution of acute HCV (Lechner *et al.*, 2000; Thimme *et al.*, 2001). CD8+ T cell responses by contrast are typically restricted in the acute phase of infection in individuals developing chronic HCV with viral mutations in the targeted epitope regions able to rapidly outpace the CD8+ T cell response (Park and Rehermann, 2014).

In chimpanzee models, secondary challenge with HCV after an initial clearance of a primary HCV infection identified a rapid production of HCV-specific T cell responses, resulting in a faster clearance of the secondary infection (Bassett *et al.*, 2001; Nascimbeni *et al.*, 2003). Studies with people who inject drugs (PWID) have also found that rates of HCV infection are lower among individuals re-infected with HCV following viral clearance than in those previously unexposed to the virus (Mehta *et al.*, 2002; Grebely *et al.*, 2006). Whilst these studies do imply that protective immunity may be generated in some cases, other studies have documented multiple HCV infection at rates of 23.3-25% suggesting it is frequently ineffective (Herring *et al.*, 2004; van de Laar *et al.*, 2009; Pham *et al.*, 2010). Studies have also suggested that protective immunity may diminish over time and among those clearing the virus, loss of memory T cells frequently occurs, indicating that re-infection is possible if re-exposed (Gerlach *et al.*, 1999).

1.3.3 Diagnosis of the hepatitis C virus

Acute HCV infection is rarely noticed by the patient and the disease often remains asymptomatic for a number of years. Consequently, infections are usually detected many years later, either incidentally when screening for other conditions or after the patient has eventually developed symptoms suggesting chronic liver disease may be an issue (World Health Organisation, 2015). Diagnostic tests for HCV can be broadly grouped into those

indirectly detecting the virus and those directly identifying the presence of a virus. Indirect assays are serological and rely on the detection of a HCV-specific antibody response. In HCV testing, sandwich enzyme immunoassays (EIA) are the most frequently used antibody test in clinical settings, indicating the presence of an immune response against HCV by the detection of anti-HCV IgG (Kamili *et al.*, 2012). Direct testing relies on the detection of a specific viral marker, usually viral antigen or viral RNA. Antigen testing uses specific detection methods, often involving fluorescently tagged antibodies, to detect the presence of a viral product. In HCV testing, core antigen assays are frequently used for this purpose as it is an easily detectable and relatively conserved antigen (Hosseini-Moghaddam *et al.*, 2012). Detection of HCV RNA in the serum has also been shown to be a predictive biomarker for viral replication within the host (Yun *et al.*, 1993; Marin *et al.*, 1994). The most widely-used assays are line probe assays (LiPA) and real-time quantitative PCR (qPCR) assays. Both assays rely on amplification of the viral genomic material and then the detection of a specific nucleotide sequence, often with a fluorescently tagged probe.

In the UK, diagnostic testing for HCV is offered in a range of different clinical settings including GP surgeries, sexual health clinics and drug treatment facilities. Standard diagnostic testing in the UK requires two tests, an antibody test and a PCR test (NHS, 2004). Two tests are used as antibody responses can take 2-8 weeks to develop in individuals after the acute phase of HCV infection and are considered poor biomarkers for acute HCV infection. In contrast, viral RNA can be detected 1-3 weeks after initial infection (Chevaliez, 2011). As a result, patients in the early stages of HCV infection may be antibody negative but PCR positive. The majority of commercially available assays target the 5'UTR or core regions of the virus (Cobb, Heilek and Vilchez, 2014). These regions of the virus are relatively well-conserved due to significant secondary structure and tests targeting these regions can be used to confirm the presence of multiple HCV genotypes. qPCR assays also have applications in the quantitative measurement of viral load in the serum which can be useful for the monitoring of treatment responses (Cobb, Heilek and Vilchez, 2014). Reliably differentiating between acute and chronic HCV is often difficult as the presence of IgM (generally considered a biomarker of acute infections) is highly variable during acute HCV and remains elevated in many individuals with chronic HCV (Farci *et al.*, 1992; Kao *et al.*, 1997). Chronic HCV is often indicated by a review of the patient history, indicating no recent exposure to risk factors, or progressive liver disease.

Clinically, the determination of HCV genotype is an important prognostic indicator as HCV genotype is influential when considering the most appropriate therapy for patients as numerous treatments are known to have genotype-specific efficacies (Manns, Wedemeyer and Cornberg, 2006; Gottwein *et al.*, 2011). Distinct inflammatory gene transcription profiles in the liver have also been described for genotypes 1 and 3 (Robinson *et al.*, 2015) and higher rates of steatosis, cirrhosis and HCC have been linked with gt3 infection (Leandro *et al.*, 2006; Kanwal *et al.*, 2014). Genotyping assays need to target regions of the virus that have inter-genotypic variation but are relatively well conserved at the genotype level. The 'gold standard' approach involves sequencing the NS5B region of the virus and phylogenetic analysis of the sequence alongside reference sequences (Richter, 2002). Clinically however, this is expensive and time consuming and therefore alternative approaches using LiPA assays and genotype-specific qPCR assays are used (González *et al.*, 2013). Whilst the use of such assays is generally considered satisfactory, limitations including intermediate results in approximately 5% of samples, a lack of differentiation for many of the HCV subtypes and an inability to resolve mixed genotype HCV reactions have been widely documented (Chevaliez *et al.*, 2009; Larrat *et al.*, 2013; Cobb, Heilek and Vilchez, 2014).

1.3.4 Treatment of hepatitis C

Until 2011, treatment options for HCV were limited to a 48-week treatment of pegylated interferon alpha (PEG-IFN-α) and ribavirin. The drugs were difficult to tolerate, causing numerous side effects and sustained virologic response (SVR) rates were relatively low with 40-60% individuals clearing the virus as a result of treatment (Manns, Wedemeyer and Cornberg, 2006). Response rates were generally more favourable among younger patients, Caucasians and individuals infected with genotypes 2 or 3 (McHutchison *et al.*, 2000; Manns *et al.*, 2001; Andriulli *et al.*, 2008; Rodriguez-Torres *et al.*, 2009). Individuals with a favourable IL28B polymorphism also had improved responses to therapy (Li *et al.*, 2011). In 2011 however, the first of a new range of direct-acting anti-virals (DAA) specifically targeting HCV were licensed for clinical use. The first drugs to be approved, Telaprevir and Boceprevir, were NS3-4A protease inhibitors, functioning by effectively restricting the viral capacity for post-translational cleavage (Gao *et al.*, 2010; Hofmann and Zeuzem, 2011). When used in conjunction with PEG-IFN-α and ribavirin, the new treatments increased the SVR rate to 59-75% (Sarrazin *et al.*, 2012), vastly improving on previous treatment rates.

Since the licensing of the first DAAs in 2011, 3 additional classes of DAA have been licensed, second generation NS3-NS4A protease inhibitors have been developed (Simeprevir) and numerous other treatments are in late stage clinical trials. The new classes of DAA include RNA-dependent RNA polymerase (RdRp) inhibitors, nucleoside analogues targeting NS5B such as Sofosbuvir and NS5A inhibitors including Daclatasvir and Ledipasvir (Pawlotsky, 2014). Potential new classes of therapies include cyclophilin inhibitors and miRNA 122 antagonists (Pawlotsky, 2014). Similar to approaches used with highly active anti-retroviral therapies (HAART), the expansion in treatment options has enabled the use of dual and triple therapies targeting multiple aspects of the viral lifecycle. These new treatments have largely been successful, offering IFN-free regimens with improved tolerance and SVR rates upwards of 90% frequently for patients (Pawlotsky, 2014). Many of the drugs are highly efficacious against gt1 and gt4 viruses (table 1-1) with therapeutic options for other genotypes being limited either due to less favourable cure rates or, in the case of genotypes 5 and 6, the drugs have not been widely tested and there remains a lack of data (Pawlotsky *et al.*, 2016). Gt3 strains are currently the most difficult to cure with the new DAA regimens and there remains limited treatment options for individuals infected with this genotype (Foster *et al.*, 2011; Moreno *et al.*, 2012; Goossens and Negro, 2014; EASL, 2015). Treatments have been shown to be effective in patients previously failing pegylated IFN-α and ribavirin therapies and in individuals with cirrhotic liver disease with longer treatment regimens (12 or 24 week depending on the treatment) generally being recommended in such cases (Pawlotsky *et al.*, 2016). Additionally, many of the new drugs come at a prohibitive financial cost, making them difficult to access for the majority of people requiring treatment, particularly in resource-poor regions. For example, in the early access programme, designed to enable early access to the new DAA treatments for 500 urgent cases in the UK, the cost per treatment was upwards of £30,000 (NHS England, 2014).

Table 1-1; IFN-free treatment combinations for the HCV genotypes

Table detailing the IFN-free treatment options for HCV genotypes 1-6 as detailed in the EASL Recommendations on Treatment of Hepatitis, 2016 (Pawlotsky et al., 2016). The listed treatments are suitable for both HCV and HCV/HIV co-infected patients with chronic HCV without cirrhosis. Treatments suitable for treatment-naïve patients and patients who failed on a treatment based on pegylated IFN-α and ribavirin (treatment-experienced patients).

Patients	Treatment-naïve or -experienced	Sofosbuvir/ ledipasvir	Sofosbuvir/ velpatasvir	Ombitasvir/ paritaprevir/ ritonavir and dasabuvir	Ombitasvir/ paritaprevir/ ritonavir	Grazoprevir/ elbasvir	Sofosbuvir and daclatasvir	Sofosbuvir and simeprevir
Genotype 1a	Treatment-naïve	8-12 wk, no ribavirin	12 wk, no ribavirin	12 wk with ribavirin	No	12 wk, no ribavirin if HCV RNA ≤800,000 (5.9 log) IU/ml or 16 wk with ribavirin if HCV RNA >800,000 (5.9 log) IU/ml[b]	12 wk, no ribavirin	No
	Treatment-experienced	12 wk with ribavirin[a] or 24 wk, no ribavirin					12 wk with ribavirin[a] or 24 wk, no ribavirin	
Genotype 1b	Treatment-naïve	8-12 wk, no ribavirin	12 wk, no ribavirin	8-12 wk, no ribavirin	No	12 wk, no ribavirin	12 wk, no ribavirin	No
	Treatment-experienced	12 wk, no ribavirin		12 wk, no ribavirin				
Genotype 2	Both	No	12 wk, no ribavirin	No	No	No	12 wk, no ribavirin	No
Genotype 3	Treatment-naïve	No	12 wk, no ribavirin	No	No	No	12 wk, no ribavirin	No
	Treatment-experienced		12 wk with ribavirin[c] or 24 wk, no ribavirin				12 wk with ribavirin[c] or 24 wk, no ribavirin	
Genotype 4	Treatment-naïve	12 wk, no ribavirin	12 wk, no ribavirin	No	12 wk with ribavirin	12 wk, no ribavirin	12 wk, no ribavirin	12 wk, no ribavirin
	Treatment-experienced	12 wk with ribavirin or 24 wk, no ribavirin				12 wk, no ribavirin if HCV RNA ≤800,000 (5.9 log) IU/ml or 16 wk with ribavirin if HCV RNA >800,000 (5.9 log) IU/ml	12 wk with ribavirin or 24 wk, no ribavirin	12 wk with ribavirin or 24 wk, no ribavirin
Genotype 5 or 6	Treatment-naïve	12 wk, no ribavirin	12 wk, no ribavirin	No	No	No	12 wk, no ribavirin	No
	Treatment-experienced	12 wk with ribavirin or 24 wk, no ribavirin					12 wk with ribavirin or 24 wk, no ribavirin	

Numerous studies have documented the presence of naturally-occurring resistant variants to DAA treatments (Kuntzen *et al.*, 2008; Fonseca-Coronado *et al.*, 2012; Wang *et al.*, 2013). Many of the variants described are the result of non-synonymous single nucleotide polymorphisms (SNPs) that confer changes to DAA binding or interaction sites. The majority of documented SNPs are estimated to have a low prevalence of 0.1-3.5% (Chen *et al.*, 2016). Higher prevalence rates of several SNPs have been documented however, many of which have been found to have strong genotypic associations. Notable examples in the NS3 region conferring resistance to Simeprevir include Q80K, which has been found in

37.6% of gt1a strains and S122R, which was detected in 45.1% gt2 strains (Chen *et al.*, 2016). In the NS5A region, the Q30R variant associated with resistance to Daclatasvir and Ledipasvir has been found in 29.2% gt3 sequences and 55.3% of gt4 strains (Chen *et al.*, 2016). Y93H, conferring resistance to Daclatasvir and Ledipasvir has also been widely documented in gt1 strains (Lontok *et al.*, 2015). The strong genotypic associations documented indicate that barriers to resistance may be low in some genotypes, even if the resistant variants are not present in strains prior to the initiation of treatment. A lack of studies and available sequences for *in silico* analysis are however limiting our knowledge of the topic in non-gt1 viruses (Sarrazin, 2016). It has been suggested that the use of dual therapies, targeting multiple regions of the virus lifecycle simultaneously in a similar way to highly active anti-retroviral therapies, may be effective at mitigating some of these risks (Chayama *et al.*, 2012; Gutierrez, Lawitz and Poordad, 2015). Recent treatment guidelines have indicated that resistance profiling should be explored subsequent to virological breakthrough or treatment relapse (EASL, 2015).

1.4 Sequencing

1.4.1 History of PCR & sequencing

The discovery of the structure of DNA by Watson and Crick in 1953 and an understanding importance of the DNA sequence paved the way for the subsequent development of DNA sequencing techniques (Watson and Crick, 1953). Early sequencing methods were extremely time consuming, requiring the use of both electrophoresis and chromatography techniques to identify nucleotides. Technological constraints meant the method was only suitable for very short DNA fragments but they did enable the sequencing of short operator motifs from the *E. coli* lac operon among others (Gilbert and Maxam, 1973). In 1975, Sanger developed a new method of sequencing, relying on electrophoresis to separate DNA fragments by length and the use of radio-labelled nucleotides (Sanger and Coulson, 1975). The method was a significant improvement upon previous techniques and enabled the sequencing of the bacteriophage φX174 genome in 1978, the first instance of full-genome sequencing in history (Sanger *et al.*, 1978). Sanger also pioneered 'dideoxy' chain-termination sequencing (also known as Sanger sequencing), significantly increasing the length of DNA fragments that could be accurately sequenced (figure 1-6) (Sanger, Nicklen and Coulson, 1977).

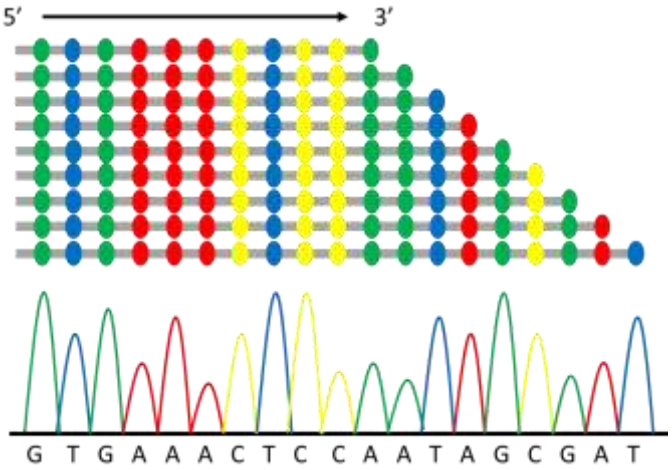


Figure 1-6; Dideoxy chain termination sequencing

An illustration of dideoxy chain termination sequencing and a typical chromatogram read out. Chain terminating nucleotides lacking a 3'-OH group are incorporated onto DNA fragments during replication with a DNA polymerase. Fragments are separated by length using gel electrophoresis and the dideoxynucleotides are labelled with fluorescent dye, enabling identification of the nucleotide at the chain end and subsequent construction of the sequence.

Dideoxy chain termination sequencing revolutionised sequencing and the study of DNA. Progress was made extending the length of fragments that could be sequenced and the technique became more accessible. The generation of increasing amounts of sequence data lead to the development of Genbank by the National Institute for Health in 1982 (Bilofsky and Burks, 1988). Genbank was established as a central repository for sequences and the researchers working with them, making the data generated highly accessible. Shortly after the establishment of Genbank, programs such as FASTA and BLAST were developed, enabling rapid comparisons between sequences and increasing the feasibility of many previously complex tasks (Pearson and Lipman, 1988; Altschul *et al.*, 1990).

The development of PCR by Kray Mullis in the early 1980s was another technique that revolutionised biomedical research (Mullis *et al.*, 1986). The PCR technique was developed to enable the rapid amplification of a specific DNA region from relatively low amounts of template. The method is based on a series of denaturation, primer annealing and extension steps which utilise repetitive thermo-cycling to generate new copies of a DNA template. Crucial to the evolution of the PCR technique was the discovery in 1976 of

the Taq polymerase (Chien, Edgar and Trela, 1976). The Taq polymerase is a DNA polymerase isolated from *Thermus aquaticus,* a thermophilic bacterial species that lives at temperatures of 50-80°C. The enzymatic activity of the Taq polymerase consequently functions well at high temperatures and this particular property was harnessed to enable the thermo-cycling of reactions during PCR. Selective targeting of the reaction to a specific region is accomplished by changing the primer sequences. The basic steps included in a PCR protocol which would typically be repeated for 20-40 cycles are listed below -

Denaturation – DNA molecule split into two template strands.
Annealing – primers are bound to the template strands of DNA.
Extension – the polymerase enzyme extends the sequence using nucleotides.

Automation of both thermo-cycling and sequencing techniques meant that increasingly larger volumes of work were possible and numerous variants of the technique have subsequently emerged. Methods enabling the conversion of RNA into complementary (c)DNA using a reverse transcriptase encoded by retroviruses (RT-PCR) were developed (Baltimore, 1970; Temin and Mizutani, 1970). RT-PCR has numerous applications in molecular biology including transcriptomics and the study of RNA viruses. Methods allowing the quantification of DNA (qPCR) during the amplification process have also been refined, using DNA probes with a reporter attached. As PCR-based amplification of the target occurs, probes binding specifically to the amplicon anneal and the amount of DNA present can be estimated by measuring the concentration of the reporter molecule. qPCR has had numerous applications including the measurement of gene expression, viral load quantification and pharmacological response monitoring.

1.4.2 Development of deep sequencing technologies

The development of capillary-based sequencing methods, using replaceable polyacrylamide matrices was a pivotal development, vastly improving the speed, resolution and separation efficiency of previous sequencing methods (Swerdlow and Gesteland, 1990; Pariat *et al.*, 1993). Building on this, the use of innovative chemistries has been harnessed to facilitate the rapid, low cost production of large amounts of data enabling the massively parallel sequencing of millions of reads. There are a number of platforms currently available, all using slightly different methods but the two market leaders have been 454 pyrosequencing and Illumina sequencing, using flow-cell technology.

454 pyrosequencing, developed by the European Molecular Biology Laboratory (EMBL), became the first commercially available deep sequencing platform in 2005 (Margulies *et al.*, 2005). During the process, fragmented single-stranded DNA is attached onto beads which then undergo emulsion-based PCR amplification. The beads are separated into wells, immobilised and nucleotides are repeatedly washed over them. Nucleotides binding to the template strands induce a light signal that is nucleotide specific. The reaction intensity is relative to the number of bound nucleotides and the emissions are recorded as a flowgram. Relatively long read lengths of 400-1000bp are possible with this platform.

In 2006 the first Illumina deep sequencing platform also became available, using bridge-based amplification within flow cells (Bennett *et al.*, 2005). In Illumina-based sequencing, adapters are ligated onto fragments of DNA which are then amplified using PCR (figure 1-7). The fragments are then immobilised into clusters on a flow cell coated in oligonucleotides that are complementary to the adapters on the sequences. The unattached end then binds to adapters that act as primers on the flow cell, creating a 'bridge'. PCR is then initiated, the clusters are amplified and sequencing by synthesis occurs. Reversible terminator nucleotides, labelled with nucleotide-specific fluorescent dyes are used, allowing the detection of each nucleotide incorporated onto the sequence. This platform was originally limited by a short read length of 2 x 50-75bp but improvements have meant that the paired-end sequencing of up to 2 x 300bp sequences is now possible.

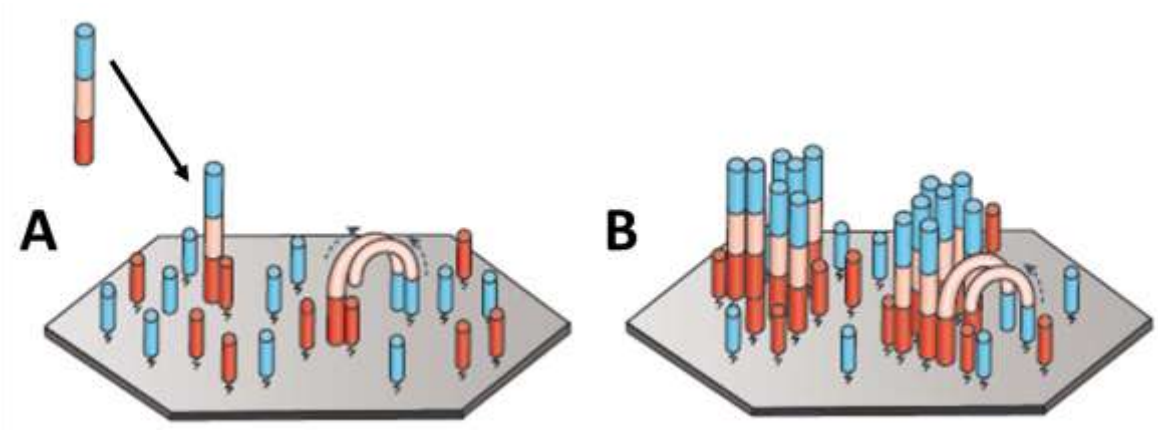


Figure 1-7; Illumina sequencing

Illumina sequencing process showing (A) the immobilisation of fragments onto flow cell using adapter sequences and the initiation of bridge-based amplification and (B) the generation of clusters for sequencing by synthesis. Figure taken from Goodwin et al., (2016)(Goodwin, McPherson and McCombie, 2016).

Since the introduction of commercially-available deep sequencing, there has been an exponential growth in the submissions of whole genome sequences to Genbank. Deep sequencing technologies are driving an 'omics' rather than a gene-centric approach to research in many areas of biology and whole genome sequencing is an increasingly affordable option for many studies. In virology, the technology offers opportunities to analyse highly diverse and rapidly evolving populations on a fine scale. New, third-generation sequencing platforms seeking to overcome the limitations of read length and accuracy are also in development. Among the most promising platforms are Helicos single molecule sequencing and Pacific Bio, designed to sequence reads of up to 10,000bp in length and offering the potential for the full genome sequencing of RNA and small DNA viruses (Korlach *et al.*, 2010; Thompson and Steinmann, 2010).

1.4.3 Applications of deep sequencing

Developments in the field of genetics have contributed towards improvements in clinical diagnostics. PCR, qPCR and multiplex PCRs are routinely used in clinical laboratories across the world to detect infections, screen donated blood and measure viral loads. Deep sequencing technologies have made the rapid sequencing of large numbers of sequences affordable and their use is becoming increasingly common within research laboratories across the world. In the near future, more advanced chemistries and the development of

improved analytical techniques offer the potential for the integration of deep sequencing into routine testing protocols in diagnostic laboratories.

Deep sequencing has enabled culture-independent virus discovery using d*e novo* sequence reconstruction techniques, often referred to as meta-genomic approaches. Reconstructed viral sequences isolated from samples can be compared with databases of known pathogens to look for similarity. If no matches are found, the genomic structure can be analysed and used to search for genera-specific patterns. This approach can be used in cases where the aetiological agent is unknown and, although a relatively high viral load is required for robust results, the method is much quicker than previously available protocols. It has been used to identify Astrovirus VA1/HMO-C in immunocompromised patients with encephalitis (Brown *et al.*, 2015), novel human papilloma viruses causing respiratory disease (Mokili *et al.*, 2013) and a new Rhabdovirus associated with haemorrhagic fever in the Congo (Grard *et al.*, 2012), among other examples. Meta-genomic techniques are frequently used in conjunction with target enrichment strategies which use custom-designed oligonucleotide probes bound to beads to selectively capture and isolate genomic fragments, increasing the concentration of the genomic material of interest (Linnarsson, 2010). The technique can also be used to selectively remove human genomic material from samples, increasing the sequencing sensitivity for low viral load samples. Metagenomics and target enrichment have applications in microbiome and virome analysis, enabling an improved understanding of the microbial diversity present within organisms. The human microbiome and its role in health and disease has now been widely studied, unravelling the complex functional roles of the microbial communities within us (Cho and Blaser, 2012; Delwart, 2013). For low-viral load samples and low-variant minor strain detection, amplicon-based deep sequencing has also been used. In order to generate amplicons for sequencing, prior knowledge of the region of interest is required and the amplicon length is constrained by the read length of the platform used. Advantages of the approach include an increased sensitivity and depth for the region of interest relative to metagenomics approaches (Lefterova *et al.*, 2015). Studies have used the technique successfully to detect drug resistant minor stains in HIV- (Simen *et al.*, 2009) and HCV-infected patients (Abdelrahman *et al.*, 2015).

The large amounts of data generated by deep sequencing allow analysis of complex microbial and viral populations, including quasispecies. Within the clinic, this technology has been successfully used to detect pre-existing drug-resistant minority variants of HIV

and HCV, informing on treatment plans for patients (Svarovskaia *et al.*, 2012; Shao *et al.*, 2013). Deep sequencing provides a useful tool for monitoring intra-host population dynamics, allowing the monitoring of disease progression and improving the understanding of the interaction between host and immune system. Deep sequencing also offers the potential to analyse virus transmission dynamics on a fine scale. In complex epidemiological studies, the genomic region sequenced and the analytical methods used can all impact on the eventual conclusions drawn (Campo *et al.*, 2015). Deep sequencing methods allow whole viral genome sequencing and the detection of minority variants, significantly improving resolution of the linkages within the transmission network. During the recent Ebola outbreak in West Africa, a portable deep sequencing kit allowed a real-time analysis of the outbreak, informing public health bodies on the effectiveness of control measures (Carroll *et al.*, 2015).

In other fields, the deep sequencing revolution has meant that the full sequencing of the human genome is now relatively affordable. Projects such as the 100,000 genomes project have been established to enable genome-wide association studies for rare disease causing variants (Siva, 2015). Transcriptome analysis has been used to develop our knowledge of disease progression and identify novel therapeutic targets (Craig *et al.*, 2013). In oncology, deep sequencing has highlighted the heterogeneous nature of many tumours, improving our understanding of the disease and its progression (Bedard *et al.*, 2013). Whole genome and transcriptome sequencing have been used to rapidly identify clinically-informative mutations in cancerous tumours, with patient treatment being tailored to the findings (Roychowdhury *et al.*, 2012). This personalised approach to oncology has the potential to improve our understanding of disease-associated biomarkers, enabling early intervention and preventative treatment in many cases.

Numerous barriers remain however before the technology can be successfully utilised in the clinic. High throughput sequencing methods are also notoriously error-prone currently, with each specific platform having particular constraints associated with it and protocols for the platforms are technically demanding. Contamination during library preparation from reagents and the laboratory environment is common and can severely impact the interpretation of the resulting data. The large amounts of data produced by high-throughput sequencing methods pose a problem, and there has been a significant acceleration in the amount of data produced in recent years. Increasingly expensive storage solutions are required to facilitate the current pace of data generation. In many cases, the biologically

and clinically relevant information may be hidden by large amounts of false positives and noise. Frequently, the development of custom pipelines by specialist bioinformaticians is required to analyse the data accurately and this again adds a significant cost to the analysis. Laboratories often lack individuals with a bioinformatic skillset, making it difficult to analyse the data accurately. There are ethical debates surrounding the use and ownership of the genetic data gathered which remain to be addressed. It is also possible that incidental findings which are clinically relevant may be uncovered during deep sequencing and this can present a challenging ethical dilemma for scientists. Understanding the limitations of these technologies and controlling for them accurately will be a vital part of adapting them for use in the clinic.

1.5 Mixed hepatitis C virus infections

1.5.1 Mixed infections

The failure of individuals infected with HCV to develop protective immunity means that patients can be infected with multiple HCV viruses, leading to mixed HCV infections. These infections have been documented at the genotype, sub-genotype and strain level. Mixed genotype HCV infections are largely grouped into two different infection processes, co-infection and superinfection (Blackard and Sherman, 2007). Co-infections occur when an individual is infected simultaneously with two or more different HCV viruses (figure 1-8). Superinfection occurs when a secondary HCV infection occurs in the background of an already present HCV infection. The two types of infection are difficult to distinguish clinically, largely due to the lack of routine serial screening of at-risk individuals. Many tests used in the clinical setting lack the sensitivity and specificity required for the diagnosis of mixed infection, therefore these infections are rarely detected (Cunningham *et al.*, 2015). Indirect evidence of mixed genotype HCV infections can be observed in reports of inter-genotypic recombinant HCV strains that have been detected (Du *et al.*, 2012; Hedskog *et al.*, 2014). Whilst recombination is thought to be a rare occurrence due to the replication strategy of the virus, the detection of recombinants does point to the simultaneous presence of two viruses in an individual.

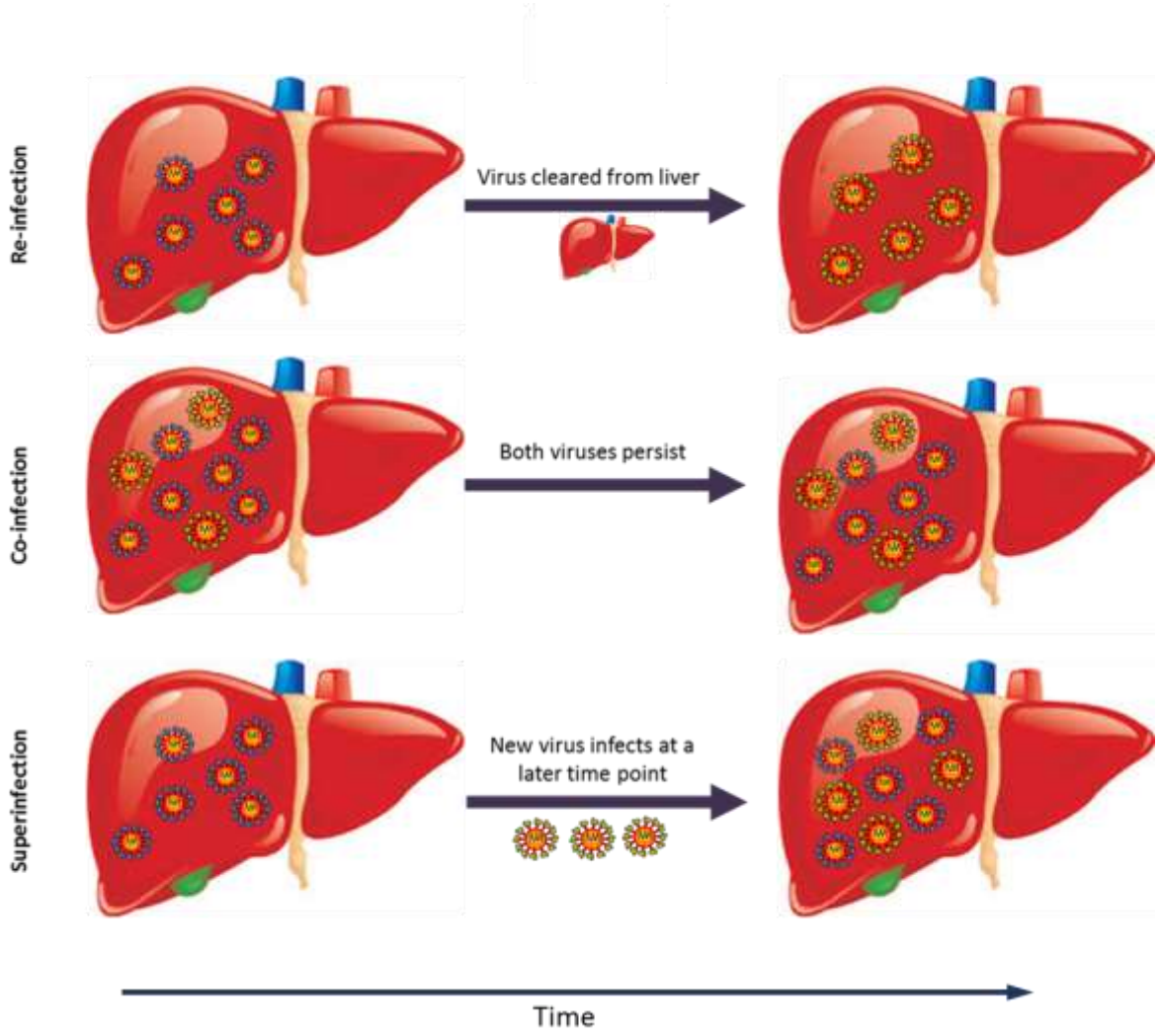


Figure 1-8; Re-infection, co-infection and superinfection

Illustration highlighting the differences between re-infection, co-infection and superinfection with HCV. Figure adapted from Blackard *et al*., (2007) (Blackard and Sherman, 2007).

As mixed infections are rarely detected clinically and cohorts of them are challenging to curate for longitudinal studies, information on the clinical outcomes of those with mixed infections is sparse. The numbers of mixed infection positive patients identified in studies are also frequently too low to obtain meaningful statistical power (table 1-2) (Cunningham *et al.*, 2015). The natural history of mixed genotype HCV infection is poorly understood and it is unclear if such infections are transient or if both strains are able to persist throughout the lifetime of the host. Previous work on the dynamics of superinfecting HCV strains in patients undergoing liver graft surgery found that mixed infections are usually quick to resolve and that in all cases studied (n=6) the virus with the greatest diversity went on to become the dominant virus strain (Ramirez *et al.*, 2010).

1.5.2 The prevalence of mixed infection

Mixed genotype HCV infections have been reported in the literature previously and there have been numerous studies to assess their prevalence across numerous geographical regions. A large amount of variation is observed in the prevalence rates of mixed infection reported in these studies with estimates of between 1.0-25.3% being reported (Sereno, Perinelli and Laghi, 2009; Pham *et al.*, 2010). Taking an average of the studies identified, the average prevalence of mixed HCV infections was estimated to be 9.1 ± 0.07%.

Table 1-2; Published studies on the prevalence of mixed HCV infections

Selected publications highlighting the significant variation among mixed HCV infection prevalence estimates. PWID denotes people who inject drugs. ND indicates the dominant genotype was not declared in the study. Where appropriate, reinfections have been excluded from the mixed HCV prevalence rates reported in the studies.

Study	Number tested	Testing method	HCV region targeted	Demographic information	HCV genotypes detected in mixed infections	Dominant genotype	Mixed HCV prevalence
(Lau *et al.*, 1996)	107	LiPA	5'UTR	USA, chronic HCV patients	1a,1b,2a,2b, 3a, 4*	ND	3.7%
(Giannini *et al.*, 1999)	213	N-PCR/RFLP	Core	Italy, chronic liver disease	1a,1b,2 *	ND	10.80%
(Natov *et al.*, 1999)	224	RFLP	5'UTR	US, renal patients	1a,1b,2a,2b,3a,4 †	ND	10%
(Blatt *et al.*, 2000)	6807	LiPA	5'UTR	USA, not selective	1,2,3,4,5,6	ND	4%
(Stamenkovic *et al.*, 2000)	117	N-PCR/LiPA	Core	Siberia, chronic HCV	1a,1b,2,3,4 *	ND	9.1%
(White *et al.*, 2000)	38	HMA	5'UTR	Australia, PWID	1a,1a,2a,2b,3a	ND	8%
(Schröter *et al.*, 2003)	600	N-PCR	5'UTR	Germany, not selective	1a,1b,2a,2b,3a,4a	ND	2.20%
(Singh, Malhotra and Sarin, 2004)	36	LiPA	5'UTR	India, chronic HCV	1a,1b,2,2b,3a,3b,4a,4e,5a †	ND	11.10%
(van Asten and Prins, 2004)	126	N-PCR	Core	Europe, PWID	1b,3a,4	ND	5%
(Herring *et al.*, 2004)	12	N-PCR	E1-E2	US, acutely infected	1a,1b,3a *	ND	25%
(Bowden *et al.*, 2005)	138	LiPA/HMA	5'UTR	Australia, PWID	1a,3a,6	ND	3.6%
(Tsatsralt-Od *et al.*, 2005)	27	PCR	5'UTR	Mongolia, liver disease	1a,1b,2a,2b,3a	ND	6.60%
(Buckton *et al.*, 2006)	44	RFLP/N-PCR	5'UTR	UK, PWID	1a,1b,2,3a †	1a (71%)	9%
(Buckton *et al.*, 2006)	37	RFLP/N-PCR	5'UTR	UK, haemophiliacs	1a,1b,2,3a,5 †	1b (50%), 3a (50%)	19%

Study	Number tested	Testing method	HCV region targeted	Demographic information	HCV genotypes detected in mixed infections	Dominant genotype	Mixed HCV prevalence
(Li *et al.*, 2008)	374	N-PCR/HMA	Core & E1-E2	US, PWID	1a,1b,2a,2b,3a	ND	7.50%
(Micalessi *et al.*, 2008)	98	LiPA	5'UTR	Belgium, PWID	3,4	ND	1%
(Yun *et al.*, 2008)	107	LiPA	5'UTR	Korea, PWID	1a,2a,2c *	ND	5%
(Ayesh *et al.*, 2009)	92	Real-time PCR	5'UTR	Gaza strip, chronic HCV	1,4	ND	7.6%
(Sereno, Perinelli and Laghi, 2009)	200	LiPA	5'UTR	Italy, PWID	1b,3a	ND	1%
(van de Laar *et al.*, 2009)	59	N-PCR	NS5B	Amsterdam, PWID	1a,1b, 2a, 2b,3a.4a,4d * †	ND	23.80%
(Pham *et al.*, 2010)	87	Real-time PCR	Core	Australia, PWID in prison	1a,1b,2a,2b,3a	1a (33%), 3a (38%)	25.30%
(Butt *et al.*, 2011)	22,125	PCR	5'UTR & Core	Pakistan, nationwide survey	1a,1b,1c,2a,2b,3a,3b,4,6a *	ND	5.50%
(Du *et al.*, 2012)	164	N-PCR	E2 & NS5B	China, PWID & non-PWID	1a,3a,3b *	ND	1.20%
(Grebely, Pham, *et al.*, 2012)	163	Real-time PCR	E1-E2	Australia, treatment cohort	1a,1b,2a,2b,3a *	3a (67%)	16.50%
(Hairul Aini *et al.*, 2012)	40	N-PCR	5'UTR & NS5B	Malaysia, haemophiliacs	1,3,4	ND	12.50%
(McNaughton *et al.*, 2014)	47	N-PCR	E1-E2	UK, PWID	1a, 3a	3a (100%)	10.6%
(Gowin *et al.*, 2016)	1159	LiPA	5'UTR	Poland, Chronic HCV	1a,1b,3a,4a *	ND	2.2%

Testing methods: HMA – Heteroduplex mobility assay, RFLP – Restriction fragment length polymorphism, LiPA – Line probe assay, N-PCR – Nested PCR

HCV genotypes detected: *Indicates inter-genotype infections were detected, †Indicates subjects infected with ≥3 genotypes detected

Studies also indicated that mixed HCV infections can be found globally and that a wide range of genotypes are involved. The greatest prevalence rates were associated with studies sampling from incarcerated, IDU and haemophiliac populations. Among incarcerated populations, exposure rates can be extremely high, caused by a combination of widespread drug use and tattooing and a poor availability of clean needles (Hellard, Hocking and Crofts, 2004; Hellard, Aitken and Hocking, 2007). Prior to the introduction of testing for HCV in 1990, haemophiliacs were at particular risk for multiple exposures as the clotting factor transfusions required to treat the disease were produced using components from multiple donations. Improved testing protocols and increased awareness of this issue has drastically reduced transmission risks in this group of patients. The high variability in mixed infection prevalence rates is likely due to differences in the study design, populations sampled and the detection methods used. Studies also showed discrepancies in the level of classification used with some examining mixed infections at a genotype level and others investigating at the sub-genotype or strain level. Studies of acutely-infected individuals also tended to report relatively high rates of mixed HCV infection. A recent study using Bayesian modelling methodologies has suggested that mixed genotype infections may be occurring more frequently than previously thought during the acute infection phase with unobserved rapid spontaneous clearance of superinfecting strains (Sacks-Davis *et al.*, 2015). This implies that these infections are relatively common but that by the time most individuals are diagnosed after chronic HCV has established, these mixed infections have resolved, leaving a single dominant strain.

1.5.3 Clinical consequences of mixed infection

As mixed genotype HCV infections are rarely detected in the clinical setting, data on the clinical implications of these infections is sparse and the impact of such mixed infections on disease progression, treatment outcome and quasispecies evolution is poorly understood. Numerous reports of partially resolved mixed infections and minor strain outgrowth during the treatment of mixed HCV infections have been documented (Schröter *et al.*, 2003; Grebely, Pham, *et al.*, 2012; Abdelrahman *et al.*, 2015). Given that many of the new DAAs for HCV have genotype-specific efficacies, our lab group at the Centre for Virus Research in Glasgow has highlighted concerns that the treatment of mixed genotype infections with such regimens may only partially resolve infections and patients will appear to have switched genotype (McNaughton *et al.*, 2014). Clinically such presentations may also lead clinicians treating individuals with HCV to suspect patients have been re-

exposed to HCV and this may lead to distrust between the clinicians and patients, which may also be detrimental to recovery.

There has been suggestions in the literature that having mixed HCV infections can lead to increased liver damage relative to individuals infected with a single virus (Kao *et al.*, 1994; Widell *et al.*, 1995). Studies that documented super-infections found that infection with a secondary HCV virus is frequently accompanied by a transient elevation in alanine transaminase (ALT) levels, suggesting damage to the liver may occur (Widell *et al.*, 1995; Kao *et al.*, 1996; Grebely, Pham, *et al.*, 2012). Some studies have also suggested that mixed infection is a relatively transient process with one virus rapidly becoming dominant and the other clearing (Laskus *et al.*, 2001; Fan *et al.*, 2003). Work by Toyoda *et al.* (1998) also identified increased quasispecies diversity among individuals with mixed infections relative to those with a single-strain infection, implicating differences in the immune response to mono- and multiple infections (Toyoda *et al.*, 1998).

The presence of multiple HCV strains within an individual also indicates that viral recombination is a possibility. Whilst recombination among HCV viruses is thought to be rare, there have been several strains reported in the literature, involving recombinants derived from multiple different HCV genotypes (Lee *et al.*, 2010; Du *et al.*, 2012; Hedskog *et al.*, 2014). It is also probable that detection of recombination is underestimated as few diagnostic protocols analyse multiple regions of the HCV genome. In the reported studies, break points frequently occur at the NS2-NS3 junction although they have also been mapped within the structural genes (Cristina and Colina, 2006; González-Candelas, López-Labrador and Bracho, 2011). Several naturally occurring genetic polymorphisms conferring resistance to DAA have been reported and there is the potential for previously sensitive viruses to acquire resistance via recombination (Bartels *et al.*, 2008; Lopez-Labrador, Moya and Gonzalez-Candelas, 2008). With the expanded use of DAA treatments, HCV viruses are likely to be forced down alternative evolutionary pathways and it is possible that increases in the numbers of treatment-resistant recombinant strains may be observed in the future.

1.5.4 Superinfection exclusion

Reports of recombinant HCV strains contrast with previous studies indicating that HCV is capable of excluding secondary strains from infecting an already-infected hepatocyte, a

process referred to as super-infection exclusion (Schaller *et al.*, 2007; Tscherne *et al.*, 2007). Super-infection exclusion in hepatocyte-derived cells has been demonstrated *in vitro* for HCV and Webster *et al.* (2013) published evidence that the process occurs within the hepatocyte during RNA replication (Webster *et al.*, 2013). Data from the study indicated that different viral replicons were likely to be confined within separate replication complexes within the hepatocytes and that mitotic cell division was a key step in the exclusion process, possibly due to resulting disruption of the membranous web (Webster *et al.*, 2013). Relatively minor mutations occurring at NS5A C2447R and within the poly U/UC tract have however been shown to overcome these replicative constraints and enable the replication of multiple HCV strains within the same hepatocyte (Webster, Ott and Greene, 2013). Work by Evans *et al.* (2004) also found that the presence of a fit, stably replicating HCV strain can impair the replication capacity of a secondary, less fit strain (Evans, Rice and Goff, 2004). Other studies have suggested that the down-regulation of CLDN1 and occludin observed after HCV infection functions to block superinfection of secondary strains (Liu *et al.*, 2008).

Whilst evidence of superinfection exclusion has been demonstrated *in vitro*, evidence is hard to obtain *in vivo*. Nonetheless, studies on super infection exclusion raise interesting questions on the viral population structure within individuals with mixed infections. HCV has been shown to form distinct foci within the liver of infected individuals (Stiffler, Nguyen, J. a Sohn, *et al.*, 2009). Within the context of mixed infections it is not clear whether the different HCV genotypes would form distinct foci or if multiple genotypes or strains would be present within the same foci. The identification of recombinant strains suggests that, at least at some point, more than one strain must be present within the same cell, otherwise these recombinants could not have arisen. It may be the case that exclusion of closely related virus strains is possible but that at the genotype level, competition between the viruses present is more complex. Studies have also hypothesized that superinfection exclusion may provide a mechanism for the survival of less-fit viral strains, ensuring the propagation of a large pool of variants within infected hosts (Tscherne *et al.*, 2007).

1.5.5 Co-infections with other pathogens

In the majority of individuals infected with HCV a chronic infection establishes which is able to persist throughout the lifetime of the host. As a consequence of this, any infections

which occur subsequent to the original HCV infection occur in the context of an established infection. Interactions can occur between these co-infecting pathogens but the implications of this are poorly understood. Numerous viruses, including HIV and HBV, share similar transmission routes with HCV and as a result, these infections are common in individuals with HCV. As a result, healthcare professionals routinely screen patients for multiple blood-borne viruses (BBVs).

The presence of HIV in individuals infected with HCV has been reported to be associated with higher HCV viral loads, a rapid progression of liver disease and poor responses to treatment (Eyster *et al.*, 1994; Merchante *et al.*, 2006; Hua *et al.*, 2013; Lo Re *et al.*, 2014). Even in patients with well-controlled HIV and good CD4+ T cell counts, liver disease outcomes for HIV-HCV co-infected patients still remain poor (Lo Re *et al.*, 2014). Numerous pathogenic mechanisms of HIV which may be influencing the course of HCV infection have been identified including the promotion of fibrogenesis, the induction of hepatocyte apoptosis and the attenuation of CD8+ T cells responses (Kim *et al.*, 2005; Lin *et al.*, 2008; Babu *et al.*, 2009). HIV-HCV co-infection is a particular problem among MSMs and in sub-Saharan Africa where HCV is becoming a leading cause of death among the HIV-positive population (Rao *et al.*, 2016).

Co-infections of HCV with other icterogenic viruses including hepatitis A virus (HAV), hepatitis B virus (HBV) and hepatitis E virus (HEV) have all been reported in the literature (Devalle *et al.*, 2003; Tsatsralt-Od *et al.*, 2005; Bayram *et al.*, 2007). HBV shares a common parenteral transmission route with HCV whereas both HAV and HEV are transmitted by enteric routes. HCV and HBV co-infection has been associated with the development of fulminant hepatitis and increased rates of cirrhosis and HCC (Benvegnu *et al.*, 1994; Wu *et al.*, 1994; Zarski *et al.*, 1998; Chiaramonte *et al.*, 1999). Replicative interference has also been documented between HCV and HBV with evidence suggesting HCV has a strong inhibitory effect on HBV replication (Jardi *et al.*, 2001). Studies of co-infection with HAV have reported increased cases of jaundice, fulminant hepatitis and increased mortality rates (Vento *et al.*, 1998; Devalle *et al.*, 2003). In contrast to this, other studies have observed increased rates of spontaneous clearance of HCV subsequent to HAV infection and it has been hypothesized that HAV may suppress HCV replication (Deterding *et al.*, 2006; Cacopardo, Nunnari and Nigro, 2009). Cases of spontaneous clearance of HCV following co-infection with HBV have also been documented (Kong *et al.*, 2014; Yu *et al.*, 2014; Bulteel *et al.*, 2016). It may be that HAV or HBV superinfection

induces a powerful immune response within the liver resulting in the spontaneous clearance of HCV.

Epstein-Barr virus (EBV)-encoded nuclear antigen 1 (EBNA1), a multifunctional viral protein associated with extrachromosomal replication, has been shown to upregulate the rate of HCV replication *in vitro* by heterologous transactivation (Y Sugawara *et al.*, 1999). Given that EBV has been shown to infect the liver and EBNA1 is frequently detected in HCC, this does suggest a possible role for EBV in the development of liver disease during HCV infection (Markin, 1994; Yasuhiko Sugawara *et al.*, 1999). Co-infection with cytolomegavirus (CMV), a closely related virus to EBV, has also been correlated with poor treatment outcomes and increased rates of fibrosis among individuals with HCV (Bader el-Din *et al.*, 2011). Studies have also indicated that co-infection with non-viral pathogens, including schistosomiasis and *Helicobacter pylori,* may lead to increased liver disease (Kamal *et al.*, 2006; Esmat *et al.*, 2012).

1.6 Intra-host diversity of hepatitis C

1.6.1 Hepatitis C virus evolution within the host

The majority of mutations occurring in the HCV genome originate from errors made by the NS5B RdRp during replication as the protein lacks a proof-reading mechanism. The RdRp encoded by HCV is particularly error prone and the virus is estimated to make at least one error every time the genome is copied (Guedj *et al.*, 2010). The high error rate of the RdRp has been found to be a crucial virulence factor for HCV with increases or decreases to the RdRp error rate proving detrimental to the pathogenicity of the virus. This error-prone replication combined with the relatively fast replication rate of the virus, which is estimated to produce around 10^{12} virions per day, generates a highly diverse population of viruses within infected individuals (Neumann *et al.*, 1998). Selection within the population occurs as a result of extrinsic factors including the host immune system and treatment with anti-virals and the intrinsic functional constraints of viral gene products that limit the tolerance for mutations (Lohmann *et al.*, 2001). Antigenic variation within intra-host populations have been noted to be relatively convergent in contrast to the extreme divergence observed at the genetic level suggesting that whilst the viruses explore a large amount of sequence space, functional limitations are highly restrictive *in vivo* (Campo *et al.*, 2012).

Divergent intra-host evolutionary rates have been identified within the genotype 1 viruses, with gt1a viruses evolving at a faster rate than gt1b viruses (Gray *et al.*, 2011). Differences have also been noted in the rate of evolution of genotypes 2 and 5 relative to the other genotypes, possibly arising as a result of unusual epidemiologic trajectories (Salemi and Vandamme, 2002). Genotypic differences in the rate of evolution among HCV viruses are likely to be conferred by adaptation of the viruses to transmission within specific communities. Gray *et al*., have hypothesized that faster rates of evolution would be selected for within populations with high rates of transmission, such as PWID (Gray *et al.*, 2011). Likewise, among populations where individuals are infected for a long period of time prior to transmission into a new host, there would be less selection for rapidly evolving viruses. This suggests that within individuals with mixed genotype infections, viruses of different genotypes would evolve at varying rates. A study by Culasso *et al.*, following a patient infected with gt1a, gt2a and gt3 over a 13-year period, found that the three viral lineages followed distinctly different evolutionary trajectories (Culasso *et al.*, 2014). This suggests that within the same host, under the same selective pressures, the different HCV genotypes have the potential to diversify along divergent evolutionary pathways and the clinical implications of this for the patient remain unclear.

Differences in the rate of evolution observed within HCV viruses from intra-host and among-host datasets have been observed with an increased evolutionary rate present in intra-host studies (Gray *et al.*, 2011). Successful infection following transmission between hosts requires viral adaptation to a new environment and it is likely that host-specific adaptations conferring fitness in one individual may not be useful in other individuals. Studies have identified transmission-associated bottleneck events for HCV, implying that initial infection is established with a limited number of viral strains (Ou *et al.*, 2011; Wilke *et al.*, 2012). Ou *et al.* (2011) also identified a second population decline, occurring during the acute phase roughly 100 days post infection (Ou *et al.*, 2011). The second decline appeared pivotal in the disease progression with diversity remaining restricted from this point among those clearing the virus and new variants emerging in those developing chronic HCV (Ou *et al.*, 2011).

1.6.2 Evolution across the hepatitis C genome

The tolerance for mutations within HCV viruses varies considerably across the genome, resulting in an uneven distribution of variation throughout the different viral products. A study by Gray *et al*. (2011) looking at gt1 viruses found that whilst the majority of the genome evolved at a relatively constant rate, evolution within the E1-E2 region occurred at a much faster rate (Gray *et al.*, 2011). Pairwise distances between and within different genotypes along the genome are reflective of variation in these evolutionary rates (figure 1-9). Untranslated regions of the genome which have been shown to have complex secondary structure are highly conserved among HCV viruses with the 5'UTR showing up 90% homology between the different genotypes (Bukh, Purcell and Miller, 1992). Structural conservation within the 5'UTR is linked to its role in the formation of the IRES, a region noted for its extensive secondary structure, which plays a central role in the initiation of CAP-independent translation (Honda, Brown and Lemon, 1996). The core region of the virus (nucleotides 342-915) is also well conserved relative to other viral regions and again, RNA secondary structures have been identified within this region (Tuplin, Evans and Simmonds, 2004; Mauger *et al.*, 2015).

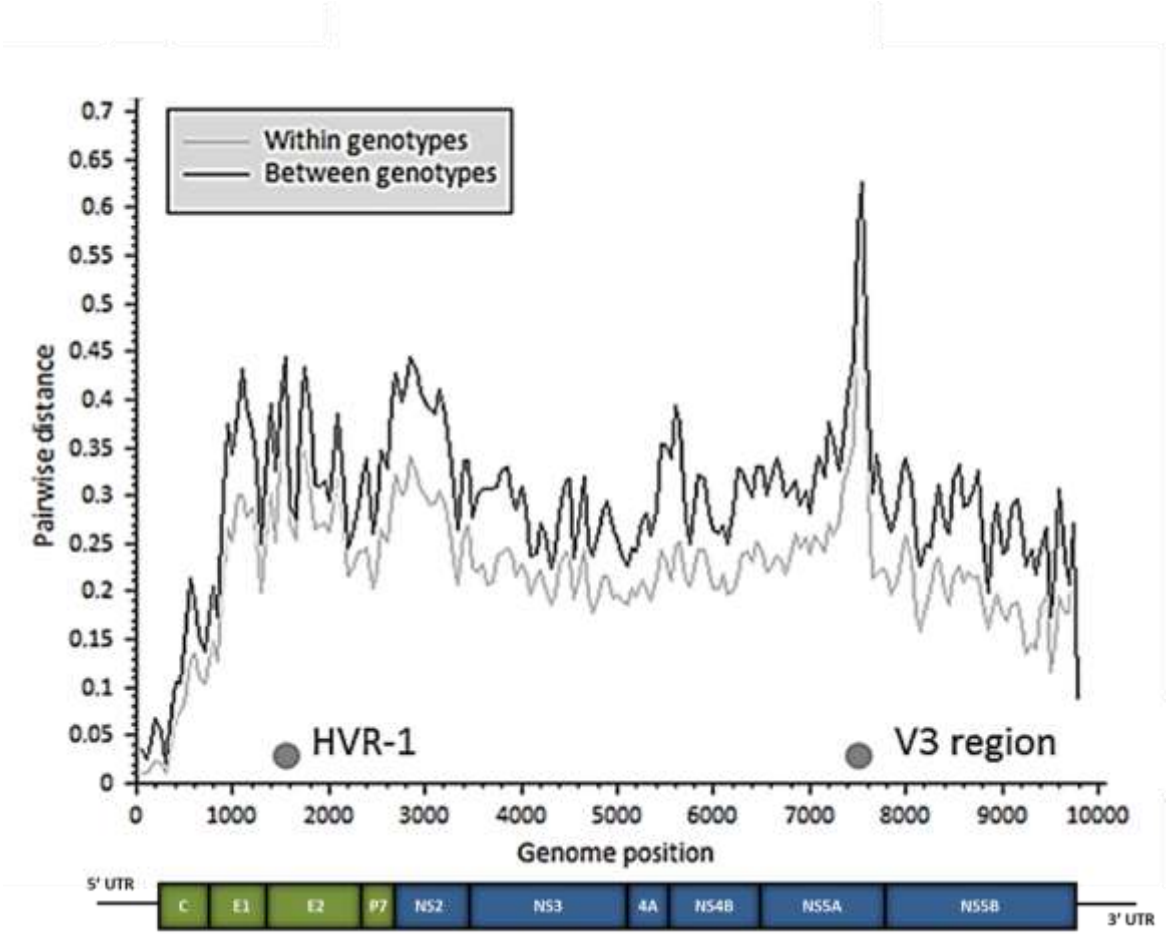


Figure 1-9; Within and between genotype diversity of HCV

p-distance scans of full-length HCV genomes comparing within and between genotype diversity, generated from 150 bp windows across the genome using SSE software (Simmonds, 2012). A schematic diagram of the HCV genome is shown underneath (adapted from figure 1-2).

The most divergent regions of the HCV genome are located within the envelope genes (E1 and E2) and NS5A (figure 1-9). Whilst a large amount of inter-genotypic NS2 diversity is also observed, this is largely between genotypes and the within- genotype diversity in the region is considerably lower. The V3 region located within NS5A is near to the interferon sensitivity-determining region and studies have correlated mutations in V3 with interferon resistance and poor treatment outcomes (Puig-Basagoiti *et al.*, 2005; Yuan *et al.*, 2010). Numerous B-cell and T-cell epitopes have been identified in NS5A, suggesting the region is under considerable immune selection (Zhang *et al.*, 1994; Dou *et al.*, 2002). Researchers have suggested that the high diversity observed in the V3 region is a result of different genotypic adaptations to the evasion of intracellular host defences (Simmonds, 2004). Within E2 there are two highly variable regions, HVR-1 and HVR-2. Antibody, CD4+ and CD8+ T cell epitopes have all been identified within the HVR-1 region, suggesting

diversity within the region is largely driven by immune-mediated pressure (Shirai *et al.*, 1999; Farci, 2000; Sarobe *et al.*, 2006). Numerous studies have also correlated high intra-host variation of HVR-1 sequences with increased likelihood of progression to chronic disease (Farci, 2000; Chen and Wang, 2005; Thomson *et al.*, 2011). Likewise restricted intra-host diversity has been associated with increased rates of spontaneous clearance during acute HCV (Farci, 2000; Thomson *et al.*, 2011). A putative role for HVR-2 mediating the structural stability of the E1-E2 heterodimer has been proposed (McCaffrey *et al.*, 2011).

These observed differences in variability have important implications in the design of studies involving sequence analysis and genotyping. The high rates of evolution within the E1-E2 region mean that whilst the region is useful for intra-host studies, the large number of mutations within the region make epidemiological inference at the population level difficult (Goncalves Rossi, Escobar-Gutierrez and Rahal, 2016). Highly conserved regions such as the 5'UTR and core show little discrimination between sub-genotypes and evolve slowly over time making them unsuitable for evolutionary studies. Detection of the 5'UTR and core regions are however useful biomarkers indicating the presence of an active HCV infection and are widely used in diagnostic laboratories to diagnose infection and for genotyping.

1.6.3 Quasispecies

Rates of evolution of HCV within an infected individual are estimated to be 10^{-2} to 10^{-3} substitutions per site per year, placing HCV amongst the fastest mutating RNA viruses (Domingo, 2007; Lutchman *et al.*, 2007). Consequently, the viral population in individuals infected with HCV is composed of many closely related but genetically heterologous variants and is often referred to as a quasispecies (Holland, De La Torre and Steinhauer, 1992; Bukh, Miller and Purcell, 1995). Within a quasispecies, viral variants are linked within a mutational network and the diversity maintained within the population is a crucial survival mechanism, enabling the virus to rapidly adapt to dynamic environmental conditions. The high mutation rate of many RNA viruses is widely considered to be a key virulence factor for these pathogens, without which they would struggle to establish productive infections. Numerous studies have shown that increasing the fidelity of RNA virus polymerases can severely attenuate the pathogenicity of a viral population *in vivo* (Pfeiffer and Kirkegaard, 2005; Vignuzzi *et al.*, 2006). In quasispecies theory, and contrary

to typical Darwinian models of evolution, selective pressures are thought to act on the population as a whole, rather than acting on individual variants within it and fitness within a quasispecies population is often considered in terms of 'survival of the flattest' rather than the more typical 'survival of the fittest' (Wilke *et al.*, 2001). 'Survival of the flattest' describes a fitness model where, among rapidly producing viruses, it is considered advantageous to occupy a large amount of sequence space within a fitness landscape rather than be highly adapted (or 'fit') to a particular niche within the landscape. By exploring a large amount of sequence space, viruses are able to mitigate for their high mutation rates and remain highly adaptable to environmental changes, therefore maintaining their genetic robustness.

In HCV, the highly heterogeneous nature of the quasispecies has been shown to have a vital role in viral persistence, disease progression and response to treatment (Domingo, Sheldon and Perales, 2012). Diversity within the population has been shown to be predictive of progression towards chronicity (Farci, 2000). During HIV co-infection where the immune response is impaired, HCV viral diversity is reduced, likely a consequence of the reduced selection for immune escape variants (Shuhart *et al.*, 2006). This implies that the immune system is a strong selective factor for heterogeneity within the viral population. It is thought that the high mutation rate observed within HCV enables continual evasion from neutralising antibodies, providing an effective immune escape mechanism for the virus. Current understanding of the viral population structure supporting the observed diversity during HCV remains incomplete however. The majority of studies in the field characterise diversity based on samples collected from a select sub-set of the total viral population (often the blood serum), and evidence suggests that models based on such inference provide inadequate interpretations of intra-host viral evolution (Gray *et al.*, 2012). Within the liver HCV has been shown to replicate within discrete foci. Viral populations within the foci have been shown to be substantially divergent from the viral population circulating in the serum (Sobesky *et al.*, 2007; Stiffler, Nguyen, J. A. Sohn, *et al.*, 2009). An *in vitro* study by Leitch *et al.*, has also indicated that discrete viral populations can be detected at the cellular level, implying that the networks underlying the diversity observed within the liver are complex (McWilliam Leitch and McLauchlan, 2013). HCV has also been shown to compartmentalise during infection and genetically distinct viral populations have been isolated from numerous non-hepatic sites including peripheral blood mononucleocytes (PBMCs), lymphocytes and brain tissue (Laskus *et al.*, 2000, 2007; Radkowski *et al.*, 2002; Roque-Afonso *et al.*, 2005). Whilst the replicative

capacity of these viral sub-populations remains unclear, the detection of multiple viral reservoirs within HCV-infected individuals indicates the presence of a highly heterogeneous population structure capable of maintaining a complex quasispecies.

1.7 Dinucleotide frequency & codon use bias

1.7.1 Dinucleotide bias

Compositional biases among the dinucleotide frequencies of vertebrate genomes were first noted in the 1960s (JOSSE, KAISER and KORNBERG, 1961; Russell *et al.*, 1976). Since then, species-specific dinucleotide frequency signatures have been described for many organisms, suggesting that observed biases are reflective of the selective pressures placed upon organisms (Karlin *et al.*, 1995; Karlin and Mrazek, 1997). Among the observations noted, TpA (or UpA in mRNA and RNA genomes)has been noted to be repressed in almost all organisms studied and CpG is frequently repressed in vertebrate genomes (Karlin *et al.*, 1995). TpA dinucleotides are found within a number of key regulatory motifs including TATA box sequences and two of the three stop codons (Karlin and Ladunga, 1994). Repression of TpA in the genome is therefore thought to reduce the risk of inappropriate binding of transcriptional activators, maintaining the regulation of gene expression (Karlin and Ladunga, 1994; Karlin *et al.*, 1995). UpA and UpU sequences in mRNA have been shown to be preferential targets for ribonucleases, particularly endonuclease L (Beutler *et al.*, 1989). In transcriptionally active regions of the genome that are destined for expression within the cytosol via mRNA, TpA motifs are highly repressed, reducing the risk of degradation and extending the half-life of the mRNA. Avoidance of TpA dinucleotides therefore reduces the risk of nonsense mutations and increases the stability of RNA. Additionally TpA dinucleotides have the lowest thermodynamic stacking energy of all the dinucleotides, implying that the repression of TpA motifs may also function to facilitate an increase DNA stability (Breslauer *et al.*, 1986).

Within vertebrate genomes, CpG dinucleotides are frequently methylated, particularly in transcriptionally active regions. CpG methylation is thought to predispose the site to spontaneous deamination whereby the cytosine (C) residue is converted to a thymine (T) residue, resulting in point mutations across the genome (Bird, 1980). As a result, CpG dinucleotides are depleted over time and replaced with TpG dinucleotides resulting in a marked suppression of CpG nucleotides among many vertebrate genomes (Karlin *et al.*,

1995). Further suppression of CpG within the genome may also function to mitigate the risk of mutational damage to the genome. Among species that do not methylate their genomes, including insects and fungi, CpG repression is not observed (Karlin and Mrazek, 1997). Many viruses, and particularly RNA viruses with cytosolic replicative phases, have been observed to have similar dinucleotide biases to those of their hosts (Jenkins and Holmes, 2003; Greenbaum *et al.*, 2008). Among dsRNA viruses, where replication occurs in discrete intracellular compartments isolating the dsRNA from the cytosol, CpG dinucleotides are generally not suppressed (figure 1-10) (Simmonds *et al.*, 2013). Numerous studies have hypothesised that dinucleotide mimicry of the host occurs among viruses in order to avoid immune detection (Karlin, Doerfler and Cardon, 1994; Belalov and Lukashev, 2013; Cheng *et al.*, 2013; Atkinson *et al.*, 2014). Unmethylated CpG motifs have been found to be immunostimulatory and are recognised by toll-like receptor 9 (TLR-9) (Uden and Raz, 2000; Weeratna *et al.*, 2000). Work by Atkinson *et al*., has suggested that further, uncharacterised sensors associated with the PKR pathway may play an additional role in the detection of RNA with an abnormally high UpA or CpG content (Atkinson *et al.*, 2014). Intriguingly, work on the 1918 pandemic influenza and the recent H5N1 strain has suggested that these viruses have a comparatively high CpG content relative to other strains. It has been hypothesised that this high CpG content may trigger stronger than usual innate immune reaction to these viruses, partially explaining the high mortality rate associated with these viruses (Chan *et al.*, 2005; Kobasa *et al.*, 2007; Greenbaum *et al.*, 2008). Studies have also suggested that the frequency of CpG and UpA dinucleotide pairs in the RNA genome may also influence the replicative capacity of the virus with studies in Theiler's murine encephalomyelitis virus (Atkinson *et al.*, 2014) and other viruses (Karlin, Doerfler and Cardon, 1994) indicating that viruses with lower than expected CpG and UpA frequencies were able to replicate faster and to greater levels than other strains.

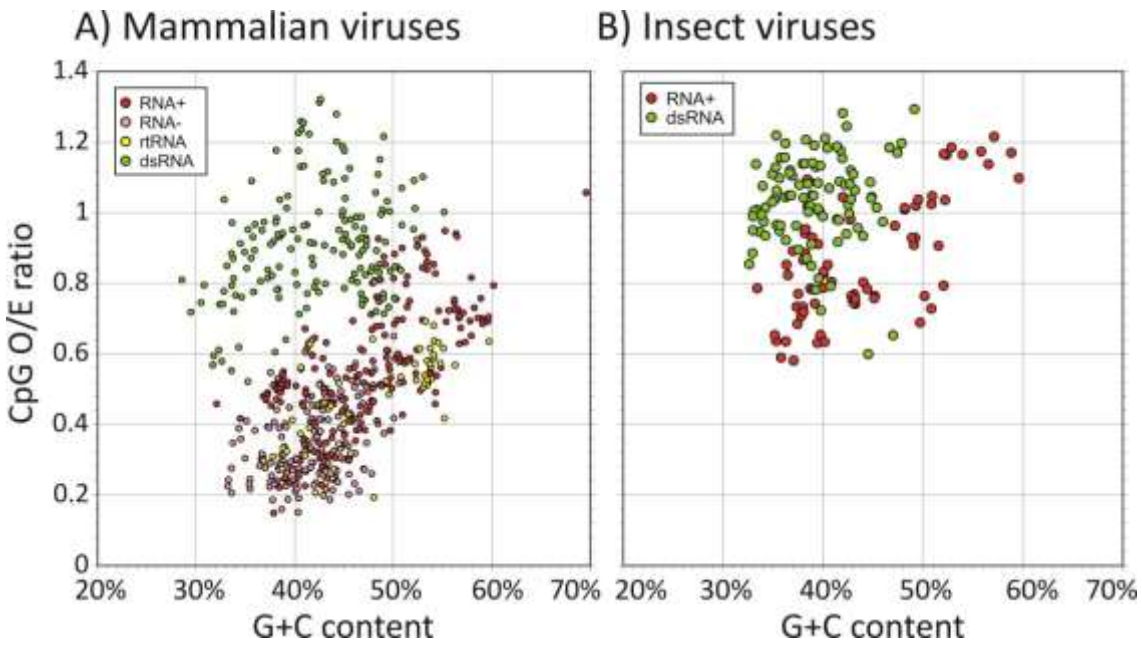


Figure 1-10; CpG expression in viruses infecting mammals and insects

CpG observed amount/expected amount ratio (O/E) among viruses infecting (A) mammalian and (B) insect species. Viruses have been categorised according to their genome type: RNA+, positive sense single-stranded RNA; RNA-, negative sense single-stranded RNA; rtRNA, retrovirus; dsRNA, double-stranded RNA. Illustration published by Simmonds et al. (2013) (Simmonds *et al.*, 2013).

1.7.2 Codon Use Bias

Shortly after the deciphering of the genetic code, it was realised that there was considerable redundancy within the system with 64 possible codons encoding just 20 amino acids (Lagerkvist, 1978). Such redundancy enables multiple codons to encode the same amino acid with degeneracy frequently observed in the second and third codon positions (Lehmann and Libchaber, 2008). Codon use bias within genomes is thought to evolve as a result of the translational and mutational pressures placed upon them (Plotkin and Kudla, 2011). Among many viral species, similarities to the host codon bias has been demonstrated and the replacement of preferred codons with randomised alternatives has been shown to severely attenuate viruses (Zhou *et al.*, 1999; Mueller *et al.*, 2006; Greenbaum *et al.*, 2008; Lobo *et al.*, 2009; Belalov and Lukashev, 2013). Codon de-optimisation has also been shown to be an effective method of generating attenuated strains for live viral vaccines (Baker, Nogales and Martinez-Sobrido, 2015; Cheng *et al.*, 2015; Diaz-San Segundo *et al.*, 2015). It is unclear how exactly codon de-optimisation functions to reduce viral infectivity. Given that viruses are reliant upon host cell proteins in order to replicate, numerous studies have suggested that reductions in translational

efficiency play role in viral attenuation by codon de-optimisation (Karlin, Blaisdell and Schachtel, 1990; Mueller *et al.*, 2006; Yang *et al.*, 2013). Recent studies by Kunec and Osterrieder (2016) and Tulloch *et al.*, (2014) have provided evidence that codon de-optimised viral vaccines are effective because of inadvertent increases in CpG (and to a lesser extent UpA) dinucleotides that occur during de-optimisation, not because of changes to the codon bias (Tulloch *et al.*, 2014; Kunec and Osterrieder, 2016). Kunec and Osterrieder (2016) also published evidence indicating that whilst viral dinucleotide frequency mimicked that of the host, the codon use bias frequently did not, appearing instead to be driven by the maintenance of the dinucleotide frequency (Kunec and Osterrieder, 2016).

1.8 Aims and objectives of the project

Within the UK, the two most frequently diagnosed genotypes are gt1a and gt3, each infecting around 45% of the total infected population (Public Health England, 2015). As both genotypes co-circulate and are transmitted within the IDU population in the UK it is likely that multiple genotype infections are occurring within this population. The only prevalence study within the UK to date identified mixed infections in 9% (4/44) of PWID (Buckton *et al.*, 2006). Mixed genotype infections are rarely identified in the clinic as most currently employed diagnostic protocols only detect the major genotype present. Many of the newly licensed DAA treatments have been documented to have genotype-specific efficacies that are particularly adept at treating gt1 infections but have poorer efficacies against gt3 infections (EASL, 2015). We have previously hypothesized that in regions such as the UK where non-gt1 viruses are highly prevalent, undiagnosed mixed genotype infections may respond poorly to some DAAs, possibly resulting in minor strain outgrowth and genotype switching (McNaughton *et al.*, 2014). Given the recent increase in HCV treatment options, a larger study, encompassing a wider range of patients, may be informative for clinicians and researchers within the field.

The aims for the project were, therefore –

1) To screen a large cohort of gt1a and gt3 HCV positive patient samples in order to determine the mixed gt1a/gt3 infection prevalence within the UK population. This will be achieved by developing and optimising highly sensitive and specific PCR-based methods that can be used for the detection of mixed gt1a/gt3 infection. Sanger-based sequencing will be used to perform phylogenetic characterisation of the mixed genotype infections identified within our cohort.

2) To quantify the relative proportions of gt1a and gt3 present in samples with mixed infections. Genotype-specific qPCR assays for gt1a and gt3 will be developed for this purpose.

3) To explore the potential of deep sequencing based methodologies for the diagnosis of mixed gt1a/gt3 infections in the clinic. In order to achieve this, pan-genotypic PCR-based assays compatible with the Illumina MiSeq platform will be developed targeting highly informative regions of the viral genome that can used for viral sub-

typing. The efficacy of the deep sequencing approach for the identification of mixed genotype infections will be compared to the classical PCR method outlined in aim 1.

4) To correlate viral sequence data in conjunction with linked clinical data to investigate differences between gt1a and gt3 viruses.

5) As a part of a general comparative analysis of viral strains involved in mono and dual genotype HCV infections, we investigated dinucleotide frequencies and unexpectedly identified interesting variances in CpG expression. This led to an additional project aim to investigate dinucleotide frequencies and RSCU in HCV and related flaviviruses. This included an analysis of the genomes, genomic regions and temporal divergence of the of HCV and flavivirus host range.

Chapter 2: Materials & Methods

2.1 Materials

2.1.1 Chemicals

Table 2-1; Chemicals

Chemical	Supplier
Agarose	Life Technologies
Ampicillin	Sigma-Aldrich
Ethanol	VWR International
Glycerol	Sigma-Aldrich
Isopropanol	VWR International
SYBR safe – DNA gel stain	Life Technologies
Nucleic acid loading dye	Life technologies
Water, distilled	Filtered in-house
Water, nuclease free	Qiagen

2.1.2 Solutions

Table 2-2; Solutions

Solutions	Components	Supplier
LB broth	2% tryptone, 0.5% yeast extract, 10mM NaCl, 2.5mM KCl,	Thermo Scientific
Luria-Bertani (LB) agar	1% peptone, 0.5% yeast extract, 0.5% NaCl, 1.2% agar	Thermo Scientific
RNA storage solution	1 mM sodium citrate, pH 6.4 (+/- 0.2)	Life Technologies
SPRI buffer	2.5 mNaCl, 20% PEG	KAPA Biosystems
Super Optimal Broth with Catabolite repression (SOC)	10mM $MgCl_2$,10mM $MgSO_4$ 20mM glucose	Sigma Aldrich
TAE (x50)	40 mM Tris-acetate, 1 mM EDTA, pH 8.3 when diluted to x1 solution	Life Technologies
Tris buffer	10mM Tris, pH 8.5	Sigma Aldrich

2.1.3 Enzymes

Table 2-3; Enzymes

Enzyme	Supplier
KOD Polymerase	Merck Millipore
RNAsin	Life Technologies
RNase H	Life Technologies
RNase OUT	Life Technologies
Proteinase K	Sigma-Aldrich
DNase I	Life Technologies
Illustra Exostar	Jencons
End repair enzyme	KAPA Biosystems
A-tailing enzyme	KAPA Biosystems
Mung bean nuclease	New England Biolabs (NEB)
Platinum Taq Polymerase	Life technologies
SuperScript III Reverse Transcriptase	Life Technologies
T4 Ligase	Thermo Scientific
T7 Transcriptase	Promega
Xba1 restriction enzyme	Promega

2.1.4 Kits

Table 2-4; Kits

Kit	Manufacturer
Ampure XP DNA beads	Agencourt
CloneJET PCR cloning kit	Thermo Scientific
High Sensitivity D1000 ScreenTape, reagents and ladder	Agilent technologies
Illumina Mi Seq V3 300bp paired end kit	Illumina
KAPA HiFi PCR kit	KAPA Biosystems
KAPA Library Preparation kit	KAPA Biosystems
KOD Hotstart PCR kit	Merck Millipore
Mini-prep kit	Qiagen
NEBNext® Multiplex Oligos for Illumina® (Index Primers Set 1)	New England Biosciences
Platinum Taq PCR kit	Life Technologies
QIAamp Viral RNA Mini Kit	Qiagen
Qubit dsDNA high sensitivity assay kit	Thermo Scientific
RNeasy mini kit	Qiagen
SuperScript III reverse transcription kit	Life Technologies
T7 RiboMAX Express Large Scale RNA Production System	Promega
TaqMan® Fast Advanced Master Mix	Life Technologies

2.1.5 Primers and probes

Primers and probes used for nested PCRs and qPCR reactions (table 2-5, primers A-H and table 2-8) were sourced from Eurofins and were manufactured using a desalting

purification. Primers used for deep sequencing (table 2-5, primers I-L) were PAGE-purified and obtained from Integrated DNA Technologies (IDT). Blunt-ended dsDNA fragments used to generate controls were ordered from and synthesized by Life Technologies. The numbering of the HCV primers and controls have been discussed based on the standardised numbering system suggested by Kuiken C *et al.*, (2006) using H77 (accession number AF009606) as a reference (Kuiken *et al.*, 2006).

2.1.5.1 PCR primers designed for this project

The pan-genotypic primers (primers I-L, table 2-5) that were used to generate the amplicons for deep sequencing were modified by the addition of a phosphorothioate bond between the last two nucleotides at the 3' end. This rendered the primers resistant to the 3' to 5' endonuclease activity of the KAPA HiFi proofreading enzyme.

Table 2-5; Primer sets

Genotype specificity	Region	Ref	Primer type	Sequence (5' → 3')	Tm (°C)	GC (%)	Length	Position (H77)
Primers for Sanger Sequencing								
Gt1a	E1-E2	A	OS	CAT ATA ACG GGY CAY CGC ATG G	54.8-58.6	50-59	22	1275-1296
		B	IS	ATG ATR ATG AAC TGG TCY CCY AC	51.7-57.1	39-52	23	1305-1327
		C	IAS	TYG TCC TYA AYA ACA CYA GRC C	49.2-58.6	36-59	22	1972-1951
		D	OAS	TGG TTY GGY TGY ACY TGG ATG AA	51.7-58.8	39-57	23	2008-1986
Gt3a	E1-E2	E	OS	TTY CTY GTG GGR CAA GCC TTC A	53-58.6	45-59	22	1203-1224
		F	IS	TTY AGA CCY CGY CGC CAT CA	51.8-57.9	50-65	20	1227-1246
		G	IAS	CAG AYG TGT TCY TGC TGR AGT C	53-58.6	45-59	22	1993-1972
		H	OAS	CCT YTW CTG CCC CAC YGA CTG	56.3-60.2	57-67	21	2143-2123
Primers for Deep Sequencing								
Pan-genotypic	E1-E2	I	S	GC NTG GGA YAT GAT GAT GAA YTG	51.7-57.1	39-52	23	1296-1318
		J	AS	GDG SGT ART GCC ARC ART ANG	50.5-60.2	43-67	21	1812-1792
Pan-genotypic	NS5B	K	S	ATG AYA CCM GVT GYT TYG ACT C	49.2-58.6	36-59	22	8257-8278
		L	AS	TAY CTS GTC ATA GCY TCC GTG A	53-56.7	45-55	22	8636-8617

2.1.5.2 Primers used for deep sequencing

Table 2-6; Dual indexing primer sets, New England Biolabs (NEB)

NEB Index	Sequence (5'→3')
NEBNext i501 Primer	TATAGCCT
NEBNext i502 Primer	ATAGAGGC
NEBNext i503 Primer	CCTATCCT
NEBNext i504 Primer	GGCTCTGA
NEBNext i505 Primer	AGGCGAAG
NEBNext i506 Primer	TAATCTTA
NEBNext i507 Primer	CAGGACGT
NEBNext i508 Primer	GTACTGAC
NEBNext i701 Primer	ATTACTCG
NEBNext i702 Primer	TCCGGAGA
NEBNext i703 Primer	CGCTCATT
NEBNext i704 Primer	GAGATTCC
NEBNext i705 Primer	ATTCAGAA
NEBNext i706 Primer	GAATTCGT
NEBNext i707 Primer	CTGAAGCT
NEBNext i708 Primer	TAATGCGC
NEBNext i709 Primer	CGGCTATG
NEBNext i710 Primer	TCCGCGAA
NEBNext i711 Primer	TCTCGCGC
NEBNext i712 Primer	AGCGATAG

Table 2-7; Adapter sequences (NEB)

NEB adapter sequence	Sequence (5'→3')
Adapter (read 1)	AGATCGGAAGAGCACACGTCTGAACTCCAGTCA
Adapter (read 2)	AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT

2.1.5.3 Primers used for qPCR

Table 2-8; Primer and probe sets used for qPCR.

*Primers published by Jones *et al*., (2010) (Jones *et al.*, 2010). All other primers were designed specifically for this study.

Genotype	Region	Ref	Primer type	Sequence (5' → 3')	Tm (°C)	GC (%)	Length	Position (H77)
Pan-genotypic	5'UTR	A	S (JFH-1 16)*	TCT GCG GAA CCG GTG AGT AC	55.9	60	20	333-352
		B	AS (JFH-1 17)*	GCA CTC GCA AGC GCC CTA TC	57.9	65	20	500-481
		C	Probe	FAM^{TM}-AAA GGC CTT GTG GTA CTG-MGB	48	50	18	460-477
GT1a	NS5B	D	S	CTG TCG AGC CGC AGG GCT C	59.7	74	19	8507-25
		E	AS	GCT CCA AGT CGT AYT CTG GYT GBG	57.9-62.0	50-63	20	8686-62
		F	Probe	FAM^{TM}-CCT CCG TGA AGG CTC TCA GGY TCG CYG CG-MGB	68.6-71.4	66-72	29	8625-597
GT3a	NS5B	G	S	GGA ACC CGG ACT TYC TYG TCT G	56.7-60.4	55-64	22	8527-48
		H	AS	CTC AAG GTC RTA GGT RGG CTG YGG	57.1-62.4	54-67	23	8684-61
		I	Probe	FAM^{TM}-CGA CGC CRT CAC TCT CRG CCA CCA CRA CYA G-MGB	67-72.3	58-71	31	8589-559

2.1.5.4 Synthetic DNA oligonucleotides

Table 2-9; Synthetic DNA oligonucleotides

Accession number	Genotype	Region	Position	Length
AF009606	1a	E1-E2	1128-2054	927
AF009606	1a	NS5B	8166-9137	972
GQ356206	3a	E1-E2	1167-2147	998
GQ356206	3a	NS5B	8217-9083	956

2.1.6 Bacteria

NEB 5-alpha competent *E. coli* cells (New England Biolabs) were used when required for the propagation of plasmids in cloning procedures.

2.2 Patient cohort

2.2.1 Ethics statement

Ethical approval for the use of patient samples in this project was granted by the NHS Research Scotland Greater Glasgow and Clyde Biorepository, application number 89. Ethical approval for the project allowed concurrent collection of anonymised linked clinical data for patients diagnosed as HCV positive by WSSVC during the period August 2013 to March 2014. The data was collected from the clinical portal system, providing data on patient contact with secondary care providers in the region. The information was anonymised by a University of Glasgow medical student, Harriet Mei Tan.

2.2.2 Sample collection

Sample collection for the project occurred in two phases. The initial phase, in 2011, involved the collection of samples from the West of Scotland Specialist Virology Centre (WSSVC) and the Lothian Specialist Virology Testing Centre (BBV Edinburgh). 25 gt1a and 25 gt3 HCV-positive serum samples were collected at each site from individuals with a history of injecting drug use. Between August 2013 and March 2014, an additional 411 gt1a and gt3 samples were collected from the WSSVC to expand the screening for mixed genotype HCV infections. For the second collection, only HCV-positive serum samples

that had been recently genotyped and had >140µl remaining were collected for the cohort. All samples were genotyped prior to collection by the diagnostic laboratory using an in-house method. The genotyping method in use at the WSSVC involved a multiplex q-PCR targeting the 5'UTR region with specific probes for gt1, gt2 and gt4/5 (see appendix, figure 8-3 for result interpretation panel). Diagnosis of gt3 was based on the absence of a specific result for the other genotypes and a further assay was used to differentiate gt1a and gt1b strains. The WSSVC were informed of all samples that were identified as having a mixed infection.

2.2.3 Statistical analyses

Clinical data linked to samples obtained from patients with HCV was analysed using SPSS software (IBM analytics) in conjunction with Excel (Microsoft) and GraphPad Prism software (GraphPad Software, Inc.). Differences in the means and distributions of data comprising two groups were compared using the independent samples t-test function within the SPSS or GraphPad Prism. For data where there were three or more groups, the ANOVA analysis option was used to compare the means using the type III sums of squares method. The significance of differences in the distribution of categorical data was determined using Chi-squared tests. The univariable and multivariable analyses were both performed using SPSS software. For the univariable analysis, the dependent factors were age, gender, HCV genotype or liver disease and the co-morbidities were analysed as discrete fixed factors. Statistical significance for all analyses was determined with a cut-off of $p < 0.05$ and values of $p < 0.1$ were interpreted as tending towards significance.

2.3 Methods

2.3.1 Control transcripts

Previous studies have shown that control RNA transcripts derived from synthetic dsDNA based on patient-isolated sequences can be used effectively as controls for PCR assay optimisation (Schwaiger and Cassinotti, 2003; Dierssen *et al.*, 2008; McLeish *et al.*, 2012). We developed a similar method to produce control RNA transcripts for HCV genotypes 1a and 3a for both the E1-E2 and NS5B regions. The method is outlined in figure 2-1. Blunt ended dsDNA fragments used to generate the controls were obtained from Life Technologies. Both the E1-E2 and NS5B controls were based on patient-derived sequences

with the gt1a controls based on H77 (accession number AF009606) and the gt3a controls based on 3a.GB.2005 (accession number GQ356206) (Kolykhalov *et al.*, 1997; Humphreys *et al.*, 2009). The criteria used for selection of the controls are discussed in chapter 4 section 4.2.1. Transcript controls were used throughout the project for assay optimisation, relative quantification during qPCR and fidelity assessments.

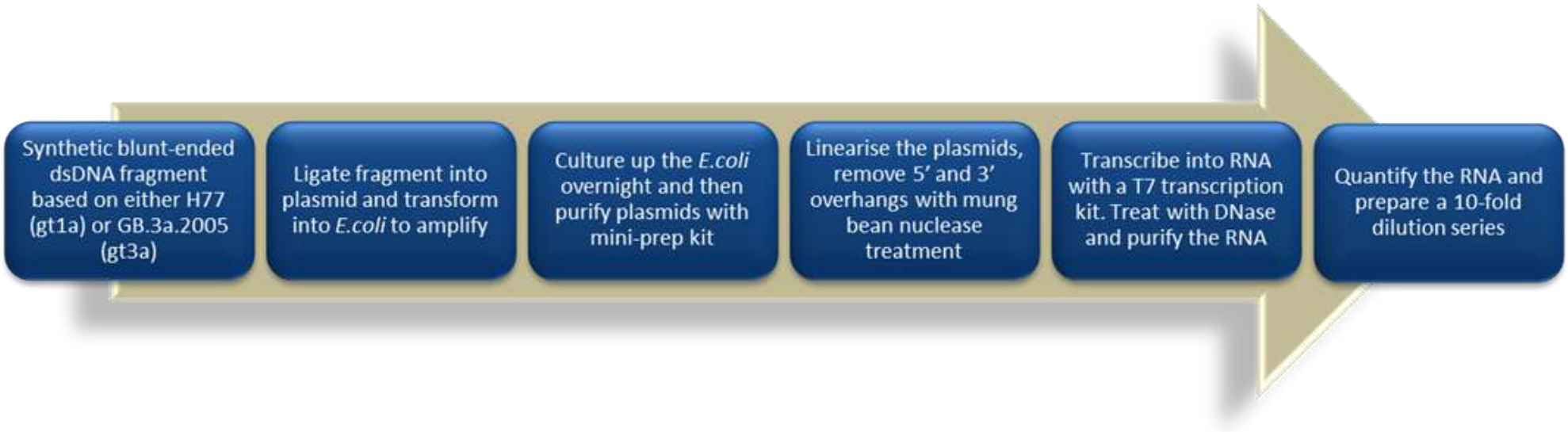


Figure 2-1; Workflow developed for the production of control transcripts

2.3.1.1 Plasmid preparation

Cloning was performed using the CloneJET PCR cloning kit using the method for fragments of 1kb or smaller. Blunt-ended DNA fragments were prepared by the addition of dH_2O as indicated by the manufacturer. The DNA fragment (0.15 pmol) was added to 0.05 pmol pJET1.2 (blunt cloning vector) along with 10 µl reaction buffer, 1 µl T4 ligase and nuclease free water (up to 19 µl). The reaction mixture was briefly vortexed and then placed on ice for 5 minutes (min) before being used directly for transformation.

2.3.1.2 Cloning

NEB 5-alpha competent *E. coli* cells were thawed on ice for 10 min. The transformation reaction was initiated by adding 1-5 µl of the ligation reaction (containing 1 pg-100 ng of plasmid DNA) to the cell mixture and gently flick mixing it. The mixture was placed on ice for 30 min and then heat shocked at exactly 42°C for 30 seconds in a water bath. Following incubation on ice for 5 min, 950µl of room temperature SOC media was added and the mixture incubated at 37°C in a shaking heat block set at 250 rpm for 60 min. The cells were thoroughly mixed during the preparation of 10-fold dilutions in SOC and 50-100 µl of each dilution was spread onto LB agar (+100 µg/ml Ampicillin) selection plates pre-warmed to 30oC. The plates were then incubated at 30oC for 48 h or37ºC for 16 h. The lower incubation temperature was used when high fidelity copying of the clonal plasmid inserts was required.

2.3.1.3 Mini-preps

Bacterial colonies that had successfully grown on LB agar (+100 µg/ml Ampicillin) were picked and individually inoculated into 5ml aliquots of LB broth (+100µg/ml Ampicillin). They were grown overnight at 37°C for 16hrs. Plasmids were purified from 1.5ml aliquots of the inoculated broth using a column mini-prep kit. Bacterial cultures were centrifuged at 6800 x g for 3 min and the supernatant discarded. The pellets were resuspended in 250 µl of pre-prepared buffer P1, 250 µl of buffer P2 was added and the tube was inverted 4-6 times. A further 350 µl of buffer N3 was then added to the solution and the tube again inverted 4-6 times. The mixture was centrifuged at 17,900 x g for 10 min, during which time a white pellet consisting of cell debris formed. The supernatant containing the plasmids was applied to a

QIAprep 2.0 spin column which was centrifuged for 1 min at full speed and the filtrate was discarded. The column was washed by the sequential addition of 500 ul of PB and 750 ul of buffer PE, centrifuging for 1 min and discarding the filtrate at each step. To remove any residual buffer, the column was centrifuged for 1 min to remove any residual buffer and then placed into a fresh Eppendorf. In order to elute the plasmids, 50 µl of buffer EB was added to the column and left to stand for 1 min. After being centrifuged for 1 min, the column was discarded and the plasmids contained in the filtrate were retained.

2.3.1.4 Restriction digests

The purified pJET1.2 plasmids were linearized by a restriction digest with Xba I (Promega). The plasmid construct (40 ul) was incubated with 5 µl Xba I and 5 µl Buffer D at 37^{o}C for 2 hours.

2.3.1.5 Mung bean nuclease treatment

Mung bean nuclease treatment was used to remove 5' and 3' overhangs from linearized plasmid constructs prior to transcription of the fragments into RNA. The reaction was set up by adding 5 µl of mung bean buffer (10x) to 2 µl of mung bean nuclease and 43 µl of sample. The mixture was then incubated for 1h at 30^{o}C. Blunt-ended dsDNA fragments, ready to undergo transcription, were generated.

2.3.1.6 T7 transcription

The following reaction was set up using the T7 RiboMAX kit (Promega) in 1.5 ml eppendorfs; T7 RiboMAX 2x buffer (10 µl), linear DNA sample (2 µl), water (6 µl) and T7 RiboMAX enzyme mix (2 µl). The contents were well mixed and incubated at 37^{o}C for 1 hour. After incubation, the samples were treated with 1 Unit of RNase-free DNase per µg of DNA and incubated for 15 min at 37^{o}C to remove remaining DNA templates.

2.3.1.7 Gel electrophoresis of RNA

1.5% agarose gels were prepared by mixing 1.5 g of agar with 100 ml of TAE preparation. The mixture was heated until boiling and then left to cool to approximately 45oC. When

cooled, 10 µl of SYBR safe stain was added to the mixture prior to pouring into a gel cast with a comb. Once set, samples were mixed with a loading dye, loaded onto the gel and run along a current gradient. A ladder was run alongside the samples for size comparison. The gels were analysed by UV light on a transilluminator and a photograph was taken of the gel.

2.3.1.8 RNA quantification

The concentrations of synthetic RNA transcripts were quantified with a Nanodrop ND-1000 spectrophotometer, which measures the optical density of nucleic acid samples at 260nm. A minimum of 3 readings/sample were taken and the average calculated. The Endmemo website (http://www.endmemo.com/bio/dnacopynum.php) was then used to convert the concentration (ng/µl) into transcript copy number (copies/µl).Stock controls consisting of suitable ten-fold serial dilutions of the RNA controls were made and stored at -80°C. The controls were aliquoted for single use and only thawed immediately prior to use to prevent degradation of RNA through freeze/thaw cycles.

2.3.2 RNA extraction

2.3.2.1 RNA extraction

Viral RNA from samples was extracted using a QIAamp viral RNA mini kit. Lysis of the HCV virions was carried out under tightly controlled conditions in an enhanced bio-containment facility. A 140 µl aliquot of patient sera was added to 560 µl of AVL buffer containing 5.6 µl of carrier RNA (supplied with the kit). The sample and buffer were vortexed briefly and the mixture was incubated at room temperature for 10 min. After briefly pulse centrifuging the samples, 560 µl of 100% ethanol was added and the samples were vortexed for 15 seconds. Following this, 630 µl of the solution was applied to a QIAmp mini column (in a 2 ml collection tube) which was then centrifuged at 6000 x g for 1 min. The filtrate in the collection tube was discarded and the column placed into a fresh collection tube. The remaining 630 µl of the sample was applied to the column and the centrifugation step repeated. The column was placed in a fresh collection tube and 500 µl of buffer AW1 was applied before centrifuging at 6000 x g for 1 min. The filtrate was discarded and the column placed into a fresh collection tube. Following this, 500 µl of buffer AW2 was added

and the sample centrifuged at 20 000 x g for 3 min. The filtrate was then discarded; the column placed in a fresh centrifuge tube and the sample was spun at 20 000 x g for a further 1 min. The column was then placed into an Eppendorf tube and 60 µl buffer AVE (elution buffer) applied to the columns. The samples were incubated at room temperature for 1 min before being centrifuged at 6 000 x g for 1 min. The columns were then discarded and the RNA was stored at -70ºC until required.

2.3.2.2 cDNA synthesis

Extracted RNA was transcribed into cDNA using SuperScript III reverse transcriptase as per manufacturer's instructions. Initially, 10 µl RNA was mixed with 1.0 µl dNTPs, 0.08 µl random primers and 1.92 µl dH_2O. The reaction mixture was incubated at 65°C for 5 min and then cooled on ice for 1 min. After this, a master mix comprising of 4.0 µl buffer, 1.0 µl SuperScript reverse transcriptase, 1.0 µl RNaseOUT and 1.0 µl DTT was prepared, 7 µl of which was added to the reaction cooled on ice, creating a total reaction volume of 20 µl. The reaction was then incubated at 25 °C for 5 min, 50 °C for 60 min and finally, 70 °C for 15 min before being held indefinitely at 4 °C. RNaseOUT was used to selectively inhibit ribonucleases and reduce any possible RNA degradation. The cDNA libraries were stored at -80ºC prior to testing.

2.3.3 Screening of mixed infections

2.3.3.1 Primer design

Unless otherwise stated, primers were designed manually using alignments of reference sequences downloaded from the Los Alamos HCV database (Kuiken *et al.*, 2005). The alignment included sequences for all 67 characterised sub-types of HCV (see table 10 for accession numbers). Alignments were supplemented with additional gt1a and gt3a sequences as required to ensure that the developed primers would be capable of identifying a diverse range of strains. The exact requirements of each primer set varied but in general, they were designed to be 18-24bp in length, have a GC content of 40-60% and a Tm of 55-60°C (although this varied depending on the enzyme used). Primer pairs were designed to have a similar Tm and designs indicating a strong likelihood of hairpin formation or primer dimerization were avoided. Degenerate positions were incorporated where appropriate to

accommodate strain variation between viral sequences. Primer properties were assessed using Oligo Calc (www.basic.northwestern.edu/biotools/OligoCalc).

2.3.3.2 Nested PCR

Sub-type specific nested PCR assays were developed and their annealing temperatures optimised to screen samples for the presence of mixed gt1a/gt3a infections. The first round PCR was performed using the Platinum Taq PCR System and the secondary PCR was carried out with the KOD Hot Start PCR system, both as per manufacturer's instructions. Primers A-D (table 2-5) were used to screen for genotype 1a strains and primers E-H (table 2-5) used to screen for genotype 3a strains. For the first round reaction, the following were mixed together in a total reaction volume of 20 µl; 0.4 µl DNTPs, 0.6 µl $MgCl_2$, 2.0 µl buffer, 0.4 µl each of the sense and anti-sense primers (using 10µM working stocks), 2 µl cDNA, 0.08 µl platinum Taq and 14.12 µl dH_2O. Cycling conditions were as follows: initial denaturation, 94°C for 2 min and 40 cycles of 94°C for 30 seconds, 55°C (gt1a) or 56°C (gt3a) for 30 seconds and 72°C for 1 min; the reaction was then held at 4°C. For the second round reaction, the following reaction mixture was prepared in a total volume of 20 µl; 2.0 µl DNTPs, 1.2 µl $MgSO_4$, 2.0 µl buffer, 0.6 µl of both the sense and anti-sense primers, 1 µl from the first round reaction, 0.4 µl KOD Taq and 12.2 µl dH_2O. Cycling condition were as follows: initial denaturation, 95°C for 2 min and 25 cycles of 95°C for 20 seconds, 55°C (gt1a) or 56°C (gt3a) for 10 seconds and 70°C for 15 seconds; the reaction was then held at 4°C. Products from the reactions were run on a 1.5% agarose gel to confirm the presence of a product of the expected size. All screening reactions had a positive control (an aliquot of 10^4 copies/µl of the control transcript of the appropriate genotype) and a negative control (nuclease free water) that were amplified alongside the samples.

2.3.3.3 Gel electrophoresis of DNA

The protocol for the electrophoresis of DNA was identical to that of the protocol for RNA gel electrophoresis (see section 2.3.1.6).

2.3.3.4 Sanger sequencing

Samples for sequencing were treated with Illustra Exostar (Jencons) to remove excess primers and dNTPs from the solution prior to sequencing. A 2 µl aliquot of Illustra Exostar (prepared as instructed) was added to 5 µl of sample. The mixture was then incubated at 37°C for 15 min and 80°C for a further 15 min. If greater than 7 µl was required, two reactions were run and the samples pooled. Samples were prepared and pre-mixed with primers as required by Eurofins. Samples were sent to Eurofins for Sanger sequencing. All the sequenced amplicons were checked against alignments of frequently used laboratory strains to screen for contamination events.

2.3.4 qPCR

2.3.4.1 5'UTR

A qPCR targeting the 5'UTR was used throughout the project to quantify the HCV viral load (primers A-C, table 8) (Jones *et al.*, 2010). The qPCR was performed in 96-well plates with the TaqMan fast 7500 system using TaqMan fast reagents in accordance with the manufacturer's instructions. Pre-prepared cDNA (section 2.3.2.2) was amplified for 40 cycles in a reaction that contained 2µl of cDNA, 900 nM of each primer, 250 nM of probe and 10 µl TaqMan fast master mix. The reaction volume was adjusted to 20 µl by the addition of dH_2O. Cycling conditions were 95°C for 20 seconds and then 40 cycles of 95°C for 3 seconds and 60°C for 30 seconds. A dilution series prepared from known concentrations of JFH-1 replicon transcripts were used as controls for comparative quantification.

2.3.4.2 NS5B

Genotype-specific primers and probes suitable for use with the TaqMan fast system (Life Sciences) were designed for HCV genotypes 1a and 3a (primers D-F and G-I, table 8). Reactions were set up in an identical manner to that described in section 2.3.4.1. Genotype-specific controls derived from synthetic blunt-ended dsDNA were used for comparative quantification (see methods 2.3.1).

2.3.4.3 Deep sequencing

A PCR-based Illumina deep sequencing approach targeting partial E1-E2 and NS5B regions of HCV was developed that was compatible with the Illumina MiSeq platform. Primers for the PCR were designed to work in conjunction with the V3 MiSeq reagent kit, allowing the sequencing of 2 x 300bp paired-end reads. An overview of the protocol used is shown in figure 2-2.

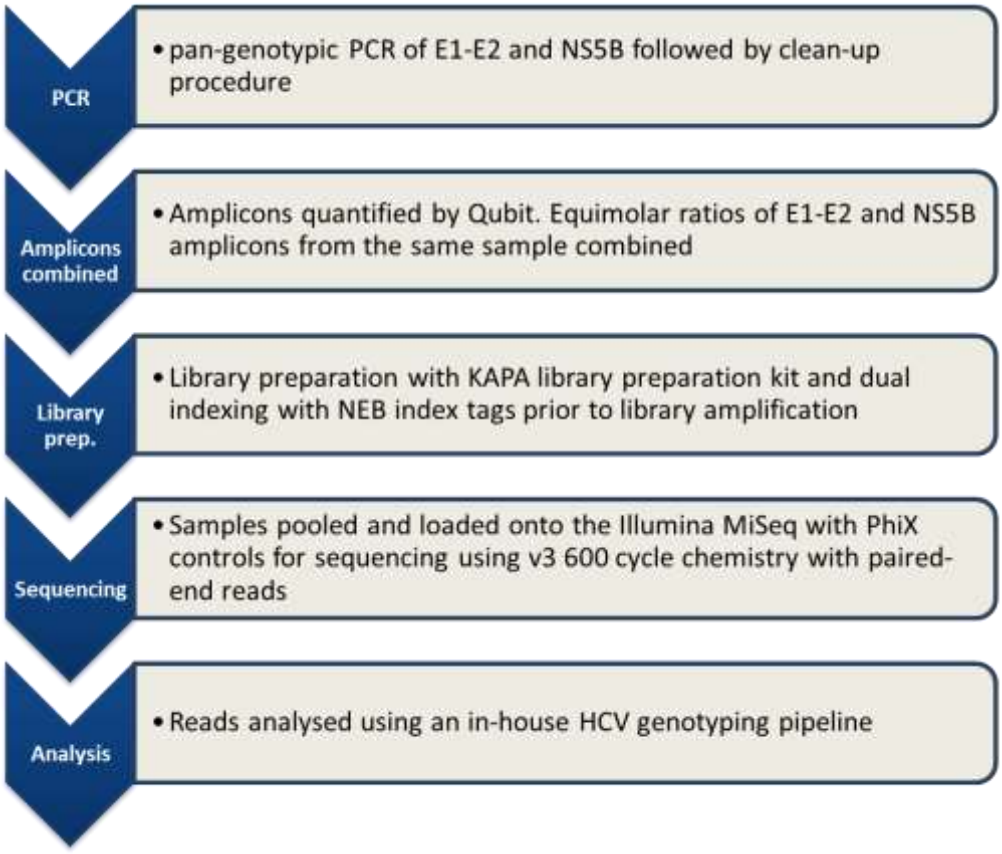


Figure 2-2; Workflow developed for the deep sequencing protocol

In order to limit the risk of contamination, samples were split into separate runs on the basis of their genotype. Samples with mixed genotype HCV infections were placed on the run of their major genotype as determined by qPCR. In cases where this information was not known, the genotype recorded by the diagnostic laboratory was presumed to be the major genotype in the sample.

2.3.4.4 PCR

To generate amplicons for deep sequencing, samples underwent two separate PCR reactions using the E1-E2 (table 5, primers I-J) and NS5B (table 5, primers K-L) pan-genotypic primers. The KAPA HiFi PCR kit was used to generate the amplicons as per manufacturer's instructions. The following reaction mixtures were prepared; 4 µl buffer, 0.6 µl DNTPs, 0.6 µl of both the sense and the anti-sense primers, 0.4 µl KAPA HiFi enzyme, 4 µl cDNA and 9.8 µl dH_2O to create a total reaction volume of 20 µl. Cycling conditions were as follows: initial denaturation, 95°C for 3 min and 30 cycles of 95°C for 20 seconds, 61°C (gt1a) or 64°C (gt3a) for 25 seconds and 72°C for 40 seconds, followed by a final extension at 72°C for 1 min. The reaction was then held at 4°C.

2.3.4.5 Clean-up of amplicons

Agencourt AMPure XP magnetic beads were used for the purification of amplicons after PCR reactions. AMPure XP beads exploit the negative charge of DNA and are able to reversibly bind to DNA at high concentrations of polyethene glycol (PEG) and salt. PCR reactions were diluted with Tris to a total volume of 50µl. Depending upon the size of products being selected, 1x – 1.4x volume of AMPure XP beads was then added. Smaller volumes of beads select for larger DNA fragments and the removal of primer-dimers was obtained by the use of 0.85x volume of beads. Samples were mixed well with the beads by pipetting and left to stand at room temperature for 10 min before being placed on a magnetic rack for 5 min. The supernatant was then removed and the beads were washed twice in 200 µl of 70% ethanol. The beads were left to air dry for 5-10 min and then resuspended in 8-15 µl Tris. Some in-house protocols have been developed whereby the beads are kept in the samples during enzymatic reactions to limit the DNA loss during preparation (Gavin Wilkie, CVR, unpublished). In cases where this protocol was used, solid-phase reversible immobilization (SPRI) buffer was used for the DNA clean-up in identical volumes to the beads.

2.3.4.6 DNA quantification with Qubit

The amount of DNA was quantified with the Qubit dsDNA High-Sensitivity (HS) kit as per manufacturer's instructions. A working solution was prepared by adding HS reagent to buffer

in a 1:200 dilution. For each sample, 1-2 µl of PCR product was added to 198-199 µl of working solution and then vortex mixed. Samples were run on the Qubit against standards and DNA was quantified in ng/µl.

2.3.4.7 Library preparation

The protocol followed for the library preparation is an in-house protocol based on the instructions provided with the KAPA library preparation (for Illumina) kit and adapted for use by Gavin Wilkie (high-throughput sequencing facility manager, MRC-CVR).

After production of the E1-E2 and NS5B amplicons with pan-genotypic primers, the DNA was purified using 1x AMPure XP magnetic beads and then quantified using Qubit. The E1-E2 and NS5B amplicons for each sample were then combined in equimolar ratios. After mixing the amplicons, the volume was adjusted to 25 µl with Tris and the samples then underwent library preparation. The ends of the PCR fragments were repaired by adding 2 µl of end repair enzyme in 3µl of end repair buffer to 25 µl of sample. The reaction was incubated for 30 min at 20°C and then the volume was increased to 50 µl using 10 mM Tris. DNA was purified using 1x AMPure XP magnetic beads and the beads were retained within the samples after elution to minimise sample loss. Samples were eluted into 21 µl of 10 mM Tris. A single 3' Adenosine residue was then added to each fragment in an A-tailing reaction by adding 21 µl of end-repaired DNA to 1.5 µl of A-tail enzyme in 2.5 µl of buffer and incubating at 30°C for 30 min. The volume was increased to 50 µl using 10 mM Tris and the DNA purified by adding 1x SPRI. Samples were eluted into 15 µl of 10 mM Tris and DNA concentration was quantified using Qubit.

The amount of DNA present was then converted to pmols using the following equation –

$$pmols = \frac{Total\ amount\ of\ DNA\ (ng)}{[average\ size\ fragments\ (kb) \times 660]}$$

As the E1-E2 and NS5B fragments were mixed together in equimolar ratios, the size of the fragments was assumed to be the average of their two lengths. The average molecular weight of a nucleotide base is considered to be 660 g/mol. Based on the calculation, adapters were

then ligated onto the ends of the A-tailed DNA in a reaction with a 20:1 molar excess of adapters to DNA fragments. The amount of adapter required was calculated with a minimum of 0.5 µl and a maximum of 5 µl adapter being used. Following this, 14 µl of A-tailed DNA was added to 1 µl of ligase in 5 µl of buffer with 0.5-5 µl NEB adapter. The total reaction volume was adjusted to 25 µl using dH_2O and incubated at 20°C for 1 hour. Each reaction was incubated for 15 min with 1 µl of USER enzyme at 37°C in order to remove unpaired uracil bases in the adapters and generate the free ends required for the indexing PCR reaction.

The volume was adjusted to 50 µl and the DNA was purified using x0.85 volume SPRI. The samples were eluted into 15 µl Tris and each sample was assigned a unique combination of an i5 and i7 indexed primer for dual index tagging. For the PCR reaction, 10 µl of DNA was added to 12.5 µl of 2x KAPA HiFi Hotstart ready mix with 1.25 µl of each assigned primer tag. The reaction then underwent PCR with the following cycling conditions; 95°C for 3 min and 12-16 cycles of 95°C for 20 seconds, 65°C 15 seconds and 72°C for 30 seconds, followed by a final extension at 72°C for 1 min. The reaction was then held at 4°C indefinitely. In the majority of cases, the PCR reactions were cycled 12 times and 16 cycles were used when the starting amount of DNA present was known to be low (based on Qubit quantification). After the PCR reaction, the volume was adjusted to 50µl, the DNA was purified with x0.85 volume of SPRI and the samples were eluted into 15 µl of Tris. Library concentrations were quantified by Qubit.

2.3.4.8 TapeStation analysis

Library preparations were analysed on TapeStation to assess the size of the products. A 2µl aliquot of sample was mixed with 2µl tape station buffer. Samples were then run on the TapeStation with a sizing ladder using a D1000 HS cassette (Aligent) for analysis. Using the library concentration obtained by Qubit and the size determined by TapeStation, the molarity of each library DNA was calculated with the following equation –

$$Molarity\ (nM) = \left[\frac{DNA\ concentration\ (ng/\mu l)}{[peak\ size\ (bp) \times 660]}\right] \times 1000000$$

Using the calculated molarity, samples were pooled together in equimolar ratios prior to running them on the MiSeq. Guidance calculating the aliquots required for equimolar pooling was provided by Gavin Wilkie.

2.3.4.9 Deep sequencing

Libraries were denatured prior to sequencing by treatment with freshly prepared 0.2 M NaOH. DNA library pools were diluted to 4 nM and 5 µl of the pools was added to 5 µl of 0.2 M NaOH. The mixture was vortexed, briefly centrifuged and then incubated for 5 min at room temperature. 10 µl of the denatured DNA was then added to 990 µl of pre-chilled hybridization buffer (provided with Illumina cartridges), resulting in a 20 pM denatured library in 1 mM NaOH. This mixture was then diluted further using the hybridization buffer to give 600µl of a desired final library concentration of between 6-20 pM. This was then loaded onto the MiSeq cartridge, which in turn was loaded onto the MiSeq sequencer. As our runs were sequencing E1-E2 and NS5B amplicons, diversity among our samples was low. This lack of variation can cause issues with the interpretation of the sequencing data, and samples were therefore run at a relatively low cluster density with 5-10% PhiX controls in order to compensate.

2.4 Phylogenetic and sequence analysis techniques

2.4.1 Alignments

All analysed sequences were examined and aligned using the MUSCLE (v3.8) program embedded within Simmonics Sequence Editor (SSE) (Edgar, 2004; Simmonds, 2012).

2.4.2 Maximum-likelihood analysis

Maximum likelihood phylogenetic analyses were carried out on alignments using MEGA 5.0 (Tamura *et al.*, 2011). Bootstrap re-sampling with 1000 replicates was used to assess the statistical support for the branches within phylogenetic trees. Bootstrap values ≥70% were considered significant and indicated on the trees. When discussing sequences, position within the genome is always quoted relative to the HCV strain H77 (Kuiken *et al.*, 2006). Reference

sequences for analysis were downloaded from either NCBI Genbank or the Los Alamos HCV Database (Kuiken *et al.,* 2005; Benson *et al.,* 2013).

2.4.3 Deep-sequencing analysis

Deep sequencing data from the Illumina MiSeq was analysed using an in-house Unix-based pipeline. Sequencing adapters were removed and low-quality reads were removed using Fast QC and Trim Galore! (Babraham Bioinformatics). High-quality paired-end reads then underwent a number of analyses for genotyping, quasispecies diversity and dinucleotide frequency.

Genotyping of the deep sequencing reads was done using an in-house bioinformatics pipeline developed by Sreenu Vattipally (Research bioinformatician, CVR). The pipeline worked by bioinformatically excising deep sequencing reads into 37bp k-mers and comparing against pools of genotype-specific k-mers. An outline of the pipeline is shown in figure 2-3. Several thousand HCV sequences were used as reference sequences and bioinformatically chopped to generate the pools of unique genotype-specific k-mers. K-mer identity matching between the reference k-mers and the read k-mers had to be 100% to allow genotyping. Consensus sequences were generated from the genotype reads and compared with previously obtained sequences using maximum likelihood phylogenetic analysis.

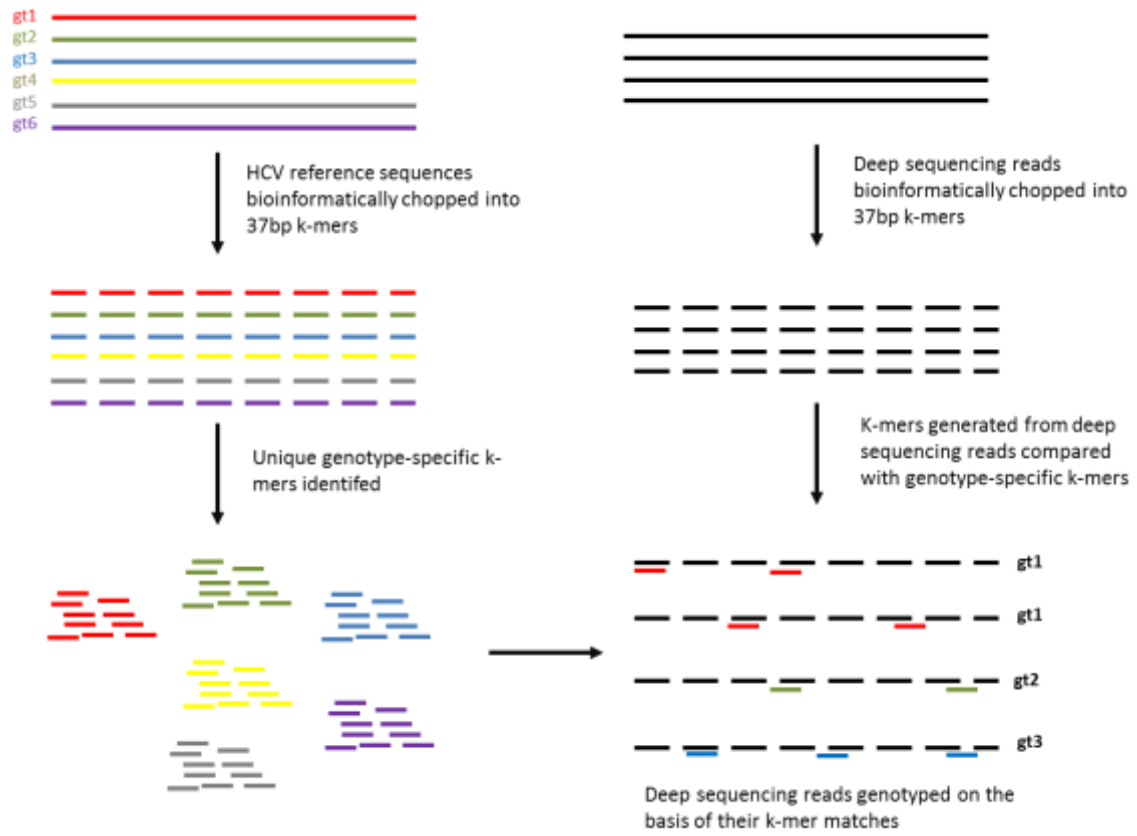


Figure 2-3; K-mer bioinformatic analysis

Outline of the K-mer bioinformatic analysis developed by Sreenu Vattipally and used to genotype the deep sequencing reads

For other comparative sequence analyses, Tanoti was used (Vattipally, 2014). Tanoti was used to compare deep sequencing reads against reference genomes and Sanger sequences from previous work. Consensus sequences were generated from the Tanoti files and compared with previously obtained products sequenced by the Sanger method using maximum likelihood phylogenetic analysis (section 2.4.2).

2.4.4 Diversity analysis

The deep sequencing data was used to assess diversity present within the viral populations. The diversity analysis was carried out with assistance from Richard Orton (Research Associate, MRC-CVR) and Sreenu Vattipally. CD-Hit (http://www.bioinformatics.org/cd-hit/) was used to perform a cluster-based analysis of the sequencing data (Li and Godzik,

2006; Fu *et al.*, 2012) on a Linux-based system available at the Centre for Virus Reseach. Using a fasta format sequence database, the CD-Hit analysis functions by utilising a greedy incremental algorithm that takes the longest read in a dataset and compares all the other reads in the dataset to it (Fu *et al.*, 2012). The sequences were initially sorted by length and the longest sequence was taken to represent the first cluster. Subsequent reads were compared to the longest read using a short word filtering algorithm. Reads that were above a user-defined threshold of similarity to the longest read were grouped into a cluster together. Reads that were not similar were taken as representative of new clusters and subsequent sequences were compared to all the clusters generated. Clusters with >500 reads associated were considered to be real variants. A threshold of 0.9 was selected for this study, meaning that each cluster detected was at least 10% divergent at the nucleotide level from the longest read and other clusters. The numbers of clusters of >500 and >1000 reads were recorded in order to gauge diversity present within the samples in this study. It should be noted that for this analysis the reads from each sample were previously separated by genotype using Tanoti (Vattipally, 2014) and that in samples with multiple genotypes detected within them, clustering analysis has been performed on each genotype separately.

2.4.5 Dinucleotide frequency analysis

The expected and observed ratios of CpG and UpA dinucleotide occurrences were calculated using a program within Simmonics Sequence Editor (Simmonds, 2012). The amount of each nucleotide (A, T, C or G) in a given sequence was counted and the expected frequency of each dinucleotide calculated based on these values. The observed frequency of dinucleotides was then calculated and used to calculate the ratio of observed:expected (O/E) for each of the dinucleotide pairs.

$$O{:}E\ Ratio = \frac{Observed\ number\ of\ dinucleotides\ in\ a\ sequence}{Calculated\ expected\ frequency\ of\ dinucleotides\ in\ a\ sequence}$$

Data was analysed using Excel (Microsoft) and SPSS software.

2.4.6 Relative synonymous codon usage

The relative synonymous codon usage (RSCU) for coding sequences were calculated using a program within Simmonics Sequence Editor (Simmonds, 2012). Alignments were checked to ensure that they were in frame prior to calculating the RSCU. The RSCU was calculated by first counting the frequency of each amino acid. The relative frequency of each codon used was then recorded and the relative usage of each codon for its respective amino acid was calculated. Base composition and dinucleotide frequency bias were not taken into account during the analysis. Data was analysed using Excel (Microsoft) and SPSS software (IBM).

2.4.7 Viral sequences used for analysis

Viral sequences used for analyses throughout the project were downloaded from the Genbank (Benson *et al.*, 2013) and Los Alamos databases (Kuiken *et al.*, 2005). Details of the sequences and the analyses they were used in can be found in the appendix, table 8-1.

Chapter 3: Analysis of linked clinical data

3.1 Introduction

Chronic infection with HCV is associated with a highly variable clinical outcome with many individuals suffering from the progressive development of liver fibrosis, cirrhosis and HCC. Cirrhosis develops after 20-30 years in approximately 10-20% of individuals with HCV (Thein *et al.*, 2008; Westbrook and Dusheiko, 2014) and numerous risk factors have been linked with an increased risk of fibrotic progression including male gender, age >40 years, excessive alcohol intake, immunosuppression, diabetes and co-infection with HIV or HBV (T Poynard, Bedossa and Opolon, 1997; Ramos Paesa *et al.*, 1998; Minola *et al.*, 2002; Gaeta *et al.*, 2003; Ratziu *et al.*, 2003; Hutchinson, Bird and Goldberg, 2005). Extra-hepatic manifestations and co-morbidities are also frequently reported amongst patients infected with HCV and include a wide range of conditions, such as cryoglobulinemia, insulin resistance, immunological disorders, depression and gastrointestinal disorders (Louie *et al.*, 2012; Cacoub *et al.*, 2014). Whilst infection rates in many nations are currently in decline, the rates of HCV-related morbidity and mortality are increasing and are causing significant financial strain on health systems globally (Wong *et al.*, 2000; Deuffic-Burban *et al.*, 2012). Similar trends are seen in the UK including Scotland, where rates of HCV-related mortality have increased significantly in recent years (McDonald *et al.*, 2010; Public Health England, 2015).

The analysis of linked clinical data from patients with HCV is a valuable resource and can highlight trends in transmission, treatment and patient prognosis, suggesting where resources may be best focused. Previous studies have indicated that there are genotype-specific differences in the immune response to HCV (Robinson *et al.*, 2015) and that gt3 infection is associated with a more rapid progression towards cirrhosis and HCC relative to gt1 infection (Leandro *et al.*, 2006; Kanwal *et al.*, 2014). In this chapter, the data collected on the cohort

was analysed by genotype, to compare the clinical and demographic characteristics of patients infected with gt1a and gt3 infections. Further analysis looking at the gender and the severity of liver disease among patients in the cohort was also performed where appropriate, providing an insight into the demographics of individuals infected with HCV in the geographical region.

3.1.1 Data collection

During the period August 2013-march 2014, 1132 HCV positive samples were genotyped at the West of Scotland Specialist Virology Laboratory (WSSVC) based at the Glasgow Royal Infirmary. Generally, HCV positive patients in the region are genotyped shortly after initial diagnosis and prior to treatment. For this study, anonymised data was collected on postcode, age, and the viral genotype of the patients providing the 1132 samples. Over the period August 2013-march 2014, any sample that had been genotyped and had ≥140 μl serum available was also collected for the mixed gt1a/gt3 screening (chapter 4). A total of 205 gt1a samples and 206 gt3 samples were collected. An additional 48 gt1a and 47 gt3 samples collected as part of an earlier study were also screened for mixed gt1a/gt3 infection (chapter 4). In addition to this, for a subset of 590 HCV-positive samples from patients living within the Greater Glasgow and Paisley postcode districts (PC) who underwent genotyping at the WSSVC between August 2013-march 2014, detailed data on co-morbidities, liver disease and viral load was collected regardless of whether or not a sample was collected concurrently. Enhanced data was only available for this subset of patients as ethical approval for the project was granted by the NHS Greater Glasgow & Clyde Biorepository and linked data for the remaining 542 patients was held by other health boards. Clinical data on the cohort was sourced from the clinical portal system, which provides data on patient contact with secondary care providers in the region and is considered a relatively robust source for data collection.

3.2 Characteristics of the population sampled

The WSSVC is a part of The Scottish National Blood Borne Virus Specialist Testing Service, providing specialist testing for blood borne viruses (BBV) in Scotland. Consequently, samples are sent to the WSSVC from throughout Scotland and the samples we collected are

reflective of this. Sampling from the Greater Glasgow region (including Motherwell and Paisley PC areas) is likely to be over-represented. Estimates of prevalence Scotland-wide have established that 39% of all HCV cases are within the Greater Glasgow and Clyde regions (Codere *et al.*, 2015) ; in our study, 51% of cases were sampled from this region (figure 1A). Greater Glasgow does however represent a significant proportion of the Scottish population and is known to have relatively high rates of IDU and HCV infection. Awareness of this issue is good among healthcare providers within the region and testing rates are likely to be high in the region as a result. This is reflected in the data with 42% of the samples collected coming from Glasgow PC areas. As the WSSVC is located within Glasgow, a prolonged retention of samples from the region relative to other samples may also have affected sampling. Additionally, Edinburgh has its own specialist BBV testing centre and samples were collected from there once, during the initial phase of sample collection for the project, and were not included in the linked clinical data collected. As a result, sampling from the Edinburgh PC area is likely under-represented within this cohort.

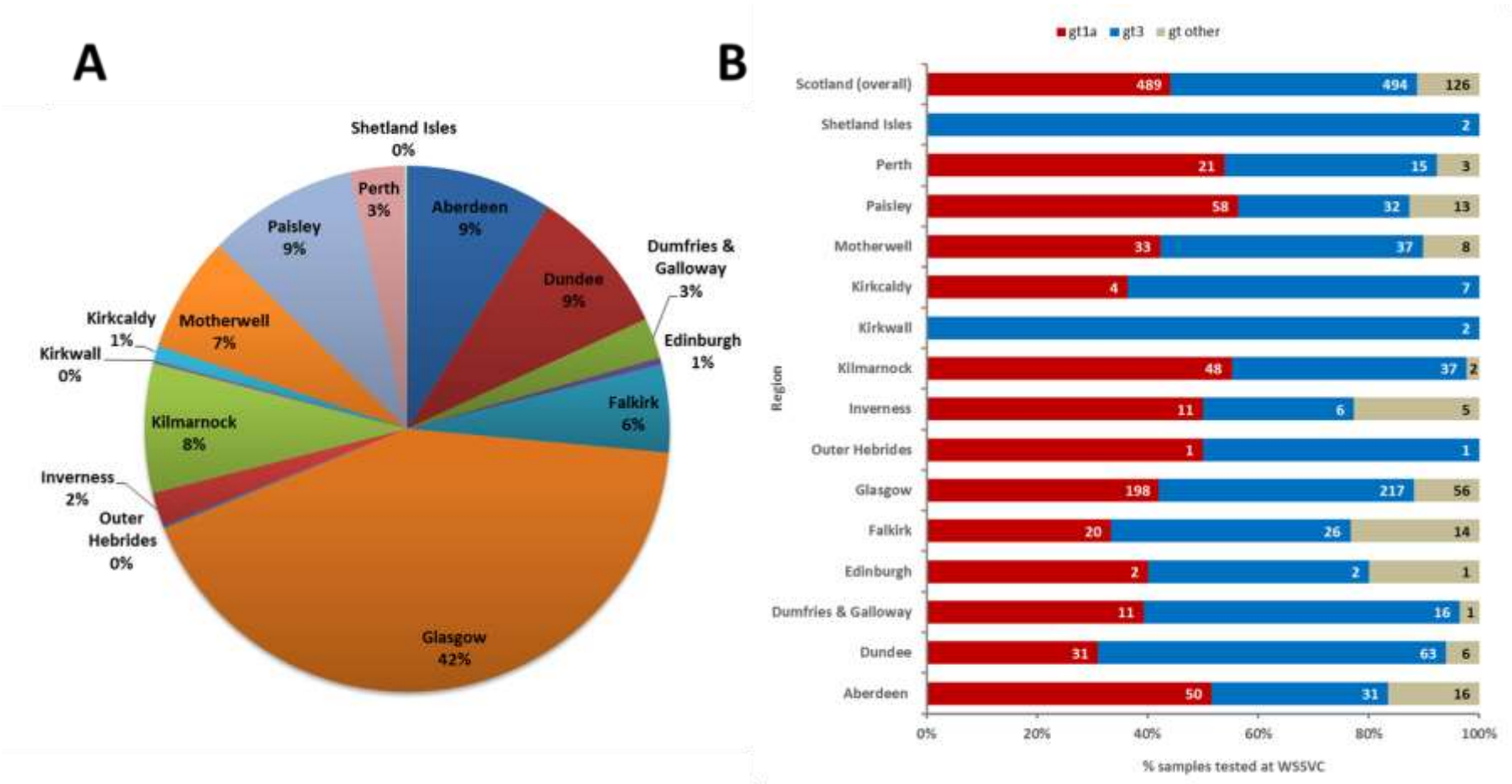


Figure 3-1; Regional prevalence of samples in the cohort

(A) Postcode (PC) regions of Scotland from where the 1132 samples were collected. (B) Regional prevalence of HCV genotypes diagnosed at the WSSVC. The numbers of samples from each region are overlaid on the chart. From these figures, it can be seen that samples and data from the Glasgow and Paisley PC areas comprise ~50% of the cohort.

Between July 2013 and March 2014, 42.3% of the samples tested at WSSVC (n=1132) were gt3, 41.8% were gt1a and 10.8% were diagnosed as having another genotype (figure 3-1, B). The sample was either insufficient or the viral load was too low to get an accurate result in 5.1% of cases. No cases of gt5 or gt7 were diagnosed in the region during the sampling period. Genotypic distribution patterns were fairly consistent throughout the different PC regions of Scotland, with most of the regions having similar prevalence patterns to the overall prevalence rate. Exceptions were seen in regions with low sampling rates including, the Outer Hebrides (n=2), Kirkwall (n=2) and the Shetland Isles (n=2). The Aberdeen, Edinburgh, Falkirk and Inverness PC areas had considerably larger numbers of HCV infections that were not gt1a or gt3 than other regions. Aberdeen and Edinburgh PC areas are documented to have the highest rates of non-Scottish residents and both regions have strong international links.

3.2.1 Characteristics of the Glasgow and Paisley PC area cohort

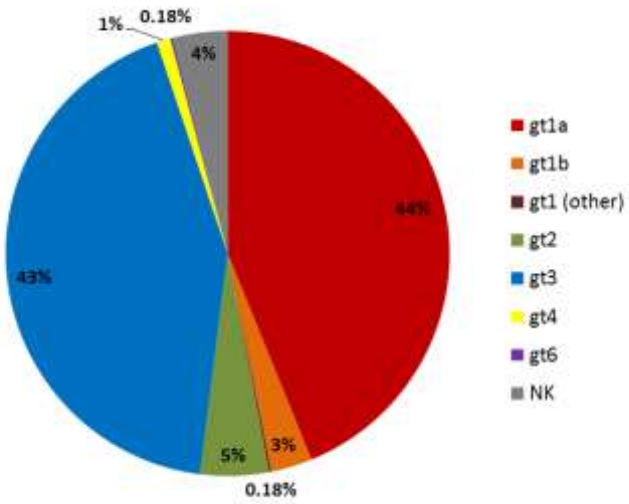


Figure 3-2; Genotypes of HCV positive patients in the Glasgow and Paisley areas

The genotypes of 590 HCV positive patients in the Glasgow and Paisley postcode districts, as determined by the WSSVC. NK includes samples where genotype was not known, due either to insufficient sample or sample not being tested. No gt5 or gt7 samples were detected within the cohort by the WSSVC.

The most common HCV genotypes diagnosed within the Glasgow and Paisley PC area were genotype 1a (44%) and genotype 3 (43%) (figure 3-2). All other genotypes were detected at relatively low frequency and the genotype in 4% of cases was not determined. This is

comparable with the reported genotype prevalence observed across the UK, with genotypes 1a and 3 being the most commonly reported, especially among populations with high numbers of IDUs (Public Health England, 2015). HCV infection was more common in males with 74.2% of the population we sampled being male. This is commonly seen among the HCV positive community, with IDU and risky behaviours generally being more common among males than females (Public Health England, 2015).

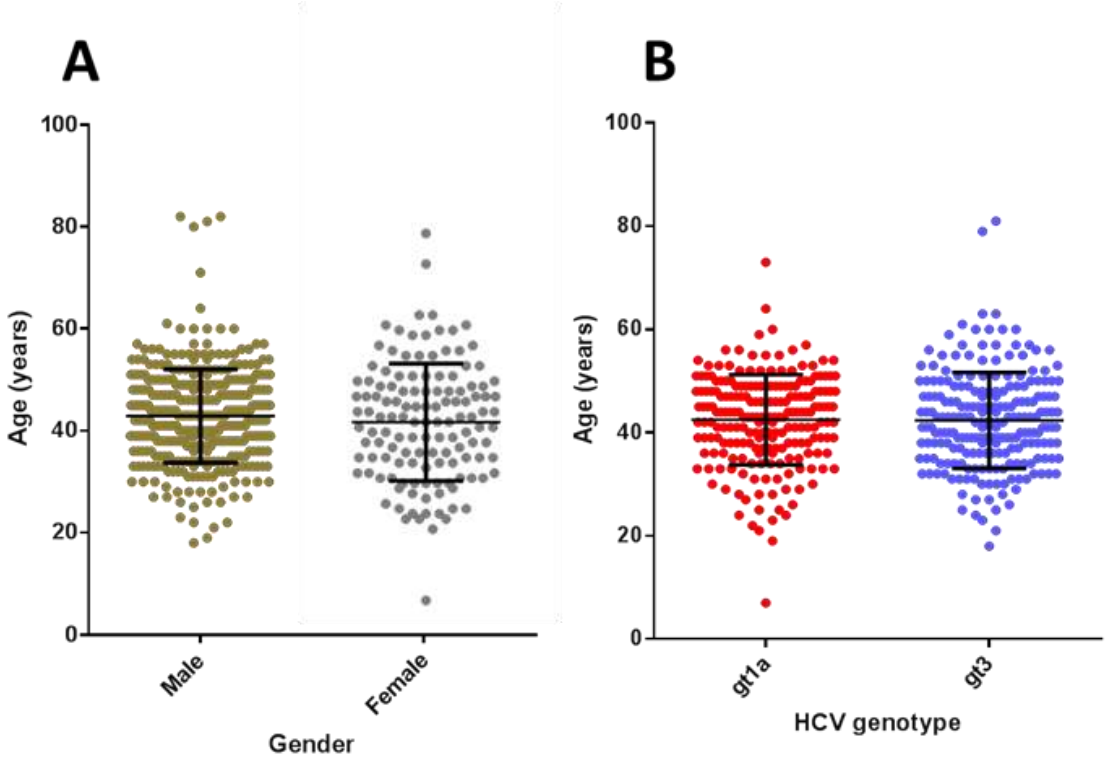


Figure 3-3; Ages of HCV-infected individuals

(A) Dot plot showing the distribution of ages of 1132 HCV-infected individuals by gender. (B) Dot plot showing the distribution of ages of HCV-infected individuals by HCV genotype (gt1a or gt3).

The average age of individuals sampled was 42.6 ± 9.8 years, with a range of 7-82 years. 97.3% of individuals in the study were aged between 18 and 60 years and the interquartile age of the cohort was 36-49 years. The range of ages observed was comparable between genders (p=0.85) with females having an average age of 41.9 ± 11.5 years and males having an average age of 42.8 ± 9.2 years. A greater proportion of females (p=0.02) were diagnosed under the age of 30 (15.1%) than males (8.7%), suggesting that females either are infected with HCV at a younger age or that they are more likely to seek medical treatment at an earlier stage than their male counterparts. The average age of individuals infected with genotype 1a

was 42.5 ± 8.8 years and this was similar (p=0.22) to the average age of those with genotype 3 infections (42.4 ± 9.3 years).

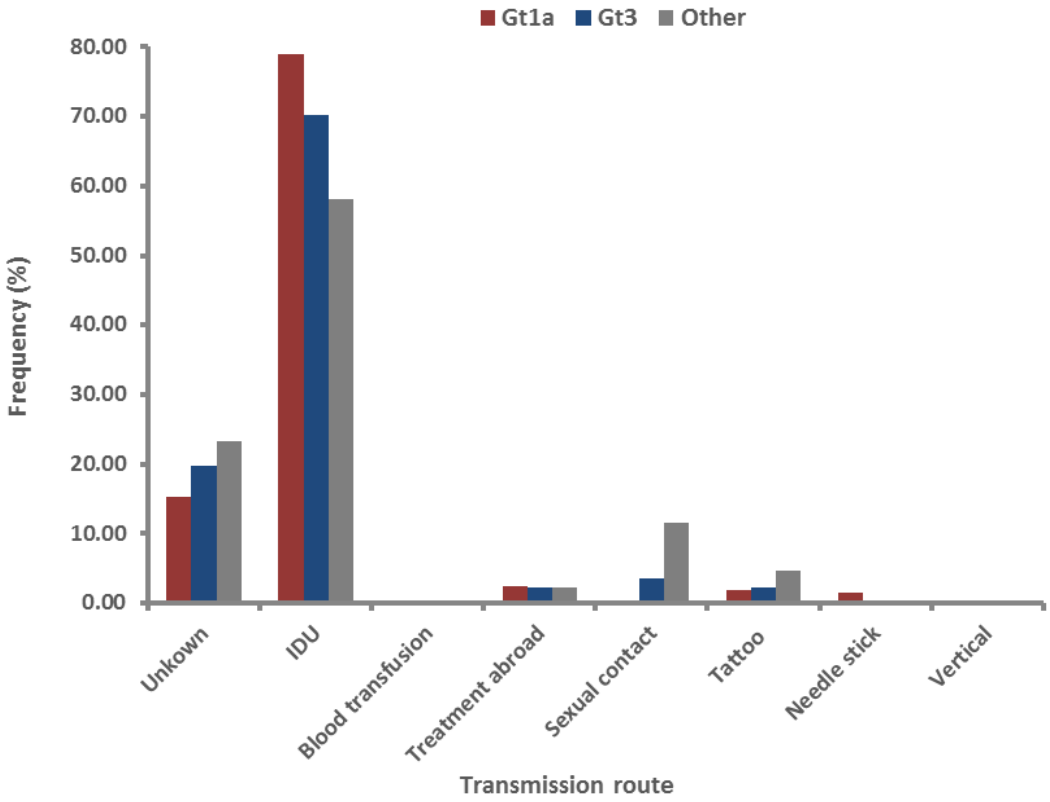


Figure 3-4; Risk factors for HCV infection

The risk factors for HCV infection for 590 HCV positive patients in the Glasgow and Paisley postcode regions. Patients have been stratified by genotype and 'other' contains patients infected with gt1b, gt2, gt4 and gt6.

Within the cohort, injecting drug use (IDU) was the major HCV risk factor for all genotypes, with 60-80% of patients recorded as having a history of IDU (figure 3-4). IDU was more common in individuals infected with gt1a or gt3 than in individuals infected with other genotypes of HCV. Other risk factors were infrequent, although sexual contact was identified as a potential route of transmission for 11.6% patients not infected with gt1a or gt3. A review of these patients (n=5) showed that they were infected with either gt1b (n=4) or gt6 (n=1). Among individuals not infected via IDU, the proportion of females in the cohort was 44.2%, 18.4% higher than the overall female representation. In 20.3% of patients, the potential route of HCV transmission was either not recorded or unknown. Whilst in a number of cases, this has likely not been recorded or the patient may have omitted to acknowledge certain risky behaviours, for a large proportion of these individuals the route of HCV transmission will be

genuinely unknown. It is also worth noting that in patients where the risk factor is unknown, renal disorders are listed as a co-morbidity in 9% cases. This is almost double the frequency of renal disorders across the whole dataset ($p = 0.08$) and whilst infection from inappropriately sterilised dialysis machines is rare in the UK, it is possible that treatment abroad may have contributed to a number of these infections. It should however be noted that renal disease is relatively common among HCV positive patients, often associated with liver failure, cryoglobulinaemia and other co-morbidities (Perico *et al.*, 2009) making it difficult to draw robust conclusions. The risk factor was listed as unknown for all individuals under the age of 20 in our cohort (n=3) and no individuals were listed as having contracted HCV infection through either vertical transmission or via a blood transfusion, implying these are uncommon events within the UK. The average age of individuals whose risk factor for infection with HCV was IDU (n=373) was significantly younger (41.7 ± 8.2 yrs) than other individuals (n=139) in the study (45.1 ± 12.9 yrs) ($p<0.01$). When individuals with an unknown risk factor for HCV transmission were excluded from the analysis (n=96), the age of individuals was higher (47.1 ± 13.3 yrs) and the difference was still found to be significant ($p<0.05$). The range in ages was also less in those who had contracted HCV via IDU (22-71 years) relative to other individuals in the cohort (7-82 years).

Using either a Fibroscan result of ≥14 kPa (Coco *et al.*, 2007) or 'cirrhosis' listed in the medical history of patients, 26.1% individuals in this study were found to have cirrhosis. Little difference was found in the rates of cirrhosis in females (21.4%) compared to males (27.7%) ($p = 0.21$). Patients with gt3 infections had higher rates of cirrhosis (30.2%) than those with gt1a infections (24.8%) ($p = 0.06$). Whilst neither of these findings are significant, there is a suggestion that gt3 infection may influence progression towards cirrhosis. Rates of fibrosis were also higher in male patients and in those with gt3 infections although Fibroscan has been shown to be less accurate at staging liver disease in patients with F1-F2 grading than it is at staging those with F3-F4 grading (Wong, 2013). Fibrotic staging by Fibroscan can also require adjusting in patients known to have an HIV-HCV co-infection and ≥12.5kPa is used as the cut-off for cirrhosis in these cases rather than ≥14kPa. Rates of known HIV-HCV co-infections were low in our cohort (n=8) and no patients had Fibroscan results that required adjustment. Rates of cirrhosis were particularly high however with 4/8 HIV co-infected individuals having cirrhotic livers. This is consistent with previous studies showing that HIV co-infection with HCV leads to worse outcomes for patients (Sulkowski, 2001; Reiberger *et al.*, 2010).

Within this cohort, 28.2% of patients had started treatment when the clinical data was collected, information on treatment status was not available for 7.5% patients and the remaining 64.3% patients had not been treated. There was only a short duration however between collection of the samples and subsequent collection of the linked clinical data. Genotyping assays are usually only done at the point of initial diagnosis and many patients sampled here are therefore likely to be in the early stages of clinical referral. This has likely contributed to the low treatment rates observed in this cohort. Of the patients that were treated, 10.9% had achieved SVR by the time of sampling and 50.2% of patients had not yet completed their therapy. Within the cohort, 30.9% individuals had failed to achieve an SVR, and this would be considered a good SVR rate for a cohort treated with PEG-IFN-α (Manns, Wedemeyer and Cornberg, 2006). However, it is not known what this cohort was treated with and it is probable that a subset of patients within the cohort were treated with DAAs such as telaprevir or Boceprevir and this has contributed to the high SVR rate observed.

3.3 Co-morbidities associated with hepatitis C infection

Within the cohort, 18.7% of individuals infected with HCV had no other recorded factors affecting their health. Consequently, the majority of individuals infected with HCV have additional complications in addition to their HCV infection. Among the other 81.3% individuals with HCV, the average number of co-morbidities occurring per person was 1.69 (± 1.38). Individuals most commonly had 1 single complication and the range was 0-7 co-morbidities per person. Complications are likely to be a combination of pre-existing medical conditions and additional morbidities that have developed as a result of infection with HCV. The most common conditions recorded, affecting ≥10% of the population sampled were, psychiatric disorders (44.6%), trauma, stabbing or assaults (15%), excessive alcohol intake (30.8%), neurological disorders (10.3%) and gastrointestinal complications (14.6%). The high levels of psychiatric disorders, trauma, stabbings, assaults and excessive alcohol consumption are patterns of behaviour reported by many patients within this cohort and are suggestive of chaotic lifestyles, which are frequently associated with IDU. Conditions affecting the respiratory, circulatory, digestive and central nervous systems are all reported at rates >5% and rates of co-infecting pathogens including HIV (1.6%), HBV (0.8%) and TB (0.4%) were found to be low in this cohort.

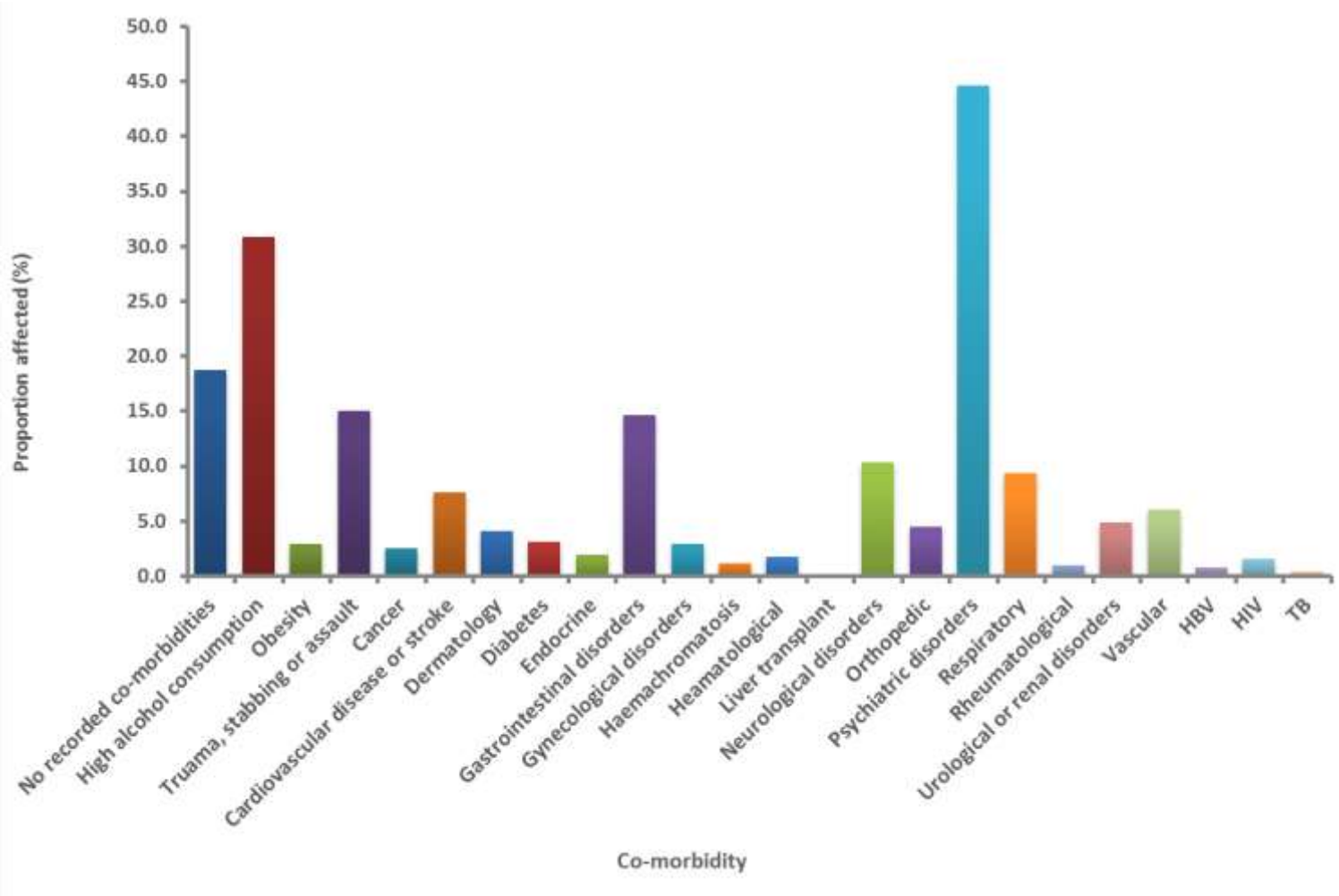


Figure 3-5; Co-morbidities

The frequencies of co-morbidities affecting HCV positive individuals in the cohort.

The rates of co-morbidities affecting the cohort were subsequently analysed after stratification by gender, liver disease severity, HCV genotype and age. Liver disease was defined as cirrhotic based on either a Fibroscan result of ≥14 kPa or 'cirrhosis' listed in the medical history of patients and patients without data recorded on liver disease were excluded from the analysis. Only individuals with either HCV gt1a or gt3 were included in the genotypic analysis and all other genotypes were excluded as numbers were low. For the age-based stratification, 45 years of age was selected as a threshold for the analysis based on United Nations age classifications, representing the division between middle adulthood and older adulthood and the end of reproductive age in females (United Nations, 1982). Univariable and multivariable analyses were performed on the dataset and the results are presented in tables 3-1 and 3-2.

In general, there was a good correlation between the univariate and multivariate analysis with the same co-morbidities being identified as significant in the majority of cases. In the univariate analysis, five significant differences were identified in the co-morbidities affecting male and female patients with HCV. Female patients were more likely to have psychiatric disorders, respiratory complications and, as would be expected, gynaecological conditions. Males were more likely to have been involved in a trauma, stabbing or assault and were almost twice as likely to have gastrointestinal complications. In the multivariable analysis, only three factors were found to be significant between male and female patients. Psychiatric disorders and gynaecological conditions were significantly more common among females infected with HCV whereas gastrointestinal conditions were more common in males with HCV. One male patient was identified with gynaecological complications and this is likely to be a mistake by the clinician filling out a sample collection form.

Table 3-1; Univariable analysis of the recorded co-morbidities

Results of univariable analyses of the co-morbidities affecting individuals infected with HCV analysed by gender, severity of liver disease and HCV genotype. P values ≤0.05 were considered statistically significant and have been highlighted. Values tending towards significance (between p>0.05 and p ≤0.1) have been underlined.

	Gender					Liver health					Genotype				
Co-morbidity	**M (n=438)**	**%**	**F**	**(n=152)**	**%**	**p**	**Not Cirrhotic (n=379)**	**%**	**Cirrhotic (n=134)**	**%**	**p**	**Gt1a (n=249)**	**%**	**Gt3 (n=247)**	**%**	**p**
No recorded co-morbidities	88	20.1	25	16.4	0.28	82	21.64	14	10.45	0.62	51	20.48	47	19.03	0.60
High alcohol consumption	134	30.6	42	27.6	0.96	100	**26.39**	58	**43.28**	**<0.01**	77	30.92	69	27.94	0.18
Obesity	10	2.3	6	3.9	0.16	9	2.37	6	4.48	0.49	9	3.61	5	2.02	0.37
Truama, stabbing or assault	73	**16.7**	17	**11.2**	**0.02**	57	15.04	20	14.93	0.35	37	14.86	41	16.60	0.49
Cancer	10	2.3	3	2.0	0.78	7	1.85	6	4.48	0.43	6	2.41	5	2.02	0.51
Cardiovascular disease or stroke	32	7.3	11	7.2	0.98	26	6.86	13	9.70	0.73	23	9.24	12	4.86	0.10
Diabetes	11	2.5	6	3.9	0.90	9	2.37	7	5.22	0.08	7	2.81	8	3.24	0.80
Dermatology	18	4.1	6	3.9	0.51	13	3.43	8	5.97	0.12	10	4.02	9	3.64	0.94
Endocrine	5	1.1	5	3.3	0.09	8	2.11	2	1.49	0.98	4	1.61	5	2.02	0.64
Gastrointestinal disorders	67	**15.3**	13	**8.6**	**0.05**	37	**9.76**	38	**28.36**	**<0.01**	28	11.24	34	13.77	0.24
Gynecological disorders	1	**0.2**	19	**12.5**	**<0.01**	12	3.17	3	2.24	0.49	6	2.41	5	2.02	0.89
Haematological	9	2.1	2	1.3	0.92	4	**1.06**	5	**3.73**	**0.01**	5	2.01	6	2.43	0.63
Hemochromatosis	4	0.9	2	1.3	0.53	3	0.79	3	2.24	0.19	3	1.20	1	0.40	0.36
Liver transplant	1	0.2	0	0.0	0.72	1	0.26	0	0.00	0.26	1	0.40	0	0.00	0.26
Neurological disorders	48	11.0	12	7.9	0.85	39	10.29	14	10.45	0.48	18	**7.23**	32	**12.96**	**0.05**
Orthopaedic	21	4.8	11	7.2	0.28	16	4.22	7	5.22	0.79	17	6.83	11	4.45	0.16
Psychiatric disorders	172	**39.3**	86	**56.6**	**<0.01**	164	43.27	65	48.51	0.31	108	43.37	110	44.53	0.90
Respiratory	35	**8.0**	21	**13.8**	0.17	35	9.23	13	9.70	0.75	22	8.84	20	8.10	0.90
Rheumatological	3	0.7	2	1.3	0.49	4	1.06	1	0.75	0.72	2	0.80	1	0.40	0.83
Urological or renal disorders	21	4.8	7	4.6	0.76	12	**3.17**	13	**9.70**	**0.03**	12	4.82	8	3.24	0.36
Vascular	24	5.5	11	7.2	0.74	17	**4.49**	14	**10.45**	**0.01**	12	4.82	20	8.10	0.14
HBV	4	0.9	0	0.0	0.59	2	0.53	2	1.49	0.43	3	1.20	1	0.40	0.49
HIV	8	1.8	1	0.7	0.55	4	1.06	4	2.99	0.49	5	2.01	2	0.81	0.37
TB	3	0.7	0	0.0	0.27	2	0.53	0	0.00	0.60	2	0.80	0	0.00	0.14

Table 3-2; Multivariable analysis of the recorded co-morbidities

Results of a multivariable analysis of the co-morbidities affecting individuals infected with HCV analysed by gender, severity of liver disease and HCV genotype. P values ≤0.05 were considered statistically significant and have been highlighted. Values tending towards significance (between p>0.05 and p ≤0.1) have been underlined.

	Gender					Liver health					Genotype				
Co-morbidity	**M (n=438)**	**%**	**F (n=152)**	**%**	**p**	**Not Cirrhotic (n=379)**	**%**	**Cirrhotic (n=134)**	**%**	**p**	**Gt1a (n=249)**	**%**	**Gt3 (n=247)**	**%**	**p**
No recorded co-morbidities	88	20.1	25	16.4	0.69	82	**21.64**	14	**10.45**	**<0.01**	51	20.48	47	19.03	0.97
High alcohol consumption	134	30.6	42	27.6	0.34	100	**26.39**	58	**43.28**	**<0.01**	77	30.92	69	27.94	0.31
Obesity	10	2.3	6	3.9	0.19	9	2.37	6	4.48	0.22	9	3.61	5	2.02	0.34
Truama, stabbing or assault	73	16.7	17	11.2	0.11	57	15.04	20	14.93	0.97	37	14.86	41	16.60	0.87
Cancer	10	2.3	3	2.0	0.84	7	1.85	6	4.48	0.10	6	2.41	5	2.02	0.69
Cardiovascular disease or stroke	32	7.3	11	7.2	0.99	26	6.86	13	9.70	0.29	23	9.24	12	4.86	0.10
Dermatology	18	4.1	6	3.9	0.85	13	3.43	8	5.97	0.20	10	4.02	9	3.64	0.53
Diabetes	11	2.5	6	3.9	0.60	9	2.37	7	5.22	0.10	7	2.81	8	3.24	0.66
Endocrine	5	1.1	5	3.3	0.07	8	2.11	2	1.49	0.66	4	1.61	5	2.02	0.80
Gastrointestinal disorders	67	**15.3**	13	**8.6**	**0.04**	37	**9.76**	38	**28.36**	**<0.01**	28	11.24	34	13.77	0.29
Gynecological disorders	1	**0.2**	19	**12.5**	**<0.01**	12	3.17	3	2.24	0.58	6	2.41	5	2.02	0.94
Haemachromatosis	4	0.9	2	1.3	0.66	3	0.79	3	2.24	0.18	3	1.20	1	0.40	0.29
Heamatological	9	2.1	2	1.3	0.32	4	**1.06**	5	**3.73**	**0.04**	5	2.01	6	2.43	0.67
Liver transplant	1	0.2	0	0.0	0.56	1	0.26	0	0.00	0.55	1	0.40	0	0.00	0.30
Neurological disorders	48	11.0	12	7.9	0.40	39	10.29	14	10.45	0.96	18	**7.23**	32	**12.96**	**0.03**
Orthopedic	21	4.8	11	7.2	0.30	16	4.22	7	5.22	0.63	17	6.83	11	4.45	0.90
Psychiatric disorders	172	**39.3**	86	**56.6**	**<0.01**	164	43.27	65	48.51	0.30	108	43.37	110	44.53	0.95
Respiratory	35	8.0	21	13.8	0.19	35	9.23	13	9.70	0.87	22	8.84	20	8.10	0.49
Rheumatological	3	0.7	2	1.3	0.46	4	1.06	1	0.75	0.75	2	0.80	1	0.40	0.53
Urological or renal disorders	21	4.8	7	4.6	0.52	12	**3.17**	13	**9.70**	**<0.01**	12	4.82	8	3.24	0.55
Vascular	24	5.5	11	7.2	0.65	17	**4.49**	14	**10.45**	**0.01**	12	4.82	20	8.10	0.12
HBV	4	0.9	0	0.0	0.24	2	0.53	2	1.49	0.28	3	1.20	1	0.40	0.29
HIV	8	1.8	1	0.7	0.40	4	1.06	4	2.99	0.12	5	2.01	2	0.81	0.22
TB	3	0.7	0	0.0	0.41	2	0.53	0	0.00	0.40	2	0.80	0	0.00	0.15

Among individuals with cirrhotic livers the frequencies of excessive alcohol intake, gastrointestinal complications, urological and renal conditions, vascular complications and haematological disorders were all found to be more common in both the univariate and multivariate analysis. The majority of these complications may occur as a direct result of having poor liver function apart from excessive alcohol intake which is likely to be a causative association. The multivariate analysis additionally identified that having no other co-morbidities in addition to HCV infection was significantly more common in individuals without cirrhotic livers.

After stratification by genotype, only a single co-morbidity was found to have a significantly different distribution between the two genotypes analysed within the cohort. Neurological disorders were found to be significantly more common in patients infected with gt3 (12.96%) in both the univariate and the multivariate analysis than in patients infected with gt1 (7.23%). The cohort was also analysed after stratification by age and this data is shown in the appendix, table 8-3. In the univariate analysis, older patients (≥45yrs old) were found to have significantly more complications associated with cardiovascular disease or stroke, cancer and gastrointestinal disorders. Additionally, in the multivariate analysis rheumatological and haematological disorders were also found to be significantly more common among individuals with HCV over the age of 45. Many of these disorders are more frequently observed in older populations and may not have occurred as a direct consequence of HCV infection.

3.4 Liver disease and viral load

HCV viral load is considered to be an important prognostic indicator and can be used as a marker of how effectively the virus is replicating within an infected individual. Studies have shown that patients with high viral loads (>800,000 IU/ml) may respond poorly to treatment (Zeuzem *et al.*, 2006; Poordad *et al.*, 2012). Significant drops in HCV viral load (>2logs) in the early phases of treatment are also considered indicative of a likely positive response to treatment and a high chance of achieving SVR (Trepo, 2000). Viral load means and distributions were analysed in our cohort using t-tests for analyses of two groups and ANOVA when three or more groups were analysed. Data on viral load was available for all 1132 patients tested at the WSSVC during the period August 2013-March 2014 (figure 3-6 A) and additional information on the severity of liver disease was

available for 590 patients form the Glasgow and Paisley postcode regions (figure 3-6 B and figure 3-7). All samples in the cohort had recently undergone genotyping at the WSSVC. Viral genotyping is usually preformed once, shortly after initial diagnosis and not commonly repeated. As a result, the majority of samples analysed subsequently should be pre-treatment viral loads. The results are shown in figure 3-6. Within this cohort, the average HCV viral load was 5.6 ± 1.0 $\log_{10}$ IU/ml. Viral loads in patients ranged from 1.39 – 7.30 $\log_{10}$ IU/ml and the ranges were comparable for both genotype 1a and genotype 3 patients. Subjects infected with genotype 3a had a similar average viral load (5.6 ± 0.9 $\log_{10}$ IU/ml) to subjects infected with genotype 1a (5.7 ± 0.8 $\log_{10}$ IU/ml). After stratification by genotype, females infected with HCV had, on average, a lower viral load (5.4 ± 1.0 $\log_{10}$ IU/ml) than that of their male counterparts (5.7 ± 0.9 $\log_{10}$ IU/ml) (p=0.01). When analysed by both HCV genotype and gender, a marked pattern in viral load emerged. Females with genotype 3 infections were found to have significantly lower viral loads (5.2 ± 0.9 $\log_{10}$ IU/ml) than males infected with either genotype 1a (5.8 ± 0.9 $\log_{10}$ IU/ml) or genotype 3 (5.7 ± 0.8 $\log_{10}$ IU/ml) (p = 0.002 and 0.015 respectively).

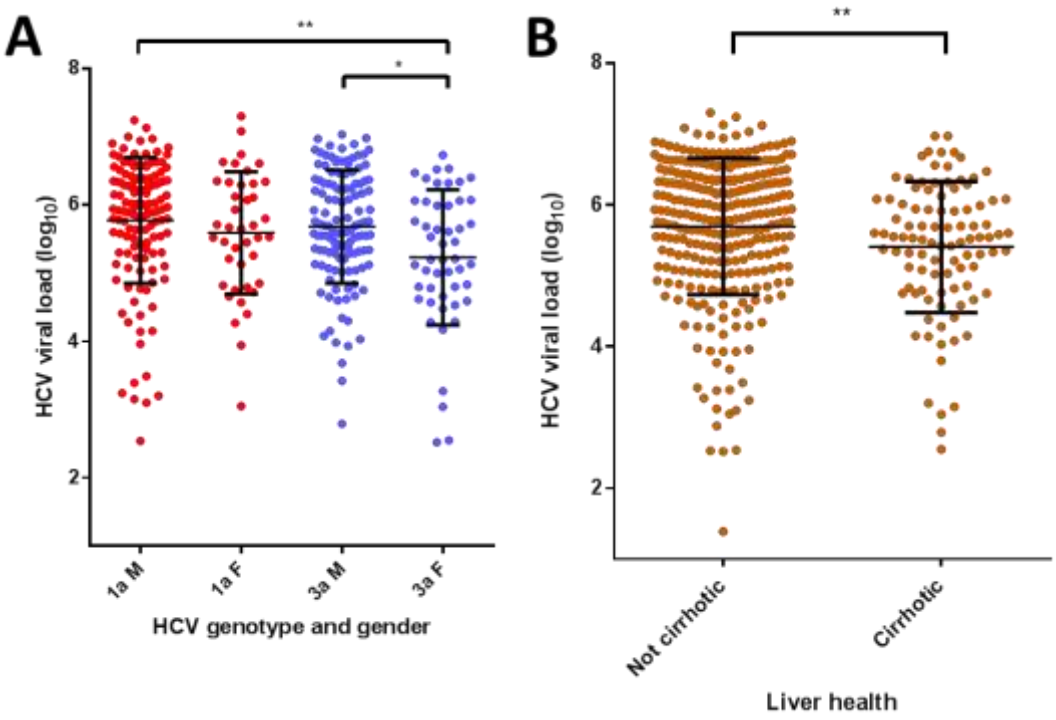


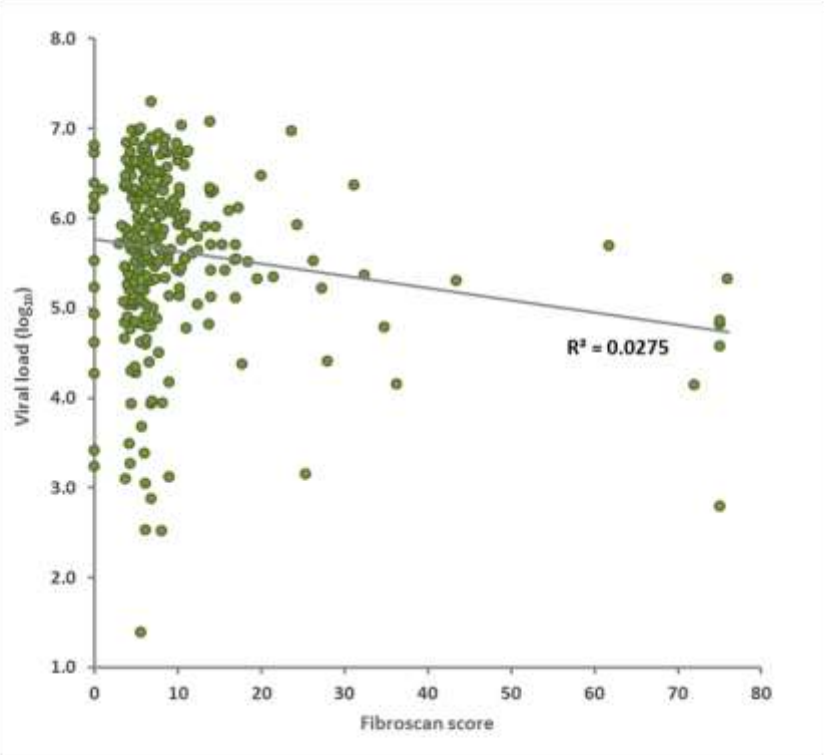
Figure 3-6; Viral load

(A) Viral load distribution by gender and HCV genotype. (B) Viral load distribution by severity of liver disease

The average viral load of individuals with cirrhotic livers was also found to be significantly lower (5.4 ± 0.9 $\log_{10}$ IU/ml) than the viral load of individuals who did not have cirrhotic

livers (5.7 ± 1.0 log_{10} IU/ml) (p = 0.009). This trend was explored further using patient data on the severity of liver disease and viral load. An analysis of viral load and Fibroscan results (available for 67.6% subjects, n=347) showed a poor correlation between liver disease and viral load within our cohort (figure 3-7, A). However a Chi-squared analysis, using a high viral load cut-off of ≥5.9 log_{10} (equates to 800,000 IU), indicated that high viral loads were significantly less common among those with cirrhotic livers (figure 3-7, B). These results suggest that the HCV virus is less capable of replicating in a diseased liver.

A



B

	Non-cirrhotic (n=304)	%	Cirrhotic (n=98)	%	p
High viral load	148	48.7	33	33.7	0.01

Figure 3-7; Viral load and liver disease

(A) Correlation between HCV viral load and Fibroscan results with a trend line and 95% confidence intervals. (B) Results of a univariable analysis of the frequency of high viral loads (VL ≥ 5.9 log_{10}) in patients according to liver cirrhosis status.

3.5 Chapter conclusions

The ages of individuals infected with HCV ranged from 7-82 and all genotypes, apart from gt5 and gt7, were identified within the population. Whilst this does suggest some diversity within the HCV-positive population in Scotland, the majority of HCV infections in the population sampled were either gt1a or gt3, This is similar to the distribution of HCV genotypes observed across the UK although there appears to an increased prevalence of gt3 infection in Scotland (Public Health England, 2015). Most of the cohort comprised individuals with a history of injecting drug use which is known to have been to have been a significant problem in Scotland, particularly in Glasgow, during the 1990s (Frischer *et al.*, 1993; Hay *et al.*, 2009). Glasgow has been documented to have had one of the highest rates of HCV infection among IDUs in Western Europe and consequently rates of HCV infection remain high within the region (Hutchinson *et al.*, 2002; Roy *et al.*, 2007). Previous work has highlighted longer injecting careers, relatively high rates of incarceration and the delayed provision of needle exchange services in the region as contributing factors (Hutchinson *et al.*, 2002; Parsons *et al.*, 2002; Judd *et al.*, 2005). Males comprised almost three-quarters of this cohort which is consistent with higher rates of IDU typically being reported among males (2, 11). Other routes of transmission were rare in the population and typically associated with individuals of an older age. This older age may reflect a number of transmissions via infected blood product transfusions prior to introduction of HCV testing. The rates of blood-transfusion associated HCV infection were fairly low in our cohort and the number of transmissions from an unknown source relatively high. The reasons for this particular finding were unclear, although they may be associated with the study focus on gt1a and gt3. HCV genotypes gt1a and gt3 are strongly associated with PWID and other genotypes, such as gt1b and gt2 are more frequently observed in older patients with have a history of blood transfusion (Smith *et al.*, 1997). The low numbers of gt1b and gt2 infected patients in the study may therefore have contributed to the low numbers of blood-transfusion associated HCV infection observed.

Rates of treatment among our cohort were not high and reasons for this are unclear. Among those that were treated, SVR rates were low but a majority of subjects were still undergoing treatment and the non-SVR rates suggested eventual outcomes were good for a cohort most likely treated with PEG-IFN-α (Manns, Wedemeyer and Cornberg, 2006). Both male gender and cirrhosis have been previously correlated with poor responses to treatment (Veldt *et al.*, 2007; Innes *et al.*, 2012). In our cohort, rates of cirrhosis were

higher among those not achieving SVR (37.5%) than in those who did achieve SVR (16.7%). In the cohort, 81.25% of individuals not achieving SVR were male, also suggesting this may be a factor in the poor SVR rates observed as studies have previously identified higher SVR rates among women, particularly for gt3 viruses (Manns *et al.*, 2001; Innes *et al.*, 2012).

Data from the cohort suggested that a number of factors may contribute to a more rapid progression of liver disease. Rates of cirrhosis were higher in both males and individuals with gt3 infections compared to females and those with gt1a infections. Being of male gender has been previously correlated with an increased risk of progression towards cirrhosis (Thierry Poynard, Bedossa and Opolon, 1997). Studies have suggested that this may be attributable to the protective effects of oestrogen (22) or findings that females are less likely to be heavy drinkers (23). There was some evidence of this in our cohort with higher rates of excessive alcohol intake observed in males (32.6%) relative to females (27.6%). Genotype 3 HCV infections have previously been correlated with a more rapid disease progression and increased rates of steatosis and fatty liver disease relative to other genotypes (Poynard, 2003; Probst *et al.*, 2011). *In vitro* studies have implicated a probable role for the gt3 core protein in triggering increased triglyceride accumulation and upregulating fatty acid synthase activity (Abid *et al.*, 2005; Jackel-Cram, Babiuk and Liu, 2007). Prolonged accumulation of fatty acids within the liver occurs as a direct consequence and this has been associated with the upregulation of inflammatory markers and reduced metabolic competence (Anderson, 2008). The high numbers of both males and gt3 infections in our Scottish cohort suggest that increased understanding of these factors and their contribution to HCV progression will be important in improving patient outcomes in the region.

High rates of comorbidities were observed within our cohort and this is consistent with previous studies indicating that rates of hospitalisation for any cause were significantly higher among HCV-positive individuals relative to the rest of the population within Scotland (McDonald *et al.*, 2011). The range of additional comorbidities identified among individuals with HCV also indicates that whilst the virus primarily infects the liver, chronic HCV infection is a complex systemic disease, associated with numerous extra-hepatic manifestations. Our findings here point towards a complex mixture of factors that are likely linked to both the high rates of IDU in the population and occurring as a direct result of HCV infection. Previous studies have correlated high rates of substance abuse with

mental health (Regier *et al.*, 1990; Crawford, Crome and Clancy, 2003). IDU and addiction are frequently recorded clinically as psychiatric disorders, and therefore observing high rates of psychiatric disorders in a population with high rates of transmission via IDU was an expected correlation. Infection with HCV has also been correlated with a marked cognitive impairment and depression in many patients and this may be contributing to the high rates of psychiatric disorders recorded (Forton *et al.*, 2002; Hilsabeck *et al.*, 2003). Gastrointestinal complications and neurological disorders have both been previously linked with HCV infection (Louie *et al.*, 2012; McCarthy and Ortega, 2012). The central role of the liver in both digestion and metabolism means that liver dysfunction consequently causes numerous gastrointestinal disorders and it is therefore no surprise that high rates are observed in HCV-positive individuals. In terms of neurological disorders, numerous studies have linked HCV to poly- and mono-neuropathies, transverse myelopathy and cognitive impairment (Heckmann *et al.*, 1999; McCarthy and Ortega, 2012). Withdrawal from alcohol is known to cause seizures (Delanty, Vaughan and French, 1998) and, in our cohort, 54.7% of patients with neurological disorders were also found to have a history of excessive alcohol consumption. This is significantly higher than the rate of alcoholism among individuals without neurological disorders which was 29.1% ($p<0.01$), suggesting that this is likely to be a contributing factor to the high levels of neurological disorders. In the cohort, 72.7% of subjects had a history of injecting drug use and the high rates of trauma, stabbing and assault observed may be associated. Individuals participating in injecting drug use are often described as having chaotic lifestyles (March, Oviedo-Joekes and Romero, 2006) and among those with a history of trauma, stabbing and assaults, it was more common to find a history of both injecting drug use (89.6%) and excessive alcohol consumption (61.3%).

The univariable and multivariable analyses showed good correlation and highlighted a number of interesting trends within the data. They indicated a wide range of interacting factors making the management of patients with HCV a complex process involving multiple specialities. In both analyses, psychiatric disorders were significantly more common in women. It has previously been reported that higher rates of eating disorders, depression, anxiety disorders and mood disorders are often observed in female substance abusers (Brooner *et al.*, 1997; Chander and McCaul, 2003).. Histories of gynaecological treatment and obstetric complications have been highlighted in the literature as triggers for the development of substance abuse in females (Busch, McBride and Benaventura, 1986). Rates of cirrhosis were higher among male patients than female patients and this may

partially explain the high rates of gastrointestinal disorders observed among males. The large number of comorbidities associated with cirrhosis highlights the central role of the liver in interactions with other organ systems. Successful treatment of HCV has been shown to significantly slow progression towards fibrosis and cirrhosis, alleviating many of the associated comorbidities (George *et al.*, 2009; Pearlman and Traub, 2011; Hsu *et al.*, 2014). Ageing itself has been associated with a decline in liver function and an increased prevalence of co-morbidities (Schmucker, 2005). In this study, rates of both gastrointestinal and haematological disorders were increased in patients aged 45 and over, but without a HCV negative cohort for comparison, the significance of this finding is unclear. Rates of cirrhosis among older patients were also elevated at 34.6%. Similar findings have been recorded in previous studies (Gramenzi *et al.*, 2010). With HCV-infected individuals frequently remaining asymptomatic for more than a decade, many of these individuals are unaware of their infection status until they become ill and consequently, HCV is rapidly becoming a disease of the elderly. Treatment and disease management is often more complex in older patients and those over the age of 65 frequently have contradictions for therapies (NIH Consensus Development Panel, 2002; Mindikoglu and Miller, 2009). These findings underscore the importance of early diagnosis and intervention whilst there is a strong chance of improving patient prognoses.

The only significant differentially occurring co-morbidity identified between gt1a and gt3 patients was neurological disorders which were more common in gt3 patients. Previous studies have linked the HCV core protein with neurotoxic effects associated with upregulation of extracellular signal-related kinase (ERK) via TLR-2 signalling (Paulino *et al.*, 2011). The gt3a HCV core protein has been previously shown to induce distinct effects on fatty acid metabolism (Jackel-Cram, Babiuk and Liu, 2007) and therefore it is possible that it is also causing additional neurotoxic effects. The significance of this finding is unclear however as data was not available on the exact nature of the neurological disorders occurring amongst patients in the cohort. A range of neurological disorders are found in HCV–positive patients including neuropathies, encephalitis, myelitis and cognitive impairment (McCarthy and Ortega, 2012). Neurological disorders, such as encephalomyopathy, are also common amongst patients with histories of excessive alcohol consumption and cirrhosis, regardless of whether or not they are HCV positive. Both excessive alcohol consumption and cirrhosis were frequently reported in the patient histories within this cohort, making it difficult to infer if the neurological disorders recorded are specifically related to HCV. In order to further confirm the significance of

these findings, it would be useful to repeat the analysis after excluding patients with a history of excessive alcohol consumption and see if the increased rate of neurological disorders in gt3 patients is still evident.

A number of significant trends were associated with viral load. The viral load was on average lower in females infected with HCV gt3 relative to other patient groups. Previous studies have also correlated gt1 infection with higher viral loads relative to gt3 and other genotypes (Berger *et al.*, 1996; Soriano *et al.*, 2008; Rong *et al.*, 2012). Females with gt3a infections have the best response rates to IFN-based therapy (Manns *et al.*, 2001; Innes *et al.*, 2012), and rates of spontaneous clearance are higher among females (Wiese *et al.*, 2000; Barrett *et al.*, 2001; Micallef, Kaldor and Dore, 2006), suggesting that immune responses against HCV in females are more effective than in males. Studies have suggested these findings are linked to the protective effects of oestrogen although the exact mechanisms involved remain to be elucidated (Baden, Rockstroh and Buti, 2014). In this study, HCV viral load was additionally found to be significantly lower in individuals with cirrhotic livers. A study by Duvoux *et al*. (1999), documented similar findings, showing that lower viremias were detected in patient with more severe cirrhosis (Duvoux *et al.*, 1999), Intriguingly, the study also found that viral loads rapidly increased in patients after liver transplants, indicating that something in the cirrhotic liver was likely to be restricting viral replication (Duvoux *et al.*, 1999). During liver cirrhosis, the hepatic environment becomes highly dysregulated, the number of hepatocytes is reduced and there is an accumulation of fibrotic tissue. It is possible that these factors are interfering with the viruses' ability to replicate and transmit via cell-to-cell routes.

Whilst this chapter has highlighted a number of interesting trends among the HCV-infected population in Scotland, there are a number of limitations to the study. Our study cohort was sampled from patients residing only in the Glasgow or Paisley postcode district areas and therefore, a number of the trends we identified may be region-specific and not as common elsewhere in the UK. Studies utilising wider geographical areas would be required in order to confirm or refute this issue. Patient sampling was also through clinical contact, hence our study was likely biased towards older individuals with chronic HCV infection. The asymptomatic nature of acute HCV and chaotic lifestyles of many IDUs means that interaction with healthcare providers is often poor. Sampling throughout this study is, therefore, likely biased towards recovering PWID rather than active drug users and individuals sampled are likely to have more stable lifestyles as a result. Patients within

our cohort will also be more likely to engage with clinical services if they are in ill-health (HCV-related or otherwise) and this may have inadvertently increased the levels of co-morbidities observed within our cohort. This study relied heavily on clinical coding and it is possible that differences in disease categorization and biases of medical professionals may have unduly influenced the data. Human error, when recording and transferring data, may also have affected small numbers of samples, such as the male patient identified as having a gynaecological condition. As mentioned previously, it was not clear within our cohort whether or not the factors identified were a direct consequence of established HCV infection and this makes interpretation of the data complex. In order to understand the long-term trends associated with HCV infection, long-term prospective studies looking at larger cohorts are required. This study also lacked data on treatment outcomes. Given the recent increase in treatment options for HCV, an improved understanding of the most common co-morbidities associated with HCV and their impact on therapeutic adherence and outcomes would be of benefit to clinicians.

Chapter 4: Screening for mixed genotype infections

4.1 Background introduction

Whilst the recent licencing of DAA therapies for HCV has revolutionised treatment of the disease, many of the new drugs have been developed in the USA where gt1 is most prevalent infection, affecting just under 75% of the total infected population (Hajarizadeh, Grebely and Dore, 2013; Gower *et al.*, 2014). As a result, many of the new treatments have been shown to have genotype-specific efficacies and are particularly adept at treating gt1 infections (EASL, 2015). Trials examining treatment efficacy in other genotypes have not been widely funded and the ability of these drugs to treat other genotypes remains to be seen. Their impact on multiple genotype infections is also unclear and several studies have hypothesised that mixed genotype infections may only be partially resolved by use of inappropriate therapies (Loulergue, Mir and Sogni, 2012; McNaughton *et al.*, 2014). Previous studies examining the prevalence of mixed genotype HCV infections also show significant differences in the testing methods used and the populations sampled, leading to wide variation in the prevalence findings (table 1-2). In the UK, a relatively small study looking specifically at PWID and haemophiliacs found prevalence rates of 9% and 19% respectively (Buckton *et al.*, 2006). Given the wide availability of the new DAA drugs, it was felt that an improved awareness of mixed genotype prevalence within the UK may be useful for clinicians. In order to do this, a cohort of HCV –positive patients from across Scotland was curated in collaboration with the WSSVC and the BBV Edinburgh. Highly sensitive genotype-specific nested PCR assays were developed targeting the E1-E2 region and used to screen samples within the cohort infected with either gt1a or gt3 for the presence of a mixed genotype infections. As genotypes 1a and 3 contribute upwards of 90% of the total HCV infections in the UK, it was decided to only to screen for these genotypes as by our estimations, this should enable us to identify around 90% of the mixed infections present in the population sampled.

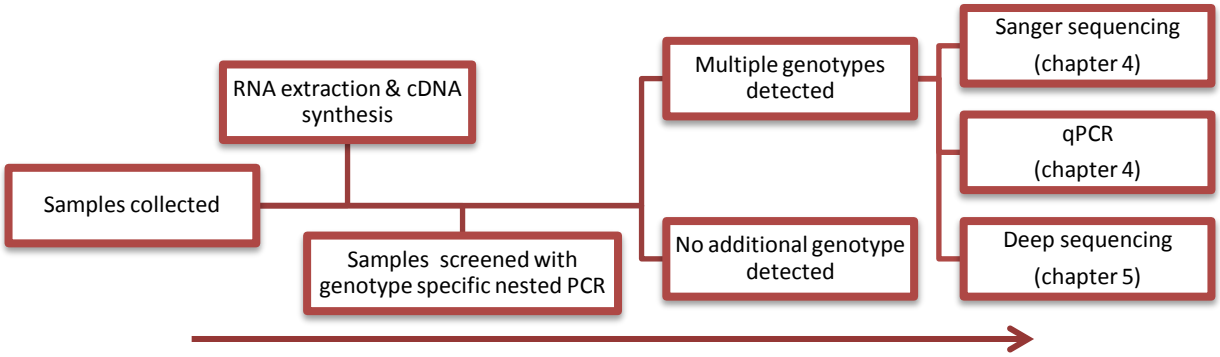


Figure 4-1; Workflow for the mixed infection screening

Workflow for the screening of mixed gt1a/gt3 infections and the subsequent testing performed on each sample

After screening of the cohort, samples with secondary infecting genotypes detected had both the gt1a and gt3 strains sent away for Sanger sequencing to confirm the presence of multiple genotypes (figure 4-1). Samples then underwent qPCR with genotype-specific primer and probe sets to examine the proportions of the different genotypes present in these individuals. Samples with mixed genotype infections present also underwent amplicon-based deep sequencing on the Illumina MiSeq platform and this is discussed in chapter 5.

4.2 Study controls

4.2.1 Selection of controls

H77 (accession number NC_004102) (Kolykhalov *et al.*, 1997) was used as the basis for the design of the genotype 1a controls. Selection of genotype 3a control was challenging as there are multiple reference strains used throughout the field with the most commonly cited strains being NZL-1 (NC_009824.1), S52 (GU814263.1) and DE (X76918.1) (Sakamoto *et al.*, 1994; Gottwein *et al.*, 2010). HCV S52 was isolated from a chimpanzee infected with HCV and therefore deemed not suitable for use as a control. The NZL-1 and DE were both found to have numerous mismatches in the primer binding regions of early assay designs. Numerous sequences were considered with a patient-isolated strain from a 2005 UK study eventually being selected as the most appropriate control (accession number GQ356206) (Humphreys *et al.*, 2009) (see appendix figure 8-1). Early assay designs were generated

using mostly locally-circulating sequences and the assays were re-designed at a later date to ensure broad specificity across the gt3a sub-type.

4.2.2 Production of controls

A protocol was successfully established that enabled the amplification of synthetic dsDNA fragments and subsequent transcription of the fragments into RNA for use as assay controls (section 2.3.1). The E1-E2 transcripts were used as controls for the nested PCR assay and the NS5B transcripts were used for relative quantification with the qPCR assay developed in section 4.5. Using this protocol, transcript controls were successfully produced for gt1a and gt3a. Fidelity of the amplification and transcription protocol was checked with Sanger sequencing of the fragments. Fragment sizes, positions in the HCV genome and any observed changes are highlighted in figure 4-2.

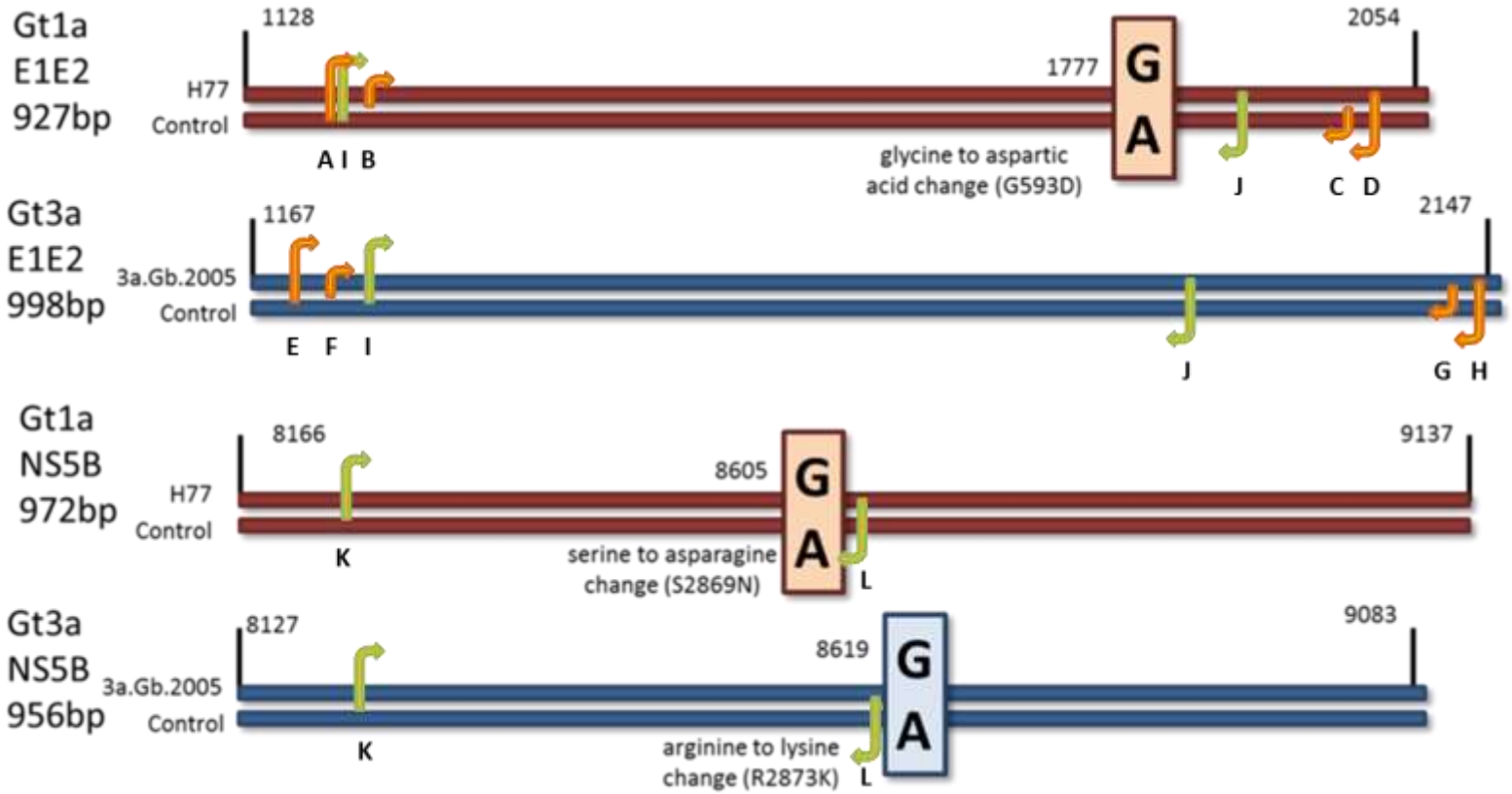


Figure 4-2; Synthetic control transcripts and primer positions

Illustration of the synthetic control fragments (not drawn to scale) designed to be used for the optimisation of mixed-genotype infection screening assays. Positions within the HCV genome are indicated and primers (A-K, table 2-5) are indicated by orange arrows (nested PCR) and green arrows (pan-genotypic primers). Discrepancies between the designed sequence and the control sequence after ligation into plasmids are highlighted with the resulting amino acid changes.

Sanger sequencing of the controls revealed single nucleotide substitutions in the gt1a NS5B control and both the E1-E2 and NS5B gt3 controls. All substitutions consisted of a G→A switch, resulting in non-synonymous changes to the encoded amino acid sequence and are indicated in figure 4-2. The substitutions are likely to have been introduced during the cloning and subsequent propagation of plasmids during production of the controls (section 2.31 for protocol). None of the sequence changes were located in primer or probe binding sites and they were therefore assumed unlikely to influence assay sensitivity. Agarose gel electrophoresis and spectrophotometer 260nm/280nm and 260nm/230nm ratios suggested the prepared transcripts were of good quality and suitable for the intended purpose. Transcripts were stored in RNA storage medium (Life Technologies) with 0.1U/ml RNasin (Promega) rather than nuclease free water as it is thought to improve the recovery of the transcripts after freeze-thaw cycles. Recovery of the transcripts by cDNA synthesis and subsequent nested PCR assays showed that the RNA produced was of a good quality and that the transcripts could be detected at low dilutions. The E1-E2 transcripts could be detected at 1 (gt1a) and 10 (gt3a) copies/µl and the NS5B transcripts were detected at 10^2 copies/µl (both genotypes) (figure 4-3).

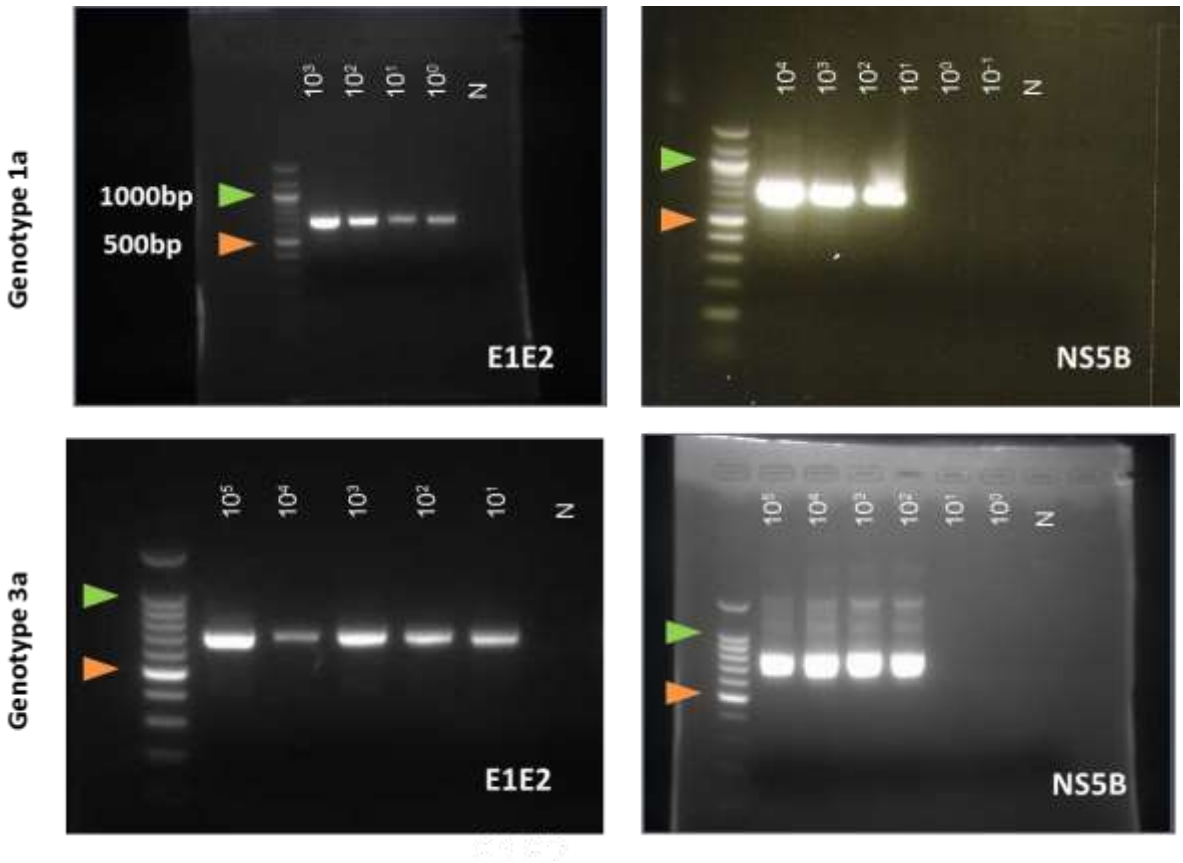


Figure 4-3; Recovery of the control transcripts by PCR

Results from PCR assays with serial dilutions of the gt1a and gt3a control fragments demonstrating the recovery of the E1-E2 and NS5B transcripts at low dilutions. The concentration of the dilutions tested is indicated along with the negative control (N). 1000bp (green arrow) and 500bp (orange arrow) are highlighted on the ladder.

4.3 Assessment of assay sensitivity

4.3.1 Probit analysis for E1-E2 assay controls

Primers for the nested PCR assays were designed in regions where there were ≥4 divergent nucleotides between the genotypes in order to confer genotypic specificity for the assays. Both the gt1a and the gt3 assay were tested with high dilutions ($\geq 10^4$ copies/µl) of the alternative control and samples of the genotype they were not specific for and no amplification was detected, indicating the assays were functioning as intended and there was no cross-reactivity.

Assay sensitivity was evaluated using known dilutions of the H77 and GB.2005 transcript controls. Dilutions of the gt1a and 3a E1-E2 transcript controls were tested in batches of 8

replicates with a negative control (figure 4). Results of the replicate testing were converted into probit values using Finney's table and assay outcomes are given in table 4-1 (Finney, 1952) (appendix table 8-3).

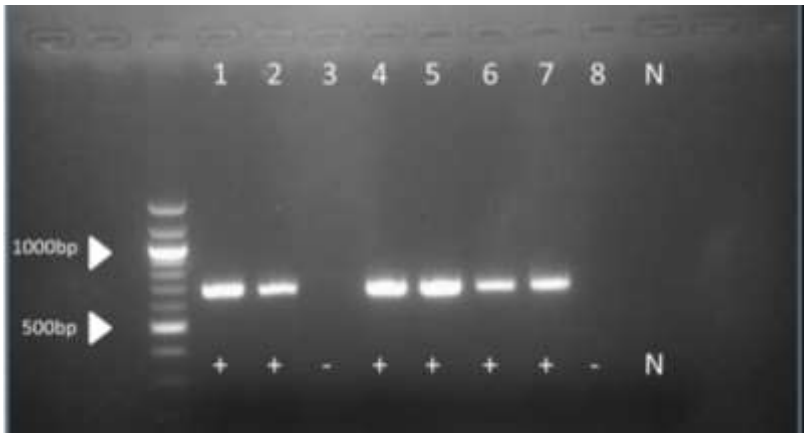


Figure 4-4; PCR replicates for probit analysis

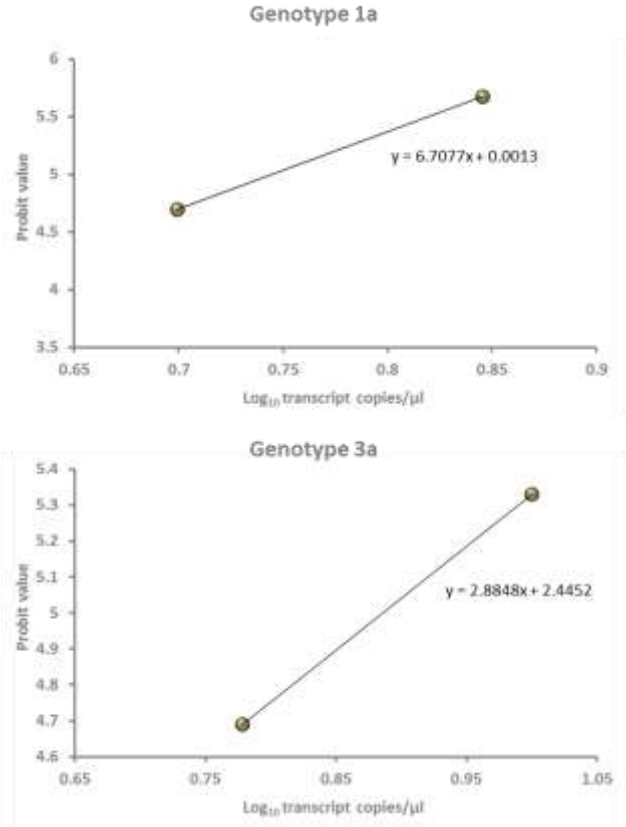
A gel of 8 nested PCR replicates of the gt1a transcript at 7 copies/µl used to calculate probit value. Positive results are indicated by + and negative results by -. A negative control (N) was used and 1000bp and 500bp are highlighted on the ladder.

Graphs were constructed by plotting probit values against Log_{10} copies/ µl of the transcript concentrations tested to obtain equations (table 4-1 and figure 4-5). The value of X was then calculated when Y=6.28 for each graph, this value being equivalent to a 90% detection rate. Using these values, a 90% detection rate of 18 copies/reaction (9 copies/µl) for gt1a, and 42 copies/reaction (21 copies/µl) for gt3a was calculated for the assays.

Table 4-1; Detection limits as determined by probit analysis

The detection limit of screening assays as determined using H77 E1-E2 (gt1a) and GB.2005 (gt3a) control transcripts and values used for determination of 90% detection rate. Log_{10} copies/ µl and the Probit value were plotted on a graph (figure 4-5) to obtain the equation used to calculate the assay sensitivity (indicated in antilog column)

Genotype	copies/µl (n)	Log_{10} copies/ µl	+ (%)	Probit value	Probit value (90% detection)	Log_{10}	Antilog
1a	7	0.84	75	5.67	6.28	0.94	8.63
	5	0.70	38	4.69			
3a	10	1	63	5.33	6.28	1.32	21.34
	6	0.78	37.5	4.69			



$$x = \frac{(y - 0.0013)}{6.7077}$$

$$x = \frac{(y - 2.4452)}{2.8848}$$

Figure 4-5; Equations used for the probit analysis

Graphs and corresponding equations used for the calculation of assay sensitivity. Values used for plotting the graph are indicated in table 4-1. Equations for x were solved for y = 6.28, equivalent to a 90% detection rate.

4.3.2 Mock mixed infections

Mock mixed infections were prepared using RNA prepared from the gt1a and gt3a transcript controls. Three different minor strain concentrations were tested in backgrounds of 10^6 copies/µl of the other genotype. Aliquots of unmixed minor strain transcripts were used as positive controls for the assays and H_2O was used as a negative control. Results are shown in figure 4-6 and table 4-2. Minor strains were consistently detected at low limits in a high background of the other genotype. Sanger sequencing was used to confirm that positive results indicated the correct genotype had been detected and there were no issues with primer specificity.

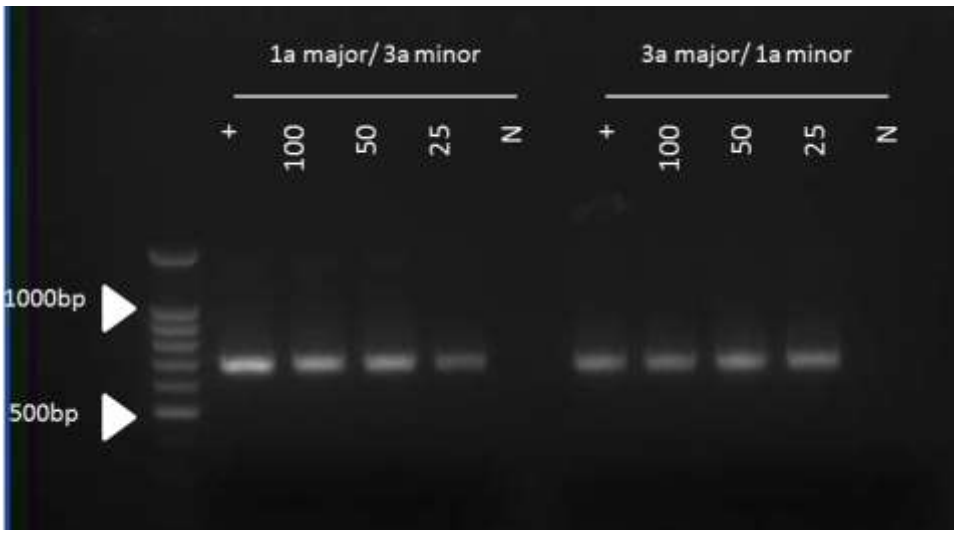


Figure 4-6; The detection of mock mixed infections

A gel of nested PCRs performed on mock mixed infections. The minor strain concentration (copies/µl) in each sample is shown on the gel. 10^4 copies/µl of the minor strain being tested for were used as positive controls (+). Negative controls (-) were run and 1000bp and 500bp are highlighted on the ladder.

Table 4-2; Assay sensitivity for mock mixed infections

Concentrations of major and minor strain spikes used to assess the specificity of the PCR assays

Minor strain	copies/µl	Major strain	copies/µl	Detected	Sequencing result
1a	25	3a	10^6	Y	gt1a
1a	50	3a	10^6	Y	gt1a
1a	100	3a	10^6	Y	gt1a
3a	25	1a	10^6	Y	gt3a
3a	50	1a	10^6	Y	gt3a
3a	100	1a	10^6	Y	gt3a

4.4 Screening

4.4.1 Prevalence of mixed gt1a/gt3 infections

A cohort of samples to screen for mixed gt1a/gt3 infection was compiled over the period August 2013-march 2014, consisting of any gt1a or gt3 sample that had been recently genotyped at the WSSVC and had ≥140 µl serum available. A total of 205 gt1a samples and 206 gt3 samples were collected and merged with an additional cohort of 48 gt1a and

47 gt3 samples collected as part of an earlier study. In order to screen for mixed genotype infections, the gt1a samples were screened using the gt3-specific nested PCR assay (section 2.3.3.2) and the gt3 samples were likewise screened with the gt1a-specific assay. Samples were screened alongside positive and negative controls to ensure the assay was functioning as intended. Positive results were based on the detection of a PCR product of the expected size (figure 4-7) and all positive results were confirmed by Sanger sequencing and BLAST analysis and phylogenetic analysis of the returned sequence. The positive result for sample G3-51 is illustrated in figure 4-7.

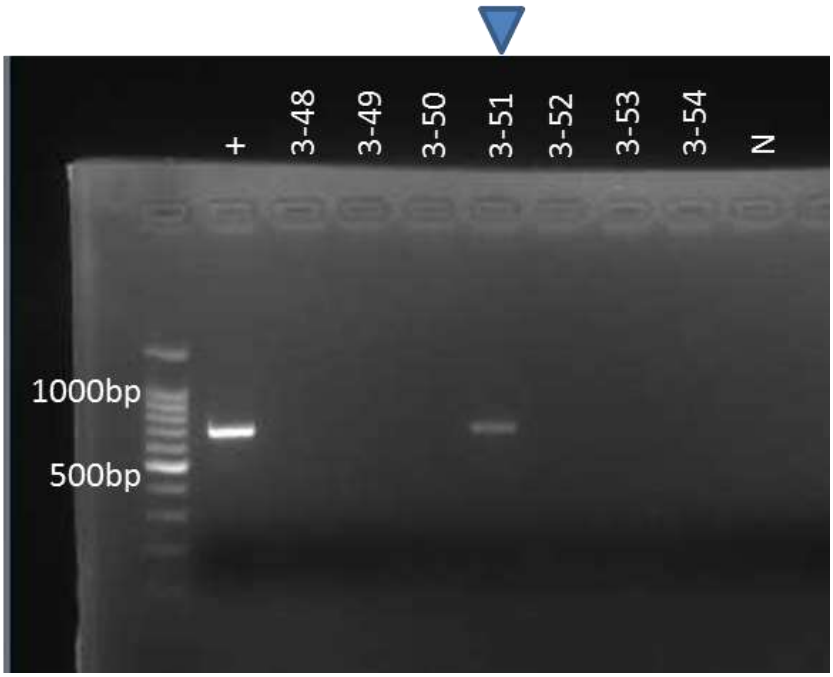


Figure 4-7; Illustration of a positive result in the mixed gt1a/gt3 screening

Results from the screening of the gt3 samples G3-48 – G3-54 for evidence of gt1a infection. A positive result for G3-51 (blue arrow) was identified during the screening of these samples. The positive control (+) was 10^4 copies/µl of the gt1a control and the negative control (N) was nuclease free water. The 1000bp and 500bp markers are highlighted on the ladder.

A total of 506 patients, clinically diagnosed with either gt1a or gt3 infection, were screened for the presence of a mixed gt1a/gt3 infection. Overall, 20 samples were identified that had a mixed gt1a/gt3 infection, but two of these samples were found to be serial samples from the same patient, taken approximately a month apart. These serial samples (G3-80 and G3-125) are discussed further in section 4.4.2 and both samples were taken forward for subsequent analysis (section 4.5 and chapter 5). In order to calculate the prevalence of

mixed gt1a/gt3 infection however, one sample was excluded on the basis that both G3-80 and G3-125 came from the same patient. Therefore, mixed gt1a/gt3 infections were detected in 19 individuals, giving a prevalence rate of 3.8% within the population that was sampled (figure 4-8). Of 253 gt3 samples screened for the presence of a co-infecting gt1a strain, 17 samples were found to have an undiagnosed gt1a strain, giving a mixed genotype infection prevalence rate of 6.7% within genotype 3-infected individuals. In the corresponding screen for gt1a diagnosed individuals, 2 out of 253 individuals were found to have an undiagnosed gt3a strain, giving a prevalence rate of 0.8%. The difference in the prevalence rates of mixed infection between the two genotypes was found to be significant using a Chi-squared test ($p<0.05$). No evidence of mixed gt1a/gt3 infections was detected in any of the 20 samples identified in this study by the WSSVC.

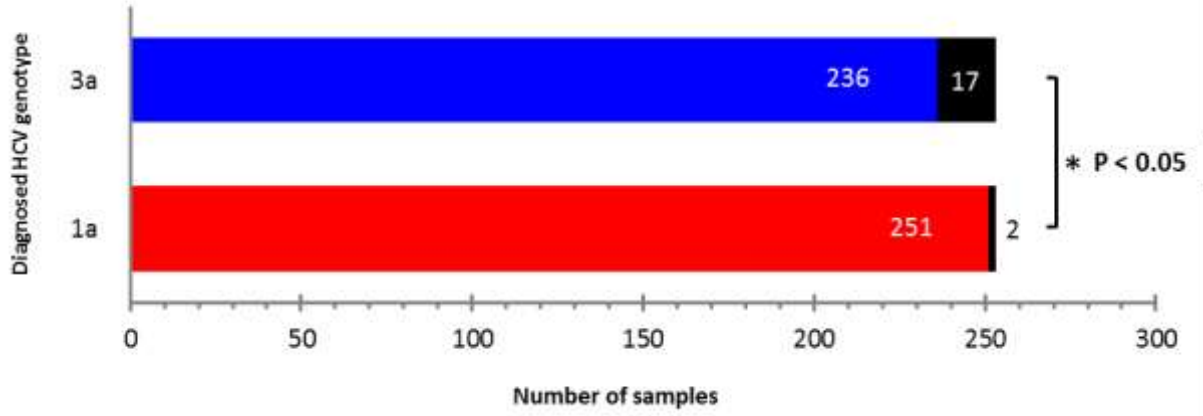


Figure 4-8; The prevalence of mixed genotype infections

The numbers of gt1a and gt3 samples with mixed gt1a/gt3 infections detected within the cohort. Samples genotyped as gt1a are shown in red and gt3 samples are shown in blue. Samples testing positive for mixed gt1a/gt3 infections are shown in black.

Mixed gt1a/gt3 HCV infections were detected in samples collected from individuals residing throughout Scotland (see figure 4-9, postcode data was not available for 7 samples) suggesting that no region is particularly affected by mixed genotype HCV infections. The average age of individuals with mixed infections was 41 ± 9.22 years which is comparable to the overall average age of the cohort sampled (42.6 ± 9.8 years). Individuals with mixed infection ranged in age from 24-61 years of age. Information was available on the viral load for 10/19 samples, and the average viral load was 5.39 ± 1.03

log10. This was similar to the average viral load observed in the cohort 5.56 ± 1.0 log10 IU/ml (p=0.66).

Figure 4-9; Postcodes of patients with mixed genotype infections

The locations of the individuals identified with mixed genotype HCV infections in Scotland.

Due to the ethical approval for collection of clinical data being restricted to the Glasgow and Paisley postcode districts, linked clinical data was only available for 4 of our subjects with mixed genotype infections. There was a mix of both male (n=2) and female (n=2) patients in the cohort and 3/4 subjects had a history of IDU. Liver disease was common amongst the subjects with 2/4 individuals having cirrhotic livers and another 1/4 having a history of fibrosis in their medical records. There was a wide range of co-morbidities in the group with an average of 2.75 co-morbidities per subject (range 1-6). All 4 of the patients had a history of psychiatric disorders. Three of the 4 patients had received treatment for their HCV infection and of these, 1/3 had not yet completed treatment and the other 2/3 had failed to achieve an SVR.

4.4.2 Phylogenetic analysis

Gt1a and gt3a isolates from samples that had mixed gt1a/gt3 infection detected in them underwent Sanger sequencing. The returned sequences were analysed in a maximum-likelihood phylogenetic tree (figures 4-10 and 4-11). Of the 19 samples identified with mixed genotype infections, paired gt1a and gt3a sequences were successfully obtained using Sanger sequencing for 19/19 samples.

Analysis of the returned sequences confirmed the presence of both genotype 1a and genotype 3a viruses in 19/19 samples (figures 4-10 and 4-11). For both genotypes, the level of variation among the strains involved in the mixed infections suggested that there are no viral-specific features in the E1-E2 region that are associated with involvement in a mixed genotype infection. The gt1a region of the tree also displayed the characteristic bifurcating topology often seen in HCV gt1a trees. Samples involved in mixed infections were found within both branches of the gt1a tree, further highlighting the diverse nature of strains detected in this study.

Table 4-3; Comparative pairwise distances of strains involved in single genotype and mixed genotype infections

Overall average pairwise distances of E1-E2 sequences from 19 samples with mixed genotype infection compared with gt1a (n=48) and gt3a (n=47) sequences isolated from individuals negative in the mixed genotype infection screen. All samples were collected from patients in a similar geographical region.

Sample group	Pairwise distance gt1a	Pairwise distance gt3a
Mixed genotype infection	0.176	0.185
Single genotype infection	0.182	0.174
Both	0.184	0.177

A pair-wise analysis of the sequences isolated from samples with mixed gt1a/gt3 infection was compared with sequences isolated from samples that were negative in the mixed infection screen (table 4-3). Comparable variation was observed within both sample groups, supporting the assertion that there a diverse range of samples involved in the mixed genotype infections. Diversity among the mixed infection gt1a sequences may have been reduced by the close relationships observed between some of the gt1a sequences. Gt1a sequences from G45 and G30 were shown to be closely related and formed a polytomy within the tree. Sequences isolated from G3-166 and G3-176 also appeared to be closely related. The bootstrapping analysis also indicated that there was more support for

the cladal relationships within the gt1a tree than there was in the gt3a tree, where only a single node had bootstrap support of ≥70%, the level generally considered significant. It is possible that some of the close relationships observed within the gt1a tree have made cladal definition easier and therefore aided reconstruction of the sequence phylogeny.

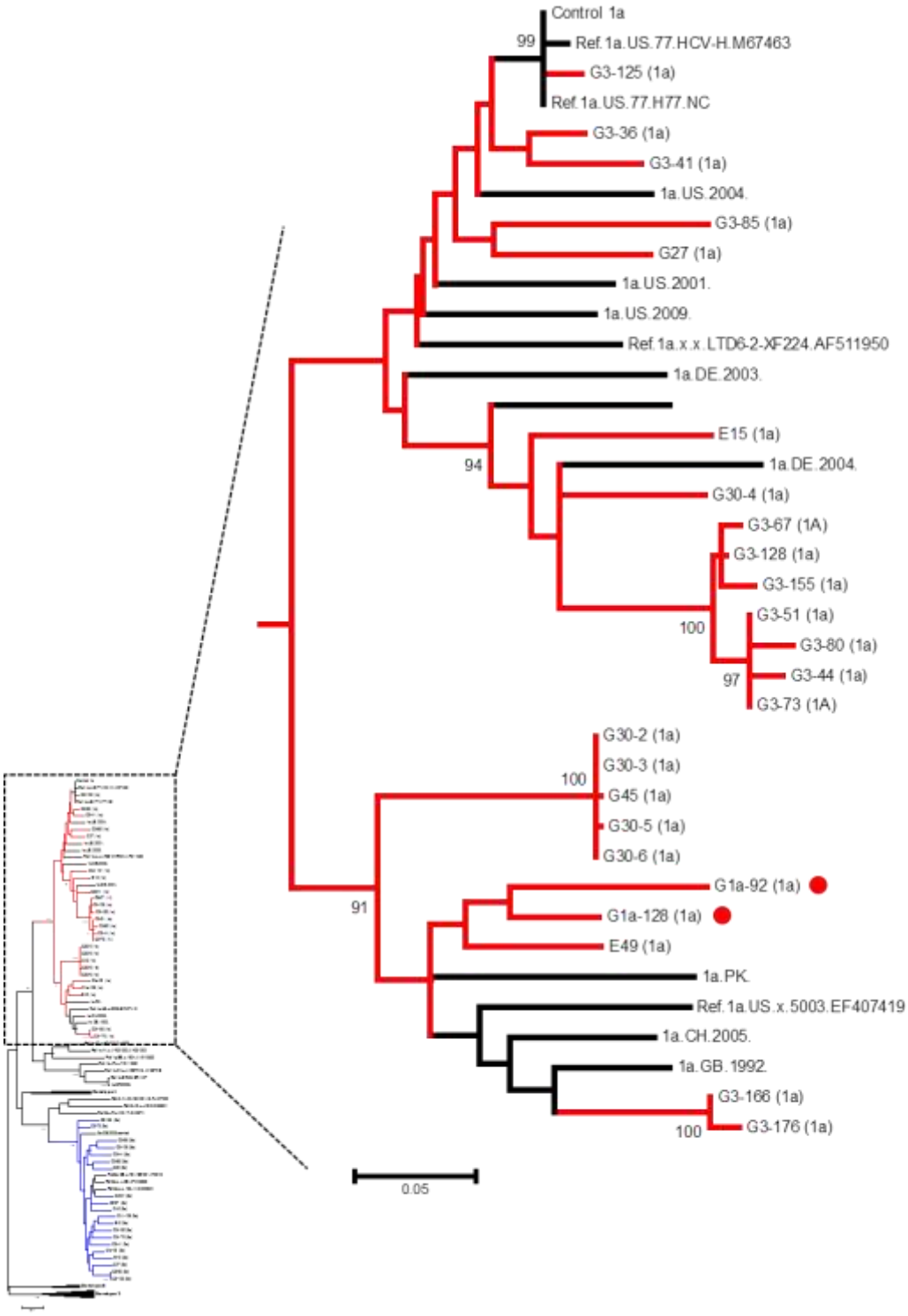


Figure 4-10; Gt1a strains isolated from samples with mixed gt1a/gt3 infections

Region of a maximum likelihood phylogenetic tree showing gt1a E1-E2 sequences (red) obtained from samples with mixed gt1a/gt3 HCV infections analysed alongside reference gt1a stains (black). Samples genotyped by the diagnostic laboratory as gt1a are indicated by ●. Paired gt3a sequences (blue) from the samples are shown in figure 4-11. Two distinct gt1a strains were isolated from sample G-30 and consequently, multiple sequences for the sample are analysed in the tree (see section 4.4.3). Bootstrap support of >70% after 1000 replicates is indicated on the tree.

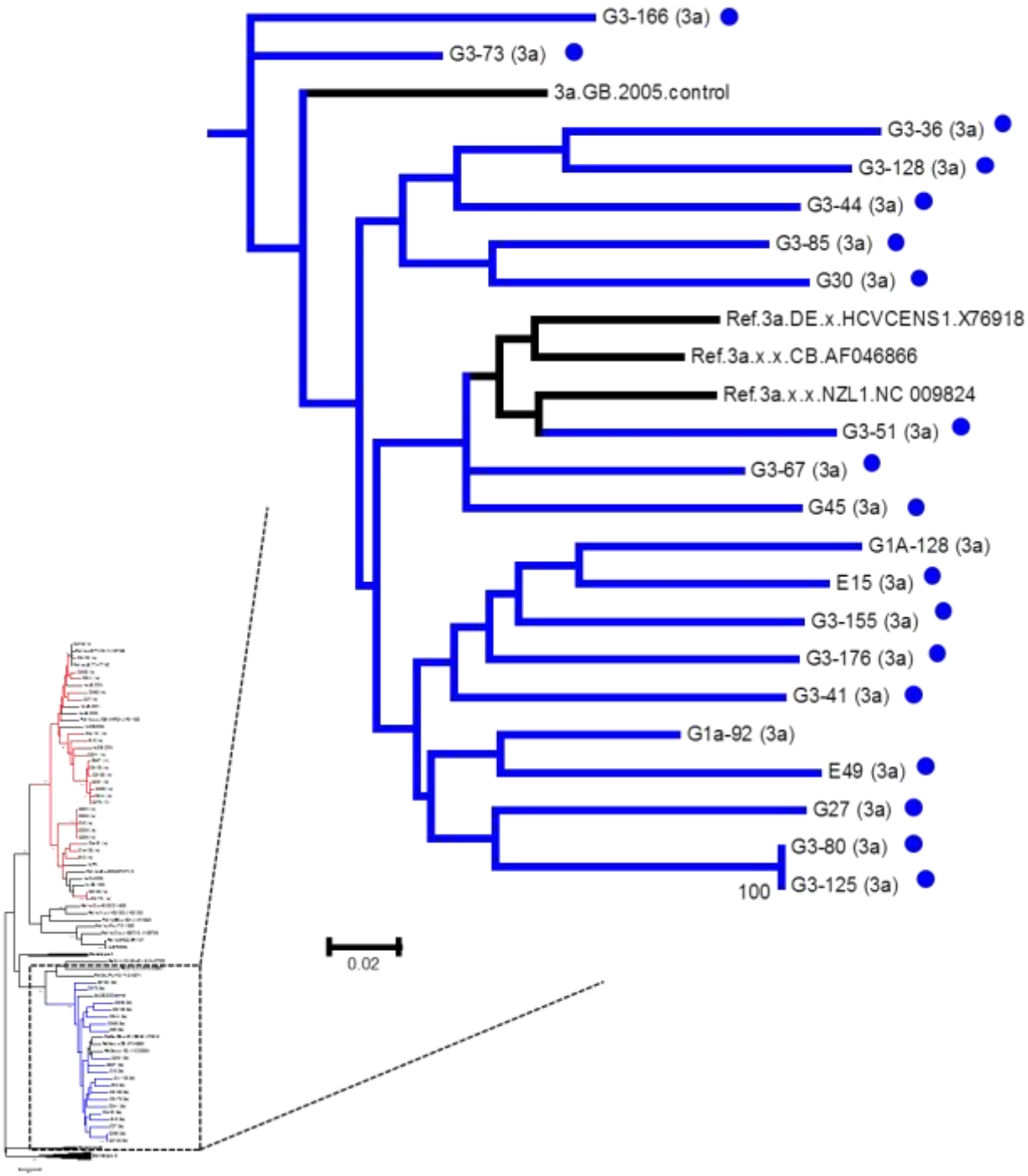


Figure 4-11; Gt3a strains isolated from samples with mixed gt1a/gt3 infections

Region of a maximum likelihood phylogenetic tree showing gt3 E1-E2 sequences (blue) obtained from samples with mixed gt1a/gt3 HCV infections analysed alongside reference gt3 stains (black). Samples genotyped by the diagnostic laboratory as gt3 are indicated by ●. Paired gt1a sequences (red) from the samples are shown in figure 4-10. Bootstrap support of >70% after 1000 replicates is indicated on the tree.

By careful analysis of the cohort data, two of the samples with mixed genotypes (G3-80 and G3-125) were identified as having originated from the same patient. Sample G3-80 was collected in October 2013, when the patient was undergoing treatment and subsequently a further sample (G3-125) was collected a month later. The genotype 3a strains sequenced from this individual were identical but differences were observed in the gt1a sequences isolated. Specifically, the gt1a strain detected in G3-125 was very closely related to the gt1a control strain (H77) and this may suggest the sample was contaminated with the gt1a E1-E2 control sequence at some point. The sequences had 15.7% sequence heterogeneity at the nucleotide level. These samples are discussed further in section 4.5.2.

4.4.3 Identification of multiple strains in a single sample

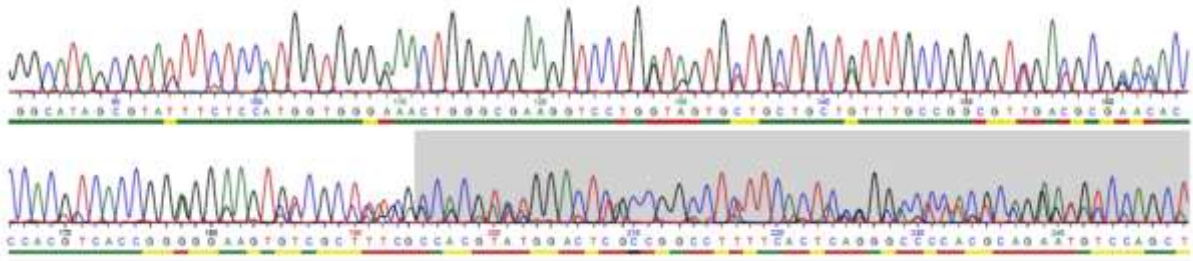


Figure 4-12; Sample G-30

Section of the chromatogram obtained from the Sanger sequencing of G-30 gt1a.

One sample, G30, had a mixed genotype infection involving multiple minor strains. The sample was diagnosed by the clinical laboratory as gt3 infection and tested positive for gt1a during the initial screening process. Despite repeated attempts to sequence the gt1a strain, the chromatograms showed mixed nucleotide peaks suggesting that multiple strains may be present (figure 4-12). Clonal analysis was performed on the sample and two distinct gt1a strains were identified in the sample in addition to the gt3a strain that was diagnosed within the clinical laboratory (figure 4-13). The two gt1a strains were determined to be distinct strains on the basis that they demonstrated >10% sequence heterogeneity at the nucleotide level, with 16% of the nucleotides being discrepant between the two strains, and they grouped into phylogenetically distinct clades with strong bootstrap support. Five sequences from the cloning of the gt1a strains were successfully sequenced with a major gt1a strain comprising 80% of the gt1a population (n=4) identified and a second strain comprising the remaining 20% of the gt1a viral population also isolated. In figure 4-9, the gt1a strains display a bi-furcating topology and the two gt1a

strains detected in sample G-30 are present in separate clades, further evidencing the differences between these two strains. This bi-furcating topology of the gt1a subtype has been well-described in previous studies (Pickett, Striker and Lefkowitz, 2011; Luca *et al.*, 2015) and the two clades have been linked with differing prevalences of NS3 resistance associated variants, particularly Q80K (Luca *et al.*, 2015).

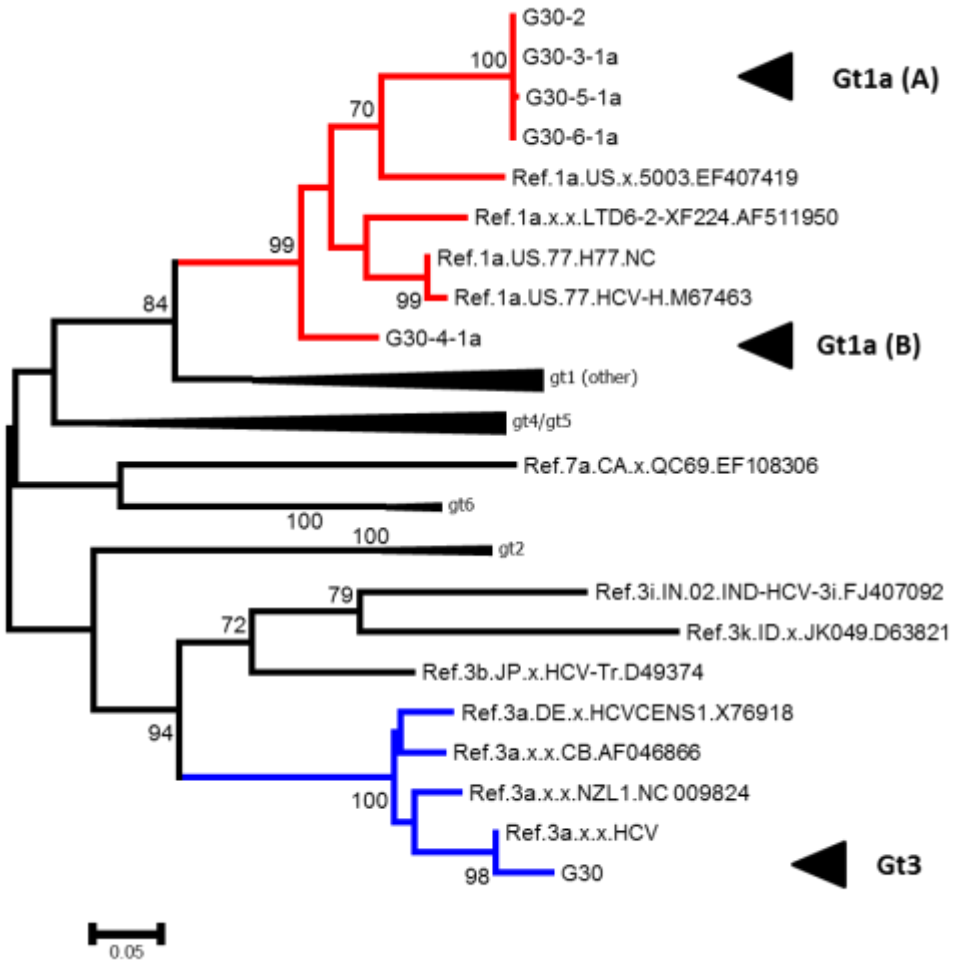


Figure 4-13; Gt3 and multiple gt1a strains detected within sample G-30

A maximum likelihood tree showing a complex mixed infection detected in sample G-30. Gt1a sequences are highlighted in red and gt3a sequences in blue. Bootstrap support of >70% after 1000 replicates is indicated on the tree.

4.5 qPCR

4.5.1 Optimisation of qPCR

Genotype-specific qPCR assays targeting the NS5B region were developed in order to quantify the relative proportions of gt1a and gt3a present in samples that were found to have mixed genotype infections. Primer and probe designs for gt1a- and gt3a-specific qPCR assays (primers D-I, table 2-8) were tested in triplicate with the H77 and the 3a.GB.2005 NS5B transcripts respectively. Consistent detection at levels of 10-100 copies/µl was obtained for both assays, suggesting they had excellent sensitivity (figure 4-14). Negative controls (dH_2O) were tested in triplicate alongside the samples and were not detected after 40 cycles of qPCR. The gt1a qPCR assay was tested in triplicate with 10^4, 10^5 and 10^6 copies/µl of the gt3a control transcript and a selection of gt3 samples and vice versa for the gt3 qPCR assay with the gt1a control and a selection of gt1a samples. No amplification was shown for either assay with the control transcripts or the samples tested indicating that there was no cross-reactivity between the assays. Adjustments to cycling temperature, primer concentration and probe concentration yielded no further increases in sensitivity.

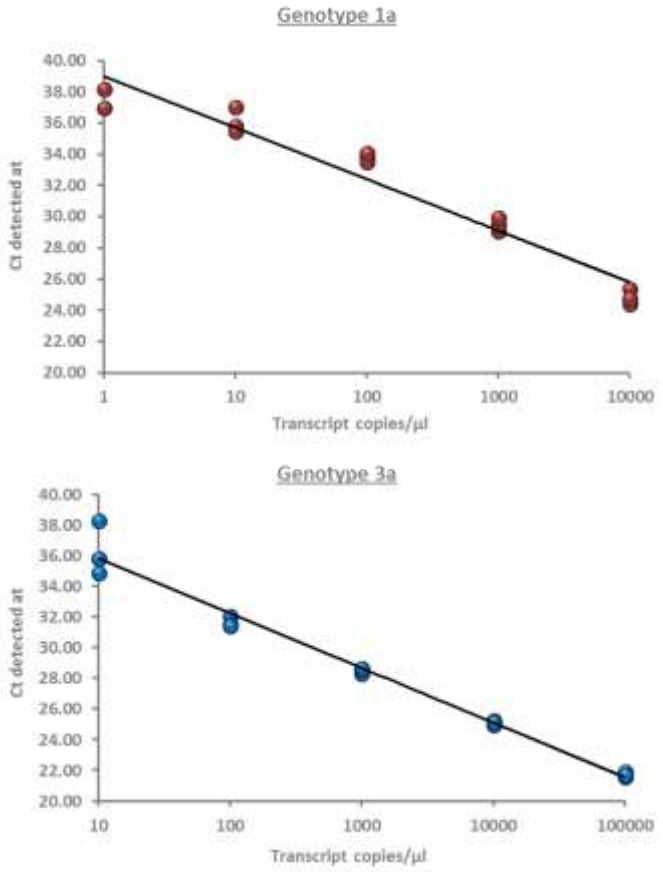


Figure 4-14; Sensitivity of the qPCR gt1a and gt3a qPCR assays

Assay sensitivity as demonstrated by qPCR of the gt1a (H77) and the gt3a (gt3a.GB.2005) control transcripts. Negative controls (H_2O) were run alongside the transcripts in triplicate and were not detected in all cases.

4.5.2 Quantification of mixed infection samples

Mixed genotype HCV positive samples were analysed with genotype-specific qPCR assays to determine the individual genotype viral loads (figure 4-15). Of the 20 samples tested, 15 samples were positive by qPCR for both genotypes. Failures were attributed to sample degradation (n=3) or primer incompatibility (n=2). The major genotype detected correlated with the clinical genotyping result of the patient in 14/15 cases. The exception was the sample G3-85 which was clinically listed as gt3 but qPCR analysis suggested that gt1a was the major viral genotype present within the sample. Diagnosis with gt3 using the in-house genotyping method at WSSVC was reliant on the absence of a result for a specific genotyping probe (see section 2.2.2 and appendix figure 8-3). This may suggest that the gt1a strain in G3-85 was not recognised by the gt1a-specific probe and, due to the lack of a specific result, the sample was misclassified as gt3.

In all gt1a-diagnosed individuals, the major genotype present was gt1a (n=2). Minor gt3a strains in these individuals were found to comprise 0.5 and 6.91% of the total viral load. In gt-3a diagnosed individuals, gt3a was the major genotype present in 92.3% of cases (12/13). Minor gt1a strains were found to comprise 0.01-21% of the total viral load. Taking into account the discrepancy with sample G3-85, where the clinically diagnosed genotype was the minor strain, gt3a was found to be the major strain present in 80% of cases (12/15). Whilst minor strains ranged from 0.02-20.63% of the total viral load, they were generally detected at low levels with two-thirds of samples having minor strains present at levels <1%.

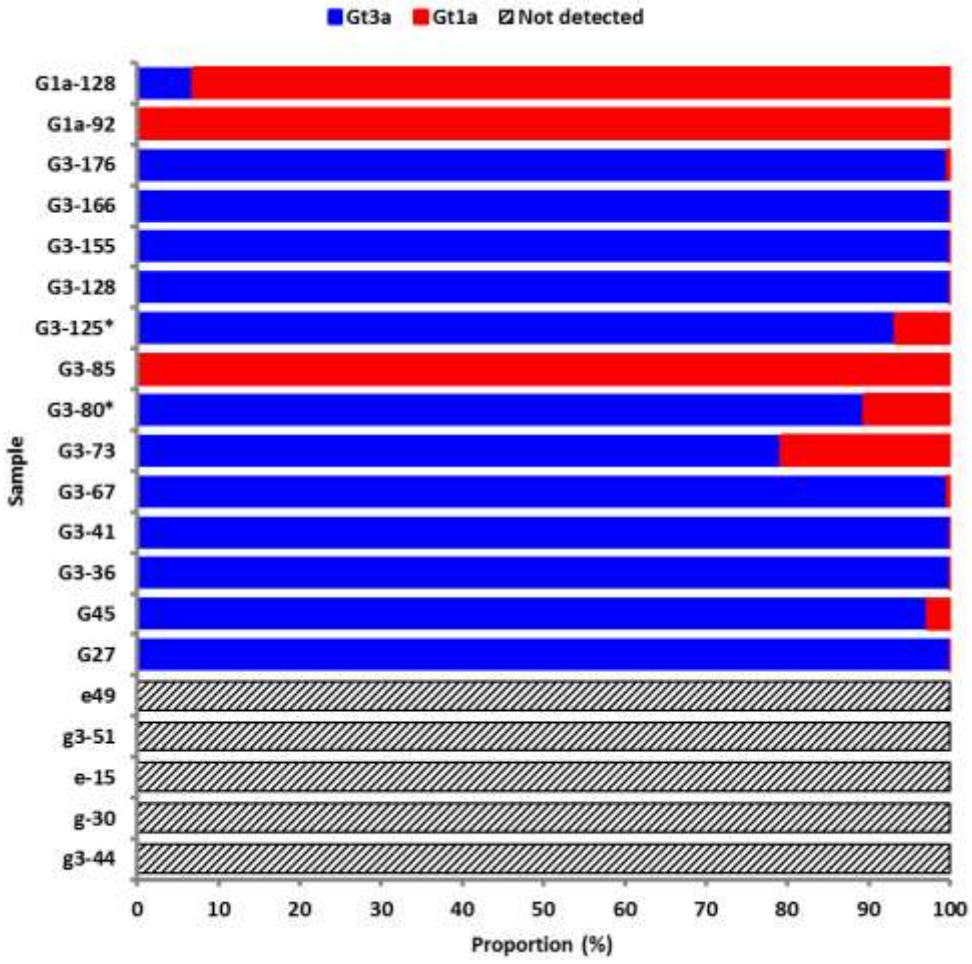


Figure 4-15; Relative proportions of gt1a and gt3a in samples with mixed gt1a/gt3 infections

The proportions of gt1a and gt3 detected in samples with mixed genotype infections. Samples G3-80 and G3-125 have been highlighted (*) as they were taken from the same patient, roughly a month apart. Samples E15, E49, G-30, G3-44 and G3-51 were not detected by the qPCR assays. See table 4-4 for further details.

The gt1a and gt3a viral loads detected by qPCR were added together to calculate the total viral load present in each sample analysed (table 4-4). Both the calculated and clinical laboratory viral load results were available for 10/20 samples with mixed genotype infections. For most of samples, the two results were comparable with 90% of samples (9/10) having the two results within 0.5 $\log_{10}$ of each other (table 4-4). This suggests that our qPCR assay has a similar sensitivity to the assays used within the clinical testing laboratory. Sample G-45 had a discordant result between labs with almost a 2 $\log_{10}$ lower calculated viral load than the clinical viral load, which was much higher. This sample was collected, processed and stored for around 18 months prior to this study and it is probable that repeated freeze-thaw cycles or suboptimal freezing conditions degraded some of the viral RNA in the sample, reducing the amount detectable by qPCR.

Table 4-4; Viral loads of samples with mixed genotype infections

Genotype-specific viral loads detected in samples with mixed genotype infections and calculated viral loads. Estimated viral loads were converted from copies/µl to copies/ml using a calculation given in the appendix adjusting for sample concentration and dilution during extraction and subsequent testing (appendix figure 8-2). Viral loads from the clinical laboratory have been given where available.

Sample	Calculated 3a VL	Calculated 1a VL	3a %	1a %	Major genotype identified	Calculated total VL (copies/µl)	Calculated VL (log_{10} copies/ml)	Clinical VL (log_{10} copies/ml)
G3-44	NK	NK	NK	NK	NK	NK	NK	5.78
G-30	NK	NK	NK	NK	NK	NK	NK	NK
E-15	NK	NK	NK	NK	NK	NK	NK	3.74
G3-51	NK	NK	NK	NK	NK	NK	NK	6.06
E-49	NK	NK	NK	NK	NK	NK	NK	5.89
G-27	5878	9.44	99.84	0.16	3a	5887	5.70	5.50
G45	393	10.65	97.36	2.64	3a	404	4.54	6.42
G3-36	27064.79	13.2	99.95	0.05	3a	27078	6.37	NK
G3-41	15479.58	19.75	99.87	0.13	3a	15499	6.12	NK
G3-67	34198.66	71.75	99.79	0.21	3a	34270	6.47	6.07
G3-73	1020.66	265.27	79.37	20.63	3a	1286	5.04	NK
G3-80	30.24	3.55	89.49	10.51	3a	34	3.46	2.93
G3-85	9.92	2781.49	0.36	99.64	3a	2791	5.38	NK
G3-125	627.31	44.37	93.39	6.61	1a	672	4.76	4.96
G3-128	20448.3	6.85	99.97	0.03	3a	20455	6.24	NK
G3-155	12557.81	2.75	99.98	0.02	3a	12561	6.03	5.99
G3-166	44811.86	13.25	99.97	0.03	3a	44825	6.58	NK
G3-176	763.48	1.67	99.78	0.22	3a	765	4.82	4.62
G1a-92	55.69	13363.43	0.42	99.58	1a	13419	6.06	6.37
G1a-128	71.79	967.27	6.91	93.09	1a	1039	4.95	5.21

The linked samples G3-80 and G3-125 (representing serial samples from a single patient) both gave satisfactory qPCR results. The first sample collected from this individual (sample G3-80) had a viral load of 2.93 $\log_{10}$ and was collected whilst the patient was undergoing treatment for HCV. One month later (sample G3-125) the viral load had increased to 4.96 $\log_{10}$. This viral increase of >2 logs is considered an unfavourable prognostic indicator. Closer examination of the data indicated that the viral loads of both genotypes had increased between sampling with the gt3a viral load increasing (x20.7) by more than the gt1a viral load (x12.5). The proportion of gt1a present in the sample increased from 6.6% to 10.6% during the testing interval. A viral load increase of >2 logs whilst on treatment is unlikely and it is possible that the patient was taken off treatment during the sampling interval without an SVR being achieved. The final outcome of this individual's treatment is unknown but the data suggests this patient did not respond favourably to treatment.

qPCR results for sample G-30, which was highlighted previously in section 4.3.3 as having two gt1a and one gt3 strains were inconclusive and neither genotype within the sample was amplified by the qPCR assays. Sample G-30 was collected in 2011 during the initial phase of sample collection and the extracted RNA was approximately two years old at the time of testing. Of the mixed infection positive samples collected in this batch, 3/5 (G-30, E-15 and E-49) yielded inconclusive PCR results. It is probable that the RNA extracted from the samples has degraded over time and it is unclear to what extent the presence of multiple gt1a strains has also contributed to the unsatisfactory qPCR result in the case of G-30.

Although the majority of samples with mixed genotype HCV infections have minor strains contributing <1% of the total viral load, there was wide variation in the relative contributions of the minor strains. All samples in which the minor strain contributed >5% of the total viral load (n=4) had viral loads $<10^4$ copies/ml (figure 4-16). The R^2 value obtained from the correlation between viral load and minor strain proportion was 0.303 suggesting that this is not a significant trend, perhaps because the numbers of samples that were analysed was low.

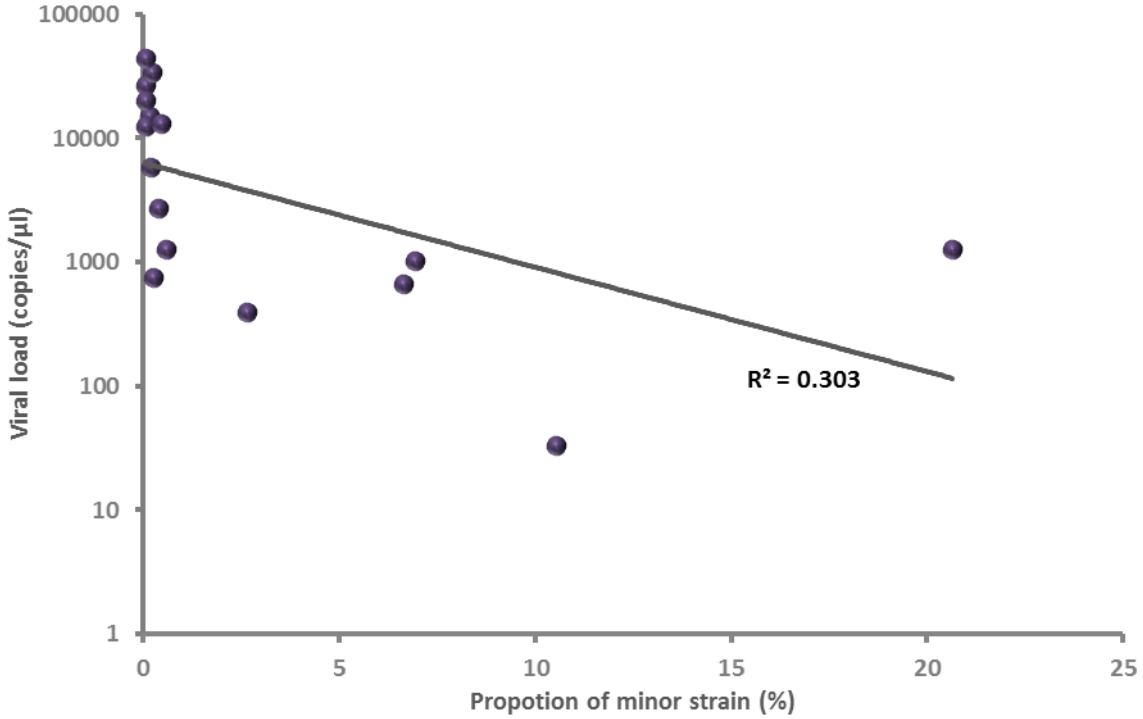


Figure 4-16; Viral load of the minor strains

HCV viral load of the minor strains plotted against the estimated minor strain proportion with a fitted trend line and R^2 value. Data taken from table 4-4.

4.6 Chapter conclusions

In this study, a genotype-specific nested PCR-based approach targeting the E1-E2 region was developed and used to screen HCV positive samples for the presence of mixed gt1a/gt3 infections. This approach was selected as it was felt that it offered both a high degree of sensitivity and specificity. After screening 506 samples, 20 samples were identified as having mixed gt1a/gt3a infections. By targeting the relatively conserved regions flanking the hypervariable regions, a highly specific assay can be developed and subsequently used to detect secondary infecting genotypes in a high background of the primary genotype. A review of the previous literature found that screening of the E1-E2 region had only been used for 5/27 of the previously published studies (Herring *et al.*, 2004; Li *et al.*, 2008; Du *et al.*, 2012; Grebely, Pham, *et al.*, 2012; McNaughton *et al.*, 2014) on mixed genotype infection identified, with the 5'UTR and core regions being favoured (table 1-2). Whilst these regions can be used effectively for genotyping, diversity is restricted within them and they are not always suitable for sub-typing HCV strains (González et al., 2013). Samples identified as positive for mixed genotype infections were also tested with genotype-specific qPCR assays targeting the NS5B region of the virus. The nested PCR that was developed and tested in this chapter appeared to be more sensitive than the qPCR assay which yielded inconclusive results for 5/20 tested samples. Viral load results were available from the clinical testing laboratory for 4/5 of the samples with inconclusive qPCR results and, of these only 1/4 samples (E-15) had a low viral load (table 4). 3/5 samples with inconclusive results were also 12-18 months older than other tested samples and it is probable that repeated freeze-thawing of the samples or suboptimal freezer conditions has led to partial degradation of the viral RNA within the samples, reducing its quality. The qPCR assay consisted of 40 cycles of amplification, whereas the nested PCR had a first round of 40 cycles and an additional second round of 25 amplification cycles. These additional 25 amplification cycles have likely increased the sensitivity of the nested PCR, enabling more consistent detection of very low viral load samples. Previous studies have suggested that the use of nested qPCR protocols, where an additional amplification step is incorporated prior to the quantitative amplification, can significantly improve sensitivity for the detection of low copy number pathogens (Takahashi and Nakayama, 2006; Tran et al., 2014). It may be worthwhile investigating if such techniques are able to increase qPCR sensitivity for mixed genotype HCV infection.

Within the infected population that was screened, mixed HCV infections were detected in individuals at a prevalence of 3.8%. A significant difference between the number of gt1a-diagnosed individuals with mixed infection (0.8%) and the number of gt3a-diagnosed individuals (6.7%) was observed ($p < 0.05$). This overall rate is similar to the low prevalence rates observed in some previous studies (Blatt et al., 2000; Yun et al., 2008; Butt et al., 2011). The stringent focus on gt1a and gt3a and broad sampling approach used in this study may have contributed to the low prevalence rates observed. Genotypes 1a and 3 are the most common genotypes within the UK and are estimated to be responsible for an estimated 90% of all HCV infections in the region (Public Health England, 2015). Given that mixed genotype infections involving a wide range of different genotypes have been documented (van de Laar et al., 2009; Grebely, Pham, et al., 2012), this suggests the true prevalence of mixed genotype infection may be up to 10% higher if all genotypes were analysed. Whilst higher rates of mixed infection in gt3a-diagnosed individuals were observed in the original pilot study (10.6%) (McNaughton et al., 2014), that study selectively sampled individuals with a known history of IDU. Individuals with a history of IDU are likely to have significantly higher exposure to multiple HCV challenges than other members of the population, therefore mixed genotype HCV infections are likely to be more common within this subset of individuals (Herring et al., 2004; Backmund et al., 2005; Pham et al., 2010). Recent studies have also indicated that there may be a higher prevalence of superinfection and re-infection occurring within populations of PWID than is currently estimated and that factors such as long testing intervals and rapid viral clearance are hindering their detection (Vickerman et al., 2012; Sacks-Davis et al., 2013, 2015). It has also been suggested that opiate use induces immune suppression and this may be additionally contributing to the higher rates of mixed genotype infections observed in populations of PWID (Peterson et al., 1987; Wang et al., 2002; Moore and Dusheiko, 2005). Two of the twenty samples identified as mixed gt1a/gt3 positive in this study were also found to have been sampled from the same patient. It is therefore possible that other individuals may have contributed multiple samples to our original screening cohort without us being aware of it. If this is the case, it would again suggest that the prevalence rate of mixed genotype infection detected in this cohort is an underestimation of the true prevalence rate within the population.

Previous studies based on data from England have estimated the average age at which injecting habits develop to be 21-22 years of age (De Angelis, Hickman and Yang, 2004; Mcnaughton et al., 2015). In Glasgow, the average injecting career length has been

estimated to be 8 years (Sutton et al., 2012). Given that the average age of subjects within our cohort is 42.6 ± 9.8 years, this suggests a large proportion of our cohort are no longer actively injecting drugs at the time of sampling and therefore likely to have chronic HCV infections. Many multiple genotype infections have been shown to be transitory in nature and it is not clear how common chronic infection is with multiple genotypes (Osburn et al., 2010; Grebely, Prins, et al., 2012). Diagnostically, it is difficult to differentiate between acute and chronic infection and it is therefore unclear if the mixed infections detected within this cohort are transitory or if both genotypes have established a chronic infection. The average age of individuals with mixed genotype infection detected was 41 ± 9.22 years, and like the overall cohort are unlikely to be actively injecting drugs implying that at least a proportion of them are likely to be infected with chronic multiple infections. Linked clinical data from the cohort indicated that 28.2% of individuals in our study had also been treated at some point, with around 80% of those either not completing or not finishing treatment. It is possible that in a number of cases, whilst an SVR was not achieved, the treatment has cleared one or more of multiple infecting genotypes in individuals with mixed infections, again reducing the prevalence observed within the study (McNaughton et al., 2014; Abdelrahman et al., 2015).

The low prevalence rate observed in this study may also be associated with the size of our study. An analysis of previous studies on mixed genotype infection prevalence found that studies with larger numbers of samples tended to find lower prevalence rates of mixed infection, with all studies sampling over 250 individuals reporting a prevalence of <10% (n=5) (figure 4-17). As our study was one of the larger studies on mixed infection rates and our sampling method was not selective for a particular patient group, such as groups with high exposure rates, the relatively low prevalence rates observed in our study are in keeping with previous research. A review of previous studies also indicated that the method used to detect mixed genotype infections could also be a limiting factor in the detection of mixed genotype HCV infections. In general, higher rates of mixed infection were reported in studies using PCR and lower rates in studies using LiPA, RFLP and HMA based methods (table 1-2). This indicates that the testing approach can have a considerable impact of the results of the study and that PCR-based methods may be superior for the detection of mixed genotype infections.

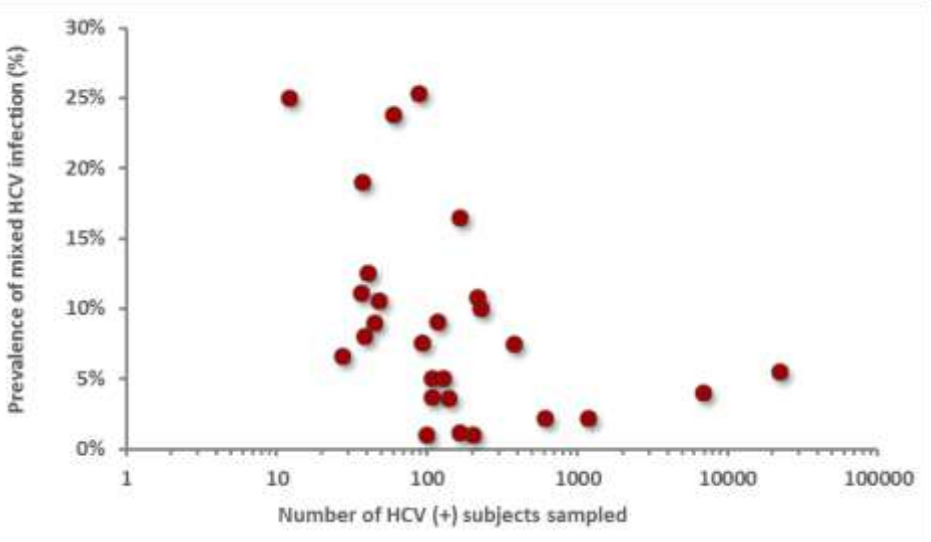


Figure 4-17; Mixed HCV infection prevalence studies

The number of patients sampled per study correlated with the prevalence rate of mixed infections detected by the study (data from table 1-2).

The results from this chapter suggest that whilst levels of mixed infection are low within the sampled population, they disproportionately affect individuals clinically diagnosed with gt3 infections. This disparity was identified in the pilot study for the project (McNaughton *et al.*, 2014) and has remained evident throughout the screening of the larger cohort. The reasons for this difference are likely complex and possible contributing factors may be viral-specific and host immunology. Work by Robinson *et al.* (2015), demonstrated that gt1a and gt3a viruses induce distinct transcriptome profiles during infection and this may be a contributing factor (Robinson *et al.*, 2015). Evidence in the literature also suggests that gt1 HCV infections can be harder to resolve with interferon-based treatments than other HCV genotypes (Zein, 2000; Harris *et al.*, 2007). It is possible that gt1a infections may be able to establish in the background of a gt3 infection but that gt3 viruses struggle to establish in the corresponding situation. Host factors may also be contributing to the differences observed. The favourable CC single nucleotide polymorphism at rs12979860 (rather than TT or C/T) in IL-28B has been shown to have strong associations with improved treatment responses and spontaneous clearance (Ge *et al.*, 2009; Thomas *et al.*, 2009; di Iulio *et al.*, 2011; de Castellarnau *et al.*, 2012). It would be interesting to investigate if the frequency of these favourable genotypes is lower in individuals with mixed HCV infections. HCV infection is frequently associated with poor cross-protective immunity (Vickerman *et al.*, 2012) and infection with one genotype rarely confers protection against infection with other genotypes. Other studies have however conversely indicated that the rate of infection among PWID who test anti-HCV positive is lower than

for those who are previously unexposed, suggesting there may be some form of partial immunity (Grebely *et al.*, 2006; Dore and Micallef, 2007). It may be the case that gt1 viruses are able to confer a broader cross-protective immunity than other genotypes although this subject is not widely studied. Without long-term follow-up and sequential sampling of individuals with mixed genotype HCV infections, it is difficult to know what may be responsible for this difference.

The use of synthetic RNA transcripts based on prototype HCV sequences has proven to be a useful tool in assay development (McLeish *et al.*, 2012), enabling a standardised method of assay characterisation throughout this project. Using probit analysis, the gt3a E1-E2 assay sensitivity was less (21 copies/µl) than the gt1a assay (9 copies/µl). Whilst this may be linked to true variation in assay sensitivity, the larger difference between the lower and upper dilutions tested in the gt3a assay may also have affected the result through the creation of greater statistical uncertainty during probit analysis leading to a lower calculated sensitivity. It should be noted also that a greater number of gt1a minor infections were identified in this study (section 4.4.1), possibly suggesting that the gt3 assay was less sensitive to very low viral load minor infections. However, no difference in assay specificity or sensitivity between the genotypes was demonstrated in the detection of the E1-E2 controls in mock mixed genotype samples, and the qPCR result for the gt3 minor strain in sample G3-85 indicated that the gt3 assays were capable of detecting minor gt3 strains with very low viral loads. Many genotyping assays in use in the clinic currently rely on the use of multiplex PCRs and often the result is interpreted from a panel of results. The clinical laboratory where these samples were collected from used an in-house qPCR-based approach with multiple probes to ascertain viral genotype at the time of sampling (appendix figure 8-3 for details). Whilst gt1 diagnosis was reliant upon a specific result, gt3 was diagnosed based on a positive viral load result coupled with the absence of a specific typing probe result (appendix figure 8-3). Multiplex qPCR methods such as these are incapable of resolving mixed genotype infections and the widespread use of these methods will therefore underestimate the prevalence of mixed genotype infections in general.

HCV viral loads in individuals with chronic infections are known to fluctuate over the course of infection (Halfon *et al.*, 1998). Previous work analysing sequential samples from individuals with mixed genotype infections has documented large fluctuations (>1 log) in the viral load of infecting genotypes in the subjects studied (Pham *et al.*, 2010; Grebely,

Pham, *et al.*, 2012). For the majority of our cohort, we only have a single sample from each subject, providing us with a snapshot view of a single time point in the course of the infection. For the single subject where we had two samples collected a month apart, large increases in the viral load of both genotypes were observed over the course of a month, despite the subject undergoing treatment at the time. It is probable that both strains within this sample have increased their replicative rate after the evolutionary bottleneck effect of developing resistance to treatment and the changes observed in this sample are therefore unlikely to be associated with natural fluctuation in viral load. Previous studies have suggested that the competition between the viral strains in mixed infections is rapid and that in most cases the strain with the highest viral load becomes dominant (Laskus *et al.*, 2001; Pham *et al.*, 2010; Ramirez *et al.*, 2010). This suggests that either the viral load of one virus is higher as it is able to replicate at a faster rate than the minority strain or that the minority strain is constrained by a stronger immune response (Freeman *et al.*, 2003; Alizon and van Baalen, 2008). Studies have also indicated that in the case of HCV superinfection, the primary strain generally remains persistent and that the super-infecting strain is transitory in nature (Pham *et al.*, 2010). Currently, without more patient history, there is no method available to differentiate between coinfection and superinfection and it is therefore unclear in which order the viral strains infecting our subjects were contracted. In the case of super-infecting HCV strains, evidence has also suggested that HCV is capable of establishing extra-hepatic sites of replication and it has been posited that minority strains may survive by being better adapted at replication within these extra-hepatic sites (Blackard, Kemmer and Sherman, 2006). It would be interesting to compare the population structure in sera, PBMCs and liver biopsy samples from individuals with mixed HCV infections to look for evidence of this.

The majority of minor strains identified in this study were found to contribute less than <1% of the total viral load present in individuals with mixed infections. Whilst this is considered relatively low, our understanding of the significance of these infections and the implications for the patients remains poor. In particular, the outcome of treatment, the progression of disease and monitoring the proportions of the genotypes present within these individuals over time requires further research. The collection of linked clinical data associated with the mixed genotype infections in this study was restricted by the large number of health boards from which the samples originated and a lack of ethical approval for samples outside of the Glasgow and Paisley postcode districts. Subsequently, there was not enough data to draw meaningful conclusions about the clinical impact of mixed

genotype HCV infections. Data was available on 4/20 samples though and analysis did suggest that liver disease was more common among these individuals relative to mono-infected individuals and that their responses to treatment were not effective. Anecdotal evidence from one individual undergoing treatment (samples G3-80 and G3-125) also supported concerns that treatment of mixed genotype infections may be difficult. A recent study has indicated that mixed genotype HCV infections may be difficult to treat with PEG-IFN-α and ribavirin, particularly in individuals also co-infected with HIV (Bagaglio *et al.*, 2015).There have been suggestions in the literature that having mixed HCV infections can lead to increased liver damage relative to individuals infected with a single genotype (Kao *et al.*, 1994; Widell *et al.*, 1995). Numerous DAAs targeting HCV have been noted to have genotype-specific efficacies; Telaprevir and Boceprevir in particular have reduced efficacy against gt3 infections (Garber, 2011; EASL, 2015). Additionally whilst PEG-IFN-α and ribavirin are not genotype-specific, treatment response rates are significantly better for gt2 and gt3 patients than for other genotypes (Zeuzem *et al.*, 2009) . It is possible that in mixed infections, undetected minority variant naturally drug-resistant genotypes will expand to replace the successfully treated majority variant genotypes, presenting in many cases as non- or partial responses to treatment (McNaughton *et al.*, 2014). In addition to following up the mixed infection cohort, it may be of interest to re-genotype a number of gt3 patients who have failed treatment with either standard of care or triple therapy to examine the frequency of genotype switching within this patient group. If genotype switching is observed, it may suggest the presence of previously undiagnosed mixed HCV infections. Increased efforts should be made to identify patients with mixed genotype infections prior to the initiation of treatment to enable assessment of the efficacy of the new DAAs on mixed infections.

Chapter 5: Analysis of mixed infection cohort by deep sequencing

5.1 Background introduction

New sequencing technologies have made the rapid sequencing of large numbers of sequences affordable, offering the potential to enhance current clinical diagnostic practises across numerous fields. Potential advantages in virology include the ability to rapidly sequence entire viral genomes, even when the pathogen is not known, and the ability to sequence to a great depth which may have applications in the detection of low-level resistance variants and mixed genotype infections. The majority of HCV genotyping assays used currently in the clinic only detect the major genotype present and minor sub-groups including co-infecting genotypes and pre-existing DAA-resistant strains are frequently not identified. Data from previous studies looking at drug-resistant minority variants (Fonseca-Coronado *et al.*, 2012; Abdelrahman *et al.*, 2015) and data from chapter 4, looking at mixed genotype infections, indicated that many co-infecting and minority variant strains can be present in very low amounts (frequently <1% of the total viral load) and therefore sequencing to a great depth may be advantageous when trying to detect them. In addition, many commercially available assays for HCV only facilitate genotypic identification and supplementary sequencing assays may be required to differentiate viral sub-types. The use of deep sequencing technologies within a diagnostic context offers the opportunity to exploit sequence-based genotyping which, as well as being cheaper and faster than currently available alternatives, are also far more discriminatory for the sub-typing of HCV strains. Many clinical decisions are currently made on genotypic-based information and increased information on viral sub-type offers the opportunity to further stratify patient populations, which given the large numbers of new DAAs available, will provide more

accurate information on the treatment specificity of many therapies, as well as providing an improved prognostic and epidemiological understanding of HCV.

In this chapter the objective was to evaluate the potential application of deep sequencing for the detection of mixed genotype HCV infections by developing a HCV genotyping assay compatible with the Illumina MiSeq deep sequencing platform. In order to accomplish this, an amplicon-based approach was selected as it was thought likely to be more sensitive than enrichment or metagenomics-based methods, especially given that two-thirds of the mixed infections identified had minor strains that contributed <1% of the total viral load (section 4.5.2). Consequently, two pan-genotypic PCR-based assays were developed targeting the E1-E2 and NS5B regions of the HCV genome. The E1-E2 region was selected in order to enable a direct comparison with the sequences isolated using the nested PCR assay developed in chapter four. Considerable intra-host diversity is also observed within the E1-E2 region (section 1.6), making it a suitable region for diversity comparisons. The NS5B region, and particularly the Okamoto region which is targeted by our assay, was selected as phylogenetic analysis of the region has been shown to reliably produce topologies comparable to those of the whole genome and polyprotein (Hraber *et al.*, 2006). This makes the region highly discriminatory for the sub-typing of HCV strains and therefore the most appropriate region for accurate typing analyses. These assays were used to amplify and sequence samples identified as having mixed genotype infections in the initial mixed genotype infection screen. Controls for fidelity, contamination and multiple genotype detection as well as gt1a and gt3 samples with single genotype infections were assessed concurrently with the samples. Results from the deep sequencing analysis were compared with the nested PCR and qPCR methods developed in chapter four and the diagnostic specificity and sensitivity of the assays evaluated. Additionally, an analysis of the inter-genotypic diversity present within the samples was also undertaken.

5.1.1 Assessment of assay sensitivity

Two pan-genotypic assays were developed targeting the E1-E2 and the NS5B regions of the virus. The E1-E2 amplicon covered 1296-1812 (516 bp) and the NS5B amplicon 8257-8636 (379 bp) (table 2-5). The sensitivity of the primers was calculated by performing

multiple PCR replicates of the control transcripts at suitable dilutions and implementing probit analysis. Results are given in table 5-1 and figure 5-1. Despite the high degeneracy of the pan-genotypic primers, initial tests on both the control transcripts and cDNA extracted from the serum of HCV positive patients demonstrated they were highly specific and only amplified the regions of interest. TapeStation analysis, which measures the length of DNA applicons, demonstrated that the generated amplicons were of a consistent size, 515-530bp for the E1-E2 amplicon and approximately 360-385bp for the NS5B amplicon. The ranges in sizes were due to technical variation inherent on individual TapeStation runs and the presence of insertions/deletions in the nucleic acid sequences of the individual virus strains that were analysed.

Table 5-1; Probit analysis of pan-genotypic assay sensitivities

The concentration of transcripts, detection rates and probit values derived from the results used to calculate the assay sensitivity for the pan-genotypic E1-E2 and NS5B PCR assays. Both the E1-E2 and NS5B assays were assessed using the gt1a and gt3a control transcripts produced in chapter 4 (section 4.2).

Genotype	Assay	copies/µl	Log_{10} copies/ µl	Positive* (%)	Probit value	Assay sensitivity (copies/µl)
1a	E1-E2	1	0	83	5.95	2.2
		0.1	-1	50	5.00	
	NS5B	25	1.40	83	5.95	22.1
		7	0.85	50	5.00	
3a	E1-E2	15	1.18	67	5.44	38.0
		10	1	33	4.56	
	NS5B	10	1	67	5.44	14.1
		7	0.85	33	4.56	

*Positive indicates the percentage of PCR replicates that were positive

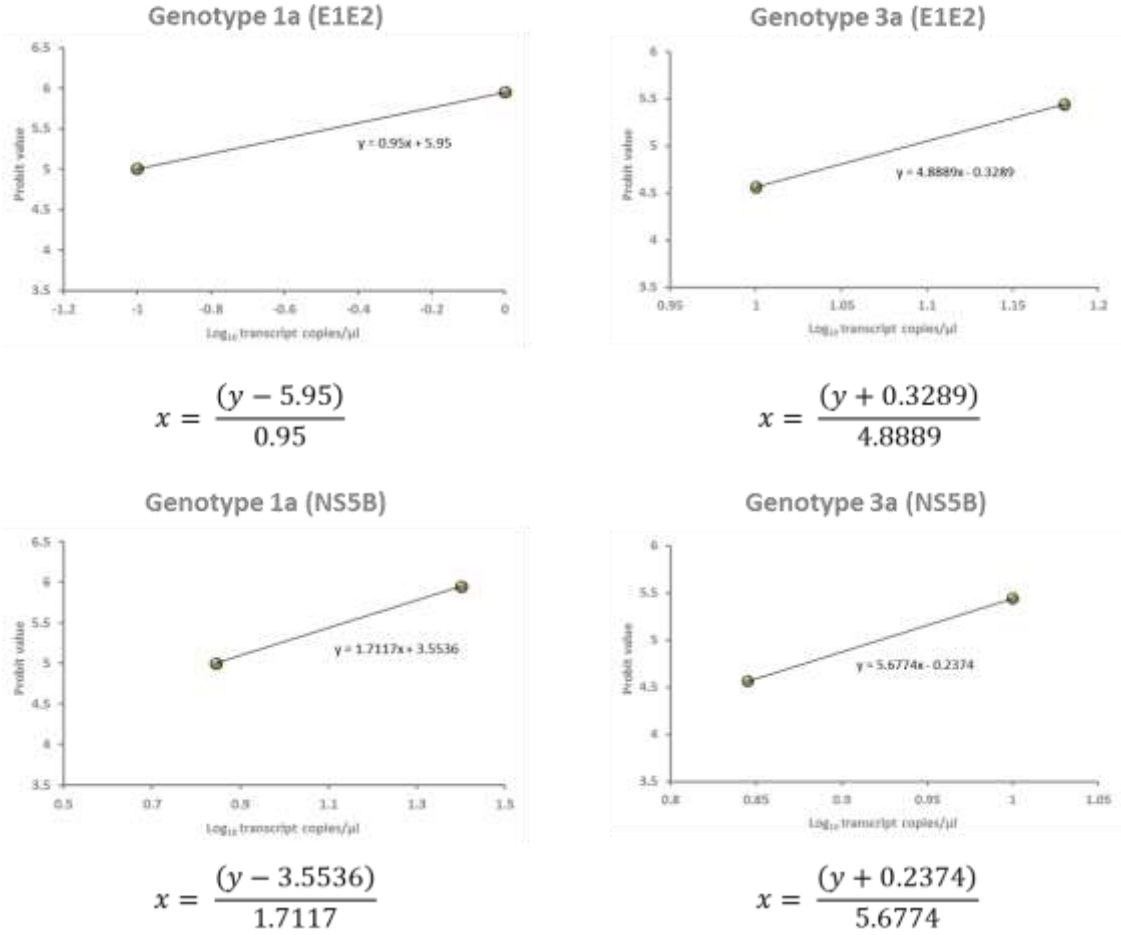


Figure 5-1; Equations used for the probit analysis

Graphs and corresponding equations used for the calculation of assay sensitivity in table 5-1. Equations for x were solved for y = 6.28, equivalent to a 90% detection rate.

Calculated using the equations generated from the probit analysis, the sensitivity of the E1-E2 assay was determined to be 2.2 copies/µl for gt1a samples and 38.0 copies/µl for gt3a samples. The sensitivity of the NS5B assay was calculated to be 22.1 copies/µl for gt1a and 14.1 copies/µl for gt3 samples. This suggested that assay sensitivity may be affected by inter-genotypic variation.

Trial runs performing library preparations on the E1-E2 and NS5B amplicons in an identical manner to the planned protocol were performed and demonstrated that the methodology worked in principle (see figure 5-2). TapeStation analysis on both the E1-E2 and the NS5B amplicons demonstrated that after library preparation the amplicons increased in size by approximately 120bp. This increase in size is consistent with the expected increase in size after successful ligation of adapters (8bp) and indexing tags (~50bp) onto the 5' and 3' ends of the amplicons.

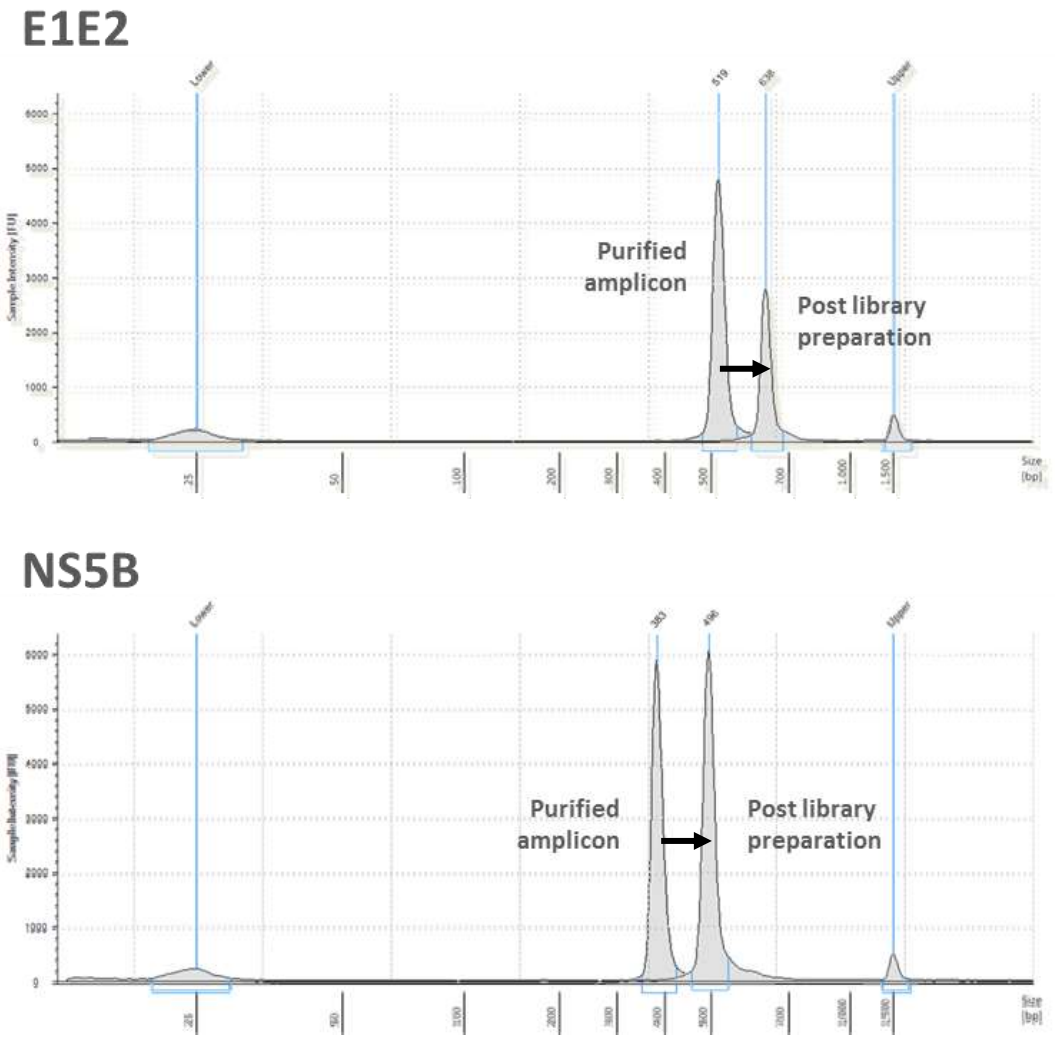


Figure 5-2; Trial library preparation of amplicons

TapeStation analysis of E1-E2 and NS5B amplicons prior to and after undergoing library preparation for deep sequencing. The amplicons were generated from randomly selected gt1a samples that were negative for mixed genotype infection during the screening process. Upper (1,500bp) and lower (25bp) peaks are also indicated.

5.1.2 Assessment of pan-genotypic specificity

The pan-genotypic primers were tested against a panel of different HCV genotypes comprising of gt1a, gt1b, gt2, gt3, gt4 and gt6 samples that was curated from an in-house sample collection. The range of genotypes included in the panel was considered representative of the genotypic diversity present in the UK currently (Public Health England, 2015). Samples underwent two rounds of PCR (40 cycles plus 25 cycles) using the same primer pairs for both rounds of amplification. Positive results were sent away for Sanger sequencing. Returned sequences were genotyped using phylogenetic analysis with a selection of HCV reference sequences available from the Los Alamos HCV Database (Kuiken *et al.*, 2004). Results from the testing are shown in table 5-2 and figure 5-3.

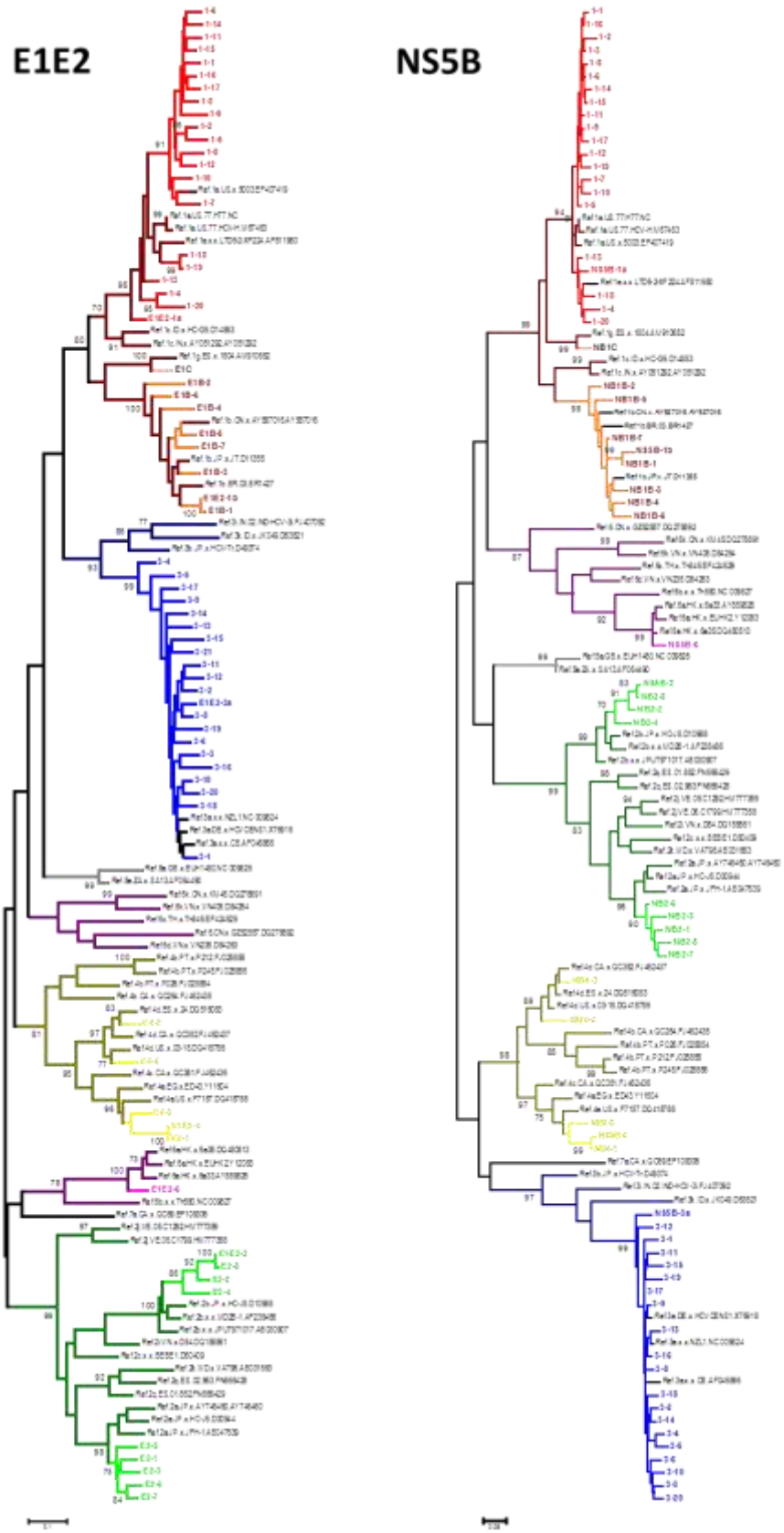


Figure 5-3; Genotyping by phylogenetic clustering

Mid-point rooted maximum likelihood trees of 64 E1-E2 and 64 NS5B sequences (coloured text, red – gt1a, orange – gt1b, green – gt2, blue – gt3, yellow – gt4, black – gt5, purple – gt6 and grey – gt7) amplified using the respective pan-genotypic primers and obtained using Sanger sequencing.. Returned sequences were analysed alongside HCV reference strains (black text) that were used to define the genotype of the isolates. Bootstrap support of ≥70% after 1000 replicates is indicated on each tree.

Table 5-2; Results of the genotyping panel

Results of the genotyping panel used to assess the breadth of the pan-genotypic E1-E2 and NS5B primers. The reported genotype from the diagnostic laboratory is given along with the typing results for the E1-E2 and NS5B amplicons generated with the primer pairs. The number of each sub-type tested is also indicated.

HCV genotype	Reported genotype	Number tested	E1-E2 typing results	Number amplified	NS5B typing results	Number amplified
1	1a	20	1a	20	1a	20
	1b	8	1b	8	1b	8
	1c	1	1g	1	1g	1
2	2	9	2a	5	2a	5
			2b	4	2b	4
3	3	20	3a	20	3a	20
4	4	5	4a	3	4a	3
			4d	2	4d	2
6	6	1	6a	1	6a	1

Both the E1-E2 and the NS5B primer sets were able to amplify 64/64 of the samples included in the test panel, giving products of the expected size in each case. The products were sent away for Sanger sequencing and sequences were returned in 64/64 samples for both the E1-E2 and NS5B assay. After analysing the sequences in maximum likelihood phylogenetic trees alongside HCV genotype reference sequences, 63/64 sequences for both the E1-E2 and the NS5B sequences gave identical genotyping results to the clinical genotyping reports. One sample was reported as a gt1c sample but phylogenetic analysis of both the E1-E2 and NS5B regions indicated it is was actually a gt1g strain (figures 5-3 and 5-4). A useful feature of the method developed here is that sequencing of these regions enables sub-typing of the HCV strains in the panel. This sub-typing result provides more information than current clinical testing protocols which only involve sub-typing of the gt1 strains. Whilst the tree topology for the E1-E2 and NS5B trees was not identical, 64/64 samples were successfully sub-typed and the results for the E1-E2 and NS5B fragments were identical for every sample. There was a discrepancy between the two trees in resolving the phylogenetic relationships relating to the gt6 E1-E2 sequences. In the E1-E2

tree, the gt6 sequences bifurcated into two distinct clades whereas, in the NS5B phylogenetic tree, they form a single group. There is known to be significant diversity within the HCV gt6 viruses and there are currently 24 subtypes assigned within the genotype (Smith *et al.*, 2014). Additionally, the E1-E2 fragment contains two hypervariable regions within it and is the most variable region of the viral genome. It is likely that the inherent high level of diversity within this region has complicated the resolving of the tree topology. Previous studies have indicated that the relationships represented by the NS5B amplicon, which contains the Okamoto region, are more likely to mirror the true viral phylogeny (Hraber *et al.*, 2006).

The single sample not matching with the original typing results was recorded as a gt1c sample by the diagnostic laboratory. Our data, from the sequencing of both the E1-E2 and the NS5B regions, suggests that it was in fact a gt1g sample and had been mistyped (figure 5-4). Both gt1c and gt1g samples are rare in the UK, and diagnosis of this genotype is likely to have required sequence analysis. The mistyping detected here may suggest that the region currently being sequenced in cases such as this by the diagnostic laboratory is not optimal.

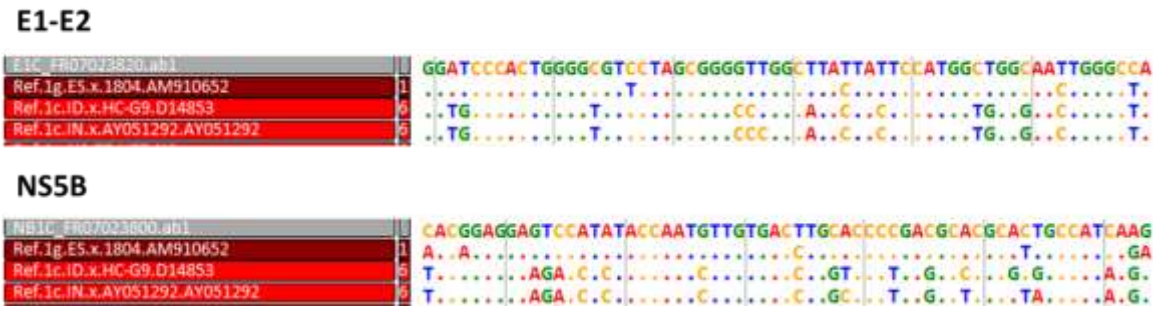


Figure 5-4; Sequence analysis of the 'gt1c' strain

Excerpts from an E1-E2 and an NS5B alignment of "gt1c" strain (grey) alongside two gt1c reference sequences (red) and one gt1g reference sequence (dark red) indicating re-classification as a gt1g sequence.

5.2 Deep sequencing

In the following sections, two runs performed on the Illumina MiSeq platform are analysed. All samples in the runs underwent identical sample preparation prior to sequencing. The samples were divided between the two runs on the basis of the presumed major genotype present to reduce the likelihood of cross-contamination. The gt3a run (run 1) consisted of 17 samples with mixed genotype infection, 20 samples identified as gt3a mono-infected, fidelity controls and mock mixed infection controls. The gt1a run (run 2) consisted of 3 samples with mixed genotype infection, 20 samples identified as gt1a mono-infected, fidelity controls and mock mixed infection controls. Both runs included a serum negative control from an individual not infected with HCV. The total number of reads generated per sample and the number of reads mapping to specific HCV genotypes in each sample are listed in tables 5-3 and 5-4.

5.2.1 Evaluation of methodology

Controls for assay fidelity, multiple genotype detection and contamination were prepared and run alongside the samples for both of the deep sequencing runs. The controls were prepared using the E1-E2 and NS5B transcripts described in chapter 4. As the control transcripts were prepared from a synthetic dsDNA fragment, any deviation from the original sequence can be presumed to be error introduced during the library preparation, the sequencing run or an issue related to the reading of the sequences.

Table 5-3; Deep sequencing reads from the control samples

Results of control samples loaded on the deep sequencing runs (raw data in tables 5-4 and 5-5). Y indicates the genotype was detected and N indicates no detection. * indicates the percentage of reads identified as either gt1a or gt3

RUN 1 (gt3)	gt1a detected	%*	gt3a detected	%*
Gt3a control	N		Y	100
Gt3a control	N		Y	100
Gt3a control	N		Y	100
99.9% gt3a:0.1% gt1a	Y	10.27	Y	89.73
99% gt3a:1% gt1a	Y	15.72	Y	83.5
90% gt3a:10% gt1a	Y	68.32	Y	31.67
H_2O neg	N		N	
Serum neg	N		N	
RUN 2 (gt1a)				
Gt1a control	Y	99.99	N	
Gt1a control	Y	99.99	N	
Gt1a control	Y	99.99	N	
99.9% gt1a:0.1% gt3a	Y	92.26	Y	7.37
99% gt1a:1% gt3a	Y	92.97	Y	7.69
90% gt1a:10% gt3a	Y	40.87	Y	59.05
H_2O neg	N		N	
Serum neg	N		N	

Results from the sequencing of the controls suggested the assay was functioning as expected and capable of detecting multiple genotypes when they were present. On the gt3 run, all three gt3 transcript controls were sequenced and all of the sequences were identified by phylogeny as gt3a. For the control spike-ins, where the minor strain was spiked in at concentrations of 0.1%, 1% and 10% of the total viral load, both gt1a and gt3 were detected in every case. Whilst lower concentrations of minor strain spike-in resulted in a lower proportion of reads mapping to the minor strain, the proportions of reads were not consistent with the calculated amount of minor strain spiked into the samples. This suggests that there is some PCR-bias occurring in which sequences from the minor strain is being over-amplified relative to the major strain. On the gt1a run, all 3 gt1a transcript controls were successfully sequenced and genotyping identified 99.99% of the sequences as being gt1a. Again for the control mixed infections, gt1a and gt3 were detected in all 3 samples tested. Both the 0.1% and the 1% controls produced similar proportions of reads (7.37% and 7.69% respectively) for the minor strain spiked into the samples. Similar to the gt3a run, the number of reads mapping to gt1a and gt3 was not proportional to the actual amount of minor strain spiked into the sample. Whilst this data demonstrates that the assays are capable of detecting multiple genotypes in the same samples, they also indicate that the assay is susceptible to PCR bias and therefore unsuitable for the quantitative estimation of the proportions of each genotype present in the samples. For both runs the serum negative controls had no reads mapping to HCV suggesting that contamination between samples was not an issue during these sequencing runs. Quantification of the DNA by Qubit for the H_2O negative control, which was prepared alongside the samples and controls, was below the detectable level, further suggesting that contamination between samples was not an issue during the library preparations.

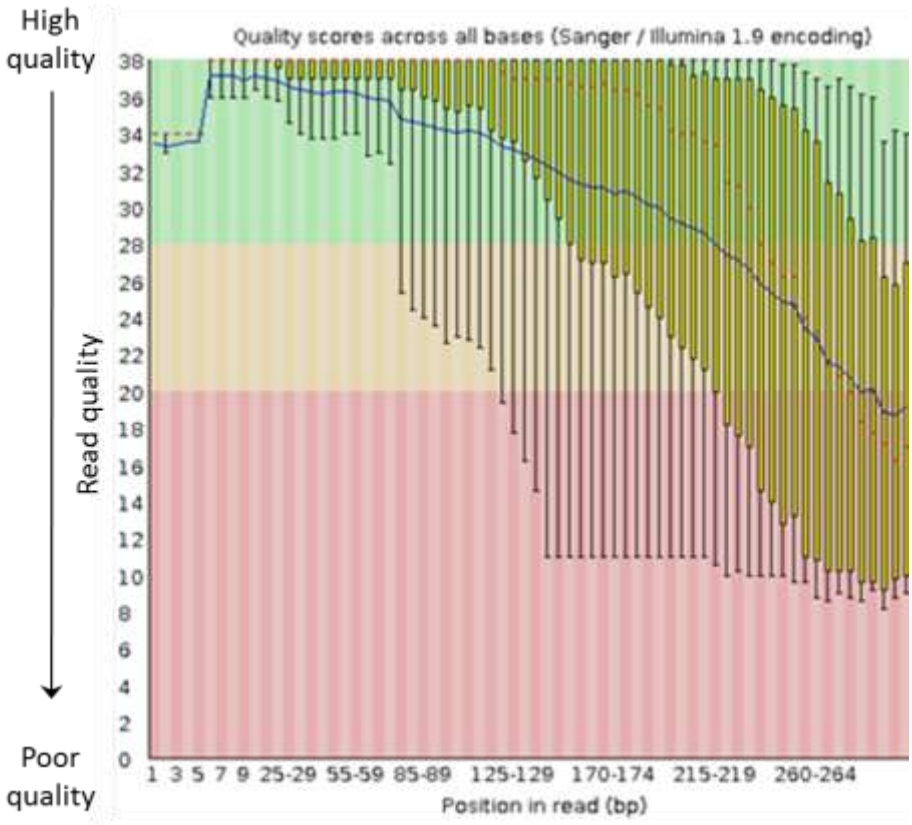


Figure 5-5; Decline in read quality along the amplicon length

FastQC plot showing read quality along the amplicon from read 1 of the gt3a major control with 0.1% gt1a . The graph illustrates the typical decline in quality score towards the end of the Illumina read that was consistently observed for samples from both runs.

Consensus sequences from the reads generated for the three control sequences in each run revealed consistent results. The gt1a E1-E2 control fragments all contained the G→A nucleotide substitution at position 1777, introduced into the transcript during production. For 1/3 controls, the sequence was 100% identical to the expected sequence. In the other two transcripts, six or seven nucleotide substitutions over positions 1552-1563 were observed. The gt3 E1-E2 controls were 100% homologous to the original control sequence, apart from in the region 1491-1570 where there were a number of errors introduced into the gt3 E1-E2 sequence and a 30bp region of the read that was either not sequenced or was unmapped. As paired-end sequencing was used, the middle of the sequence equates to the end of the individual reads of the amplicons. Read quality often deteriorates towards the end of the read, and the read quality observed throughout our data towards the end of reads was very poor (see figure 5-5). Reads were trimmed to mitigate for this but the quality of the data, particularly at the end of reads, was poor in this study and this must be considered during the analysis. It is unclear however why this issue has not affected the gt1a E1-E2 control sequences to the same degree and this may indicate that the issue is with the

bioinformatic mapping of the reads. On the whole sequencing of the NS5B region was more consistent. The NS5B amplicon was shorter (382bp) than the E1-E2 amplicon (520bp), therefore there was a greater overlap between the forward and reverse reads in the middle of the NS5B amplicon. It is probable that this has contributed to the improved consistency of the reads in this region. The gt1a NS5B controls all had a G→A nucleotide substitution at position 8605, a mutation introduced into the transcript during its production and consequently, all the sequences had 100% sequence homology with the control transcript. The gt3 controls were also found to be identical to the control transcripts apart from a G→A nucleotide substitution at position 8512 which was observed in all three gt3 controls sequenced. This substitution was not previously detected in the control transcript.

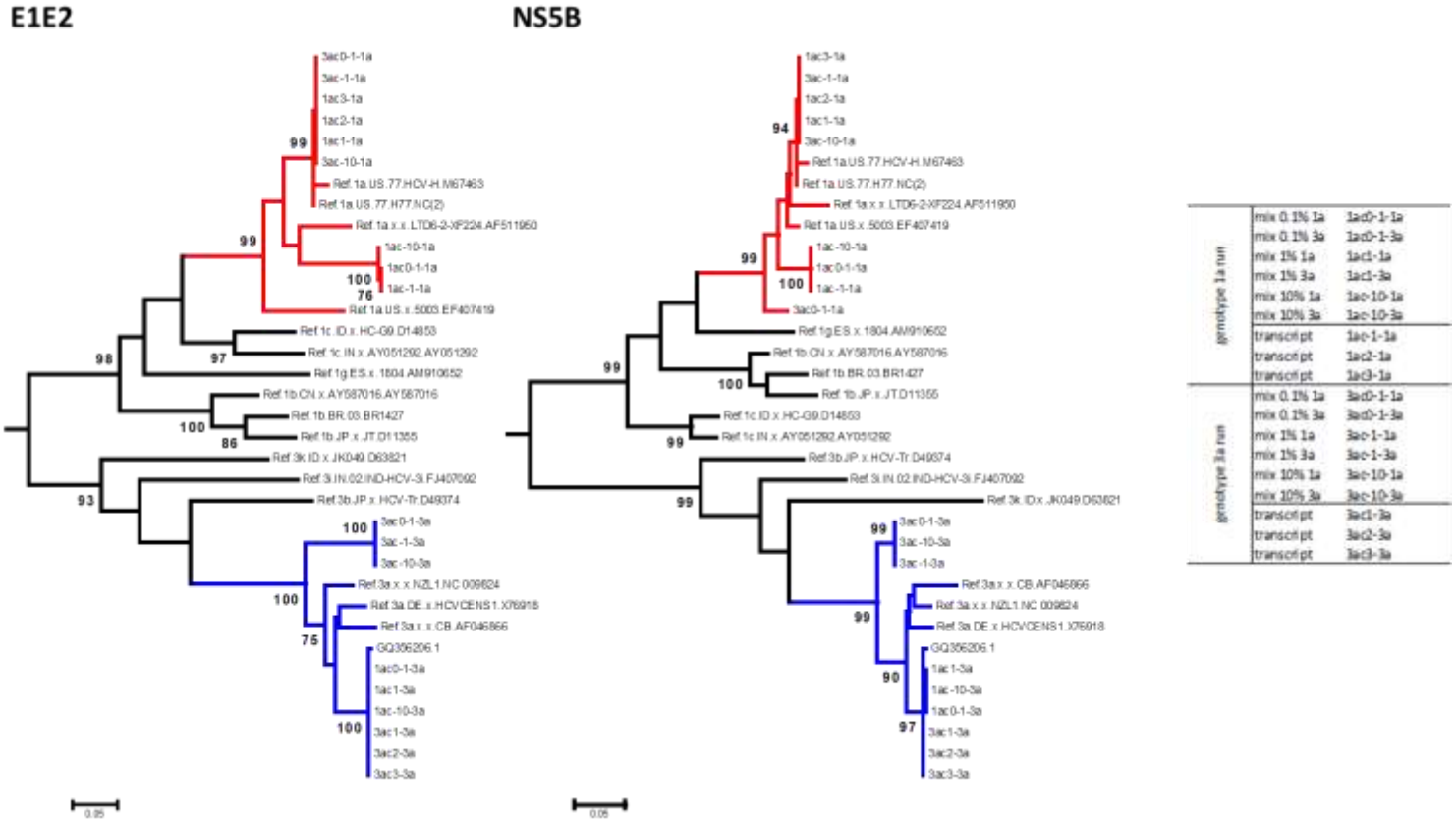


Figure 5-6; Phylogenetic analysis of the control consensus sequences

Maximum likelihood phylogenetic trees comparing control consensus sequences from the deep sequencing runs with the original transcript sequences. There is a key identifying the controls in the tree. Gt11a strains are highlighted in red and gt3a strains in blue. Bootstrap support of >70% after 1000 replicates is indicated on each tree.

Multiple sequencing of the control transcripts was used as a marker for the fidelity of the process. As the transcripts were all derived from the same sequence, any deviation from the sequence must stem from either the introduction of errors during cDNA synthesis, subsequent amplification or from an error resolving the images during sequencing. The levels of fidelity observed were not consistent across the length of the transcripts, with a number of errors being identified in the middle of the sequences. In addition to the known decline in read quality towards the end of the read, Illumina have also acknowledged issues with the chemistry of the v3 300 bp paired-end kits. The issues are known to be causing a significant drop in the quality of the reads much earlier than expected, making longer fragments over 500 bp difficult to sequence accurately. It is likely that this issue is contributing to some of the loss of fidelity observed during this study.

5.2.2 Clinical Samples

Data on the genotypes detected in each sample and the number of reads generated from the gt3a and the gt1a runs is given in tables 5-4 and 5-5 respectively. The majority of samples produced between 10,000-800,000 reads mapping to the major genotype. There was significant variation throughout the samples however with a range of 20-6,000,000 reads mapped. Despite pooling the NS5B and E1-E2 amplicons together in equimolar ratios, in the majority of samples, approximately 10-fold more reads were obtained for the NS5B amplicon compared to the E1-E2 amplicon. Generally, recovery of both amplicons was good although there was an issue with the mono-infected gt1a samples (run 2). The E1-E2 amplicon was missing for 70% of these samples during the library preparation stage. In the 30% of gt1a mono-infected samples where the E1-E2 amplicon was retained, it was detected at much lower levels than the corresponding NS5B amplicon. This issue was detected when the samples underwent Tapestation analysis prior to loading them onto the sequencing cartridge and it not known why the E1-E2 amplicons were lost during the library preparation. It may suggest that combing the amplicons in equimolar ratios was not ideal however and that if the runs were repeated, the E1-E2 and NS5B amplicons should undergo library preparation sepaerately. The E1-E2 PCR and library preparation was repeated for the affected samples and they were included in the run with different tags from the original samples.

Table 5-4; Deep sequencing results from run 1

Deep sequencing results from run 1 which included the gt3 mono-infected samples and mixed infection samples where gt3 was thought to be the major genotype present. The total number of reads is given and the proportion mapping to each genotype and amplicon. NK indicates that reads for the genotype were detected (and a consensus sequence isolated) but that the proportion of reads contributed was not quantified. Y indicates that the genotype was detected in the sample and ND indicates that the genotype was not detected. DTan* indicates that the genotype was not detected by the k-mer genotyping algorithm but was identified using Tanoti. Reads mapping to H77 (gt1) and JFH-1 (gt2) are shown; these sequences were contaminants from another source (section 5.2.3).

	Sample	Run ID	Total number of reads	Gt3a	% total reads	E1-E2 reads	NS5B reads	Gt1a	% total reads	E1-E2 reads	NS5B reads	Gt other	% total reads	E1-E2 reads	NS5B reads
Mixed infection samples	g3-36	mix-1	381980	Y	99.99	10^4-10^5	10^5-10^6	Y	0.01	NK	0	ND			
	g3-41	mix-2	258382	Y	99.29	10^4-10^5	10^5-10^6	DTan*				ND			
	g3-44	mix-3	269863	Y	99.97	10^5-10^6	10^5-10^6	ND				JFH-1	0.03	0	10^2-10^3
	g3-51	mix-4	280456	Y	86.41	10^4-10^5	10^5-10^6	DTan*				JFH-1	5.54	0	10^3-10^4
	g3-67	mix-5	257728	Y	97.15	10^4-10^5	10^5-10^6	ND				JFH-1	2.78	0	NK
	g3-73	mix-6	402438	Y	85.32	10^5-10^6	10^5-10^6	Y	8.61	0	10^3-10^4	JFH-1	6.07	0	10^3-10^4
	g3-80	mix-7	340870	Y	17.46	10^5-10^6	10^5-10^6	Y	16.17	0	10-10^2	JFH-1	66.37	0	10^3-10^4
	g3-125	mix-10	254849	Y	10.9	10^4-10^5	10^5-10^6	Y	84.44	10^4-10^5	10-10^2	JFH-1	4.66	0	10^2-10^3
	g3-128	mix-11	279970	Y	99.24	10^4-10^5	10^5-10^6	DTan*				ND			
	g3-155	mix-12	263209	Y	99.66	10^4-10^5	10^5-10^6	ND				JFH-1	0.34	0	NK
	g3-166	mix-13	264005	Y	99.94	10^4-10^5	10^5-10^6	DTan*				ND			
	g3-176	mix-14	298911	Y	86.86	10^5-10^6	10^5-10^6	ND				ND			
	g-27	mix-19	297039	Y	99.16	10^4-10^5	10^5-10^6	Y	0.84	NK	0	ND			
	g-30	mix-20	321081	Y	100	10^4-10^5	10^5-10^6	ND				ND			
	g-45	mix-21	315873	Y	100	10^4-10^5	10^5-10^6	DTan*				ND			
	e-15	mix-22	124017	Y	3.48	10-10^2	10^3-10^4	Y	25.62	10^4-10^5	10-10^2	JFH-1	70.91	10^5-10^6	10^4-10^5
	e-49	mix-23	293546	Y	99.05	10^4-10^5	10^5-10^6	Y	0.46	NK	0	JFH-1	0.49	0	NK
Gt3a only samples	3-18	gt3a-1	256068	Y	97.53	10^5-10^6	10^5-10^6	H77	2.47	10^2-10^3	10-10^2	ND			
	3-29	gt3a-2	216272	Y	99.98	10^4-10^5	10^5-10^6	H77	0.02	NK	NK	ND			
	3-35	gt3a-3	462306	Y	99.98	10^5-10^6	10^5-10^6	ND				ND			
	3-42	gt3a-4	143354	Y	99.91	10^4-10^5	10^5-10^6	ND				JFH-1	0.19	0	NK
	3-52	gt3a-5	150063	Y	100	10^5-10^6	10^5-10^6	ND				ND			
	3-60	gt3a-6	206186	Y	45.32	10^4-10^5	10^5-10^6	H77	1.01	10-10^2	10-10^2	JFH-1	53.32	0	10^4-10^5
	3-71	gt3a-7	197	ND				ND				ND			
	3-82	gt3a-8	380197	Y	99.83	10^5-10^6	10^5-10^6	H77	0.01	0	NK	JFH-1	0.16	0	NK
	3-95	gt3a-9	148143	Y	99.82	10^5-10^6	10^4-10^5	H77	0.18	NK	NK	ND			
	3-109	gt3a-10	112752	Y	97.4	10^4-10^5	10^4-10^5	ND				ND			
	3-116	gt3a-11	124717	Y	100	10^4-10^5	10^4-10^5	ND				ND			
	3-127	gt3a-12	151875	Y	99.43	10^4-10^5	10^5-10^6	ND				JFH-1	0.57	0	NK
	3-136	gt3a-13	414681	Y	99.66	10^4-10^5	10^5-10^6	H77	0.04	NK	0	JFH-1	0.31	0	NK
	3-146	gt3a-14	132191	Y	100	10^5-10^6	10^4-10^5	ND				ND			
	3-153	gt3a-15	144989	Y	100	10^5-10^6	10^2-10^3	ND				ND			
	3-164	gt3a-16	113057	Y	100	10^4-10^5	10^4-10^5	ND				ND			
	3-188	gt3a-17	854473	Y	70.86	10^4-10^5	0	H77	29.14	10-10^2	<10	ND			
	3-191	gt3a-18	135359	Y	100	10^5-10^6	10^5-10^6	ND				ND			
	3-209	gt3a-19	141103	Y	100	10^5-10^6	10^4-10^5	ND				ND			
	3-218	gt3a-20	117275	Y	100	10^5-10^6	10^5-10^6	ND				ND			
Controls	3a-transcript1	control-8	372773	Y	100	10^4-10^5	10^5-10^6	ND				ND			
	3a-transcript2	control-9	276947	Y	100	10^2-10^3	10^5-10^6	ND				ND			
	3a-transcript3	control-10	336302	Y	100	10^5-10^6	10^5-10^6	ND				ND			
	0.1% 1a-major3a	control-11	87298	Y	89.73	10^4-10^5	10^3-10^4	Y	10.27	10^2-10^3	10-10^2	ND			
	1% 1a-major3a	control-12	201247	Y	83.5	10^4-10^5	10^5-10^6	Y	15.72	10^3-10^4	10^3-10^4	JFH-1	0.77	0	NK
	10% 1a-major3a	control-13	3021262	Y	31.67	10^5-10^6	10^6-10^7	Y	68.32	10^5-10^6	10^4-10^5	ND			
	neg-serum-2	control-2	333976	ND				ND				ND			
	neg-H_2O	control-3	0	ND				ND				ND			

Table 5-5; Deep sequencing results from run 2

Deep sequencing results from run 2 which included the gt1a mono-infected samples and mixed infection samples where gt1a was thought to be the major genotype present. The total number of reads is given and the proportion mapping to each genotype and amplicon. NK indicates that reads for the genotype were detected (and a consensus sequence isolated) but that the proportion of reads contributed was not quantified. Samples E1a-1 – E1a-20 are the E1-E2 only repeats of the samples that lost their E1-E2 amplicon during the initial library preparation. Y indicates that the genotype was detected in the sample and ND indicates that the genotype was not detected. DTan* indicates that the genotype was not detected by the k-mer genotyping algorithm but was identified using Tanoti. H77 (gt1) and JFH-1 (gt2) indicate that whilst gt1 and gt2 reads were detected during the analysis, they were contamination from another source (section 5.2.3). 1c** indicates that whilst gt1c was detected in the sample using the k-mer algorithm, sequence-based analysis of the consensus sequence generated from the reads demonstrated it was in fact gt1a (section 5.1.2).

	Sample	Run ID	Total number of reads	Gt1a	% total reads	E1-E2 reads	NS5B reads	Gt3a	% total reads	E1-E2 reads	NS5B reads	Gt other	% total reads	E1-E2 reads	NS5B reads
Mixed	g3-85	51	316812	Y	99.88	10^5-10^6	10^5-10^6	DTan*				JFH-1	0.02	0	10^2-10^3
	g3-92	52	497035	Y	96.33	10^5-10^6	10^5-10^6	DTan*				JFH-1	1.98	0	NK
	1a-128	44	206780	Y	100	10^4-10^5	10^5-10^6	DTan*				ND			
Gt1a only samples	1a-1	5	214446	Y	99.57	10-10^2	10^5-10^6	ND				ND			
	1a-2	6	217702	Y	100	0	10^5-10^6	ND				ND			
	1a-3	7	180537	Y	99.98	0	10^5-10^6	ND				ND			
	1a-4	8	255150	Y	100	0	10^5-10^6	ND				ND			
	1a-5	9	147575	Y	100	10^4-10^5	10^5-10^6	ND				ND			
	1a-6	10	501101	Y	100	10-10^2	10^5-10^6	ND				ND			
	1a-7	11	477300	Y	99.98	0	10^5-10^6	ND				ND			
	1a-8	12	295762	Y	99.99	0	10^5-10^6	ND				JFH-1	0.01	0	10^3-10^4
	1a-9	13	513370	Y	99.99	10^4-10^5	10^5-10^6	ND				ND			
	1a-10	14	274704	Y	21.4	10^3-10^4	10^5-10^6	ND				JFH-1	78.6	0	NK
	1a-11	15	422191	Y	99.98	0	10^5-10^6	ND				ND			
	1a-12	16	183938	Y	98.57	0	10^5-10^6	ND				JFH-1	1.43	0	10^2-10^3
	1a-13	17	428967	Y	99.4	0	10^5-10^6	ND				JFH-1	0.6	0	NK
	1a-14	18	275951	Y	98.79	0	10^5-10^6	ND				ND			
	1a-15	19	336214	Y	99.91	0	10^5-10^6	ND				ND			
	1a-16	20	526931	Y	99.97	0	10^5-10^6	ND				ND			
	1a-17	21	181921	Y	100	0	10^5-10^6	ND				ND			
	1a-18	22	139110	Y	99.99	0	10^5-10^6	ND				ND			
	1a-19	23	264805	Y	99.96	10^4-10^5	10^5-10^6	ND				ND			
	1a-20	24	476207	Y	100	0	10^5-10^6	ND				ND			
Gt1a only samples(E1-E2 amplicons)	E1a-1	25	183504	Y	100	10^5-10^6	0	ND				ND			
	E1a-2	26	197065	Y	100	10^5-10^6	0	ND				ND			
	E1a-3	27	119708	Y	100	10^5-10^6	0	ND				ND			
	E1a-4	28	141138	Y	99.95	10^5-10^6	0	ND				ND			
	E1a-5	29	238171	Y	100	10^5-10^6	0	ND				ND			
	E1a-6	30	175808	Y	100	10^5-10^6	0	ND				ND			
	E1a-7	31	221019	Y	100	10^5-10^6	0	ND				ND			
	E1a-8	32	245537	Y	100	10^5-10^6	0	ND				ND			
	E1a-9	33	341520	Y	100	10^5-10^6	0	ND				ND			
	E1a-10	34	256385	Y	100	10^5-10^6	0	ND				ND			
	E1a-11	35	237551	Y	100	10^5-10^6	0	ND				ND			
	E1a-12	36	132433	Y	100	10^5-10^6	0	ND				ND			
	E1a-13	37	194218	Y	100	10^5-10^6	0	ND				ND			
	E1a-14	38	160106	Y	100	10^5-10^6	0	ND				ND			
	E1a-15	39	217099	Y	100	10^5-10^6	0	ND				ND			
	E1a-16	40	335389	Y	100	10^5-10^6	0	ND				ND			
	E1a-17	41	447856	Y	100	10^5-10^6	0	ND				ND			
	E1a-18	42	610952	Y	96.14	10^5-10^6	0	ND				1c**	3.86	NK	0
	E1a-20	43	183134	Y	100	10^5-10^6	0	ND				ND			
Controls	1a-transcript1	1	638640	Y	99.99	10^5-10^6	10^5-10^6	ND				ND			
	1a-transcript2	2	651035	Y	99.99	10^5-10^6	10^5-10^6	ND				ND			
	1a-transcript3	53	195539	Y	99.99	10^4-10^5	10^5-10^6	ND				ND			
	0.1% 1a-major3a	3	180305	Y	92.26	10^4-10^5	10^5-10^6	Y	7.37	10^2-10^3	10^3-10^4	ND			
	1% 1a-major3a	4	139813	Y	40.87	10^3-10^4	10^5-10^6	Y	59.05	0	10^4-10^5	ND			
	10% 1a-major3a	56	195539	Y	92.97	10^4-10^5	10^4-10^5	Y	7.69	10^5-10^6	0	ND			
	Sero neg	57	236469	ND				ND				ND			
	neg-H_2O	58	0	ND				ND				ND			

5.2.3 Contamination

Aside from the expected gt1a and gt3a reads detected in the samples, gt1c was reported in E1a-18, gt2 was reported in 21 samples and gt1a was found in 7 of the mono-infected gt3a samples. The gt1c sequence detected in E1a-18 was not reported in the corresponding NS5B sample (1a-18). A comparative analysis of the gt1a and gt1c consensus E1-E2 sequences from the sample identified that there were just 13/541 nucleotides varying between the two sequences. Phylogenetic analysis consistently grouped the gt1c sequence with other gt1a sequences, indicating that it was in fact a gt1a sequence (figure 5-7) and the 1a and '1c' consensus sequences isolated from sample E1a-18 were shown to cluster next to each other with strong bootstrap support. This misclassification suggests that there may be issues with using such a short fragment length (37bp) in the k-mer genotyping pipeline, especially in a highly variable region of the viral genome.

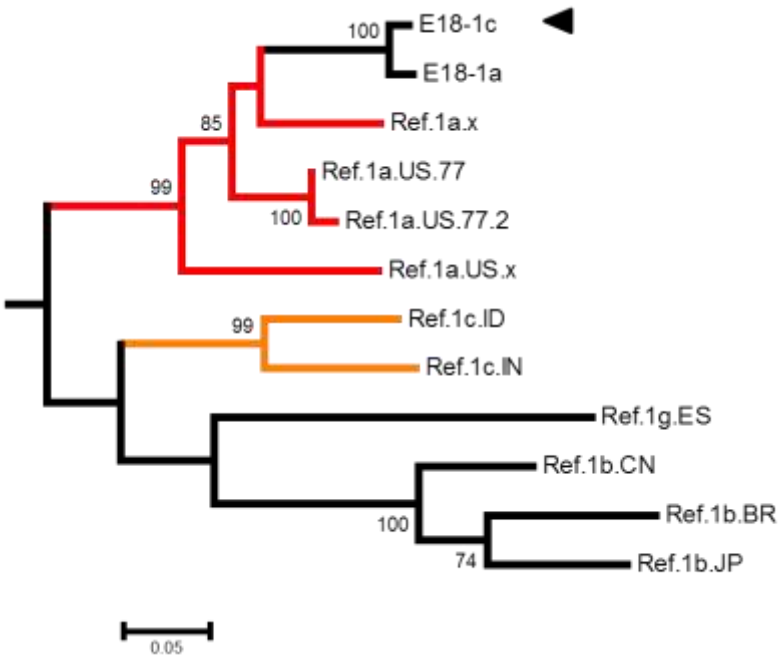


Figure 5-7; Phylogenetic analysis of gt1c strain detected in sample E1a-18

Maximum likelihood phylogenetic tree comparing the gt1c E1-E2 consensus sequence from E1a-18 (E18-1c, highlighted with black arrow) with gt1a and gt1c reference strains (red and orange respectively). The gt1a consensus sequence from E1a-18 (E18-1a) has also been included in the tree. Bootstrap support of >70% after 1000 replicates is indicated on the tree.

Genotype 2a was detected in 21 samples from both run 1 and run 2. In 20/21 cases, only the NS5B fragment was detected suggesting that a replicon may be the contaminating source.

Phylogenetic analysis indicated also that all the gt2 reads detected in the samples were highly similar to JFH-1 (figure 8). JFH-1 sub-genomic replicons, constructed from the non-structural HCV genes and often containing a lucierferase reporter element, are a widely used tool within the building where this work was carried out. The library preparations were however performed in a laboratory where JFH-1 replicons are not handled and laboratory users are required to change lab coats between the two laboratories to prevent such contamination. The detection of JFH-1 in a large number of these samples suggests that these measures have however been insufficient to prevent contamination and transfer of material is occurring between the two sites.

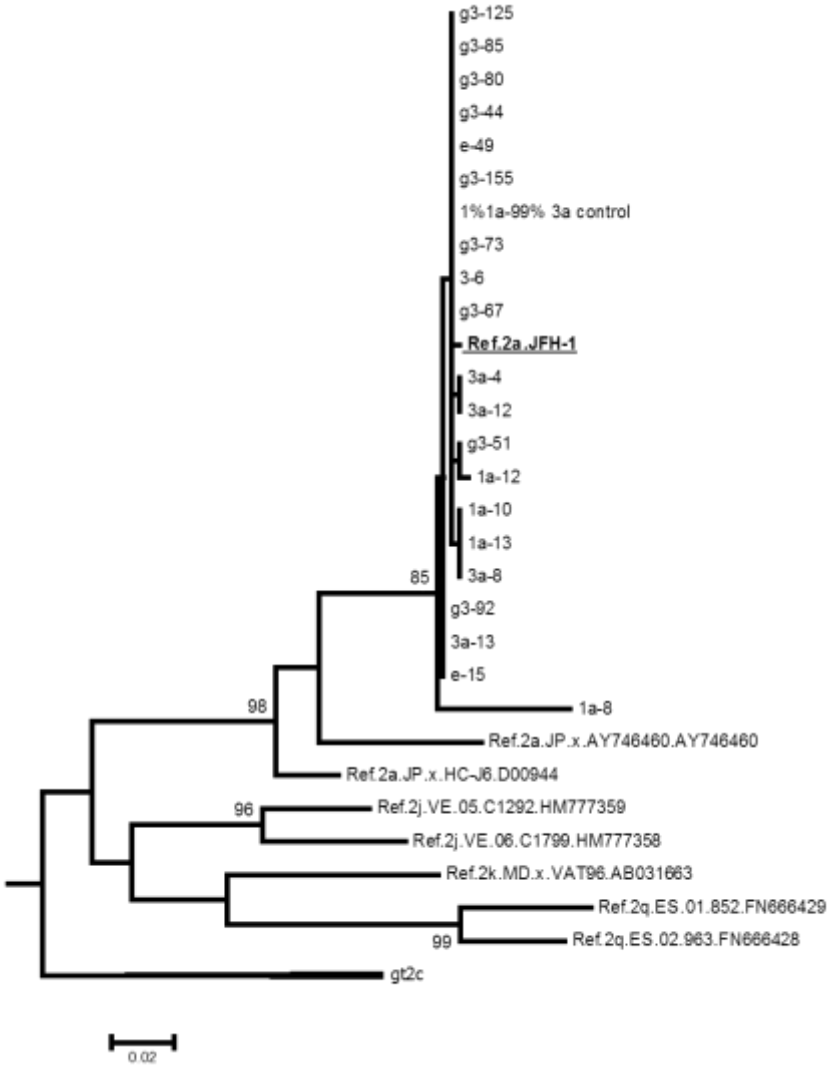


Figure 5-8; Phylogenetic analysis of gt2 sequences detected in the samples

Maximum likelihood tree of gt2a NS5B consensus sequences detected during the deep sequencing runs. Bootstrap support of >70% after 1000 replicates is indicated on the tree and the JFH-1 reference strain has been underlined.

Of the 21 gt2 NS5B consensus sequences analysed, the sequence from sample 1a-8 was found to be the most distantly related to the JFH-1 replicon with substitutions occurring at 16/385 sites. A comparative analysis of the two sequences identified a 40bp region within sample 1a-8 that was identical to H77 (figure 5-9). This suggests it may be contamination from a chimeric replicon, which is also widely used in the laboratory.

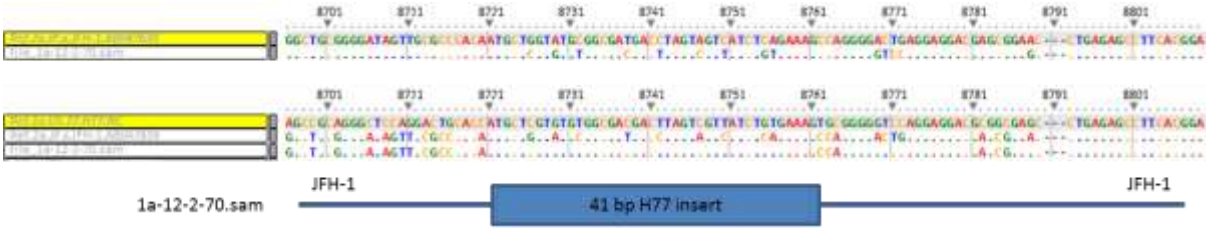


Figure 5-9; Sequence analysis of gt2 sequence detected in sample gt1a-8

Alignments and illustration indicating that the sequence gt1a-8 (1a-12-2-70.sam) is derived from JFH-1 and H77.

A gt2 sequence occurring in both the E1-E2 and the NS5B regions was detected in sample E-15. Phylogenetic analysis of the NS5B sequence indicated that it was closely related to the JFH-1 replicon strain (figure 5-8). The E1-E2 sequence was found to be missing a 73bp region in the middle of the fragment at a similar location to the gaps identified in the gt3 control sequences. Analysis of the E1-E2 sequence found that it was distantly related to JFH-1 but was closely related to HC-J6.D00944, a strain isolated from a patient in Japan in 1991 (figure 5-10) (Okamoto *et al.*, 1991). The strain is commonly used in HCV research for HCVcc culture systems and again, was used within the building where these libraries were prepared. This is further indication that the measures used whilst handling these samples on the open bench were insufficient to prevent contamination and that ideally, work of this nature should be done in allocated facilities.

Figure 5-10; Phylogenetic analysis of gt2 sequence detected in sample E-15

Phylogenetic analysis showing the distant relationship of the E-15 E1-E2 consensus sequence and JFH-1. Bootstrap support of >70% after 1000 replicates is indicated on the tree.

Both E1-E2 and NS5B gt1a sequences were detected in 5/20 mono-infected gt3 samples that had been previously screened and found to be negative for the presence of gt1a strains (figure 5-11). Additionally an E1-E2 sequence was detected in sample 3-13 and an NS5B sequence was detected in sample 3-8. All seven samples had previously tested negative for the presence of gt1a strains using the nested PCR assay developed in chapter 4. Phylogenetic analysis of the sequences indicated that the majority of the sequences detected were closely related to H77 and therefore likely to be a result of contamination. In the NS5B tree, all the sequences were closely related to H77 apart from the gt1a sequence detected in sample gt3a-2. This sequence was attributed to a single NS5B read isolated in the sample. In the E1-E2 region, there were 8 gt1a reads isolated from sample gt3a-2 and they were near-identical to the H77 control strain. The gt1a NS5B sequence from sample gt3a-17 was also found to be incomplete and had two large 60bp deletions within it. This is likely a result of the low numbers of reads contributing to the consensus sequence (table 5-4). In the E1-E2 tree, all the unexpected gt1a sequences samples were closely related to H77 except the sequences isolated from samples gt3a-1 and gt3a-6. The gt1a sequences from both samples gt3a-1 and gt3a-6 were highly similar to H77 apart from 6 or 10 nucleotides respectively in the middle of the sequences (figure 5-12). These heterogeneous regions were located within the region where gaps were identified within the gt3a control sequences. As the number of reads contributing to these sequences was relatively low, read coverage in this region is likely to be even lower still and sequence calling may have been affected by the excess of signals from other gt3 amplicons on this sequencing run. Whilst sequence clustering was reduced and a high Phi-X spike-in was used to control for issues of this nature, they may have still affected sequence calling in some cases.

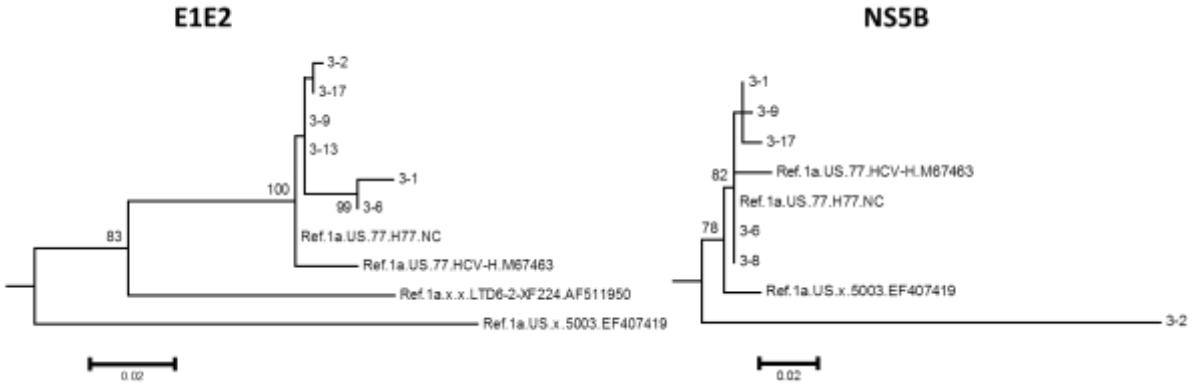


Figure 5-11; Phylogenetic analysis of gt1a sequences detected in gt3a only samples

Maximum likelihood trees showing the gt1a sequences detected in the non-mixed infection samples that were similar to H77. Samples included are gt3a-1 (3-1), gt3a-2 (3-2), gt3a-6 (3-6), gt3a-8 (3-8), gt3a-9 (3-9), gt3a-13 (3-13) and gt3a-17 (3-17). Bootstrap support of >70% after 1000 replicates is indicated on the trees.

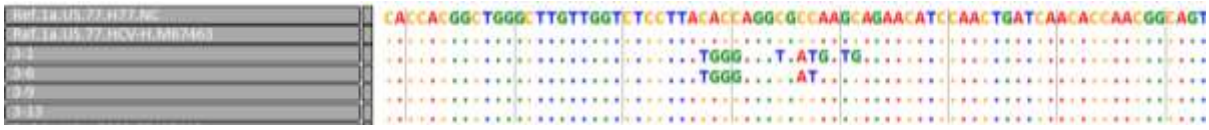


Figure 5-12; Sequence analysis of gt1a sequences isolated from gt3a-1 and gt3a-6

Inserts in the E1-E2 consensus sequences from samples gt3a-1 and gt3a-6.

The source of the H77 contamination is unclear but there are two possible sources. H77 replicons are frequently used in other laboratories in the building where this work was carried out. From the analysis of the gt2 strains, there is evidence of transfer between two laboratories and this may explain some of the H77 contamination detected. The gt1a control transcripts developed for this project in chapter 4 are also based on H77 and they included both the E1-E2 and NS5B regions targeted by the pan-genotypic assay. It is possible that some of the samples analysed here have been contaminated with the control transcripts. The detection of H77 in both the E1-E2 and the NS5B assays however suggests that if the control transcripts were the contaminating source, the samples were contaminated on two separate occasions and this is unlikely.

As a result of these analyses indicating that the additional genotypes detected were either as a result of genotype misclassification or contamination, they have been excluded from the subsequent analysis.

5.2.4 Genotype 1a and genotype 3a samples

20 gt1a and 20 gt3 samples were selected randomly from our cohort of samples previously screened for mixed gt1a/gt3. All samples had previously tested negative for the presence of mixed gt1a/gt3 infection using the nested PCR assays developed in chapter 4. Libraries were prepared from the E1-E2 and NS5B amplicons for each sample and the samples then underwent deep sequencing on the Illumina MiSeq platform. Deep sequencing reads were obtained for 20/20 gt1a samples and 19/20 gt3 samples. No reads were obtained for either the E1-E2 or the NS5B amplicons of sample gt3a-7. It is unclear why this occurred as the library concentration obtained using Qubit was satisfactory but it may indicate that the sample was loaded into the final sequencing pool at an inadequate concentration. In addition to the 20 gt1a and the 19 gt3 sequences expected from these samples, initial results indicated that a number of the samples may contain multiple genotype infections. Analysis of the sequences indicated they were closely related to either JFH-1 (gt2) or H77 (gt1a), implying that they were contaminating sequences. These sequences were therefore excluded from the analysis, meaning that no additional genotype sequences were detected in either the 20 gt1a or the 19 gt3a samples. A phylogenetic analysis of the consensus sequences obtained for these 39 samples is shown in figure 5-13. The NS5B sequence from sample gt1a-10 has been excluded from the tree as it was only 219bp in length. BLAST analysis of the sequence indicated it was gt1a and that the sequence was distinct from the H77 control. In both the E1-E2 and NS5B trees, the sequences were highly divergent from the control sequences included in the analysis and also from each another, indicating that they are derived from patient isolates and not the result of contamination.

One issue encountered during the library preparation of the mono-infected samples was the loss of the E1-E2 fragment from the gt1a samples during the library preparation (table 5-5). This occurred despite mixing the E1-E2 and NS5B fragments together in equimolar ratios, as was done for all the other samples in this study. The E1-E2 pan-genotypic PCR was repeated for these samples, library preparations were performed and they were loaded into the final sequencing pool with separate tags. It is unclear however why this issue particularly affected these samples.

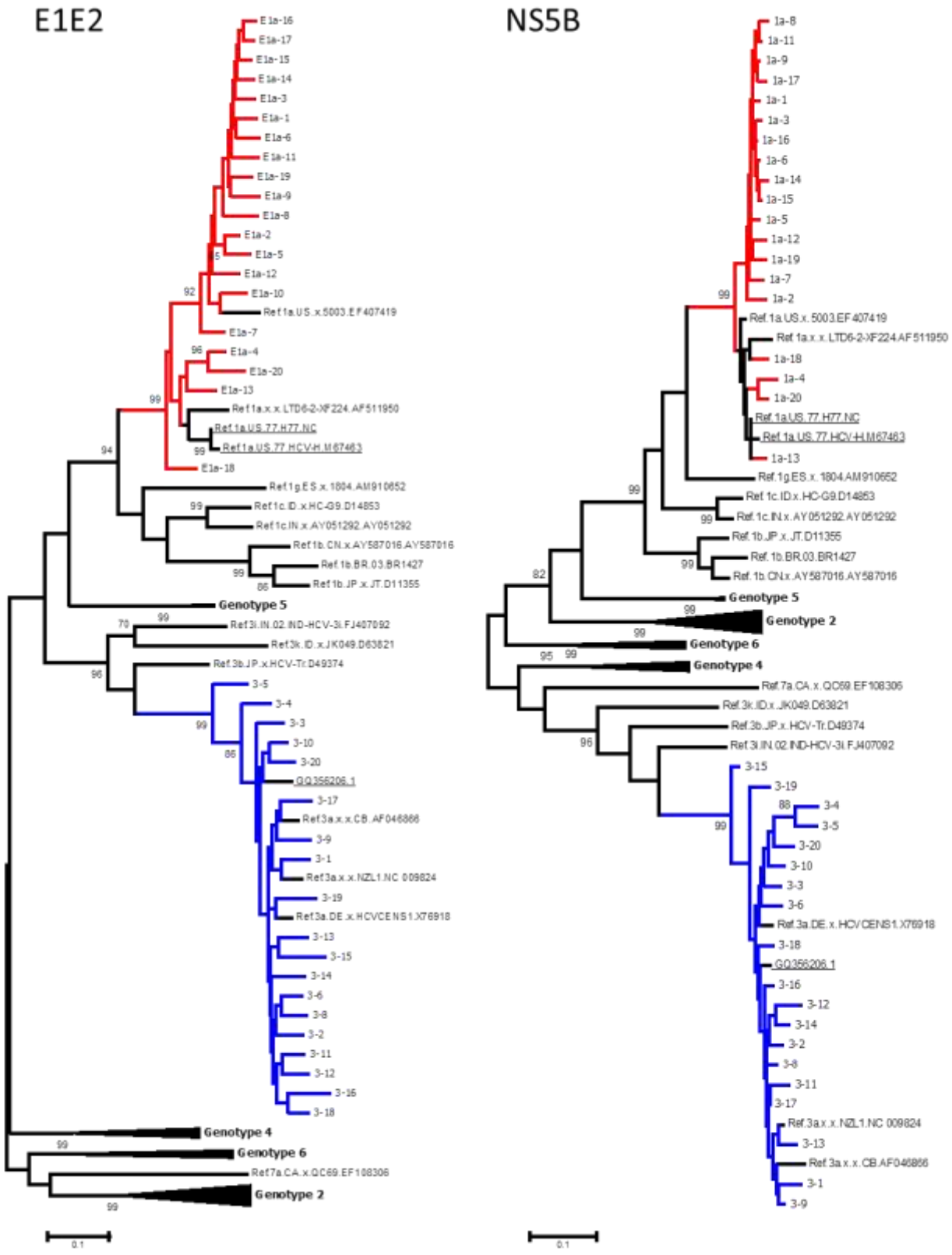


Figure 5-13; Phylogenetic analysis of sequences isolated from the gt1a only and gt3 only samples

Maximum likelihood phylogenetic trees of the consensus sequences for the single-genotype infected gt1a and gt3 samples. Gt1a sequences have been highlighted in red and gt3 sequences have been highlighted in blue. Control gt1a and gt3 sequences have also been underlined. Bootstrap support of >70% after 1000 replicates is indicated on the trees.

5.2.5 Mixed genotype 1a/genotype 3 samples

The 20 samples of mixed gt1a/gt3 infections also underwent deep sequencing analysis. E1-E2 and NS5B pan-genotypic PCRs were performed on the samples, libraries were prepared and the samples were then sequenced on the Illumina MiSeq platform. Samples where the major genotype was presumed to be gt3 were loaded onto run 1 (n=17) and samples where genotype 1a was presumed to be the major genotype were loaded onto run 2 (n=3). Reads were obtained for 20/20 samples and the results from the runs are given in the table 5-4 and 5-5. Differences were noted in the detection sensitivity between the E1-E2 and NS5B pan-genotypic assays, with the E1-E2 assay appearing to be more sensitive for the detection of mixed genotype infections (table 5-7). This issue is discussed in section 5.2.6 and the analysis here focuses on the detection of E1-E2 reads, for which there were comparative sequences available, generated by Sanger sequencing (see chapter 4). All of the major genotypes were identified by deep sequencing. An initial analysis of the minor genotype using the k-mer genotyping algorithm detected gt1a reads in 7/17 (41%) of gt3 samples and gt3 reads in 0/3 of the gt1a samples (table 5-6). Further analysis of the reads from the samples using Tanoiti (Vattipally, 2014) was able to detect gt1a reads in an additional 5/17 (29%) of the major gt3 samples (run 1) and 3/3 (100%) of the major gt1a samples (run 2). Whilst gt1a deep sequencing reads were detected for sample g3-176, phylogenetic analysis indicated that they were identical to H77 (highlighted in figure 5-14). These sequences were likely the result of contamination and the gt1a minor strain from g3-176 was classed as not detected. Overall using both the k-mer detection algorithm and Tanoti (Vattipally, 2014), mixed gt1a/gt3 infections were detected in 15/20 (75%) of the samples tested. Gt2 was also identified in 11/20 samples and analysis indicated that all the gt2 sequences detected were highly similar to JFH-1 (figure 5-8). They were excluded from subsequent analysis as they were considered to be contaminants.

Among these samples, the major genotype present was identified as gt3 in 15/17 samples from run 1 and the major genotype was found to be gt1a in 3/3 samples from run 2. Previous qPCR results from sample g3-125 had indicated that the gt1a minor strain comprised 6.6% of the total viral load. In contrast to this, just 11.4% of the deep sequencing reads were gt3a and 88.6% of the reads were gt1a, implying that gt1a was the major strain present in this sample (after exclusion of gt2 reads). In sample E-15, 88.0% of the total reads were found to be gt1a and 12.0% of the reads were genotyped as gt3 (after exclusion of gt2 reads). Previously qPCR assays on the sample had not yielded a result and subsequently, gt3 was presumed to be the major genotype present in the sample on the basis of genotyping results from the clinical

laboratory. It should be noted that results from the sequencing of the controls (table 5-3) indicated that PCR bias was an issue with assay and that the deep sequencing assays should be used solely for detection of the genotypes, and not for estimating the proportions of the genotypes present in the samples. Without comparable qPCR results, it is difficult to draw robust conclusions on the major genotype present in E-15 based on these results.

In figures 5-14 (gt1a sequences) and 5-15 (gt3 sequences), the E1-E2 sequences detected by both Sanger sequencing and deep sequencing from the major and minor strains isolated from the samples have been compared. Major genotypes for the samples (as indicated by qPCR or by the diagnostic laboratory) are listed in table 5-7. Consensus sequences generated from the deep sequencing runs in the phylogenetic trees are identifiable by the tag '.ds'. The comparison of the major strain E1-E2 sequences indicated that 18/20 (90%) of the consensus sequences generated from deep sequencing reads were identical or highly similar to the E1-E2 sequences produced by Sanger sequencing and they can be seen to cluster with high bootstrap values (highlighted in yellow). Considerable heterogeneity between the deep sequencing and Sanger sequences of the major strain was observed for g3-92 and E-15. As discussed previously, it was not clear what the major genotype present in sample E-15 was. The sample was genotyped as gt3 by the diagnostic laboratory but deep sequencing produced substantially more gt1a reads than gt3 reads. The gt1a consensus read from the deep sequencing clustered closely with the gt1a Sanger sequence whereas the two gt3 sequences did not. This may suggest that the majority stain in the sample is in fact gt1a although ideally, qPCR would be used to confirm this. Correlation between the Sanger sequences and the deep sequencing consensus sequences was not consistent for the minor strains analysed however and just 2/15 minor strain sequence sets, gt1a-128 and E-49, were found to cluster together in the analysis. The minor strain for E-49 contributed just 0.46% of the total reads and was detected using Tanoti (Vattipally, 2014), suggesting it too contributed a low amount of the total reads from the sample. The fact that sequences for these minor strains clustered together despite the low number of reads attributed to them indicates that additional factors, such as contamination, poor read quality or the presence of previously undetected strains, are also likely to be impacting clustering of the minor strain sequences within the tree.

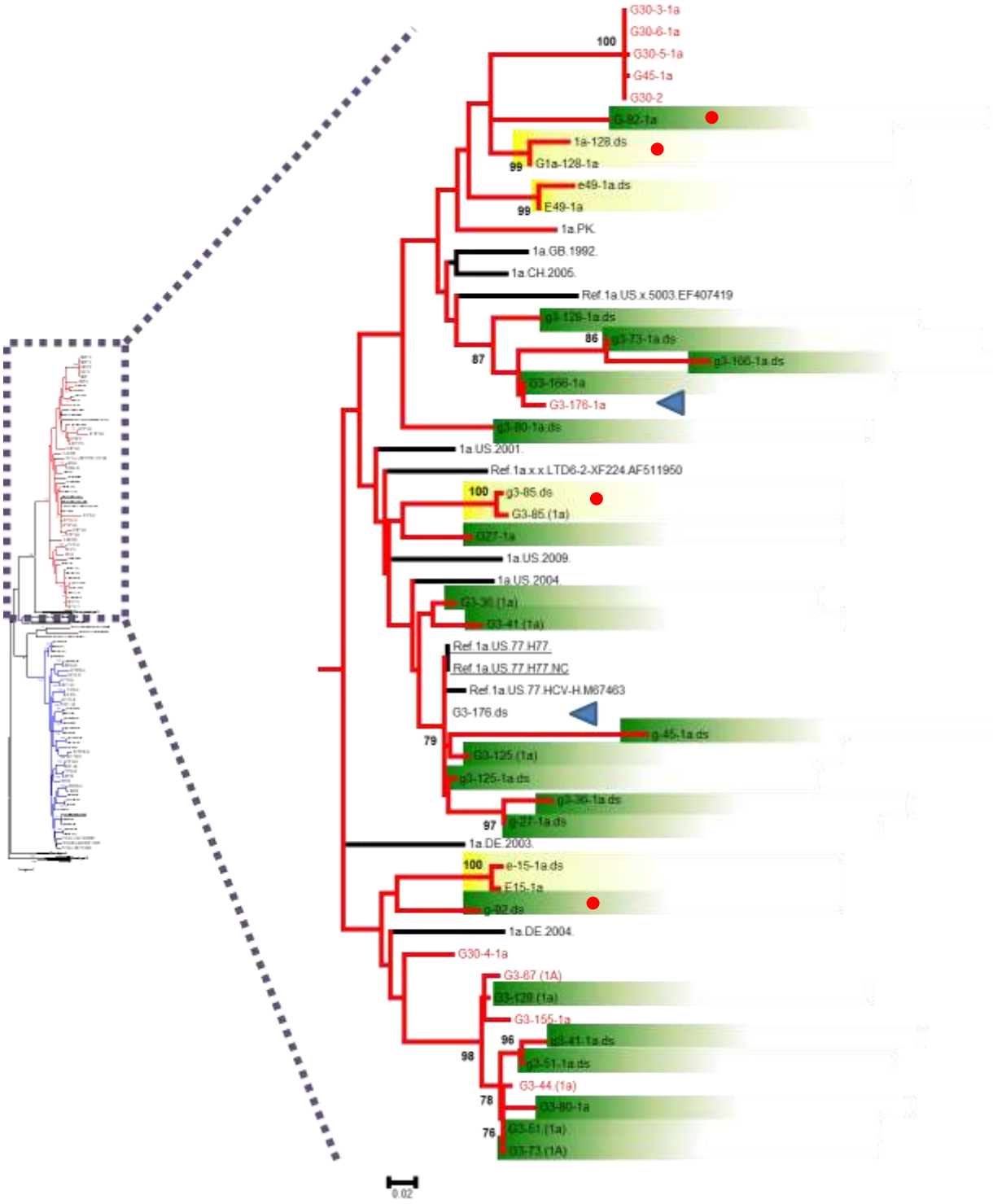


Figure 5-14; Phylogenetic analysis of gt1a Sanger sequences and deep sequencing consensus sequences

Region of a maximum likelihood phylogenetic tree comparing gt1a E1-E2 sequences of mixed infection positive samples obtained by Sanger sequencing and E1-E2 consensus sequences from deep sequencing. Control strains are underlined and samples without a corresponding consensus sequence from deep sequencing are highlighted in red. Sequences from the same sample grouping next to each other in the tree are highlighted in yellow and those grouping separately in the tree are highlighted in green. Bootstrap support of $\geq$70% after 1000 replicates is indicated on the tree. Red dots have been used to highlight the major strains (as determined by qPCR). Blue arrows have been used to highlight sequences from G3-176, where the gt1a deep sequencing reads were found be contaminants.

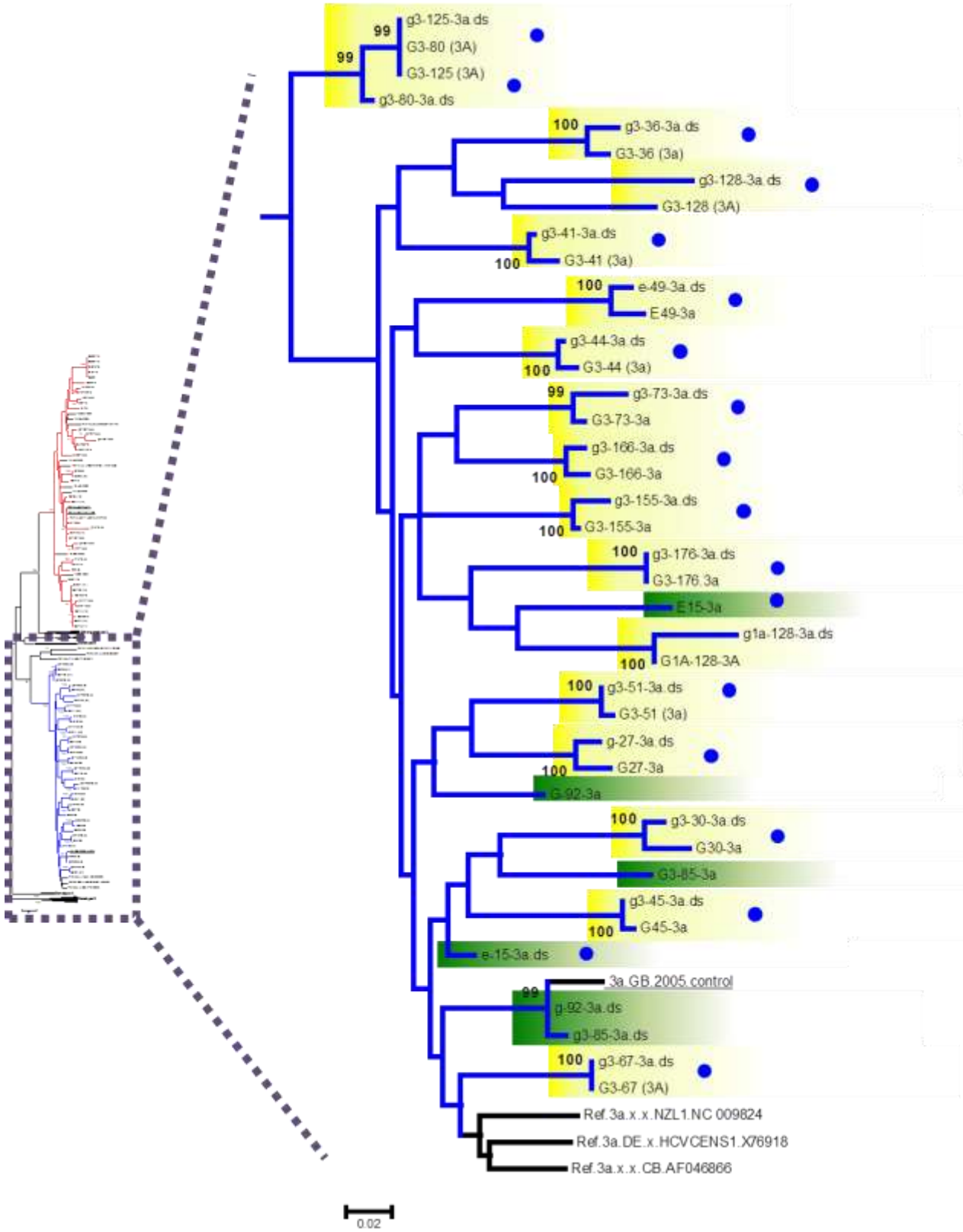


Figure 5-15; Phylogenetic analysis of gt3 Sanger sequences and deep sequencing consensus sequences

Region of a maximum likelihood phylogenetic tree comparing gt3a E1-E2 sequences of mixed infection positive samples obtained by Sanger sequencing and E1-E2 consensus sequences from deep sequencing. Control strains are underlined and samples without a corresponding consensus sequence from deep sequencing are highlighted in red. Sequences from the same sample grouping next to each other in the tree are highlighted in yellow and those grouping separately in the tree are highlighted in green. Blue dots have been used to highlight the major strains (as determined by qPCR). Bootstrap support of >70% after 1000 replicates is indicated on the tree.

A comparative analysis of the method used to detect the minor strain in the deep sequencing reads and the estimated proportion of the minor strain as calculated by qPCR (chapter 4) indicated that minor strains were more likely to be detected if they comprised >1% of the total viral load (table 4-4). Using the k-mer algorithm and Tanoti (Vattipally, 2014), all minor strains comprising >1% of the viral load were detected. Strains detected using the k-mer algorithm comprised, on average, a greater proportion of the viral load (7.59% ± 8.54) than the strains detected using Tanoti (1.54% ± 2.56) although there was considerable variation within both groups of samples. This indicates that Tanoti may be a more sensitive, albeit more time-consuming, method for the detection of minor strains. In samples in which the minor genotype was not detected by deep sequencing, qPCR indicated that all had minor strain populations comprising < 0.25% of the total viral load. Minor strains were however detected by the k-mer algorithm or Tanoti in five other samples where the minor strain comprised <0.25% of the total viral load.

Table 5-6; Proportion of the minor strain and method of detection in the deep sequencing reads

Comparison of the minor strain viral load (%) as determined by qPCR (table 4-4, chapter 4) compared with the method used to detect the minor strain in the E1-E2 deep sequencing reads. Average viral load for each detection method and the standard deviation (SD) are also given. *All samples detected by k-mer analysis were also detectable with Tanoti but have not been included in this analysis to emphasise the sensitivity of Tanoti.

Detected by K-mer algorithm		**Detected by Tanoti***		**Not detected**	
sample	%	sample	%	sample	%
E-15	NK	g3-51	NK	g-30	NK
E-49	NK	g3-85	0.36	g3-44	NK
g-27	0.16	g3-92	0.42	g3-67	0.21
g3-36	0.05	g1a-128	6.91	g3-155	0.02
g3-73	20.63	g3-128	0.03	g3-176	0.22
g3-80	10.51	g3-41	0.13		
g3-125	6.61	g3-166	0.03		
		g-45	2.64		
Average	7.59	**Average**	1.54	**Average**	0.15
SD	8.54	**SD**	2.56	**SD**	0.11

By deep sequencing of g3-80 and g3-125 both genotype 1a and genotype 3a strains were detected in both samples using the k-mer algorithm. The consensus sequences of the gt3 strains (major genotype) for both samples were highly similar and grouped together closely in the phylogenetic tree with their comparable Sanger sequences. Consistent with previous work, the gt1a strains that were identified from the samples by deep sequencing demonstrated significant heterogeneity from each other and were incongruent relative to their respective Sanger sequences. Only the major gt3 strain within the sample G-30 was detected using deep sequencing. The presence of two distinct gt1a strains with a 16% difference at the nucleotide level was detected previously in sample G-30 by clonal analysis (chapter 4).

5.2.6 Comparison of E1-E2 and NS5B assays

An analysis of the NS5B amplicon reads from the deep sequencing runs found that the major strain was identified in 19/20 (95%) samples (table 5-7). NS5B amplicon reads for the minor strains were identified in 6/20 (30%) samples using both the k-mer algorithm and Tanoti (Vattipally, 2014). These pairs are highlighted in a phylogenetic analysis of the sequences (figure 5-16). Neither the major nor the minor NS5B seqeunces were identified for sample G-30 and there were no comparable NS5B sequences available for the samples to analyse alongside the deep sequencing sequences. Relative to the deep sequencing analysis of E1-E2 where 75% of the minor strains were detected using either the k-mer algorithm or Tanoti (Vattipally, 2014), the NS5B assay appears to be considerably less sensitive with just 30% of the minor strains detected. This was also despite the fact that in 60% of the mixed infection cohort, approximately 10-fold more NS5B reads were detected than E1-E2 reads (tables 5-4 and 5-5). For those samples in which minor strain NS5B reads were detected, previous qPCR results indicated that the minor strain comprised >5% of the total viral load for 3/6 samples. For E-15, qPCR results were not available and for G3-92 and G3-85 the qPCR results indicated that the minor strain comprised <0.5% of the total viral load. There were not enough samples with a gt3 minor strain (n=3) to make meaningful comparisons about the inter-genotypic sensitivity of the either the E1-E2 or the NS5B assay.

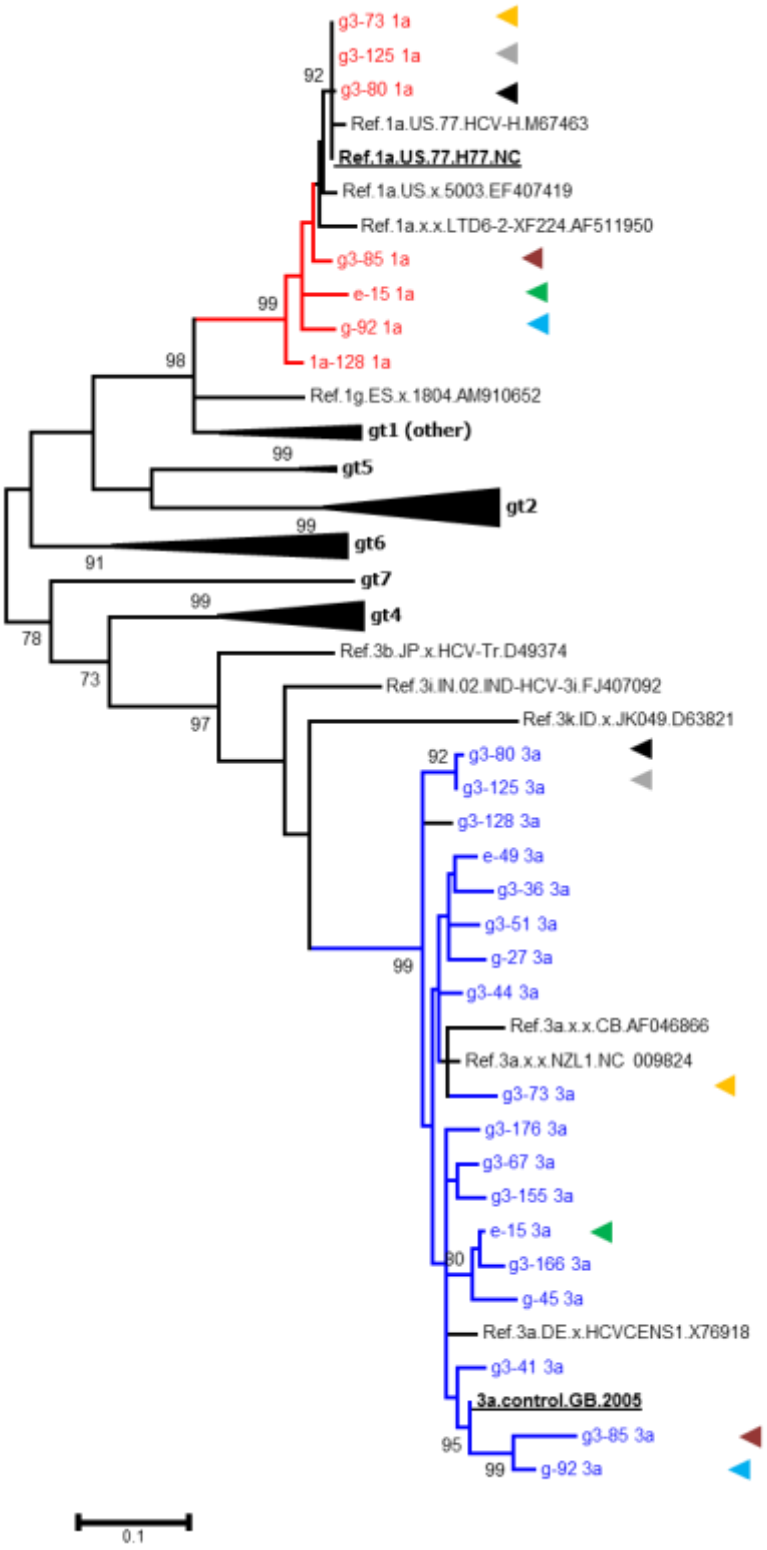


Figure 5-16; Phylogenetic analysis of consensus NS5B sequences

Maximum likelihood phylogenetic tree of consensus NS5B generated from the deep sequencing of the mixed infection positive samples. Samples for which both the major and minor strains were detected are highlighted with arrows. Bootstrap support of ≥70% after 1000 replicates is indicated on the tree.

Table 5-7; Gt1a and gt3 strains detected in the samples by nested PCR and deep sequencing

Sequences identified in mixed infection positive samples by nested PCR, E1-E2 deep sequencing assay and NS5B deep sequencing assay. The major genotype for each sample is shown. Y indicates a sequence was identified and N indicates a sequence was not detected. Y* indicates that sequences detected were similar to controls.

Sample	Major genotype	E1-E2 nested PCR		E1-E2 deep sequencing		NS5B deep sequencing	
		1a	3a	1a	3a	1a	3a
G27	3a	Y	Y	Y*	Y	N	Y
G30	3a	Y	Y	N	Y	N	N
G45	3a	Y	Y	Y	Y	N	Y
E15	3a	Y	Y	Y	Y	Y	Y
E49	3a	Y	Y	Y	Y	N	Y
G3-36	3a	Y	Y	Y*	Y	N	Y
G3-41	3a	Y	Y	Y	Y	N	Y
g3-44	3a	Y	Y	N	Y	N	Y
g3-51	3a	Y	Y	Y	Y	N	Y
G3-67	3a	Y	Y	N	Y	N	Y
G3-73	3a	Y	Y	Y	Y	Y*	Y
G3-80	3a	Y	Y	Y	Y	Y	Y
G3-85	1a	Y	Y	Y	Y*	Y*	Y
G3-125	3a	Y	Y	Y	Y	Y*	Y
G3-128	3a	Y	Y	Y	Y	N	Y
G3-155	3a	Y	Y	N	Y	N	Y
G3-166	3a	Y	Y	Y	Y	N	Y
G3-176	3a	Y	Y	N	Y	N	Y
G3-92	1a	Y	Y	Y	Y*	Y	Y
G1a-128	1a	Y	Y	Y	Y	Y	N

5.3 Diversity analysis

An advantage of the deep sequencing approach used within this chapter is that it also offers the opportunity to analyse the diversity of the strains present within the samples. There were two main objectives for this analysis, and they were firstly to examine the samples for evidence of multiple strains of the same genotype and secondly, to compare the diversity of strains isolated from mixed genotype-infected samples with strains from samples infected with a single genotype. Clonal analysis was used to identify the presence of multiple gt1a strains in sample G-30 (chapter 4). Whilst no gt1a strains were detected for G-30 during the deep sequencing runs, there may be multiple intra-genotype strains present in other samples. The second objective was to compare the diversity within the mixed genotype infected samples with the diversity of samples infected with a single HCV genotype. Numerous approaches were either attempted or considered for this analysis including, pair-wise analysis, phylogenetic analysis, Swarm (Tiffin and Ross-Ibarra, 2014) and CD-Hit (Li and Godzik, 2006; Fu *et al.*, 2012). Given the large number of reads detected per sample (frequently >100,000), both phylogenetic and pair-wise analysis were too time consuming and required a significant amount of computing power. The long length of these amplicons meant that after ligation of the adapters and sequencing primers, there was little overlap between the paired-end reads (figure 5-17). In phylogenetic and pair-wise analysis, this may have meant only being able to analyse read 1 or read 2 and therefore the loss of approximately half of the sequence and diversity information contained within it. The early testing of Swarm suggested the programme was more appropriate for quasispecies analysis but there was difficultly reducing the numbers of clusters identified by the programme. A detailed description of the pipeline used to analyse the samples is given in the materials and methods section (chapter 2).

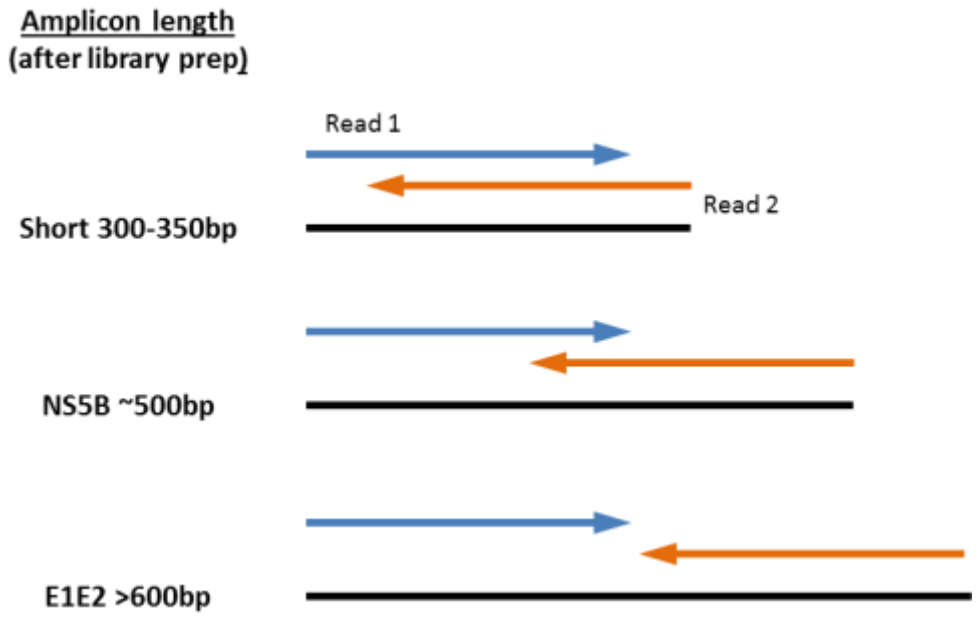


Figure 5-17; Amplicon length and read coverage

The impact of amplicon length (after library preparation) on read coverage using 2 x 300bp paired end reading. Read coverage for short fragments, the NS5B and the E1-E2 amplicon are illustrated.

CD-Hit analysis utilises a greedy incremental algorithm that works by taking the longest read in a dataset and comparing all the other reads in the dataset to it (Fu *et al.*, 2012). Reads are grouped into clusters on the basis of their similarity (user-assigned threshold) to the longest read and clusters with >500 associated reads are considered to be real variants. It should be noted for this analysis that the reads from each sample were firstly separated by genotype using Tanoti (Vattipally, 2014) and that in samples with multiple genotypes detected within them, clustering analysis has been performed on each genotype separately. Raw data from the CD-Hit analysis, after removal of the gt2 (JFH-1) and gt1 (H77) contaminating sequences, is given in the appendix, tables 8-4 (run 1) and 8-5 (run 2). A consolidated table of the data, showing the average number of clusters detected in each sample group is shown in table 5-8.

Examination of the data indicates that the analysis can be influenced by the numbers of reads produced by the samples. Particularly large numbers of clusters were observed in the 10% gt3a control (run 1) and in the gt1a NS5B and E1-E2 samples, which had the largest number of reads. An excessive amount of the 10% gt3a control (run 1) was loaded into the final sequencing pool in error and this sample produced 10-fold more reads than average in the run. On run 2, many of the twenty gt1a-only sample E1-E2 fragments were lost during the library preparation, resulting in a greater proportion of reads produced for the NS5B fragment relative to other samples. The repeated E1-E2 amplicons were prepared separately without the NS5B amplicons and consequently, the number of E1-E2 reads produced by these samples is

higher than the number produced by other samples. The increased number of E1-E2 reads in these samples is likely to have inflated the numbers of clusters detected within the samples. These factors make it difficult to draw meaningful conclusions from the gt1a-only samples as they cannot be easily compared to other samples. The control samples from both run 1 and 2, which were derived from a synthetic dsDNA fragment and should show no variation, constitute up to 30 clusters (discounting the gt3a 10% control). More clusters were also identified in controls from run 2 (particularly the gt1a transcript controls) relative to run 1. It is unclear why this should occur, as similar numbers of reads were produced for each set of controls and neither set of controls was excessively contaminated with JFH-1. PCR-deep sequencing is known to be highly error-prone as errors are introduced into the reads during the initial PCR of the amplicons, during library amplification and during the sequencing run itself. Additional errors can also be introduced when reading the sequences as strong fluorescence from many reads with identical nucleotides can mask weaker signals from other variants. Both sets of samples underwent identical library preparation however and there was no difference in the number of PCR cycles that the samples underwent. The transcript and mixed genotype transcript controls also show a similar diversity between the E1-E2 fragments and the NS5B fragments. As the E1-E2 amplicon spans HVR-1 and HVR-2, we would expect to observe higher diversity within this amplicon than in the NS5B region.

Table 5-8; Diversity analysis of the deep sequencing reads using CD-Hit

Summary statistics of a CD-Hit analysis of the deep sequencing data from runs 1 and 2 (using 0.9 similarity threshold) stratified into mixed genotype, single genotype and controls. The average number of clusters is presented for each group and the standard deviation. In mixed genotype samples and controls, only the major genotype has been analysed in this table. Many of the E1-E2 fragments were lost from the single gt1a samples in run 2 during the library preparation stage and the repeated E1-E2 libraries have been analysed separately in the single gt1a (E) group. *denotes the number of distinct clusters identified with >500 or >1000 reads within the cluster. **the large number of reads associated with the 10% mixed genotype control were confounding in this analysis and this sample has therefore been excluded.

RUN 1

	E1E2		NS5B	
Group	**>500***	**>1000***	**>500***	**>1000***
Mixed gt	13.5	7.5	11.5	6.1
SD	6.9	4.2	5.1	3.1
Single gt3	13.3	8.2	5.7	3.3
SD	6.7	3.5	7.5	4.1
Controls**	7.2	5.2	6.4	3.0
SD	1.8	1.3	3.4	1.0

RUN 2

	E1E2		NS5B	
Group	**>500***	**>1000***	**>500***	**>1000***
Mixed gt	24.0	13.0	20.0	11.3
SD	12.5	8.2	21.1	10.6
Single gt1a	2.7	1.8	24.4	12.2
SD	5.1	3.5	13.3	6.8
Single gt1a (E)	28.5	16.5	0.0	0.0
SD	14.0	7.7	0.0	0.0
Controls	14.5	7.0	18.0	9.3
SD	11.9	4.4	11.8	5.2

Examination of the data from run 1 comparing diversity indicates several interesting trends. Firstly both single genotype infection sample groups have a greater number of clusters within the E1-E2 analysis relative to the controls and the NS5B analysis, indicating that as expected, more diversity was detected within the E1-E2 reads relative to the NS5B reads. Secondly, in the NS5B analysis, more clusters were identified within the mixed infection cohort at both the >500 and >1000 thresholds (11.5 and 6.1 respectively) than were identified in the samples with a single genotype infection (5.7 and 3.3). In contrast, the numbers of clusters identified within the E1-E2 analysis were comparable for both groups of samples. This suggests that gt3 viruses, in the context of a mixed genotype infection, are more diverse than gt3 viruses that are involved in single genotype infections. Reasons for this observed difference are unclear but it may indicate either a higher exposure rate among those with mixed infections or that the gt3 viruses are subjected to the added selective pressures likely encountered in multi-genotypic infections leading to an increase in quasispecies breadth.

Caution should also be used however when interpreting the level of diversity that these identified clusters are taken to represent. It is widely accepted that distinct strains within the same HCV subtype can be confidently distinguished if they show a >10% divergence at the nucleotide sequence level whereas quasispecies diversity is usually characterised at levels of <5% divergence from the major variant(s) present. In this analysis, a similarity threshold of 0.9 was used to identify clusters of sequences showing 10% divergence from the longest read. However, given the large numbers of clusters identified in the control samples, this analysis appears to have over-estimated the diversity present within the samples. Without consensus sequences from each of the clusters, it is impossible to comment on whether these clusters indicate the presence of multiple strains or a highly divergent quasispecies or if they are simply due to the additive effect of a high error rate during library preparation and sequencing on normal levels of quasispecies divergence. Adjusting the parameters such that control samples show approximately 1 cluster may give a more accurate estimate of the number of within-subtype strains present in the patient samples.

5.4 Chapter conclusions

The regions targeted by the pan-genotypic primers in this method were selected because of the informative nature of the amplicons produced. The NS5B amplicon encompasses the Okamoto region, a small region (329nt) that is often used for HCV genotyping because of its ability to describe the entire HCV phylogeny. Phylogenetic trees of the Okamoto region have been shown to be highly similar to phylogenetic trees of entire HCV genomes, enabling consistent genotyping from a relatively small genomic region (Hraber *et al.*, 2006). Our data indicated that the NS5B amplicon produced by our assay was useful for genotyping and we were able to accurately genotype and sub-type all the samples we tested it with (n=64). The region targeted by the E1-E2 pan-genotypic primers contained the HVR-1 and HVR-2 and was therefore selected as analysis of the reads can reveal insight into the quasispecies diversity present within each sample. Regions on either side of the HVRs, particularly nucleotides 1400-1490 and 1680-1760 (absolute numbering based on H77), are however highly conserved at a sub-genotype level. The E1-E2 fragment amplified by this assay was also shown to be useful for the sub-genotypic classification of HCV strains, despite the additional variation encoded by the HVRs. Although the tree topology described by the E1-E2 and NS5B analyses were different, the clustering of the sequences alongside the reference strains was consistent for every sample tested. A systematic review by Jacka *et al*. (2013), found that the E1-E2 was rarely used in genotyping studies and that when it was used, it was frequently as part of a fragment >1000 bp in length (Jacka *et al.*, 2013). Data in this chapter indicates that when using sequencing-based genotyping, the relatively short E1-E2 region targeted by this assay can also be highly discriminatory. Additionally, the misclassification of the HCV gt1g strain as a HCV gt1c strain suggests that the region being used by diagnostic laboratories may not be the most informative region for sequence analysis. Although rates of gt1g and gt1c are very low in the UK, the misclassification of sub-types may lead to inaccurate conclusions being drawn about associated clinical morbidities and response rates to treatments. In an era when we have numerous therapeutic options available for more personalised treatments of individuals with HCV, the careful and accurate stratification of HCV genotypes will be crucial to ensure that the most appropriate treatment plan is followed.

The pan-genotypic assays and the analysis pipelines used (k-mer algorithm and Tanoti) were able to detect 75% (E1-E2) and 25% (NS5B) of the mixed genotype infections identified by

Sanger sequencing in chapter 4. Previous studies have correlated read-sequencing sensitivity with qPCR thresholds (Rosseel *et al.*, 2012; Prachayangprecha *et al.*, 2014; Thorburn *et al.*, 2015) and a similar trend was observed within this study. For both assays, sensitivity for the minor strain decreased if the qPCR results (chapter 4) indicated that the minor strain was present at levels <0.25% of the total viral load. Samples in which the minor strain contributed >2% of the total viral load were also more likely to be identified. Considerable heterogeneity between the E1-E2 minor strain sequences from samples with mixed gt1a/gt3 infection identified by deep sequencing and Sanger sequencing (chapter 4) was also observed (figures 5-14 and 5-15). This heterogeneity was not evident for the major strains and suggests that the low number of reads from which a number of the minor stain consensus sequences were derived may have been a contributing factor. This may have been exacerbated by the very poor read quality observed in the majority of sequences generated for this study (figure 5-5). For the E1-E2 amplicon, which was at the limits of maximum read length for the 2 x 300bp sequencing kits, this means both the middle and end of the sequences were likely to be either trimmed or of poor quality and this may have adversely affected the consensus sequences generated from the deep sequencing reads, particularly when they were derived from low numbers of reads. In addition to this, JFH-1 contamination was identified in a number of samples, including both mono- and mixed infected samples. Whilst not ideal, this contamination has demonstrated that amplicon-based deep sequencing pipelines can be used to detect the presence of multiple genotypes within samples. The findings indicate that the detection of multiple genotypes within a single sample is possible but that very low levels of the minor strain(s) make detection less probable. The results from this chapter also support the evidence from the qPCR results (chapter 4) indicating that in the majority of the mixed infections identified, the minor strain was present at very low amounts, with 66.7% of samples tested (n=15) having a minor strain that contributed <1% of the total viral load.

Previous studies have identified PCR-bias as a major determinant of assay sensitivity and error in deep sequencing, with rates of frequency being distorted by margins of up to 100-fold relative to the true frequency (Jabara *et al.*, 2011; McElroy, Thomas and Luciani, 2014). PCR-bias was observed in the control mixed infections which had minor strains spiked in at concentrations of 0.1%, 1% and 10%. In all samples and in both the E1-E2 and NS5B assays, the minor strain was however over-amplified relative to the spike-in and was therefore easier to detect as a result. As the bias dynamic within our assay suggests that minor strains (present at ≥0.1%) are liable to be over-amplified and therefore detectable, this again suggests that the

majority of the minor strains in the mixed infection samples are present at extremely low levels. Previous work has indicated that enzyme selection and specific primer annealing dynamics are major sources of bias in sequencing-based studies (Polz and Cavanaugh, 1998; Hong *et al.*, 2009; Aird *et al.*, 2010). The pan-genotypic primers developed for the pipeline in this study were highly degenerate in order to facilitate broad coverage and consequently are liable to be a large source of bias. For the E1-E2 primers, there were 16 sense and 96 anti-sense combinations and 48 sense and 8 anti-sense for the NS5B primer set. Certain nucleotide combinations within the primers will anneal with better efficacy relative to other combinations and are therefore more likely to result in effective amplification. Concentrations of the most effective primer pairs that are able to amplify a given sequence within the master mix will also be relatively low when degenerate primers are used and this may also impact amplification biases. For enzyme selection, KAPA HiFi is widely regarded to be the leading enzyme for Illumina library preparation because of its high fidelity and improved coverage of G+C or A+T rich regions relative to most other enzymes (Quail *et al.*, 2011). KAPA HiFi was used for all amplification steps throughout this study as a result. Early testing of the pan-genotypic assays with a non-proof reading taq polymerase indicated however that the sensitivity of the assay was curtailed by the use of a high-fidelity enzyme, as might be expected.

Many deep sequencing protocols still have relatively high error rates of around 1% (Radford *et al.*, 2012; Sims *et al.*, 2014) and are clearly highly prone to contamination. Given that our qPCR data suggests that most minor strains in mixed infections comprise <1% total viral load, many of these reads would likely have been discarded as error if 5% or 1% cut-offs were used and the mixed genotype infections would not have been detected. Viral diversity would therefore likely be underestimated in many cases using deep sequencing, unless an appropriate protocol can be developed. Until deep sequencing fidelity improves and low-frequency variants can be confidently and accurately distinguished, the data in this study indicates that amplicon-based deep sequencing methods will be inferior to genotype-specific nested PCR assays (chapter 4). The bioinformatic pipeline developed for this analysis was tailored for this project (by Sreenu Vattipally, CVR bioinformatician) and the k-mer based genotyping algorithm developed for it used 37bp k-mers. In one sample, this algorithm misclassified gt1a k-mers as belonging to gt1c, over-estimating the diversity present within the sample. HVR-1 is 27 amino acids in length and evolution and diversity within the region is rarely constrained by the viral genotype. This may suggest that phylogenetic algorithms, or

a longer k-mer, would be more accurate for genotyping than the 37bp k-mer used in this pipeline, particularly for the E1-E2 reads. Longer k-mers however increase the risk of mismatches occurring and may require more computing power than smaller k-mers.

Contamination can be particularly problematic in deep sequencing studies and can complicate interpretation in studies on diversity and mixed genotype infections. This study was affected by high levels of JFH-1 contamination, particularly in run 1. Run 1 contained the majority of the mixed infection samples which were particularly badly affected by the JFH-1 contamination. As these samples underwent more tests than other samples, they were handled more and consequently have had a higher exposure to any contaminating sources. Measures were in place to reduce the risk of contamination including clean room facilities for reagents, the separation of extraction facilities from amplification facilities and a strict clean-down protocol with diluted bleach. Despite rigorous adherence to these practices, contamination was still a serious problem in this study, and is likely a consequence of the fact that library preparations for the samples were performed on an open bench as there was a lack of alternative options. Ideally, the library preparations for a study of this nature should be processed in dedicated facilities, where the exposure of samples to possible contaminants can be controlled for. New facilities recently installed at the Centre for Virus Research (where this work was undertaken) provide a discrete facility for handling clinical samples prior to reverse transcription and for handling cDNA prior to down-stream amplification. Run 2 samples E1a-1 to E1a-20 were some of the first samples to be handled in this facility and no evidence of JFH-1 contamination was detected in these samples. If deep sequencing protocols similar to this are to be developed for use in clinical environments, the implementation of a multi-step clean room system incorporating the separation of procedures and facilities according to contamination risk will be a vital part of the quality assurance management.

An analysis of the diversity present within the samples suggested that individuals with mixed genotype HCV infections may harbour more diverse viral populations than individuals infected with just a single genotype. This is in agreement with a previous study by Toyoda *et al.*, which observed that a greater quasispecies diversity was present in samples with mixed genotype infection (Toyoda *et al.,* 1998). In this study it was however difficult to draw conclusions about whether this inferred the samples were more likely to be infected with multiple strains of the same genotype or if it indicated that the quasispecies within the mixed infection samples was more diverse. Additionally no gt1a reads were detected in sample G30

by PCR-NGS, from which two distinct gt1a strains were previously identified by PCR methods. If reads had been available for this sample or any other sample with multiple inter-genotypic strains were analysed, it would be expected that this greater divergence would be identified by cluster analysis. The result does however suggest that individuals with mixed genotype infection either have had higher HCV exposure rates relative to those with single genotype infections or that the viruses within these individuals have evolved a more diverse quasispecies. Given the high IDU rates observed within our study population (>70%) and the diverse range of ages observed in our mixed infection cohort, a significant difference in exposure rates is unlikely. The diversity observed within the controls indicated that CD-Hit was over-estimating the diversity in all the studied samples making it more plausible that the increased diversity observed in the mixed infection samples is reflective of an increased quasispecies breadth. Over-estimation of diversity in the samples may also have been associated with the poor data quality observed in this study which indicated that sequencing errors were likely to be a problem in reads generated. The detection of differences in the viral intra-genotype diversity between mixed and mono infected individuals implies that there may be immunological characteristics specific to mixed genotype infection. Previous studies have identified genotype-specific transcriptional profiles with HCV gt1 inducing an elevated type 1 IFN response relative to gt3 (M. Robinson *et al.*, 2015; M. W. Robinson *et al.*, 2015). This may suggest that the increased diversity observed during mixed infection is a viral response to increased and/or different immune pressures and intra-viral competition for resources. Data from the diversity analysis was also confounded by the fact that samples with larger numbers of reads generated larger numbers of clusters during the analysis. Future analysis of this nature should digitally normalise the numbers of reads in each sample, reducing the numbers of duplicate reads in the samples and providing a similar of number of reads from each sample for downstream analysis.

Amplicon-based deep sequencing was used for this study as it offers a vastly increased depth relative to other approaches. Despite this increased depth, the very low viral loads of many minor strains and PCR-bias have meant that the detection of mixed infections was poor relative to the Sanger-based genotype specific assays. Numerous methods to reduce PCR-bias have been developed and the inclusion of target-enrichment protocols or linear amplification strategies such as multiple displacement amplification, rather than amplicon-based approaches may have been superior methods to try (Dean *et al.*, 2002). The inclusion of such steps into a protocol would however incur longer preparation and testing turn-around, in

addition to further financial cost. When these factors are all considered, deep sequencing approaches are currently not competitive or cost-effective when compared with current Sanger-based techniques for the detection of mixed genotype infections. In the short term, the incorporation of more sequence-based genotyping methods into clinical diagnostics will improve the diagnosis of mixed genotype infections, leading to better patient stratification for clinical trials and follow-up studies as well as providing data of interest to public health bodies. The two pan-genotypic assays developed for this study were shown to be more informative than currently used typing methods and were effective at amplifying and typing gt1, gt2, gt3 and gt4 strains and one gt6a isolate. It would be of interest to test the assay against genotypes rarely seen in the UK including other gt3 subtypes, gt5 and more gt6 isolates to ascertain the suitability of this assay for clinical diagnostics in other regions. HCV genotypes 4 and 6 are particularly diverse with 14 and 24 subtypes respectively (Smith *et al.*, 2014) and it would be important to demonstrate that there is sufficient diversity in the E1-E2 and NS5B regions to resolve the typing of these entire genotypes. Data collected during the project (during assay optimisation) suggested that the primers were able to function effectively with at least 1-2 known mismatches suggesting that there are unlikely to be any significant issues amplifying these strains.

Chapter 6: Analysis of the dinucleotide biases in hepatitis C viruses and other *Flaviviridae*

6.1 Background introduction

Differences between the HCV genotypes may reach as much as 25-30% at the nucleotide level and distinct geographical and clinical associations have been attributed to the different genotypes. The implications of infection with multiple HCV genotypes are poorly understood, both in terms of the viral evolution and the clinical prognosis. It is possible that unique selective pressures are placed on the viruses during mixed infections. Previous work during this PhD identified a number of samples which had mixed gt1a/gt3a infections. In most of the mixed infections identified, gt3 was the major strain present (figure 4-14) and the reasons for this were unclear. It may be the case that gt3 strains have a competitive advantage relative to gt1a strains, enabling them to more readily adapt *in vivo* during infection and replicate faster. Reductions in both CpG and UpA motifs in RNA viruses have previously been shown to confer an increased replicative rate to other +ssRNA viruses *in vitro* (Atkinson *et al.*, 2014). In this chapter, an in-depth analysis of differences in the dinucleotide expression and codon bias between mixed and mono-infected samples, and the different HCV genotypes is presented. A previous study by Kapoor et al. has also highlighted the potential of nucleotide composition analysis for the inference of probable viral host species (Kapoor *et al.*, 2010) and the application of this approach to explore the evolutionary history of HCV is considered in this chapter. The evolutionary origins of hepatitis C virus are relatively unknown and, unlike HIV, a cross-species transmission event from a progenitor virus has not been identified. It remains unclear if a single ancestral virus developed the ability to infect humans from which multiple genotypes subsequently evolved or, similar to HIV, if multiple

ancestral viruses evolved the ability to infect humans separately and each gave rise to a different genotype.

Developing this analysis further, the dinucleotide biases of the *Flaviviridae* were analysed and factors that may be shaping observed patterns were investigated. The diverse range of hosts infected by the *Flaviviridae* and the use of arthropod vectors by a number of viruses mean that these similar viruses have had to adapt to a range of different environments and that the evidence of this should be visible within their genomes. In general, among +ssRNA viruses, the CpG and UpA dinucleotide pair frequency within the genome mimics that of their host genome, suggesting there is some form of recognition capacity within the host (Greenbaum *et al.*, 2008; Lobo *et al.*, 2009). The *Flaviviridae* encompasses a diverse family of around 70 viruses that are sub-divided into 4 genera, *flaviviruses*, *pegiviruses*, *pestiviruses* and *hepaciviruses*. The family includes a large number of important veterinary and medical pathogens and are spread via mosquito-borne, tick-borne, respiratory or parenteral transmission routes. They are found globally, infecting a wide range of both vertebrate and invertebrate species. Given that the main drivers of selection for dinucleotide pair frequencies and RSCU are thought to be either the host immune system or the viruses' ability to translate within the host, this variation in both host and transmission route make the *Flaviviridae* an interesting family to investigate dinucleotide pair frequencies and relative synonymous codon usage (RSCU) within.

6.2 Dinucleotide frequencies in mixed genotype infections

6.2.1 CpG and UpA in mixed genotype infections

Working under the hypothesis that differences in the CpG or UpA expression between the genotypes may be giving gt3 viruses a competitive edge during infection, an analysis of the CpG and UpA dinucleotide frequencies in the E1-E2 Sanger-sequenced isolates from both the gt1a and gt3a strains within the mixed gt1a/gt3 positive samples was undertaken. This data was compared with identical analyses from other gt1a and gt3a E1-E2 sequences from samples that were not associated with mixed genotype infections. The results are shown in figure 6-1.

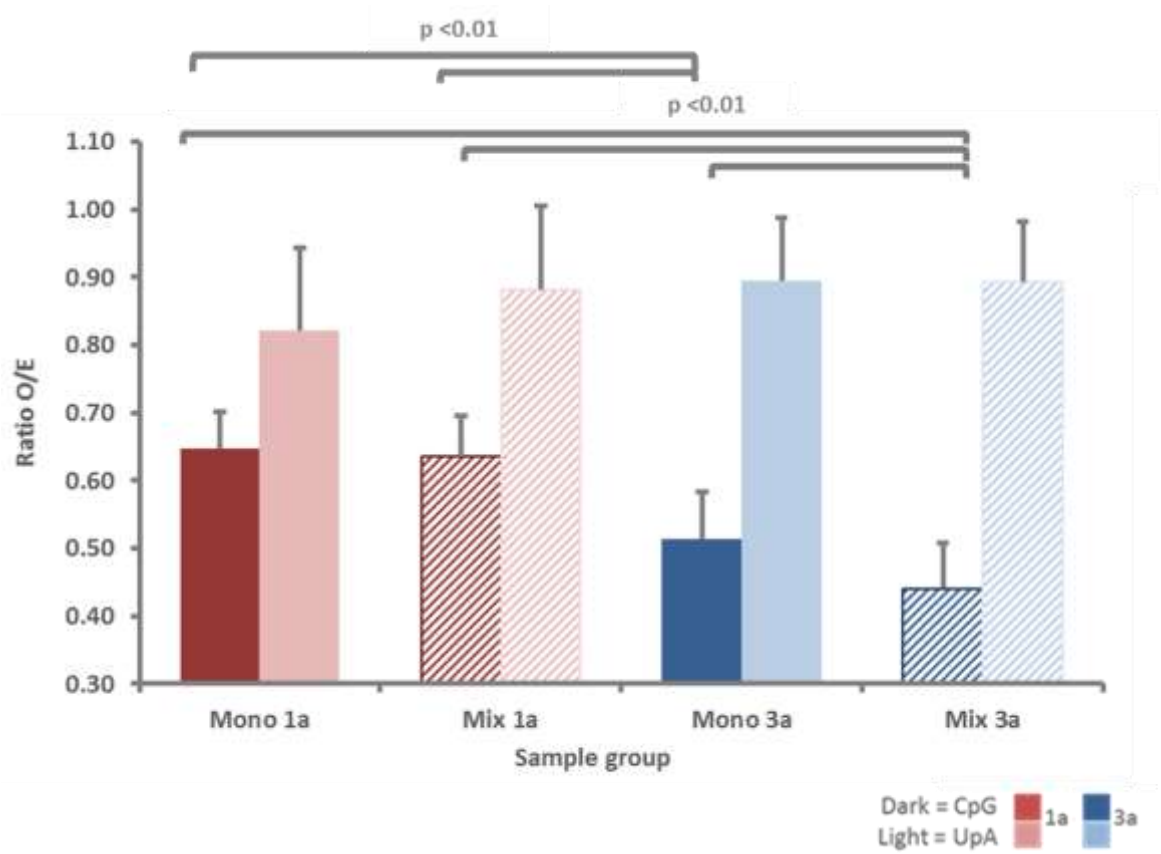


Figure 6-1; CpG and UpA expression within mixed infection positive samples

CpG and UpA expression within the E1-E2 region of gt1a and gt3a Sanger-sequenced isolates involved in both mixed (striped colour) and mono (solid colour) infections. In samples with mixed genotype infection, both the gt1a and gt3a sequences isolated from the samples have been analysed separately. Statistical significance was calculated using t-tests.

In the E1-E2 sequences isolated from mono-infected individuals, CpG levels were significantly lower in gt3a strains (0.51 ± 0.07) than in gt1a strains (0.65 ± 0.05). A similar trend was observed in the sequences isolated from the mixed gt1a/gt3 infected samples with the CpG levels in gt3 strains (0.44 ± 0.07) being significantly lower than those in the gt1a strains (0.64 ± 0.06). Interestingly, whilst the CpG expression in gt1a strains from both mixed and mono-infected samples was similar (0.64 ± 0.06 and 0.65 ± 0.05 respectively), the sequences isolated from gt3a viruses in mixed gt1a/gt3a infections expressed significantly lower CpG (0.44 ± 0.07) than gt3 sequences isolated from mono-infected samples (0.51 ± 0.07). Additionally, no differences were identified in the UpA levels between any of the sample groups analysed.

In section 4.5, qPCR assays were used to quantify the relative amounts of gt1a and gt3a virus in the mixed gt1a/3a infections. Where data was available, the estimated proportion of the minor strain present was plotted on a graph with the level of CpG and UpA expression in the gt3a E1-E2 sequences isolated from the same sample (figure 6-2). No correlations between the proportion of the minor strain and the relative levels of CpG or UpA were identified.

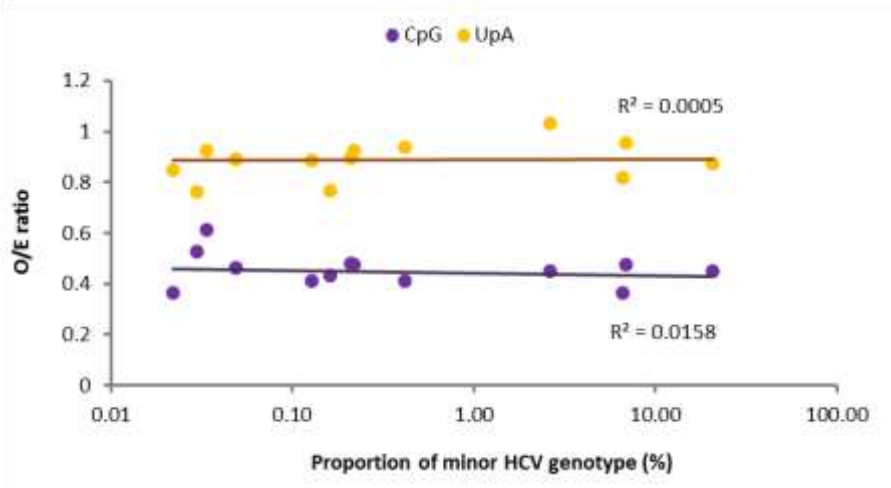


Figure 6-2; Proportion of the minor strains and their CpG and UpA expression

Correlation between the proportion of the minor HCV genotype and the O/E ratios of CpG and UpA expression in gt3a E1-E2 sequences. Trend lines are shown with calculated R^2 values.

Consensus sequences were available for gt3a NS5B sequences from 15/20 mixed infection samples that underwent deep sequencing. Dinucleotide frequency analysis was carried out on these samples and compared with an analysis of gt3a NS5B sequences (figure 6-3). A subset of the samples used for the E1-E2 analysis were used in the NS5B analysis since a number of samples had to be excluded due to having only partial genome sequences available. There were not enough gt1a NS5B sequences obtained from the mixed infection samples during the deep sequencing to enable a similar comparative analysis.

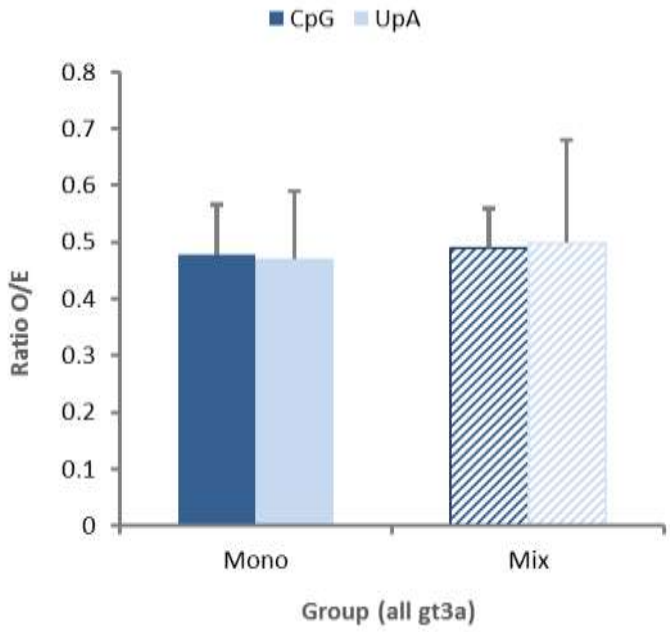


Figure 6-3; CpG and UpA expression in the NS5B region of mixed infection positive samples
CpG and UpA expression within a 390bp region of NS5B in genotype 3a sequences involved in both mixed gt3a/gt1a (striped colour) and single genotype (solid colour) infections.

Both CpG and UpA were repressed within the NS5B region analysed, with O/E ratios of <0.5 calculated for each dinucleotide pair in both groups. In contrast to the E1-E2 region however, no difference was observed between the CpG levels in the two sample groups analysed. This suggests that the decreased CpG levels observed during mixed infections may only occur within selected genome regions. There were no differences observed between the expression of CpG and UpA in the NS5B region (0.01 and 0.01 respectively) unlike in the E1-E2 region where differences ranged from 0.17 (gt1a mono infection) to 0.45 (gt3a mixed infection).

6.2.2 CpG and UpA frequency in other mixed genotype infection studies

In order to further investigate the reduced CpG expression in gt3a viruses in mixed infections, a literature search was carried out to identify publications on mixed genotype HCV infections that also published sequences from the study in Genbank. Once identified, the sequences were downloaded, aligned and the CpG and UpA levels within them were analysed. Results from the study are shown in figures 6-4 and 6-5. The accession numbers of analysed sequences are listed in table 8-1.

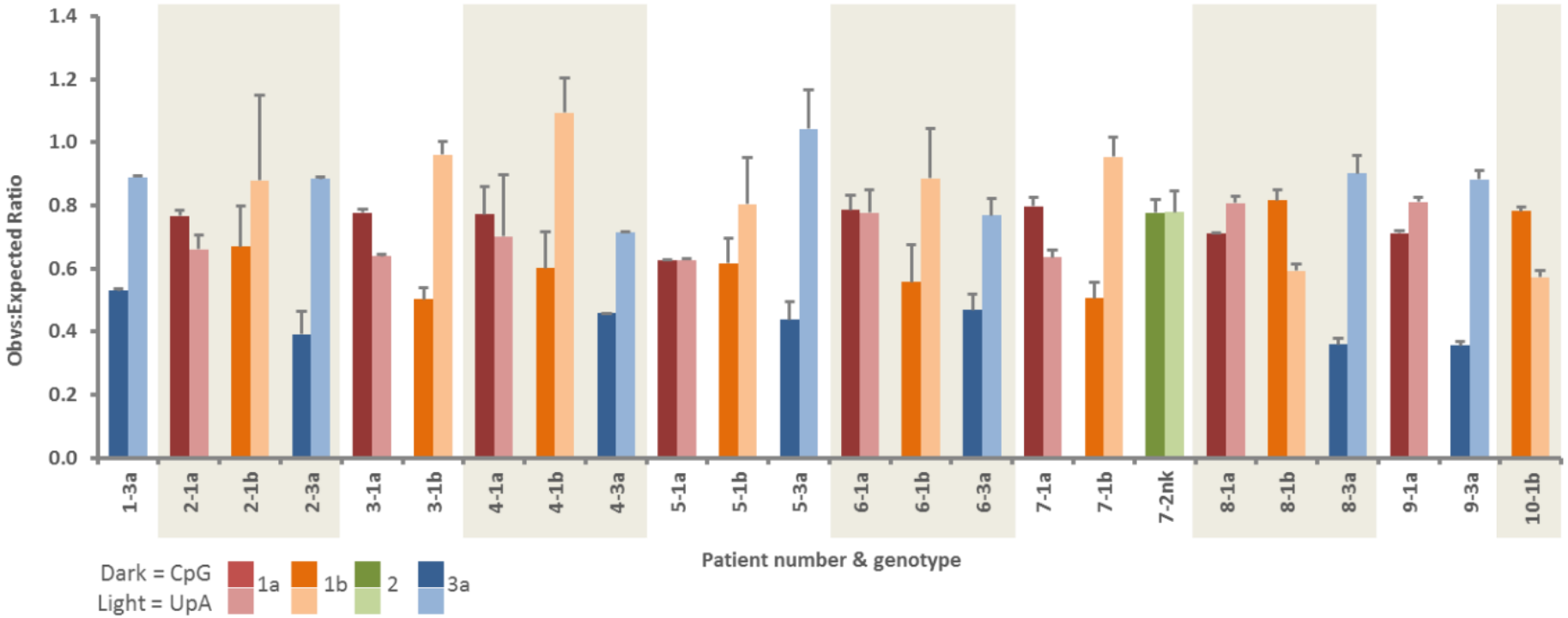


Figure 6-4; CpG and UpA expression in E2, sequences from Smith *et al.* (2010)

CpG and UpA expression in a 225bp region of E2 in patients with mixed HCV infections. Patient number is denoted before genotype and sequences were taken from Smith et al., (2010) (Smith et al., 2010).

In a study by Smith *et al.*, (2010) (Smith *et al.*, 2010), 10 patients were identified with acute mixed genotype HCV infections. Seven individuals were identified as having mixed infections involving gt3a and 6/7 of these involved at least one other genotype. An analysis of the CpG and UpA expression levels in the sequences isolated from the subjects is shown in figure 6-4. Consistent with previous observations, genotype 3a viruses were shown again to have particularly strong repression of CpG dinucleotides, having an average O/E of 0.43 ± 0.03 compared to gt1a strains (average O/E of 0.74 ± 0.03). Gt3a sequences were also the only genotype to consistently have CpG repressed to a greater extent than UpA. They also demonstrated the greatest difference between the levels of CpG and UpA expression with an average difference of O/E of 0.44 compared to 0.04 for gt1a and 0.30 for gt1b. Evidence also suggests that repression is greater in mixed infections containing at least one other genotype than in mixed infections containing multiple gt3a strains (patient 1), although there is only one example in this study. Of the 8 infections involving genotype 1a, only the sequences from subjects 8 and 9 repressed CpG to a greater extent than UpA.

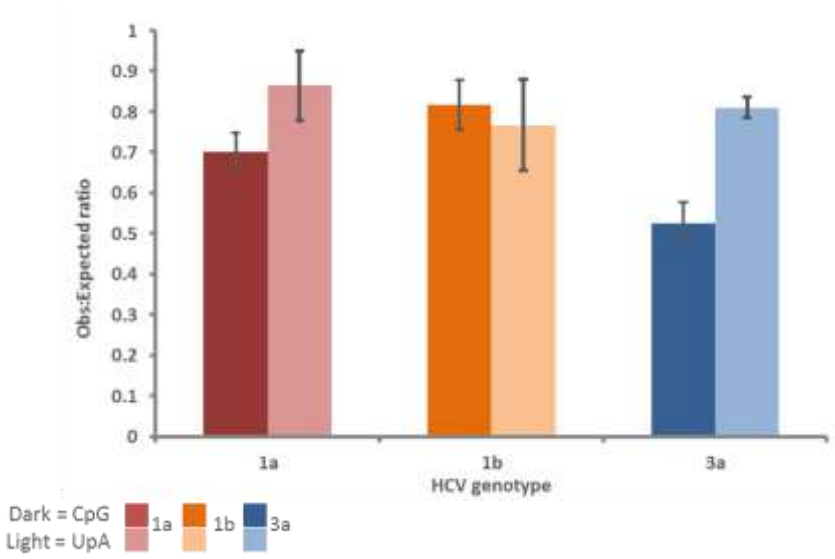


Figure 6-5; CpG and UpA expression in E1-E2, sequences from Herring *et al.* (2004)

CpG and UpA expression in a 490bp region spanning E1-E2 in samples from a cohort of injecting drug users. Sequences were published by Herring, *et al.*, (2004) (Herring *et al.*, 2004) and have been analysed by genotype.

A study by Herring, BL et al., (2004) (Herring et al., 2004) followed a cohort of injecting drug users over a period of 12 months and identified a high rate of superinfection within the cohort. A dinucleotide frequency analysis was performed on sequences published from the study and the results are presented in figure 5. Whilst it was not possible to distinguish the individual patients in this study, the gt3a strains can again be seen to have an increased repression of CpG dinucleotides (O/E 0.52 ± 0.05) relative to the other genotypes (O/E 0.70 ± 0.04 for gt1a and 0.82 ± 0.06 for gt1b). As seen previously, there was no difference in the levels of UpA expression between the different genotypes analysed.

6.3 CpG and UpA dinucleotide frequency within the hepatitis C genotypes

6.3.1 Genotypic differences

The hepatitis C viruses show a significant amount of diversity, having up to 30% sequence diversity between the genotypes at the nucleotide level. Differences in transmission associations, disease progression and response to therapy have all been documented. Given these differences, and the differences observed between gt1a and gt3 in section 6.2, it was hypothesised that there might also be differences in the relative dinucleotide expression of the other genotypes. An expanded analysis, looking at full-length HCV sequences and grouping them by genotype was undertaken. It is possible that any differences detected may improve our understanding of the evolutionary history of HCV which is still relatively unknown. All the analysed sequences are presumed to be isolated from individuals with single genotype infections. The number of sequences that were analysed is indicated in table 6-1 and the results are shown in figure 6-6.

Table 6-1; Sequences analysed during HCV genotype study

Numbers of full-length HCV sequences of each genotype downloaded from the Genbank database (October 2014) and analysed for CpG and UpA dinucleotide frequency.

HCV genotype	Number of sequences analysed
1a	373
1b and 1(other)	452
2	172
3	69
4	66
5	11
6	81
7	1

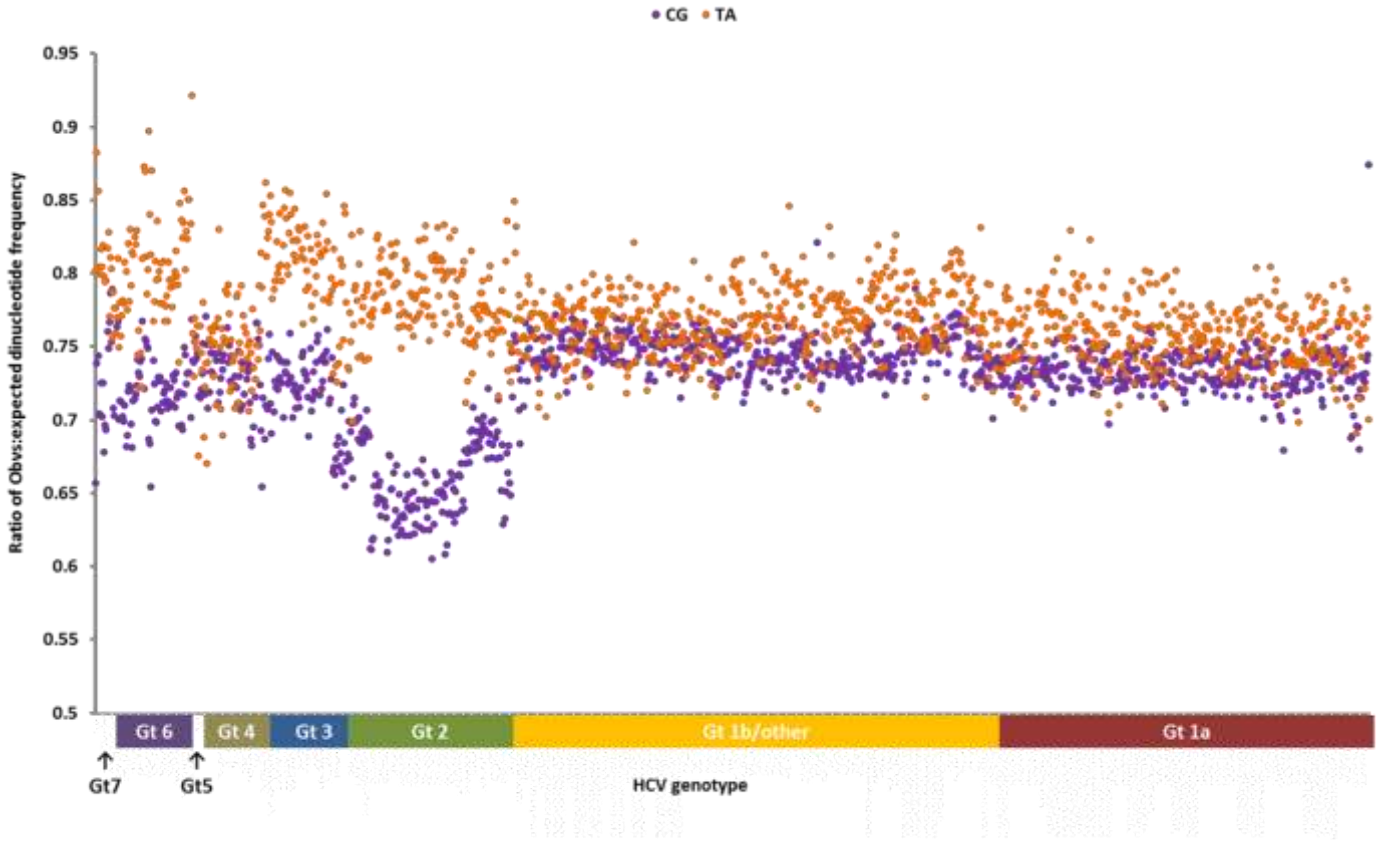


Figure 6-6; CpG and UpA expression in the HCV genotypes

O/E ratios of CpG and UpA dinucleotide frequencies in all available full-length HCV sequences downloaded from Genbank. Recombinant strains were excluded from the analysis and sequences have been grouped by genotype.

An analysis of all full-length HCV sequences available in the NCBI Genbank database indicated there was wide variation in the CpG and UpA frequency ratios between the HCV genotypes. The ratio of O/E CpG frequency averaged 0.73 ± 0.03 across the genotypes and UpA averaged 0.77 ± 0.03. In the majority of sequences, CpG was repressed to a greater extent than UpA. The lowest CpG levels were observed in gt2 sequences which were found to have an average O/E ratio of 0.68 ± 0.02. A closer look at the gt2 sequences indicated that gt2b was largely responsible for this low CpG level. HCV gt2b had the lowest levels of CpG across the genome with an average O/E of 0.65 ± 0.02. The highest levels of UpA expression were observed in gt5 and gt6 sequences although there was significant variation, particularly among the gt6 sequences. Despite there being fewer sequences available to analyse, greater variation in CpG expression was observed in gt4, gt5 and gt6 than in gt1. The average O/E for CpG in gt1a was 0.73 ± 0.01, and this was similar to the O/E calculated for gt3 sequences (0.72 ± 0.02). This indicates the large differences in CpG expression between gt1a and gt3 strains observed during mixed infections in sections 6.2.1 and 6.2.2 are likely to be region-specific. The difference between UpA and CpG expression appeared to be less for gt1 and gt4 sequences relative to the other genotypes analysed. The ratio of O/E UpA frequencies in genotype 4 was also found to be lower on average (0.75 ± 0.02) than for other genotypes. These findings correlate well with previous work by Washenberger *et al*. (2008) (Washenberger *et al.*, 2008), who found that gt2 and gt3 strains expressed greater amounts of UpA dinucleotides relative to gt1 strains.

6.3.2 Across the hepatitis C genome

Whilst the HCV genome is transcribed as a polyprotein, an analysis of the CpG and UpA dinucleotide frequency by gene may still be useful as it allows insight into the uniformity of CpG and UpA along the full length of the genome. The CpG and UpA levels across the HCV genome were analysed after grouping by viral genotype and splitting the genome up into its gene products. Results are shown in figure 6-7.

The data indicated that repression levels varied significantly throughout the genome and that differences between the genotypes were present. All the HCV genotypes showed similar trends of CpG repression across their gene products with the lowest levels of CpG expressed

within the E2 and NS4B regions. Gt2 viruses were also noted to have particularly low CpG expression in their p7 and NS2 regions. The highest levels of CpG expression were generally observed in the 5'UTR and the core region. Both these regions have been shown to have high levels of secondary structure (Mauger *et al.*, 2015) and as a result, the adaptability of these regions is constrained. In the majority of analysed regions CpG is suppressed to a greater extent than UpA, although exceptions do occur, noticeably in the core, E1 and NS5A regions. In terms of genotypic differences, the average difference between CpG and UpA repression is greater for gt2 (26%) than for either gt1a (5%), gt3a (16%) or all genotypes (7.4%). This greater difference occurs because in general, gt2 sequences had lower CpG expression across the genome and higher UpA levels relative to the other genotypes. Gt2 viruses were found to repress CpG more than UpA in 9/11 regions analysed and gt3 viruses in 8/11 regions. Among the gt1a viruses, CpG was repressed to a greater extent than UpA in 6/11 regions analysed. Gt3 viruses also had the greatest variation in levels of CpG repression across the regions with a 38% difference in the expression levels between core and NS4B. For gt1a, the maximum difference was 23%, for gt2 it was 28% and for all the genotypes it was 21%. The differences between the expression levels of UpA in different regions were similar for gt3 (46%) and gt2 (47%) viruses. Less variation in the levels of UpA was again observed in gt1a viruses (37%) and within all the genotypes (31%). Higher levels of CpG expression are observed in the E1 and E2 regions of gt3 in figure 6-7 than were observed in figure 6-1. This was a result of the fact that the sequences analysed in figure 6-1 were short regions spanning E1 and E2 (approximately nt 1200 - 1800 on H77) and did not include the full E1 or E2 sequences. This data indicates that the E1-E2 region and HVR-1 and HVR-2 have low CpG expression relative to the full E1 and E2 sequences.

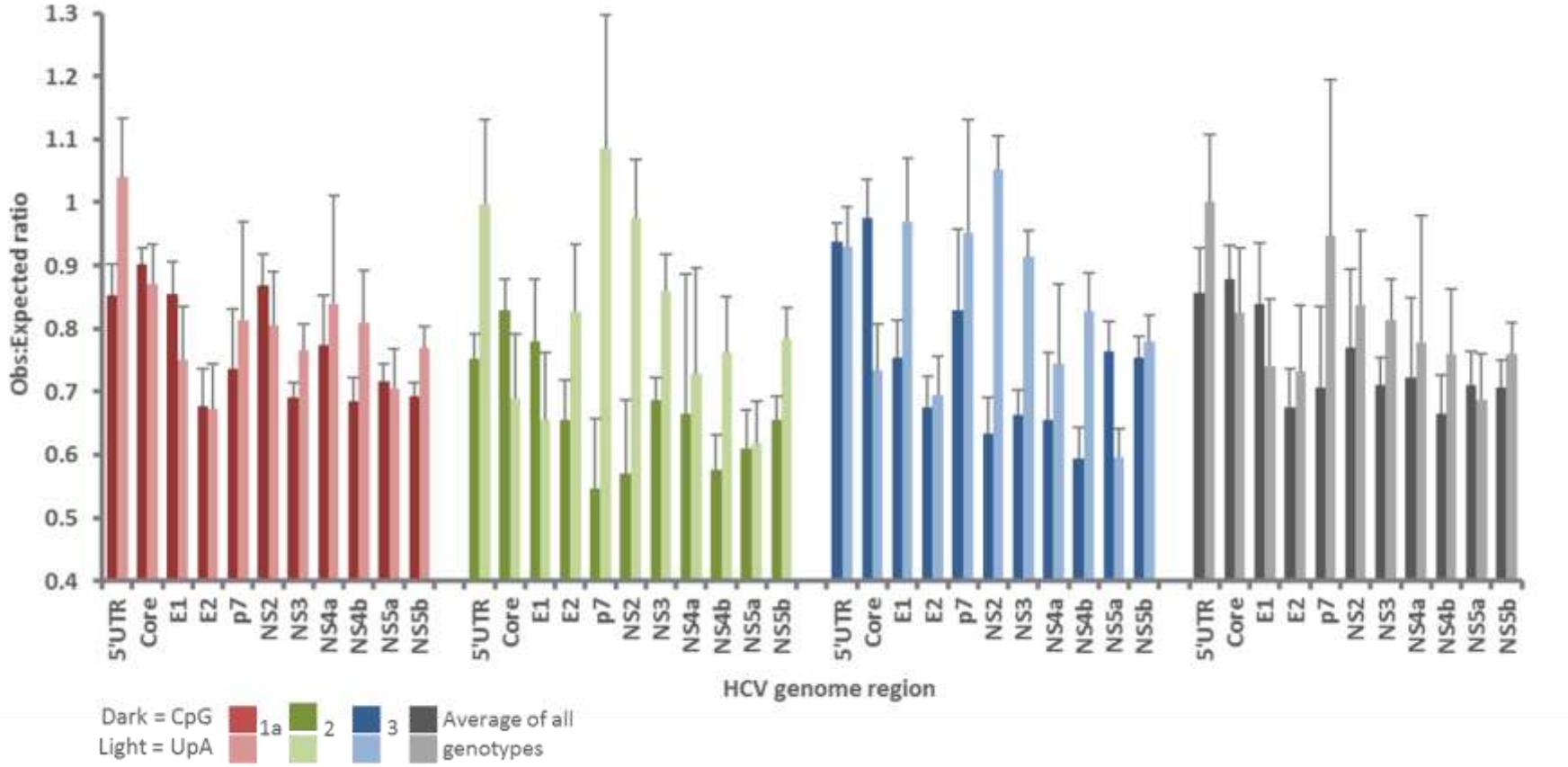


Figure 6-7; CpG and UpA expression across the HCV genome

Analysis of CpG and UpA frequencies across the HCV genome for gt1a (n=373), gt2 (n=172), gt3 (n=69) and averaged across all genotypes (n=1225). The analysis performed using all available full-length HCV sequences downloaded from Genbank (table 6-1).

6.4 Dinucleotide frequency within the *Flaviviridae*

6.4.1 Dinucleotide expression within the *Flaviviridae* genera

Full-length *Flaviviridae* reference genomes, including viruses from all 4 genera, were analysed to determine the O/E ratio of each dinucleotide pair. Results are displayed in figure 6-8. Across the *Flaviviridae*, patterns of dinucleotide pair frequency are relatively conserved with almost all the viruses repressing CpG and, to a lesser extent UpA dinucleotide pairs. Increases in other dinucleotide pair frequencies that are likely to be compensatory also appear to be conserved across the family, with TpG and CpA having O/E ratios greater than 1.2 in the majority of viruses analysed. The majority of other dinucleotide pairs clustered together with O/E ratios of between 0.8 – 1.2, suggesting there was little or no repression or selection for these combinations of nucleotides. The O/E ratio of the CpG levels ranged from 0.36 (porcine *pestivirus*) to 0.84 (rodent pegivirus). The corresponding values for UpA ranged from 0.42 (TBEV) to 0.95 (*pestivirus*). The average difference in O/E expression between the most and least expressed nucleotides was 0.75 ± 0.2. The HCV reference strains 1-6 followed the typical pattern of repression observed throughout the virus family but did not show particularly low CpG and UpA frequencies when compared to the rest of the virus family. The average difference between the O/E for the most and least expressed dinucleotides of the HCV viruses was 0.53 ± 0.04. The viruses showing the greatest difference in their expression levels (> 1.0) were Dengue virus-2, YFV and TBEV. O/E ratios within the coding regions of the human genome have been estimated to be 0.43-0.46 for CpG dinucleotides (Duret and Galtier, 2000; Witteveldt, Martin-Gans and Simmonds, 2016) and 0.51 for UpA dinucleotides (Duret and Galtier, 2000). Whilst expression levels of CpG and UpA dinucleotides among dengue viruses, TBEV, WNV and YFV are similar to those observed in human mRNA, the expression of CpG and UpA within the *hepaciviruses* and *pegiviruses* is considerably higher than that observed in humans.

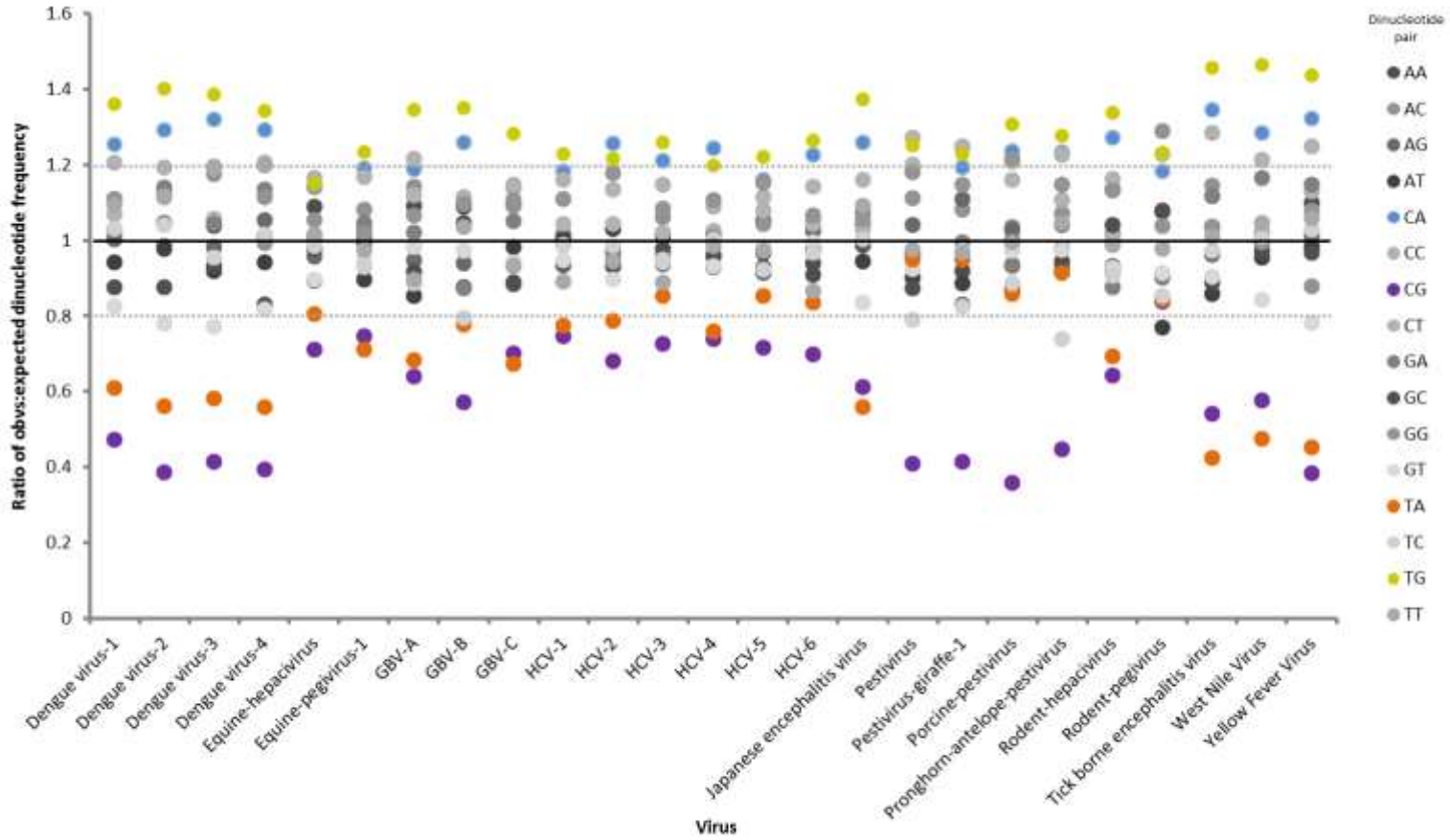


Figure 6-8; dinucleotide pair frequencies within the *Flaviviridae*

Whole genome analysis of dinucleotide pair frequencies of *Flaviviridae* reference strain sequences available from the Genbank database. The dinucleotides CpG, UpA, CpA and TpG have been highlighted in colour. Lines indicating O/E ratios of 0.8 (grey), 1.0 (black) and 1.2 (grey) have been included.

An analysis of the UpA and CpG expression patterns of 76 full-length *Flaviviridae* reference genomes indicated that the patterns of dinucleotide expression were closely associated with the phylogenetic grouping of the viruses (figure 6-9). Distinct phylogenetic groups clustered in the phylogram either on the basis of their genus or on the basis of their transmission vector. In the 56.6% of cases, CpG is repressed to a greater extent than UpA. The average difference between the UpA and the CpG O/E expression was 0.16 ± 0.17, indicating there was considerable variation of both the CpG and UpA expression levels throughout the viruses. In general, vertebrates have been shown to repress both CpG and UpA whereas invertebrates only repress UpA motifs in their genomes (Bird, 1980). Viruses infecting only insects, the insect *flaviviruses* (ISFV), showed a similar pattern to their insect hosts, repressing UpA but not CpG. The ISFV had the highest level of CpG expression with an average O/E of 0.85 ± 0.07. Similarly to work presented by Lobo *et al.* (2009) (Lobo *et al.*, 2009), viruses infecting vertebrates repressed both UpA and CpG motifs, reflecting the trends of their hosts. Exceptions were observed within the *pegiviruses* and the *hepaciviruses*, both of which had a relatively high expression of both UpA and CpG when compared with the other *Flaviviridae* and their mammalian hosts. The difference between the amount of CpG repression relative to UpA was also less for this group of viruses. Insect *flaviviruses*, insect-like *flaviviruses*, (ISFV-L) and tick-borne *flaviviruses* were also unusual as they all repressed UpA to a greater extent than CpG. The viruses showing the greatest repression of CpG were the *pestiviruses* with an average O/E of 0.38 ± 0.05. The *pestiviruses* also had the highest levels of UpA expression, with an average O/E of 0.90 ± 0.03. The expression patterns observed in the culex spp. and two aedes spp. transmitted viruses were all within a relatively similar range (O/E 0.3-0.6). Of the viruses transmitted by Culex spp., 80% repressed UpA to a greater extent than CpG whereas among the aedes-transmitted viruses only 11.8% of viruses were found to do so. This may represent evolutionary adaptations the viruses have made to their respective hosts. Interestingly, the types of infections caused by the two virus groups are clinically distinct with Culex-transmitted viruses causing mainly encephalitis-like conditions and aedes-transmitted viruses causing acute febrile fevers.

In both figures 6-8 and 6-9, large differences between the CpG and UpA expression within the HCV reference sequences for gt2, gt3, gt5 and gt6 can be observed. In the genotype reference sequences for gt1 and gt4, very little difference between the expression levels of CpG and UpA is observed. These trends within the reference sequences correlate well with the data from figure 6, analysing all full-length HCV sequences that could be identified in the

Genbank database. In figure 6-6, a low expression of CpG in the gt2 sequences can be observed and for gt3, gt5 and gt6, a higher expression of UpA is observed. The expression of both UpA and CpG is similar for both gt1 and gt4 in figure 6-6.

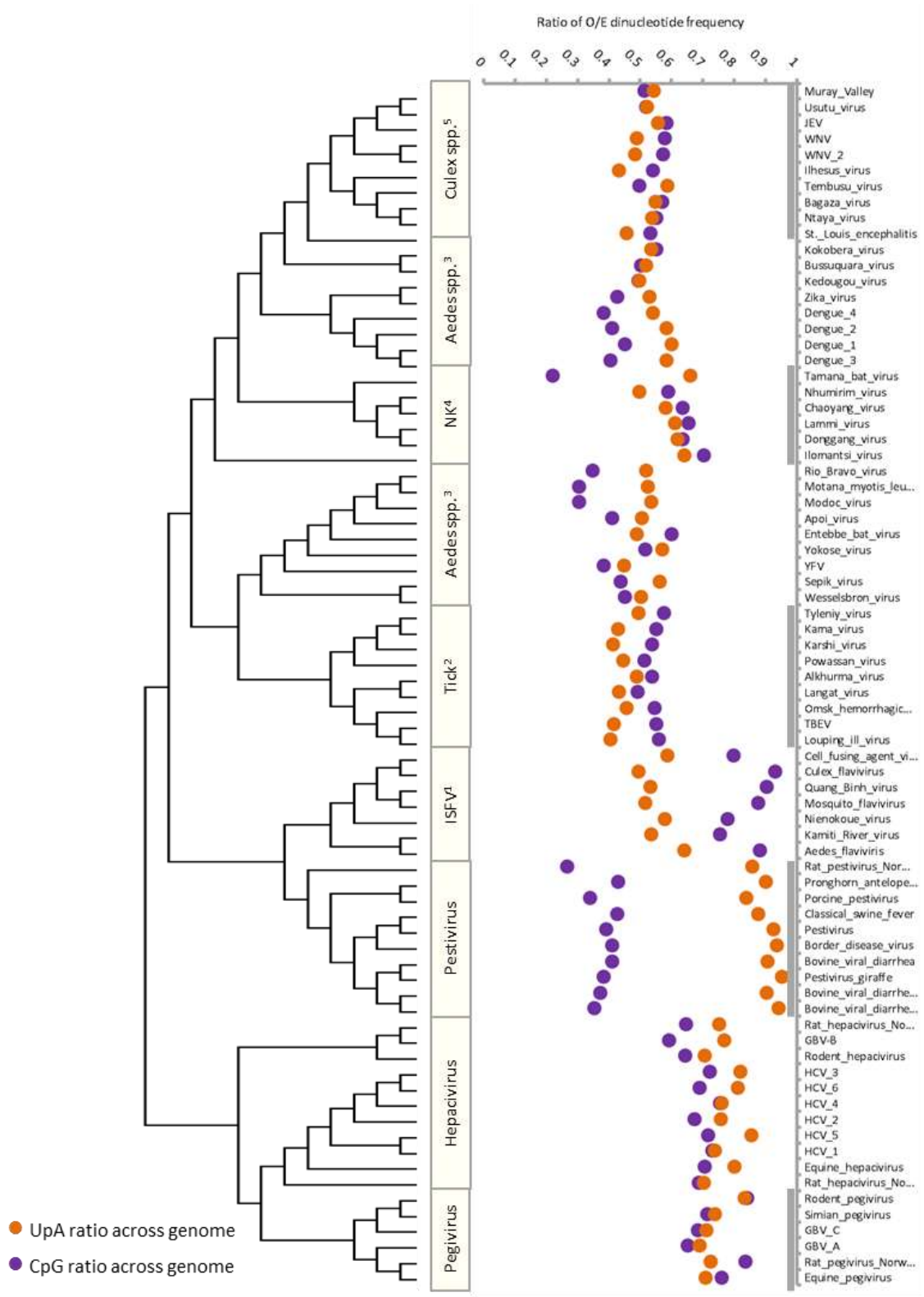


Figure 6-9; CpG and UpA expression in the *Flaviviridae*

Phylogram and graph illustrating the phylogenetic relationships of the Flaviviridae and the levels of UpA and CpG across their genomes. Strain reference sequences, available from the Genbank database, were used for the analysis. Phylogenetic ordering groups the viruses into flaviviruses, pestiviruses, pegiviruses and hepaciviruses. Using similar groupings to Moureau et al., (2015) (Moureau et al., 2015), the flaviviruses have been further grouped into (1)insect flaviviruses (ISFV), transmission by (2)ticks, (3)aedes mosquito species and (5)culex mosquito species and (4) host not known (NK). Grey bars have been used on the axis on the right to highlight the groups as indicated on the left axis.

6.4.2 The impact of viral host range within the *hepaciviruses* and the *pegiviruses*

Findings presented in section 6.4.1 indicate that viral host range may influence the relative expressions of CpG and UpA. The influence of host range was examined with an expanded analysis of hepatitis C viruses and closely related *hepaciviruses* and *pegiviruses* (figure 6-10). Phylogenetically, *pegiviruses* and *hepaciviruses* grouped together on the basis of the host they infect, suggesting that the specific adaptations required to infect and adapt to host species leaves an imprint on their viral genomes. In general, the patterns of UpA and CpG expression were reasonably consistent across the viruses, with 55.4% of viruses repressing CpG to a greater extent than UpA. Greater repression of UpA relative to CpG was more common among the *pegiviruses* relative to the *hepaciviruses*. In the majority of cases, the difference between UpA and CpG expression was not large, with the average difference in O/E for the two dinucleotides being 0.08 ± 0.08. On average, the difference between the expression of CpG and UpA was found to be similar for both the *hepaciviruses* (0.08 ± 0.06) and the *pegiviruses* (0.07 ± 0.1) although there was considerable variation within the *pegiviruses*. The difference in expression levels was greater in the *pegiviruses* and this was in large part due to the expression patterns observed in viruses infecting bats. In almost all bat *pegiviruses* analysed UpA was repressed to a greater extent than CpG. Viruses infecting bats had the highest levels of CpG expression and had the greatest difference in the expression levels of the two dinucleotide pairs.

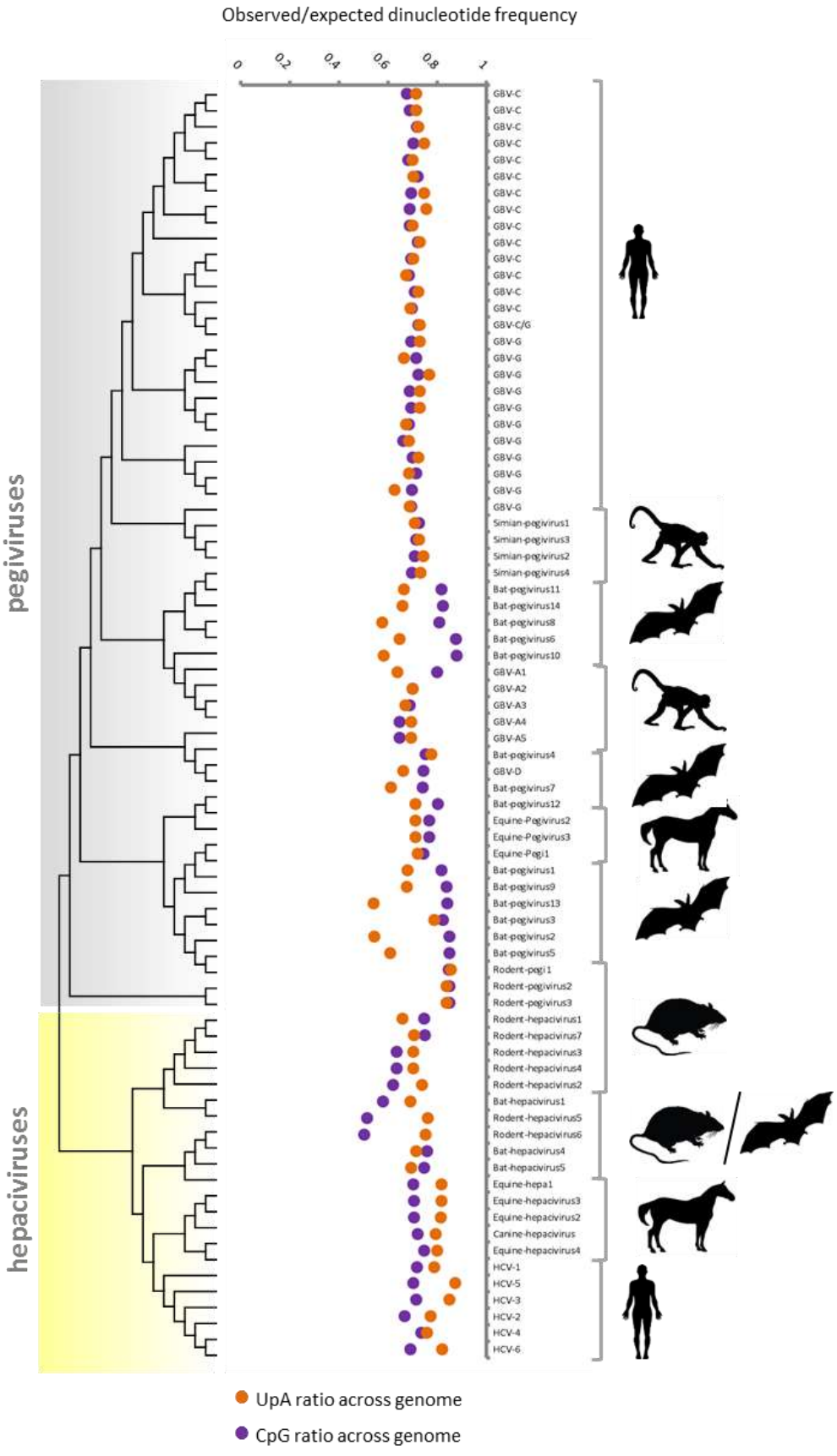


Figure 6-10; CpG and UpA expression in the *hepaciviruses* and the *pegiviruses*

Phylogram and graph illustrating the phylogenetic relationships of full-length *hepacivirus* (highlighted in yellow) and *pegivirus* (highlighted in grey) genomes and their relative expression of UpA and CpG. Phylogenetic ordering can be observed to group the viruses by host (illustrated) in the majority of cases.

6.5 Relative synonymous codon usage

There are 20 amino acids and 64 possible codons, therefore many amino acids can be encoded by multiple codons. This degeneracy in the genetic code enables organisms to select for and against certain codons. Numerous factors are thought to influence the codon use bias in organisms including translational efficiency, genetic drift and RNA stability. Selection for and against specific dinucleotide pairs within the genome will additionally impact on the codon usage of organisms. In the following sections, the synonymous codon usage within the hepatitis C virus genotypes and then within the *Flaviviridae* is analysed.

6.5.1 Relative synonymous codon usage within the hepatitis C viruses

The RSCU of 1273 full-length HCV genomes were analysed and then an average calculated for each genotype. The results are given in figure 6-11 (gt1a to gt7) and the numbers of HCV sequences for each genotype are given in table 6-1. An overall suppression of CpG and UpA-containing dinucleotides was observed and, for 12/14 of the amino acid- encoding codons, compensatory increases were observed in codons without these dinucleotide signatures. The results indicated a particularly strong suppression of UpA-containing codons in the HCV genome. A particularly strong repression of TTA (L) was identified in all the genotypes analysed. There was also a strong bias against the use of GTA (V) in all genomes, and this appeared to drive strong selection for the GTG (V) codon in its place. In the case of tyrosine (Y) where both codons contain UpA motifs, there was a marked preference among all genotypes for TAC rather than TAT. The ATA (I) codon for isoleucine (I) was also repressed in all genotypes and a compensatory preference for the ATC (I) codon was consistently observed in the sequences. Whilst repression of CpG-containing codons was also observed, it was not consistent. GCG (A) was not repressed in 5/8 of the analysed genotypes despite there being 3 other codons encoding the amino acid. A strong bias against CGA (R) was also evident in all the genotypes but CGC and CGG both appeared to be used preferentially despite containing CpG and there being two other codons available that don't include a CpG motif. All the genotypes displayed a strong repression of CpG-containing codons for proline (P) (CCG), serine (S) (TCG) and threonine (S) (ACG), and compensatory preferences for CCC (P), TCA (S), TCC (S) and ACC (T) were observed.

Differences were identified between the genotypes in the extent to which some codons were favoured relative to others. Gt1a was observed to show the strongest biases in selection for isoleucine, leucine and valine codons. Gt3 sequences showed less codon bias relative to the other genotypes. The O/E ratio for many of the CpG- and UpA-containing codons was 10-20% higher for the gt3 sequences when compared with other genotypes. Preferential selection for alternative codons was also less for gt3 sequences. This suggests that the g 3 viruses are less constrained by codon bias although the implications of this are unclear. Gt5 sequences showed similarity with the RSCU observed in genotype 3 strains although it should be noted there were only 11 genotype sequences available for this analysis.

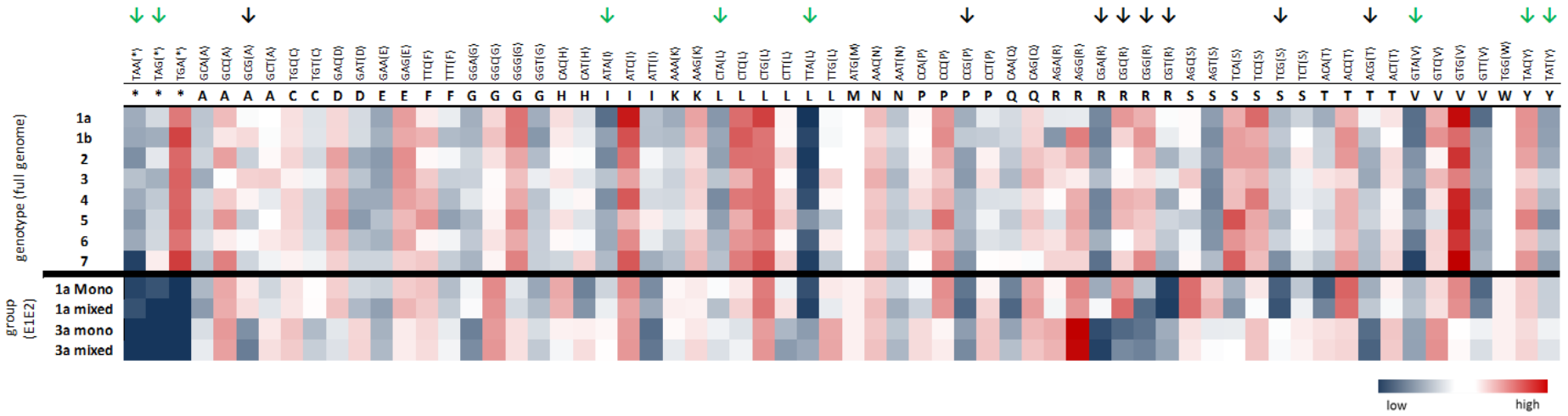


Figure 6-11; Relative synonymous codon usage of the HCV genotypes

RSCU values were determined for gt1a, gt1b, gt2, gt3, gt4, gt5, gt6 and gt7 as an average calculated from all available full-length HCV genomes (table 6-1 and table 8-1). Codons containing TpA (UpA) and CpG nucleotides have been highlighted with green and black arrows respectively. Codon use bias was also examined in the E1-E2 sequences from samples with mixed infections (mixed groups) and compared with samples of the same genotype not involved in mixed infections (mono groups). Lower than expected codon use is highlighted in blue and higher than expected codon use in red..

The second part of figure 6-11 shows the RSCU within the E1-E2 sequences isolated from individuals with mixed gt1a/3a infection compared with sequences from individuals with either a gt1a or a gt3a infection (mono groups). This is an identical sequence set to the one previously analysed in figure 6-1. In these sequences, there are marked differences between the two genotypes. Although gt3 (7/8) and gt1a (6/8) sequences demonstrated a selection bias against a similar number of CpG-containing codons, the selection bias was stronger in general for the gt3a sequences. In gt3a sequences, there was a strong selection bias for the two Arginine (R) codons that did not contain any CpGs whereas all four of the CpG-containing codons were strongly repressed. In gt1a sequences, only 3 of the 4 CpG-containing arginine codons were selected against with positive selection for CGC (R) observed. Of the other two codons for arginine, only AGG showed a positive selection bias. Gt3a strains showed a selection bias against CGC (A) which was not observed in gt1a sequences. These differences observed in the RSCU patterns between the two genotypes suggest that these viruses may adapt differently to pressure from the host immune system within this region. Sequences from individuals with mixed infections showed the same RSCU biases as those from individuals with a single genotype infection. The selection against CpG-containing codons in mixed gt3a infections was comparable to the selection biases observed in the mono-infected groups. This suggests that the difference observed in figure 6-1, whereby overall CpG is repressed to a greater extent in mixed compared to mono-infected gt3a samples, is the result of CpG repression in multiple codons rather than one particular codon. The repression of UpA-containing codons was comparable for both gt1a and gt3 with both genotypes repressing 7/8 UpA-containing codons. The exception was TAC (Y) which was actively selected for by both the genotypes in the E1-E2 region in preference to TAT, the only other triplet encoding tyrosine.

As the HCV genome is translated as a single polypeptide it is unusual that stop codons were encountered in the analysis of the full-length genomes and also that repression of the two UpA-containing codons occurred alongside an increase in the only non-UpA stop codon. All genotypes preferentially used the TGA stop codon rather than TAA or TAG which both contain UpA dinucleotides. Evidence for the presence of stop codons within the genome of a virus that is translated as a single polypeptide was unexpected. Given that the virus has a 3'UTR at the end of the genome, a single stop codon at this position would be expected but the data presented here indicating a preferential use of the TGA stop codon implies that there must be multiple stop codons within the genome. In the analysis of the E1-E2 sequences from

mixed and mono-infected individuals, all sequences displayed a strong repression of all the stop codons. Given that the sequences analysed are from a region of a virus that translates a polyprotein, we would not expect to find any stop codons within this region and the data supports this. It is possible that a number of stop codons may be present within the 5' and 3' UTRs in the full-length genome analysis as they would have a negligible effect on the translation of the genome in this position. It is also probable that a number of stop codons may be the result of sequencing errors although the preferential use of the non-UpA codon would suggest they are occurring naturally within the genome.

6.5.2 Relative synonymous codon usage within the *Flaviviridae*

The synonymous codon usage of full-length reference genomes of the *Flaviviridae* were analysed and the results are shown in figure 6-12. The results indicate that within the *Flaviviridae* there is a strong repression of codons containing either UpA or CpG, with the possible exception of TAT and TAC, perhaps because these triplets are the only ones encoding tyrosine. In viruses where there was a strong overall repression of either CpG or UpA, the codons containing those dinucleotides were strongly repressed. Likewise, in viruses where relatively little CpG or UpA repression was observed, the codon use patterns are also less biassed. In particular, strong selection against UpA containing stop codons and CpG containing arginine codons is evident throughout the family. The three most repressed codons, with an average O/E of <0.6 are TTA (L), CGA (R) and GTA (V). Amino acids with a single codon, M and W, show exactly the expected codon frequency (O/E = 1.0 +/- y), as viruses are unable to preferentially favour an alternate codon for these amino acids.

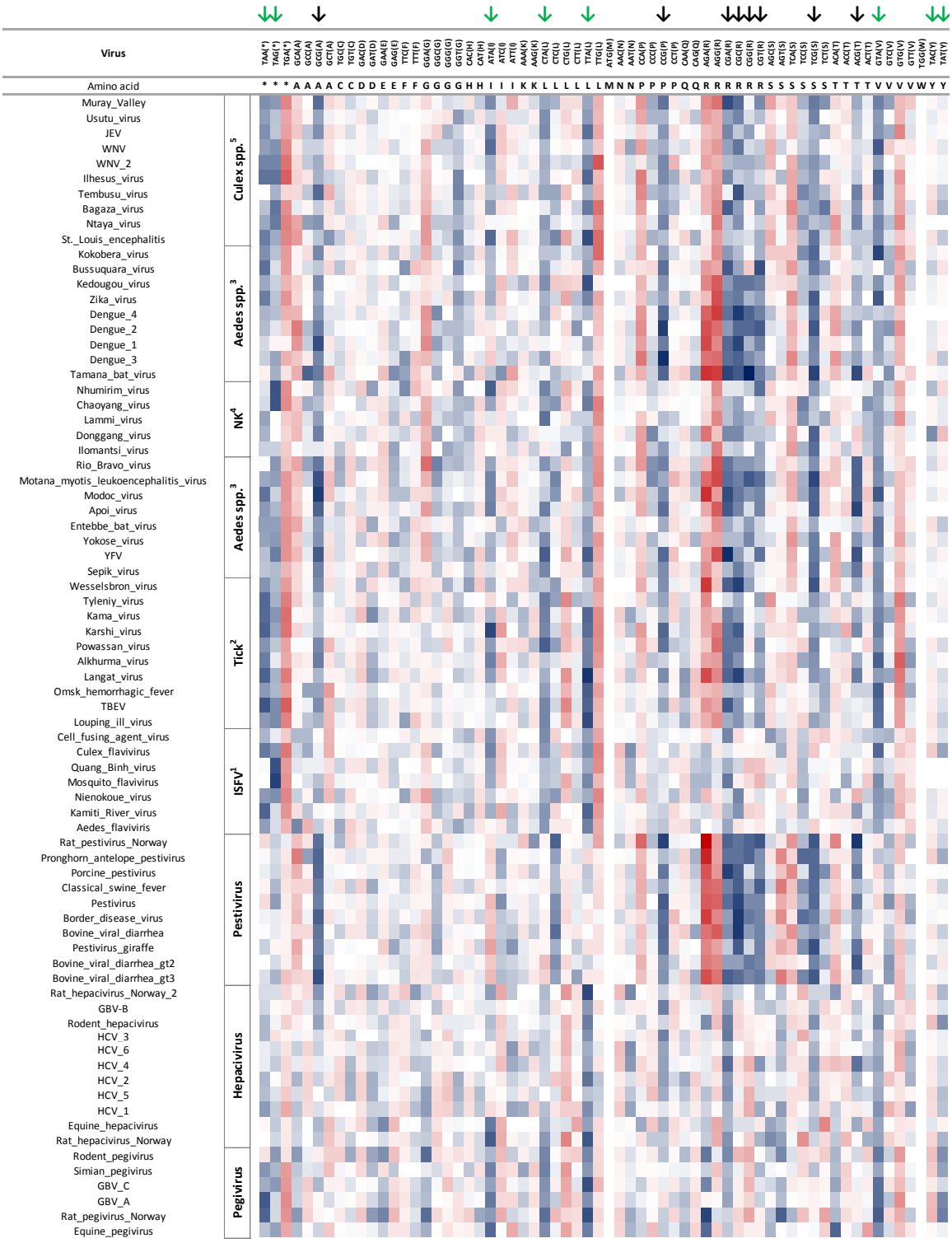


Figure 6-12; Relative synonymous codon usage within the Flaviviridae

A heat map of the RSCU within the *Flaviviridae*. Codons containing TpA (UpA) and CpG nucleotides have been highlighted with green and black arrows respectively. Lower than expected codon use is highlighted in blue and higher than expected codon use in red. Phylogenetic ordering groups the viruses into *flaviviruses*, *pestiviruses*, *pegiviruses* and *hepaciviruses*. The *flaviviruses* have been further grouped by transmission vector as ([2])tick, ([3])aedes mosquito species and ([4])culex mosquito species. Remaining viruses have been grouped as either ([1])insect *flaviviruses* (ISFV) or ([4]) host not known (NK).

Many of the observations in this study mirror those found previously in the dinucleotide frequency analysis. ISFV show a bias against UpA-containing codons but do not appear to select against CpG-containing codons. As seen previously, the RSCU of the ISFV-L viruses is also more similar to viruses infecting vertebrate hosts than it is to other ISFV. Within the *pestiviruses*, an extremely strong repression of codons with CpG is observed and this is particularly evident for the CpG-containing arginine and serine codons. Likewise, there is a strong compensatory selection for AGA (R) and AGG (R) among the *pestiviruses*. *Pestiviruses* also showed little repression of UpA dinucleotides and that is reflected here with several UpA-containing codons having O/E values >1.0. Strong biases against CpG-containing codons were also observed in the aedes-transmitted *flaviviruses*, particularly in the Dengue viruses. There is strong evidence to suggest that there is a common pattern of codon usage within the vertebrate-infecting *flaviviruses* in this analysis. All these viruses show strong biases against CpG- and UpA-containing codons. The RSCU observed in *hepaciviruses* and *pegiviruses* showed bias against UpA-containing codons but not against CpG-containing codons. The overall patterns observed within the *hepaciviruses* and *pegiviruses* were remarkably similar to those observed within the ISFV group. It is of interest to note that like *hepaciviruses* and *pegiviruses, some* ISFV have been documented to persist throughout the lifetime of their mosquito hosts (Sang *et al.*, 2003). It is probable that in order to establish these chronic infections, all these viruses groups have evolved immune escape mechanisms to allow them to persist thus reducing the viruses' reliance on mimicry of host codon usage to avoid detection.

6.6 Chapter conclusions

In this chapter, the HCV genotypes have been shown to express CpG and UpA dinucleotides at a considerably higher level than is observed within the coding regions of the human genome. Lobo *et al.*, (2009) have hypothesized that this relatively high expression of CpG and UpA may be linked to the viruses' ability to modulate both the innate and the adaptive immune responses during infection (Lobo *et al.*, 2009). Differences were also noted in the CpG and UpA expression between the different HCV genotypes, and there is an indication that these differences have evolved as a result of immune-mediated pressure on the genotypes. Further work investigating the CpG and UpA expression within the *Flaviviridae* indicated that there is wide variation in the extent of CpG and UpA repression throughout the family with patterns generally being associated with phylogenetic groupings and common host tropisms. This suppression is also evident within the RSCU patterns observed within the viral family which, depending upon the viral host, often show marked preferences for codons not containing CpG and UpA.

This study found evidence suggesting that gt3a viruses display an enhanced repression of CpG motifs in the E1-E2 region when involved in mixed genotype infections with gt1a. This observation was made using sequences from the data set presented in section 6.2 (figure 6-1) and comparable results indicating a very low expression of CpG dinucleotides in gt3 strains involved in mixed infections were found after analysing published sequences from other studies (Herring *et al.*, 2004; Smith *et al.*, 2010). The reasons for this reduced CpG expression are unclear. Reductions in both the CpG and UpA motifs in RNA viruses have been shown to confer an increased replicative rate to the viruses studied (Atkinson *et al.*, 2014). Previous work has also shown that low CpG and UpA poliovirus mutants can accumulate at greater titres in the brain (Lauring *et al.*, 2012). These studies suggest that low CpG and/or UpA may confer a fitness advantage in RNA viruses. The low CpG observed in this study in gt3a viruses may confer a replicative advantage relative to the co-infecting gt1 viruses and this correlates well with our earlier qPCR study demonstrating that gt3a viruses were most often the major viral genotype present in the mixed infections we analysed. Given that the E2 region of gt1 viruses has the unique property of protein kinase R (PKR) inhibition a key anti-viral protein (Atkinson *et al.*, 2014), changes to this sequence may result in the loss of this function in this genotype, making the virus more interferon-sensitive (Lloyd *et al.*, 2007). The gt3a E2 region is not constrained by such a function and this may explain the differences in the biases observed between the two

genotypes. It is also possible that the reduced gt3a CpG expression is a response to strong immune responses triggered by the gt1a viruses. Additionally, there is evidence to suggest that HCV viruses compartmentalise during infection with distinct viral populations detectable in the liver, serum and PBMCs among others (Navas *et al.*, 1998; Di Liberto *et al.*, 2006). It may be that reduced CpG enables gt3a viruses to compartmentalise into other tissues during mixed infections, enabling survival during mixed infection.

This study was constrained by the limited availability of sequences from samples involved in mixed genotype infections. Studies in the field are often focused on the prevalence of mixed genotype infection and therefore sequences from identified samples are rarely published. In studies where sequences have been published, there is considerable variation in the region of the virus genome that was sampled with studies using the 5'UTR (Tsatsralt-Od *et al.*, 2005), E1-E2 (Smith *et al.*, 2010) and NS5B (van de Laar *et al.*, 2009) regions to identify mixed genotype infections and in several studies, the sequences are poorly annotated making it difficult to decipher the sequences of interest. Coupled with the low number of studies, this variation in methodology makes it difficult to infer robust conclusions from the available data. In both the mixed infection samples from this research and the sequences from the literature, genotype 1a persistently co-infected with gt3a. In order to investigate if gt3a sequences also show a reduced CpG expression when in co-infection with other genotypes, further work would be required to identify individuals with such rare infections. The majority of sequences analysed in this study were taken at a single time point and previous studies have shown fluctuating viral loads over the course of mixed infections (Pham *et al.*, 2010; Grebely, Pham, *et al.*, 2012). It would be interesting to analyse dinucleotide frequencies from samples at these different time points and investigate if there is any correlation with the changes in viral load. A recent breakthrough enabling the *in-vitro* culturing of clinical HCV isolates (Saeed *et al.*, 2015) will also offer the opportunity to study mixed genotype infections *in-vitro* for the first time. In addition to facilitating the culturing of isolates from patients with mixed genotype infections, this will also make the propagation of mock mixed infections, involving the co-culture of isolates from multiple patients possible. Whilst not providing an ideal model of the *in-vivo* infection process as many hepatoma cell lines lack key immune signalling pathways, these models will allow researchers to study how competition between the genotypes occurs. It would be interesting to take regular samples from these mixed isolate cultures and look for evidence of changes in the CpG levels of the virus over time as this may provide further evidence to support the observations made in this study.

A previous analysis of the codon usage of HCV by Zhou *et al.*, (2013) demonstrated that HCV has a tendency to select against CpG- and UpA-containing codons (Zhou *et al.*, 2013). Our study demonstrated a similar trend among the hepatitis C viruses and builds on the work of Zhou *et al*. (Zhou *et al.*, 2013), by analysing a considerably larger dataset of viral sequences and analysing the different HCV genotypes separately. The study presented in this chapter found evidence of distinct differences in dinucleotide frequency and RSCU between the HCV genotypes. Previous work has suggested that the host immune system is a key driver of divergence among the HCV genotypes (Pang, Planet and Glenn, 2009). We found evidence supporting this with patterns in dinucleotide frequency and codon use among the HCV genotypes correlating negatively with interferon sensitivity. Genotype 2, which is the most sensitive to interferon-based therapies, expresses the lowest amount of CpG. The most interferon resistant genotype, gt1, and the other genotypes with intermediate interferon sensitivities express greater levels of CpG dinucleotides. Whilst these do suggest genotypic differences in interferon sensitivity, it is difficult to extrapolate if this variation in response to therapies indicates an accurate measure of *in vivo* endogenous interferon sensitivity. *In vitro* studies have suggested that CpG detection may be mediated by PKR and gt2 the is only HCV genotype whose NS5A is unable to inhibit PKR (Noguchi *et al.*, 2001). PKR inhibition in genotype 1 viruses is uniquely mediated by both the E2 and NS5A proteins, resulting in a strong resistance to interferon (Noguchi *et al.*, 2001; Lloyd *et al.*, 2007). Studies have identified increased interferon induction in patients infected with gt1 (Robinson *et al.*, 2015; Robinson *et al.*, 2015). It is possible that the repression of CpG motifs in other genotypes may reduce the levels of interferon being activated which would be advantageous given they are more sensitive to its effects. The properties of these genotypes correlate well with dinucleotide biases observed in the genotypes in this study. In the dataset analysed, however, there was an over-representation of gt1 sequences (n=825) relative to the other genotypes (n=400). The number of studies contributing non-gt1 sequences was also low for some genotypes, meaning that the diversity of the isolates analysed was low in a number of cases. This bias within the sequences may have affected interpretation of the results and the dataset should be supplemented with full-genome sequences from other genotypes when they become available. A larger number of non-gt1 sequences within the dataset could provide more diversity for this analysis and increase the support for the observations made in this study.

One of the major selective factors for dinucleotide frequency patterns among RNA viruses is thought to be the mimicry of host genomes (Greenbaum *et al.*, 2008). Previous work by

Lobo *et al*. (2009), has examined this subject in detail and the data form this study supports the findings of the 2009 study (Lobo *et al*., 2009). The dinucleotide frequencies observed in the analysed viruses generally demonstrated a good correlation with the patterns observed in their host genomes. The majority of the studied viruses ultimately infect vertebrates in which the CpG residues are frequently methylated, increasing the risk of spontaneous deamination (Kow, 2002). Relative expression of CpG dinucleotides in vertebrates is low as a result and consequently, in order to avoid invoking immune responses against their genomes, many vertebrate-infecting RNA viruses mimic their hosts. UpA expression is also less in many of these viruses although generally to a lesser extent than CpG. Invertebrate genomes do not methylate their CpG residues and they have relatively high levels of CpG expression as a result (Bird, 1980). This is mirrored in the viruses infecting them, with *flaviviruses* that only infect insects (ISFV) exhibiting strong UpA suppression but little repression of CpG motifs. Interestingly, patterns of dinucleotide expression among the *flaviviruses* with unknown hosts (NK) were markedly different with greater suppression of CpG observed relative to the ISFV. These differences were also reflected in the RSCU of the NK viruses. A number of these viruses were isolated from mosquitoes and their hosts remain unknown (Takhampunya *et al.*, 2014; Pauvolid-Correa *et al.*, 2015). Work by Kapor *et al*. (2010) has suggested that dinucleotide composition analysis can be used to infer the most probable host for some virus species (Kapoor *et al*., 2010). The conserved repression of CpG among the NK viruses suggests that these viruses are likely to infect as-yet unidentified vertebrate hosts. Viruses with ticks as intermediate vectors were also interesting, displaying a tendency to repress UpA to a greater extent than CpG. Ticks are known to have relatively long lifespans and may carry viruses for long periods of time, effectively acting as reservoirs as well as vectors. Viruses carried by them may as a result have increased adaptations to enable replication within insect hosts rather than vertebrates (Parola and Raoult, 2001). The *pestiviruses* also exhibited unusual trends, having a particularly low expression of CpG and a relatively high expression of UpA. Uniquely, *pestiviruses* have been shown to be able to inhibit the induction of type-1 IFN by dsRNA and subsequently inhibit the induction of RNAse L (Schweizer and Peterhans, 2001; Matzener *et al.*, 2009). RNAse L activation has also been shown to specifically target UpA and UpU motifs in viral genomes. As *pestiviruses* are able to inhibit activation of RNAse L regardless, there is a reduced selective pressure against UpA in these viruses. Whilst it is unclear why CpG suppression is so marked among this genus, it is possible that the suppression is associated with the host range of the viruses. The *pestiviruses* transmit directly between vertebrate hosts in herds via nasal secretions without an insect vector

(Lindenbach and Rice, 2007). Given that the *pestiviruses* are not reliant upon an insect vector for transmission, the viruses do not need to adapt to replication within insects and this may partly explain why the CpG expression is so low within the genus.

This study adds to previous work by Lobo *et al.* (2009) (Lobo *et al.*, 2009) by including an analysis of the dinucleotide frequencies and RSCU within the *pegiviruses* alongside the other three genera of the *Flaviviridae*. Both *hepaciviruses* and *pegiviruses* were shown to have higher levels of CpG and UpA in their genomes relative to other viruses within the *Flaviviridae*. All known *hepaciviruses* and *pegiviruses* infect vertebrate hosts and many have been shown to be capable of chronic infection in their relative hosts. A key requirement in the establishment of chronic infection is the modulation of both the innate and adaptive immune responses. It has been suggested that the ability of these viruses to repress immune function reduces the necessity to mimic host genomic patterns (Lobo *et al.*, 2009). Among these viruses however, viruses infecting bats were found to have marked differences relative to other viruses, with relatively high CpG and low UpA expression. Bats are an ancient lineage of mammals, having evolved separately from other mammals over 50 million years ago (Simmons *et al.*, 2008) and there are distinct differences in their immune systems relative to other mammals. In particular, bats lack the PYHIN gene family, many of which have key functions in DNA sensing and are known to encode the toll-like receptor group of proteins (Zhang *et al.*, 2013). This suggests they are lacking a receptor capable of detecting unmethylated CpG motifs, therefore viruses infecting these species are not required to select against CpG to the same degree as viruses infecting other mammalian species.

A study by Kapour *et al.* (2010) used nucleotide composition analysis to ascertain the likely hosts of three novel picorna-like viruses (Kapoor *et al.*, 2010). This type of work is of particular interest as there remains considerable debate within the field on the evolutionary origins of HCV. The data presented in figure 6-9 indicate that the CpG and UpA profiles of the *hepaciviruses* and the *pegiviruses* are much more similar to the vector-transmitted viruses than they are to other groups of viruses included in the analysis. Viruses utilising alternative routes of transmission including the pestiviruses, which are mucosally transmitted among mammals, and the insect-only flaviviruses, show markedly different UpA and CpG expression relative to the *hepaciviruses* and the *pegiviruses.* This compositional similarity to vector-transmitted viruses may indicate that the evolutionary history of both the *hepaciviruses* and the *pegiviruses* has at some point involved vector-

borne transmission phases. Research by Pybus *et al*. (2007) (Pybus *et al.*, 2007) has modelled the mechanical transmission of HCV by biting arthropods and demonstrated theoretically that it could contribute to the sustained endemic transmission of HCV. Pybus *et al*. also reason that in regions where HCV is endemic, such as West Africa and South East Asia, humans are subjected to relatively high biting rates by a wide range of arthropods, many of which are theoretically capable of facilitating the mechanistic transmission of HCV (Pybus *et al.*, 2007). The findings of this study, showing the compositional similarity of HCV with vector-borne flaviviruses indicates that further research into the plausibility of arthropod transmission is warranted. The study by Kapoor *et al*. used a more rigorous discriminatory analysis to determine the most likely host incorporating information about mononucleotide and all 16 dinucleotide frequencies (Kapoor *et al.*, 2010). It would be of interest to apply a similar analysis to a more extensive collection of *Flaviviridae* sequences and to determine if there is good discrimination between the different host ranges and transmission routes within the family. Our current knowledge of the *hepaciviruses* and the *pegiviruses* is expanding with numerous new species discovered in recent years (Pfaender *et al.*, 2014). Our understanding of the evolutionary history of these viruses is likely to be further improved by the recent discovery of human hepegivirus 1 (HHpgV-1) (Kapoor *et al.*, 2015). HHpgV-1 shares many features of both the *hepaciviruses* and *pegiviruses* and phylogenetic analysis has indicated that the virus clusters within a *pegivirus* clade that is distinct from other known human *pegiviruses* (Kapoor *et al.*, 2015). Future work incorporating an analysis of the dinucleotide frequencies and RSCU of HHpgV-1 into this study may also be informative from an evolutionary perspective.

Chapter 7: Discussion

7.1 Cohort characteristics

In order to study the prevalence of mixed genotype HCV infections in Scotland, a cohort of patient-derived samples collected from the infected population in the region was successfully curated, comprising samples and anonymised clinical data from 512 patients, all of which underwent HCV genotyping during the period August 2013-March 2014. Clinical data was available for a further 620 patients who had a HCV genotyping test during the same period and of these 1132 records, detailed clinical information was available for a subset of 590 patients residing within the Glasgow and Paisley postcode districts. Whilst there was some bias within the cohort resulting in patients from Edinburgh being underrepresented and those interacting with clinical services being overrepresented, the cohort was considered illustrative of the current HCV-infected population in the region. Linked clinical data for the study was collected from the clinical portal system, which can includes details on patient contact with secondary care providers and can therefore be considered a relatively robust data set.

An analysis of the cohort indicated that injecting drug use was the most common risk factor for infection, with 72.7% of subjects reporting a history of injecting drugs. Genotypes 1a and 3 were the most prevalent genotypes within the cohort, both infecting an estimated 45% of the population studied and almost 75% of the cohort was male. In line with a previous study by McDonald *et al*. (McDonald *et al.*, 2011), the rates of co-morbidities were high in our population. The most common co-morbidities experienced by patients within the cohort were similar to those reported in previous studies and included psychiatric disorders, trauma, stabbing or assaults, excessive alcohol intake, neurological disorders and gastrointestinal complications (McDonald *et al.*, 2011). Further analysis revealed several previously unreported trends within our cohort. HCV viral load levels were noted to be significantly lower among female subjects infected with gt3 viruses when compared with male subjects infected with either gt1a or gt3 viruses. Previous research has

suggested that immune responses against HCV among females are more effective than in males, with improved responses to IFN-based therapy (Manns *et al.*, 2001; Innes *et al.*, 2012) and higher rates of spontaneous clearance being reported in females (Wiese *et al.*, 2000; Barrett *et al.*, 2001; Micallef, Kaldor and Dore, 2006). These associations are not apparent in postmenopausal females, suggesting that oestrogen may have a protective effect although the exact mechanisms involved remain to be elucidated (Baden, Rockstroh and Buti, 2014).

Patients infected with gt3 viruses had a significantly higher prevalence of neurological disorders (13.0%) than subjects infected with gt1a viruses (7.2%). Few studies analysing HCV-associated co-morbidities have stratified their analysis by viral genotype and as a result, this finding has not previously been reported in the literature. A previous study identified increased rates of depression and anxiety in gt3 patients relative to those infected with other genotypes (Ashrafi *et al.*, 2012) however, a patient history of psychiatric disorders was recorded separately in this study and, no difference was observed in the prevalence of psychiatric disorders between the genotypes in this study. Further information on the nature of the neurological disorders experienced by individuals in this study is required to confirm the significance of this finding as several factors may be confounding. Neurological disorders can be related to excessive alcohol intake and cirrhosis and, as both were frequently reported in patient histories within our cohort, they are likely confounding factors in the analysis. A repeat of the analysis, after exclusion of individuals with a history of excessive alcohol intake may be useful in establishing the significance of this finding. If the finding remains significant then the higher prevalence of neurological disorders among gt3 patients may indicate that there is either an aspect of the gt3 viral life cycle that results in increased neurotoxic activity or, that gt3 viruses are more likely to compartmentalise into the central nervous system. Intriguingly, studies have detected negative strand HCV genomes in the brain, suggesting that extra-hepatic viral replication can occur within the brain (Radkowski *et al.*, 2002; Vargas *et al.*, 2002). Radkowski et al. (2002) also identified two patients with different genotypes in serum- and brain-derived samples (Radkowski *et al.*, 2005). The study detected gt1a in the brain-derived sample of one patient and gt3 in the other with gt1b isolated from the serum of both patients, indicating that gt3 viruses are capable of compartmentalisation into the brain during multi-genotypic infection can occur (Radkowski *et al.*, 2002).

It was intended that the clinical history of patients with mixed genotype infections would be compared with the history of those infected with a single genotype in order to investigate potential differences in clinical presentation and treatment outcomes. Partway through the study, it was realised that the specialist nature of the WWSSV meant that samples from across Scotland (excluding Lothian) were sent there and our ethical approval for the collection of anonymised clinical data only covered the Greater Glasgow and Paisley regions. Only 4/19 mixed infection positive samples were identified as belonging to patients from within the Greater Glasgow and Paisley regions therefore ultimately we were unable to draw any robust conclusions about the clinical impact of mixed genotype HCV infections. The difficulties encountered in forming cohorts of individuals infected with mixed genotype infection suitable for longitudinal studies have restricted the understanding of their prognosis. A well-planned cohort study comparing the health of those with mixed infections to those with a single genotype infection would be immensely valuable to the HCV field. Questions such as does liver disease progress faster during mixed infection, do viral loads fluctuate over time under these constraints, how responsive to DAA treatment are these patients and are there any sub-populations (aside from PWID) who are particularly prone to mixed genotype infection remain to answered.

7.2 Mixed genotype infections

Our research group has previously suggested that undiagnosed mixed genotype infections may respond poorly to some DAAs, possibly resulting in minor strain outgrowth and genotype switching (McNaughton *et al.*, 2014). Within the UK, mixed genotype infections are rarely diagnosed and therefore little is known about their prevalence. A 2010 sentinel surveillance study of hepatitis testing encompassing data from 19 diagnostic laboratories in England found that mixed genotype infections were diagnosed in just 0.2% (Brant *et al.*, 2010). Using assays developed for mixed infection screening in gt1a and gt3 patients, an estimated prevalence rate for mixed gt1a/gta3 infection of 3.8% was observed within the study cohort. Despite only concentrating on g1a and gt3 strains, these results were comparable to those previously published in the literature (table 1-2). An evaluation of the study indicated that this was likely to be an underestimate of the true prevalence within the population (section 4.6). Given that none of the cohort samples had been previously identified by the clinic as having mixed infections prior to our study, this suggests the prevalence of mixed infection is grossly underestimated in the clinic and indicates that current methods used in diagnostic laboratories are inadequate for the diagnosis of mixed

genotype infections. A further analysis indicated there was a significant disparity in the prevalence of mixed genotype infections between the two genotypes tested and that within gt3 diagnosed patients, the prevalence of mixed infection was 6.7%. The prevalence of mixed genotype infection among gt1a diagnosed patients was significantly lower at 0.8%. This disparity was identified in the pilot study for this project (McNaughton *et al.*, 2014) and remained evident after the project was scaled up. The reason for this difference is unclear and it has not been reported in previous studies. Further qPCR analysis supported this finding, indicating that gt3 was the major genotype in 80% (n=12) of the mixed genotype infections with results (n=15). In 66.6% of cases (n=10), the minor strain detected was calculated to contribute less than 1% of the total viral load. Previous studies have discussed viral competition in mixed genotype infections with many researchers concluding that usually the virus with the highest viral load eventually becomes dominant (Laskus *et al.*, 2001; Pham *et al.*, 2010; Ramirez *et al.*, 2010). Studies in HIV-positive MSM have however shown that multiple HCV genotypes can be maintained over long periods of time (Culasso *et al.*, 2014) and that super-infecting strains have the capacity to rapidly supplant pre-existing HCV strains (Chung *et al.*, 2015), suggesting that viral dynamics during mixed infection are likely to be more complex than previously posited. Well documented studies of this nature in HIV-negative individuals are rare in the literature as many PWID have chaotic lifestyles and are considered a difficult to access population for healthcare professionals. Long-term studies of individuals with mixed genotype infections involving sequential sampling and the collection of linked clinical data are needed to evaluate the viral dynamics, the clinical prognosis of subjects and the consequences of treatment in the DAA era.

Given that gt3 viruses are already considered relatively difficult to treat with the new DAA therapies (EASL, 2015), the additional impact of mixed genotype infection on treatment outcomes for this genotype should be evaluated, even if minor strains are present at low amounts. A study by Abdelrahman *et al.* (2015) looking at HCV positive patients failing treatment with pegylated IFN-α and ribavirin has shown that minor strains can persist during treatment and supplant other drug-sensitive strains, even if they are present in very low amounts (Abdelrahman *et al.*, 2015). Research by Bagaglio *et al*. (2015) has also indicated that infection with mixed genotype HCV infection can be an important prognostic indicator for treatment failure with pegylated IFN-α and ribavirin-based therapy (Bagaglio *et al.*, 2015). Studies establishing response rates of mixed genotype infections to DAAs are lacking currently, largely because current diagnostic tests often lack the

sensitivity to detect them and individuals who are diagnosed with mixed genotype infections are also frequently excluded from clinic trials. Numerous treatment options for mixed genotype infections are available, including reputedly pan-genotypic treatments such as the sofosbuvir/ velpatasvir (marketed as Epclusa) (The Medical Letter, 2016) and combinations of other DAAs approved separately for the treatment of specific genotypes (AASLD-IDSA Guidance Panel, 2016). Studies on the effectiveness of DAAs for the treatment of mixed genotype infection should be undertaken to inform on clinical guidelines. Studies of this nature will also be of interest to public health bodies, some of whom have proposed treating injecting drug users with DAAs in order to reduce transmission within the population (Hellard *et al.*, 2015). It should be considered however that the incidence of mixed genotype infection has been shown to be considerably higher in active drug users (van de Laar *et al.*, 2009; Pham *et al.*, 2010) and unless further studies are published showing these treatments are effective, such initiatives should proceed with caution to ensure resistance to DAAs is not selected for in these populations. The inappropriate treatment of mixed genotype HCV infections with DAAs may also lead to genotype switching, with DAA-resistant minor genotypes outgrowing and supplanting successfully treated major strains (McNaughton *et al.*, 2014). Without repeated genotyping after non-responses to DAA treatment, genotype switching may be incorrectly interpreted as re-infection (McNaughton *et al.*, 2014), and public health interventions treating people actively injecting drugs should be aware of this possibility.

Informative *in vivo* studies on mixed genotype HCV infections are challenging as current diagnostic assays often lack the sensitivity to detect them and many subjects infected with mixed genotype infections are PWID. Subsequently, studies are often limited by small sample sizes, often in relatively specific populations, and incomplete longitudinal data (Cunningham *et al.*, 2015). In the study of other viruses, the use of *in vitro* model systems has been highly informative however, the study of HCV within *in vitro* model systems has largely been constrained by the limited successes in culturing patient isolates in cells. The study of mixed HCV infection in *in vitro* systems has so far been challenging and studies in the field have relied heavily on the use of replicons and culture-adapted JFH-1 strains (Schaller *et al.*, 2007; Tscherne *et al.*, 2007; Webster, Ott and Greene, 2013). Recent work by Saeed *et al.*, (2015) has shown that the addition of SEC14L2 to hepatoma cell lines can facilitate the culture of a wide range of HCV genotypes isolated from patients (Saeed *et al.*, 2015). This new method could provide a suitable *in vitro* model for the study of mixed genotype infections. Previous to this breakthrough, chimeric replicon and cell culture

studies based on JFH-1 have been the most widely available *in vitro* systems used to model mixed genotype infections. Studies in this system were a poor model for *in vivo* conditions, largely because high amounts of variability in the replicative rate of chimeric replicons and virions is observed (Imhof and Simmonds, 2010), therefore inter-genotypic comparisons were not possible. In addition to this, HCV viruses are not naturally chimeric and by using chimaeras, the action of many proteins is not studied, making these poor models for mixed genotype infections. Within the new SEC14L2 cell line model (Saeed *et al.*, 2015), viruses of different genotypes can be co-cultured and the intrinsic replicative rates of different genotypes can be directly compared. An interesting study that this model system will enable is an analysis of the viral interactions observed during co-infection and super infection. Several clinical studies have observed that the strain with the highest viral load often eventually becomes dominant (Laskus *et al.*, 2001; Pham *et al.*, 2010; Ramirez *et al.*, 2010) and, using this model (Saeed *et al.*, 2015), a robust *in vitro* study is now feasible. The replicative rates and sequences of the viruses can all be easily monitored and the importance of factors such as initial viral load and order of infection can be studied in detail. The SEC14L2 cell line model (Saeed *et al.*, 2015) also has considerable applications in the testing of DAA sensitivity for mixed genotype infections. The system could be used to observe viral dynamics under selective pressure from DAAs, suggesting which treatments are likely to be most effective *in vivo*. In individuals not responsive to treatment, isolates can also be cultured and numerous drugs can be tested against the viruses in order to find the most appropriate combination for the patient.

7.3 The potential of deep sequencing for diagnostics

Samples identified as having mixed genotype infections by traditional RT-PCR methods were further analysed by deep sequencing, utilising pan-genotypic PCR assays developed for this study targeting the E1-E2 and NS5B regions that were compatible with the Illumina MiSeq platform. Initial testing indicated that both assays were sensitive and the amplicons they generated proved to be highly discriminatory for HCV viral sub-typing, with the assays providing concordant results for 64/64 samples tested. An examination of the mixed infection samples yielded E1-E2 reads for both genotypes in 75% (n=15) of the samples tested. The data indicated that minor strains were more likely to be detected if they contributed a larger proportion of the total viral load and all minor strains contributing >0.3% of the viral load (as estimated by qPCR), were detected with the E1-E2 assay. The

NS5B assay was less sensitive than the E1-E2 assay, detecting just 30% of the minor strains in the mixed infection samples. This data indicated that the deep sequencing approach was less sensitive for the detection of mixed genotype infection than the nested PCR approach used in the initial screening process. Whilst sequence-based analysis of the NS5B region is generally considered the gold standard for HCV sub-typing (Hraber *et al.*, 2006; Smith *et al.*, 2014), this study has highlighted the unexpected suitability of the E1-E2 region for sequence-based typing analysis. Probit analysis indicated that the E1-E2 assay was considerably more sensitive than the NS5B assay for gt1a strains. In the majority of mixed genotype infections detected, the minor strain was gt1a and the increased sensitivity of the E1-E2 assay for gt1a strains may be a contributing factor to the difference in detection rates. In addition, the NS5B region has extensive RNA secondary structure (Tuplin, Evans and Simmonds, 2004) and it may be that, after conversion into cDNA, secondary structure is still present, resulting in reduced primer binding and hence decreased sensitivity. Diversity analysis undertaken on the deep sequencing data also suggested that gt3a viral populations involved in mixed genotype infections were more diverse than those involved in single genotype infections. Increased viral diversity may have important clinical consequences in terms of the disease progression and treatment responses. Few studies have investigated the diversity of strains involved in mixed genotype infections and compared them to strains from single genotype infections. Findings from this analysis are in agreement with work by Toyoda *et al.*, (1996) who used fluorescence single-strand conformation polymorphism analysis to show a greater quasispecies diversity among individuals with mixed genotype infection (Toyoda *et al.*, 1998). These findings are of interest as they provide insight into how HCV viruses adapt during mixed infection. To my knowledge this is the first study to compare viral diversity in mixed and mono-infected samples using deep sequencing methods and as a result, numerous issues were encountered in finding the most appropriate bioinformatics method for such an analysis. In particular, all methods tested were not able to confidently distinguish the presence of multiple strains of the same genotype from quasispecies diversity and this made it difficult to draw robust conclusions from the study. The number of reads generated by each sample was also found to be a confounding factor with more diversity generally observed in samples with larger numbers of reads. Future approaches should digitally normalise the numbers of reads in each sample, providing a similar of number of reads from each sample for downstream analysis. As the expansion in deep sequencing techniques continues to grow, improved bioinformatic tools for an analysis of

this nature will be developed and it may be worth re-analysing this data with new methods when they become available in the near future.

HCV viral typing in diagnostic laboratories within the UK is usually classified at the genotype level with an additional test frequently used to sub-type gt1 infections into gt1a, gt1b and other gt1strains. Other genotypes are not routinely sub-typed and there appears to be no studies examining the diversity of these non-gt1 subtypes in the region. Previous studies have shown gt3 infections in the region to be largely caused by gt3a strains (Mcnaughton et al., 2015) which have strong associations with transmission by injecting drug use. Data from this study supports this and no evidence of other gt3 sub-types present in the region was found throughout this study. The genotyping panel used to test the pan-genotypic assays (chapter 5) was selected using samples from patients within the region with results from the diagnostic lab indicating that multiple gt2 and gt4 sub-types were present within the population sampled. Genotyping results with our pan-genotypic primers from the gt2 samples (n=9) identified both gt2a (n=5) and gt2b (n=4) strains and results from the gt4 samples (n=5) distinguished gt4a (n=3) and gt4d (n=2) in agreement with the diagnostic lab. Gt5 and gt6 strains are rare in the region and only a single gt6 strain was typed during the study. Non gt1 and non gt3 infections are estimated to contribute 6-11% of the total HCV infections within the UK depending upon the region (Public Health England, 2015). The epidemiology of these strains within the UK is not widely studied and the responses to DAA treatments for gt2 and gt4 sub-types are largely presumed to be consistent within a given genotype. The considerable differences observed in the genetic barriers to resistance against DAAs between gt1a and gt1b (Cento et al., 2012; Paolucci et al., 2013) strongly suggest that this is unlikely to be the case and, whilst studies have been published examining the sub-type specific responses to treatment in other genotypes (Al Ashgar et al., 2013; Chayama et al., 2015), it remains an understudied field. Further evaluation of the pan-genotypic primers developed during this project with a wider range of HCV sub-types to assess their potential for genotyping in diagnostic laboratories would be useful. It is also probable that resistance profiling by deep sequencing will become standard practice prior to DAA treatment in the near future. The regions of the HCV genome targeted by these assays should be evaluated for their potential to accurately sub-type viruses and diagnose mixed genotype infections. The development of methods for the reliable detection of minor HCV strains has applications in DAA resistance testing and treatment monitoring, in addition to the diagnosis of mixed genotype infection. The pan-genotypic primers developed for this study demonstrated that targeted primers can be

developed that are capable of amplifying a broad range of different genotypes. Similar approaches could be developed for regions where DAA resistance associated variants can be detected and used to check for either pre-existence within the quasispecies or emergence during treatment.

The major barrier to effective detection of minor strains remains the low viral loads of many resistance-associated and secondary infecting minor viruses. Recent work by Thomson et al. (2016) using a range of different deep sequencing library preparation methods found that viral loads below 1,000 IU/ml were difficult to sequence consistently (Thomson, Ip, Badhan, Christiansen, Ansari, *et al.*, 2016) and other methods developed by Wei et al. (2016) were only able to consistently identify resistance-associated variants when they comprised greater than 5% of the total viral load (Wei et al., 2016). The method developed for this project has comparable sensitivities but qPCR results indicated that many of the minor strains detected in the mixed genotype infections were present at amounts considerably lower than 5% and consequently, the detection of mixed infections was poor relative to the Sanger-based genotype-specific approach used in chapter 4. These findings suggest that sensitivity is currently limited by the technology and that until improvements in the technology emerge, deep sequencing approaches will remain not competitive or cost-effective when compared with current Sanger-based approaches for the detection of mixed genotype infections. Recently Thomson *et al*. (2016) demonstrated that numerous library preparation methods were capable of detecting mixed genotype infections during whole genome sequencing and that the methods used could accurately quantify the ratios of each genotype present (Thomson *et al.*, 2016). In addition to the limitations imposed by the low viral loads of the minor co-infecting genotypes, we were also limited in our options for this study as carrier RNA was used in the initial viral RNA extraction process and lack of additional sample prevented a second extraction from taking place. Carrier RNA is incompatible with many whole genome sequencing library preparation methods as detection of carrier RNA rather than viral RNA requires an increase in sequencing depth that is unattainable at worst or has serious cost implications. Numerous studies have had success extracting and sequencing low viral load samples by replacing carrier RNA with linear polyacrylamide during the extraction process (Malboeuf *et al*., 2013; Naccache *et al*., 2013). The incorporation of linear polyacrylamide into extraction protocols could provide a robust extraction method suitable for low-viral load samples without restricting downstream sequencing options. In this study, the use of carrier RNA meant we were limited to amplicon-based methods and substantial PCR bias was

observed in our controls, indicating that this method cannot be used to estimate the relative proportion of a minor strain present, making it unsuitable for resistant variant detection in its current form. A recent study comparing several methods of deep sequencing for HCV indicated that target-enrichment based methods could be both sensitive and affordable, and were capable of accurately detecting the ratios of mixed genotype HCV infections (Thomson *et al.*, 2016). This suggests that target-enrichment may have been a useful alternative approach for our study although the relatively high error thresholds used are likely to reduce the sensitivity of the approach. Methods involving the incorporation of target-enrichment or linear amplification strategies should be trialled and compared directly with the approach developed for this study to determine if they improve upon current sensitivity. Linear amplification strategies may improve upon current diagnostic sensitivity by increasing the number of templates but maintaining the proportion of each strain present (Hashimshony *et al*., 2012).

7.4 Dinucleotides biases of HCV and other *Flaviviridae*

Previous studies have suggested that low CpG (and/or UpA) levels in RNA viruses may confer increased replicative rates (Atkinson *et al.*, 2014) and enable viruses to accumulate at greater rates within the brain (Lauring *et al.*, 2012). A study analysing the dinucleotide frequency expression in HCV found that gt3 viruses involved in mixed genotype (gt1a/gt3) infections expressed fewer CpG dinucleotides in their E1-E2 region than gt3 strains in single genotype infections (chapter 6). No significant differences were observed in UpA expression. The qPCR analysis of our mixed infections found that gt3a viruses were the major viral genotype present in the majority of mixed infections we analysed, supporting the possibility that the low CpG expression is conferring a replicative or immune escape advantage to the gt3 viruses over the gt1a strains. The previously mentioned SEC14L2 cell line model (Saeed *et al.*, 2015) may be a useful tool for studying this phenomenon as mixed genotype infections could be generated *in vitro* and isolates of each genotype sampled and sequenced at regular intervals. A key experiment would be to investigate competition between gt1 and gt3viruses after artificially reducing the CpG expression in gt1a viruses to see if the competitive advantage of the gt3 viruses is diminished. If the viruses replicate at similar rates after decreasing the CpG levels in gt1a viruses, it would add further weight to the findings of the *in vivo* study in this project, and support the hypothesis that reduced CpG expression is a key factor in viral competition during mixed

genotype infections. It would be interesting to investigate if there are specific regions within E1-E2 where CpG dinucleotides are selected against during mixed infections and to understand how rapidly this occurs. Whilst differences in CpG expression were detected in the E1-E2 sequences analysed in this study, no differences in expression were observed in the NS5B sequences from the same samples, implying that some regions of the HCV genome may be constrained in their ability to repress CpG. Using the SEC14L2 cell line model (Saeed *et al.*, 2015), it would also be possible to evaluate the full-length of the HCV genome for altered CpG expression and investigate if other regions of the virus repress CpG during mixed infection. A recent publication by Witteveldt *et al.* (2016) has shown that reducing the CpG and UpA dinucleotides in HCV replicons and the luciferase or neomycin genes encoded within them can drastically enhance replication (Witteveldt, Martin-Gans and Simmonds, 2016). Intriguingly and contrary to what was observed during this study, genotypic differences in the response to modifications were observed, with gt3 and gt4 replicons displaying a reduced replication enhancement relative to the gt1b and gt2a replicons (Witteveldt, Martin-Gans and Simmonds, 2016). Our study identified genotypic differences in the expression of CpG dinucleotides with gt2 viruses, and particularly gt2b strains, having the lowest expression of CpG. A recent study has implicated protein kinase R (PKR) as a key mediator of CpG detection (Atkinson *et al.*, 2014) and it is notable that gt2 the is only HCV genotype whose NS5A is unable to inhibit PKR (Noguchi *et al.*, 2001). The study by Witteveldt *et al.* (2016) (Witteveldt, Martin-Gans and Simmonds, 2016) and the findings of our study suggest that the differences observed in the dinucleotide expression between the genotypes may have functional implications for the different genotypes but what they are remains unclear.

Previous research by Lobo *et al*. (2009) (Lobo *et al.*, 2009), was expanded on during this project and a study of the dinucleotide expression and RSCU of a wide range of *Flaviviridae* was undertaken, including several pegivirus strains. Throughout the family, the evidence of viral adaptation to replication within their hosts could be observed within their genomes. The pegiviruses were shown to be remarkably similar to the hepaciviruses in this analysis and both genera exhibited compositional similarity to vector-transmitted viruses. This suggests that at some point in the evolutionary history of both the *hepaciviruses* and the *pegiviruses*, vector-borne transmission was feasible, raising interesting questions about the ancestral origins of the two genera. In addition to this, a wide range of mammalian-infecting *hepaciviruses* and *pegiviruses* have been identified through increased sampling and sequencing studies (Kapoor *et al.*, 2013, 2015; Pfaender *et*

al., 2014), increasing the plausibility of a zoonotic origin for HCV. Gaps in our knowledge of the factors contributing to the sustained endemic transmission of HCV in regions such as West Africa and South East Asia still exist, with a considerable number of transmission events remaining unexplained. The mechanistic transmission of HCV by vector has been shown to be theoretically plausible (Pybus *et al.*, 2007) and the findings of this study suggest that further research should be undertaken. The increased sampling of a wide range of biting arthropods from HCV- endemic regions and screening using sensitive sequencing methodologies may be of value in exploring this theory.

The *Flaviviridae* are responsible for considerable morbidity and mortality worldwide and, in addition to HCV and ongoing endemic Dengue transmission, there has recently been outbreaks of Zika virus in South America and Yellow Fever in Africa. Other viruses in the family, and particularly the pestiviruses, are important veterinary pathogens and have a considerable economic impact in the agricultural industry. Understanding what restricts the transmission range of the *Flaviviridae* and drives immune reactions against them is therefore of particular importance for reducing the medical and veterinary burdens caused by these viruses. An improved knowledge of the immunological pathways activated by CpG dinucleotides may be of use in the development of vaccines against the viruses. Artificially increasing the levels of CpG and UpA in Polio virus has been shown to be detrimental to the viral infectivity and authors of the manuscript postulated that it could provide an innovative approach to vaccine design for a range of viruses (Burns *et al.*, 2009) and an alternative to the live and attenuated vaccines that are currently available. In this study, *hepaciviruses* and *pegiviruses* of bats were found to express relatively high amount of CpG compared to viruses infecting other mammals. Other research indicated that bats lack the PYHIN gene family, which is known to encode the toll-like receptor group of proteins (Zhang *et al.*, 2013). This adds weight to the hypothesis that the TLR proteins (particularly TLR-9) are involved in the detection of RNA with high levels of CpG motifs (Ohto *et al.*, 2015). Intriguingly research has also been published showing that PKR, which was implicated in the characteristic low level of CpG expression of HCV gt2 strains, can be activated by TLRs (Hsu *et al.*, 2004). Further dissection of the immune pathways underlying the dinucleotide expression observed within the family may be of use for vaccine development and could be particularly useful for the improvement of veterinary vaccines for the *pestiviruses*, which have extremely low CpG expression.

7.5 Summary

Previous to the development of DAA therapies, the standard of care therapy for HCV was a non-specific treatment consisting of pegylated IFN-α and ribavirin, regardless of genotype. The introduction of DAAs, many of which are more efficacious in certain genotypes, has initiated a need to invest more research into understanding the differences between the genotypes. Over recent years studies have emerged showing different pathologies and distinct transcriptional upregulation patterns induced by gt1a and gt3 strains (Robinson *et al.*, 2015; Robinson *et al.*, 2015). Gt3 viruses have also been associated with an accelerated progression to fibrosis (Bochud *et al.*, 2009; Larsen *et al.*, 2010) and higher rates of HCC relative to other genotypes (Nkontchou *et al.*, 2011). These findings suggest that gt3 viruses are causing more damage when infecting hepatocytes relative to other genotypes. This raises the intriguing consideration that perhaps gt3 viruses are less well adapted to replication within the liver than other HCV genotypes. Gray *et al.* (2012) have previously stated that the complex viral population structures observed within HCV-infected individuals most likely arise from multiple sub-populations of HCV, including populations from both hepatic and extra-hepatic sites of replication, as other models are not sufficient to explain the diversity observed (Gray *et al.*, 2012). The data from this project indicated that during mixed genotype infection, gt3 was the most often the dominant population within the serum. Samples from other sites, including PBMCs and the liver, were not available to us for testing and it would have been interesting to investigate if the proportions of the genotypes were similar at each site. A study by Radkowski *et al*. (2002) has published evidence of the compartmentalisation of HCV genotypes into discrete sites during multi-genotypic infection (Radkowski *et al.*, 2002),. Our study also identified significantly more neurological disorders occurring in gt3 infected individuals. A recent publication has shown that distinct HCV lineages can be isolated from the CNS of individuals with cognitive impairment (Tully *et al.*, 2016).Whilst a larger study is required to confirm the association of neurological disorders and gt3, it is interesting to consider if this is evidence that gt3 viruses are more adept at replication within the CNS than gt1a viruses. Little was known about the nature of the neurological disorders experienced by individuals in this study and it would be advantageous to gain more insight into this topic.

The 2015 Public Health England Report has also documented that UK hospital admissions and mortality from HCV -associated end-stage liver disease and HCC within the UK have

increased dramatically over the past decade (Public Health England, 2015). The effective treatment of HCV, in addition to clearing the virus, has been shown to reduce mortality from all causes, reduce the co-morbidity burden and improve the overall quality of life in those achieving an SVR (Spiegel *et al.*, 2005; Backus *et al.*, 2011; Simmons *et al.*, 2015). Increased availability and access to the DAAs targeting HCV has resulted in the treatment of HCV with these drugs becoming increasingly feasible although clearance rates for individuals infected with non-gt1 (and particularly gt3) viruses remain unsatisfactory, especially when the prohibitive financial cost of the new DAAs is considered. In the UK, and particularly in Scotland, where gt3 is responsible for an estimated 45% of infections, this poses a significant challenge to clinicians trying to treat patients infected with HCV. From previously being considered the genotype with the most favourable treatment responses to PEG-IFN-α and Ribavirin, gt3 viruses are now considered the hardest to treat with the new regimens available (Goossens and Negro, 2014). Understanding contra-indications to treatment, improving the stratification of patient groups and developing our knowledge of subtle differences between the genotypes and sub-types will all be key factors in improving our capacity to successfully treat non-gt1 HCV infections. Current genotyping methods used within diagnostic laboratories often discriminate between gt1a, gt1b and other gt1 isolates but the characterisation of other genotypes remains limited to the genotypic level. Sub-type discrepancies in treatment efficacy have also been documented with a phase II trial of Ombitasvir/Paritaprevir/Ritonavir in gt2 patients noting an SVR_{24} rate of just 27% among gt2b patients compared to an SVR_{24} rate of 90% in gt2a patients (Chayama *et al.*, 2015), indicating this an area that warrants further research. Additionally, the efficacy of DAA regimens for the treatment of mixed genotype infections also remains poorly studied. There is concern in the field that ineffective treatment of undiagnosed mixed genotype infections may result in genotype switching (McNaughton *et al.*, 2014). The increased prevalence of mixed gt1a/gt3 infections observed in gt3 patients suggests that gt3-diagnosed individuals are likely to be disproportionately affected by genotype switching during DAA treatment, further complicating the treatment of individuals diagnosed with gt3 viruses. Ensuring the accurate diagnosis of mixed genotype infections, and developing more discriminatory methods HCV genotyping and sub-typing will aid the identification of groups that may be difficult to treat or are likely to fail treatment is therefore increasingly important. Sequencing-based diagnostic methods are increasing feasible in clinical virology and investment in such methods should be encouraged.

Chapter 8: Appendices

Table 8-1; Viral sequences

Viral sequences used for analysis throughout the project and the study they were used for.

Study used for	Figure	Accession numbers
CpG and UpA analysis, codon bias analysis (*Flaviviridae*)	Figure 1-1 Figure 6-8 Figure 6-9 Figure 6-12	NC_023439.1, NC_023424.1, NC_027998.1, NC_001474.2, NC_002640.1, NC_001563.2, NC_001477.1, NC_001475.2, NC_009942.1, NC_012812.1, NC_012671.1, NC_012532.1, NC_012533.1, NC_012534.1, NC_012735.1, NC_026797.1, NC_026624.1, NC_021154.1, NC_021153.1, NC_004102.1, NC_003996.1, NC_003678.1, NC_003675.1, NC_003676.1, NC_003679.1, NC_002657.1, NC_000943.1, NC_001837.1, NC_002032.1, NC_001809.1, NC_001655.1, NC_002031.1, NC_001672.1, NC_001564.1, NC_001461.1, NC_001437.1, NC_024889.1, NC_012932.1, NC_025679.1, NC_025677.1, NC_025673.1, NC_025672.1, NC_024017.1, NC_024018.2, NC_024806.1, NC_024805.1, NC_024377.1, NC_024299.1, NC_021069.1, NC_023176.1, NC_018705.3, NC_009029.2, NC_015843.2, NC_020902.1, NC_018713.1, NC_009028.2, NC_009026.2, NC_008604.2, NC_009824.1, NC_009827.1, NC_009823.1, NC_009826.1, NC_007580.2, NC_008718.1, NC_008719.1, NC_016997.1, NC_006947.1, NC_006551.1, NC_005064.1, NC_005062.1, NC_005039.1, NC_004355.1, NC_004119.1, NC_003690.1, NC_003687.1, NC_003635.1, NC_001710.1, NC_017086.1
CpG and UpA analysis (*hepaciviruses* and *pegiviruses*)	Figure 6-10	NC_020902.1, KC410872.1, NC_024377.1, NC_021154.1, KC815311.1, KF234530.1, KF234529.1, KF234528.1, KF234527.1, KF234526.1, KF234525.1, KF234524.1, F234523.1, KF234522.1, KF234521.1, KF234520.1, KF234519.1, KF234518.1, KF234517.1, KF234516.1, KF234515.1, KF234514.1, KF234513.1, KF234512.1, KF234511.1, KF234510.1, KF234509.1, KF234508.1, KF234507.1, KF234506.1, KF234505.1, KF234504.1, KF234503.1, KF234502.1, KF234501.1, KF234500.1, KF234499.1, KC796073.1, KC796093.1, KC796089.1, KC796088.1, KC796087.1, KC796086.1, KC796085.1, KC796084.1, KC796083.1, KC796082.1, KC796081.1, KC796080.1, KC796079.1, KC796076.1, KC618401.1, KC618400.1, KC618399.1, KC618398.1, NC_001837.1, U94421.1, NC_001710.1, JN127373.1, HQ331235.1, HQ331234.1, HQ331233.1, GU566734.1, U22303.1, AF023425.1, AF023424.1, AY196904.1, AF309966.1, AF104403.1, U94695.1, AY949771.1, D87255.1, D90601.1, D90600.1, U75356.1, U63715.1, AF121950.1, AF070476.1,

		AF081782.1, AF031829.1, AF031828.1, AF031827.1, AF006500.1, AB008336.1, AB008342.1, AB003293.1, AB003292.1, AB003289.1, D87263.1, D87262.1, AB013501.1, AB013500.1, AB018667.1, AB021287.1, AB003291.1, AB008335.1, D87715.1, D87714.1, D87713.1, D87712.1, D87711.1, D87710.1, D87709.1, D87708.1, AB003290.1, AB003288.1
CpG and UpA analysis (complete HCV genomes)	Figure 6-6 Figure 6-7 Figure 6-11	EU256028, EU256029, JX463525, JX463555, JX463556, JX463569, DQ889268, DQ889269, DQ889270, DQ889271, DQ889272, DQ889273, DQ889274, DQ889275, DQ889276, DQ889277, DQ889278, DQ889279, EU781751, EU781800, EU781787, EU781794, EU781795, EU781786, EU781782, EU781804, EU781789, AY956468, AY956464, AY956469, AY956466, AY956463, AY956465, EU255999, GQ149768, EU155288, EU155291, EU155294, EU482876, EU155295, EU155296, EU155297, EU155298, EU482834, EU155319, EU482835, EU155323, EU239716, EU155276, EU256008, EU256010, EU482836, EU256016, EU256018, EU482887, EU482868, EU482869, EU155241, EU529679, EU482871, EU256025, EU529680, EU256095, EU256026, EU256030, DQ430812, DQ430811, JX463526, JX463527, JX463528, JX463529, JX463534, JX463535, JX463557, JX463559, JX463560, JX463538, JX463539, JX463540, JX463541, JX463546, JX463571, DQ889283, DQ889284, DQ889286, DQ889287, DQ889288, DQ889289, DQ889290, EU781803, EU781752, EU781783, EU781750, EF032883, EU781768, EU781777, EU781774, DQ889280, EU781790, DQ889281, EU781760, DQ889282, EU781780, EU255996, EU255931, EU255997, EU256086, EU255933, EU155299, EU155309, EU862840, EU155311, EU155313, EU239715, EU155277, EU482865, EU862841, EU256020, EU256021, EU256022, EU256023, EU256056, EU660387, EU256027, EU255938, EU155245, EU256032, EU155246, DQ430814, DQ430813, JX463530, JX463536, JX463537, JX463558, JX463561, JX463562, JX463563, JX463564, JX463565, JX463566, JX463567, JX463568, JX463542, JX463543, JX463544, JX463547, JX463548, JX463549, JX463570, JX463572, JX463573, JX463574, JX463575, JX463576, JX463551, JX463552, JX463582, JX463583, JX463584, JX463553, JX463554, EF032884, DQ889291, EU660383, EU155213, EU255993, EU155215, EU255995, EU260396, EU862828, EU256039, EU482837, EU256040, EU482878, EU155310, EU482884, EU482840, EU482841, EU529676, EU155314, EU256009, EU482847, EU256011, EU256013, EU155239, EU256014, EU256015, EU155240, EU155242, EU256094, EU482870, EU256035, EU256036, EU256037, EU256038, JX463531, JX463532, JX463533, JX463585, JX463613, JX463614, JX463545, JX463550, JX463577, JX463578, JX463579, JX463580, JX463581, JX463624, EF032885, EU155247, EU260395, EU255994, EU155214, EU255932, EU569723, EU862829, FJ024282, EU256041, EU256042, EU256043, EU155265, EU256051, EU256104, EU482843, EU256004, EU660384, EU256017, EU529678, EU256058,

EU155243, EU482848, EU155244, EU256033, JX463586, JX463587, JX463588, JX463615, JX463616, JX463596, JX463597, JX463598, JX463599, JX463600, JX463625, JX463626, JX463627, JX463607, JX463608, FJ205869, EU862830, EU155248, EU155249, EU155250, EU155251, EU256060, EU482838, EU155252, EU482842, EU155266, EU155267, EU256046, EU256047, EU256048, EU256049, EU155268, EU256050, EU155269, EU155270, EU862831, EU155271, EU256067, EU256052, EU155272, EU256068, EU155273, EU155274, EU255939, EU155275, EU256002, EU482882, EU255940, EU482844, EU155278, EU482845, EU256003, EU155233, EU482861, EU155236, EU256005, EU256006, EU255934, EU482862, EU482863, EU155237, EU482864, EU482866, EU482867, EU256007, EU255936, EU256012, EU529677, EU155282, EU255941, EU155283, EU255937, EU256057, EU256024, EU256031, EU256034, JQ061779, JQ061780, JQ061781, JQ061782, JQ061783, JQ061784, JQ061785, JQ061786, JQ061787, JX463589, JX463590, JX463591, JX463617, JX463619, JX463594, JX463620, JX463601, JX463602, JX463603, JX463604, JX463628, JX463629, JX463630, JX463631, JX463632, JX463633, JX463634, JX463635, JX463636, JX463637, JX463609, JX463610, JX463638, JX463611, JX463612, FJ024087, FJ181999, FJ024274, FJ024275, FJ205867, FJ024276, FJ182000, FJ024278, FJ205868, FJ182001, FJ024280, FJ024281, FJ390399, EU255935, EU155238, EU255942, JX463592, JX463595, JX463621, JX463622, JX463623, JX463605, JX463606, JX463639, JX463640, FJ390394, FJ410172, FJ390395, AY651061, EU781829, EU781832, AB691953, EU781830, FN435993, AF054250, , D85516, AF356827, D45172, D30613, M96362, EU781827, EU781826, EU781825, AB779679, AB779562, AB080299, AF054251, AF054259, AF054252, AF054248, AF054253, AF054254, AF054249, AF054255, AF054247, AF054256, AF054257, AF054258, AF313916, EU781828, D13558, D10750, EU781831, , AJ000009, EF032892, EF032893, EF032894, EU155368, EU155359, EU529682, EU256080, EU256081, EU155366, EU155367, EU155369, EU482859, EU155356, EU155357, EU155360, EU482860, EU256084, EU256085, EU155358, EU256075, EU256076, EU155361, EU155362, EU862837, EU155372, EU256077, EU256078, EU155363, EU256079, EU482874, EU256082, EU256083, EU155370, EU155371, EU155373, EU155374, EU482875, EU155364, EU155365, EU155375, EU155376, EU255960, EU155377, AY587016, D10934, AY460204, L02836, EF638081, HQ639940, HQ639937, HQ639946, HQ639947, EU857431, KC844052, KC844051, GU451218, GU451221, GU451222, GU451219, GU451223, HQ912956, HQ912957, HQ912958, GU451220, HQ912959, GU451224, AJ238799, AJ132996, AJ132997, U45476, AJ238800, EU482833, EU155382, EU155381, JN120912, AB154185, AB154186, AB154187, AB154188, AB154189, AB154190, AB154191, AB154192, AB154193, AB154194, AB154177, AB154195, AB154196, AB154197,

AB154198, AB154199, AB154200, AB154201, AB154202, AB154203, AB154204, AB154178, AB154205, AB154206, AB154179, AB154180, AB154181, AB154182, AB154183, AB154184, AB442219, AB442222, AB016785, AY045702, M58335, D90208, D50480, D50481, D50482, D50483, D50485, D50484, AB426117, AB435162, AB119282, AF139594, D63857, S62220, AB191333, AB049087, AB049088, AB049089, AB049090, AB049091, AB049092, AB049093, AB049094, AB049095, AB049096, AB049097, AB049098, AB049099, AB049100, AB049101, D14484, X61596, D11168, D11355, AB442221, AB442220, DQ244140, DQ244141, AF165045, AF165063, AF165064, AF207752, AF207753, AF207754, AF207755, AF207756, AF207757, AF207758, AF207759, AF207760, AF165046, AF165047, AF207761, AF207762, AF207763, AF207764, AF207765, AF165048, AF207766, AF207767, AF207768, AF207769, AF207770, AF165049, AF207771, AF207772, AF207773, AF207774, AF208024, AF165050, AF165051, AF165052, AF165053, AF165054, AF165055, AF165056, AF165057, AF165058, AF165059, AF165060, AF165061, AF165062, D89872, D89815, AB429050, AB249644, U16362, U01214, AF176573, AY587844, AF483269, U89019, M84754, DQ071885, EF407459, EF407473, EF407481, EF407467, EF407458, EF407495, EF407500, EF407502, EF407491, EF407474, EF407475, EF407476, EF407493, EF407478, EF407480, EF407468, EF407496, EF407498, EF407466, EF407469, EF407486, EF407472, EF407479, EF407483, EF407470, EF407497, EF407461, EF407471, EF407463, EF407484, EF407503, EF407465, EF407462, EF407482, EF407487, EF407490, EF407499, EF407489, EF407492, EF407485, EF407501, EF407494, EF407460, EF407488, EF407477, EF407464, EF407504, GU133617, EU155333, EU482849, EU155331, EU155334, EU155337, EU155325, EU234061, EU255961, EU255962, EU155324, EU482888, EU155326, EU155327, EU155328, EU155329, EU234062, EU155335, EU155336, EU155330, KC439507, KC439508, KC439515, KC439527, EU155332, HQ110091, KC439503, KC439504, KC439505, KC439506, KC439509, KC439510, KC439511, KC439512, KC439513, KC439514, FJ478453, EU660388, EU482877, EU155301, EU155302, EU155303, EU155262, EU256102, EU155315, EU239714, EU256103, EU155279, EU155280, EU256054, EU862835, KC439516, KC439517, KC439518, KC439519, EU256001, EU155219, EU155223, EU155230, EU256089, EU256090, EU155231, EU155232, EU256098, EU256099, EU256100, EU155305, EU256101, EU155306, EU155307, EU155316, EU155317, KC439488, KC439499, KC439481, KC439482, EU256000, EU256088, EU155224, EU155225, EU155227, EU155228, EU155300, EU155304, EU155308, EU256045, EU256066, EU482881, KC439487, KC439500, KC439501, KC439483, KC439484, KC439485, EU155217, EU155220, EU155221, EU155222, EU155226, EU155229, KC439486, EU155218, EU482883, EU256059, EU155261, EU155263, EU155318, EU155281, KC439489, KC439490, KC439491,

KC439492, KC439502, EU660386, EU155253, EU256061, EU155254, EU155255, EU155256, EU256062, EU155257, EU482839, EU155258, EU482879, EU482880, EU155259, EU155260, EU256064, EU256065, EU155264, EU155234, EU482885, EU155235, EU256091, EU482886, EU256092, KC439493, KC439494, KC439495, KC439496, KC439497, KC439498, FJ024086, FJ024277, FJ024279, FJ390396, FJ390397, FJ390398, KC439520, KC439521, KC439522, KC439523, KC439524, KC439525, KC439526, AB622121, HQ537005, HQ537006, AY587845, KC844047, D14853, AY051292, KC248194, AM910652, JF735110, JF735112, JF735116, JF735117, JF735118, JF735119, KC197238, KC197236, KC197237, KC197239, AM408911, EF026073, AF169003, AF169004, NC_009823, JX014307, AF238481, AF238482, AF238483, AF238484, AF238485, AF169002, AF169005, AF177036, KF676351, KF676352, HQ639943, HQ639944, HQ639945, HQ639938, HQ639939, KC844043, KC967476, AY746460, D00944, AB047640, AB047641, AB047642, AB047643, AB047644, AB047645, AB047639, AB690460, AB690461, KF700370, AB030907, AF238486, KC844048, KC967478, KC967477, JQ745651, KC197226, AB661373, AB661374, AB661375, AB661376, AB661377, AB661378, AB661379, AB661380, AB661381, AB661382, AB661383, AB661384, AB661385, AB661386, AB661387, AB661388, AB661389, AB661390, AB661391, AB661392, AB661393, AB661394, AB661395, AB661396, AB661397, AB661398, AB661399, AB661400, AB661401, AB661402, AB661403, AB661404, AB661405, AB661406, AB661407, AB661408, AB661409, AB661410, AB661411, AB661412, AB661413, AB661414, AB661415, AB661416, AB661417, AB661418, AB661419, AB661420, AB661421, AB661422, AB661423, AB661424, AB661425, AB661426, AB661427, AB661428, AB661429, AB661430, AB661431, AB661432, AB559564, D10988, AY232748, AY232749, AY232730, AY232731, AY232732, AY232733, AY232734, AY232735, AY232736, AY232737, AY232738, AY232739, AY232740, AY232741, AY232742, AY232743, AY232744, AY232745, AY232746, AY232747, DQ430815, DQ430817, DQ364460, EU643835, D50409, KC967479, KC197228, KC197227, JX227949, JX227950, JX227951, JX227965, JX227966, JF735114, JF735120, KC844050, KC844042, KC197229, KC197231, KC197230, DQ155561, DQ155560, JF735113, KC197232, KC197233, HM777359, HM777358, KC197234, JX227953.1, JX227953, AB031663, KC197235, KC197240, JX227967, JX227968, JX227974, JF735111, FN666429, FN666428, JF735115, JF735124, GU294484, AB691595, AB691596, AF046866, D28917, D17763, NC_009824, AB792683, DQ437509, HQ639941, HQ639942, HQ912953, KC844041, X76918, GQ356217, GQ356216, GQ356214, GQ356212, GQ356211, GQ356210, GQ356209, GQ356208, GQ356207, GQ356206, GQ356205, GQ356204, GQ356203, GQ356202, GQ356201, GQ356200, HQ738645, GQ275355, JQ717260, JQ717259, JQ717254, JQ717255, JQ717257,

		JQ717258, JN714194, JQ717256, KF035125, KF035123, KF035124, KF035126, KF035127, GU814263, DQ430819, AY956467, DQ430820, JQ065709, KC844044, D49374, JF735123, JX227954, JF735121, JF735126, JX227969, JF735125, JX227955, JX227956, FJ407092, JF735122, D63821, , DQ418785, GU814265, NC_009825, Y11604, DQ988078, DQ988079, DQ988073, DQ988074, DQ988075, DQ988076, DQ988077, DQ516084, AB795432, DQ418782, DQ418783, DQ418784, DQ418788, DQ418787, DQ418789, FJ462435, FJ025854, FJ025855, FJ025856, FJ462436, EU392172, FJ462437, KC844045, DQ516083, DQ418786, EU392169, EU392170, EU392174, EU392175, EF589160, EF589161, FJ462432, JX227971.1, JX227963.1, EU392171, EU392173, FJ462438, FJ839870, JX227958.1, JX227957.1, FJ462433, JX227972.1, JX227961.1, FJ462441, JX227973.1, JX227970.1, FJ462440, JX227979.1, JX227978.1, JX227977.1, FJ462431, FJ462434, FJ462439, JX227976.1, JX227975.1, JX227962.1, FJ839869, HQ537008, HQ537009, KF373567, KC844046, NC_009826, Y13184, AF064490, KC767831, KC76783, KC76783, KC76783, KC76783, KC76782, DQ278892, KC191671, HQ639936, HQ91295, HQ912954, KC84403, KC84403, AY859526, DQ480513, DQ480516, DQ480523, DQ480514, DQ480515, DQ480518, DQ480519, DQ480520, DQ480521, DQ480522, DQ480517, DQ480524, DQ480512, Y12083, EU246930, NC_009827, D84262, EF424629, D84263, DQ314805, EU408326, EU246931, EU246932, DQ835760, DQ835764, EU246936, DQ314806, D63822, D84265, DQ835762, EU246935, DQ835770, DQ835761, DQ835769, AY878651, DQ278893, AY878650, DQ278891, D84264, EF424628, EU246933, JX183555, JX183556, DQ835767, DQ835765, DQ835766, DQ835763, AY878652, DQ278894, DQ835768, EU246937, EU246938, EF424627, EU408327, EU246934, EF424626, EF424625, EU408328, EU408329, EU246939, EF632069, EF632070, EF632071, EU408330, EU408331, EU408332, EU246940, EU798761, FJ435090, EU798760, EU158186, EU643836, EU643835, EF108306.2
Los Alamos HCV genotype reference sequences (2012) (Kuiken *et al.*, 2005)	Figure 1-4	NC_004102, M67463, EF407419, AF511950, EF032892, AY587016, D11355, D14853, AY051292, AM910652, AY746460, D00944, AB047639, D10988, AB030907, AF238486, D50409, DQ155561, HM777359, HM777358, AB031663, FN666429, FN666428, X76918, AF046866, NC_009824, D49374, FJ407092, D63821, Y11604, DQ418788, FJ462435, FJ025854, FJ025855, FJ025856, FJ462436, FJ462437, DQ516083, DQ418786, EU392172, EF589160, EF589161, EU392169, EU392170, EU392174, EU392175, FJ462432, FJ462438, EU392171, EU392173, FJ839870, FJ462433, FJ462441, FJ462440, FJ462431, FJ462434, FJ462439, FJ839869, HQ537008, HQ537009, NC_009826, AF064490, DQ278892, AY859526, DQ480513, Y12083, NC_009827, EF424629, D84263, EU408326, EU246932, DQ835764, DQ314806, D84265, EU246935, DQ835761, DQ278893, D84264, EU246933, DQ835766, AY878652, EU246938, EU408327, EU246934, EF424626, EF424625, EU408328,

		EU408329, EU246939, EF632069, EF632070, EF632071, EU408330, EU408331, EU408332, EU246940, EU158186, EU798761, FJ435090, EU798760, EU643834, EU643836, EF108306
Mixed genotype infection CpG analysis (gt1a sequences)	Figure 6-1 Figure 6-11	AB520610, AF009606, AF271632, AF290978, AF511948, AF511949, AF511950, AF529293, AJ278830, AJ557444, AY388455, AY615798, AY695436, AY885238, AY956463, AY956464, AY956466, AY956468, AY958045, D10749, DQ061300, DQ061304, DQ061308, DQ061313, DQ061319, DQ061323, DQ061327, DQ430811, DQ838739, EF032883, EF032886, EF032887, EF032891, EF032895, EF032896, EU370586, EU370616, EU370636, EU370656, EU482831, EU482832, EU482834, EU482835, EU482836, EU482837, EU482838, EU482840, EU482841, EU482842, EU482843, EU482844, EU482845, EU482846, EU482847, EU482848, EU482850, EU482852, EU482853, EU482854, EU482855, EU482856, EU482872, EU482873, EU677247, EU677252, EU687193, EU687194, EU687195, EU781823
Mixed genotype infection CpG analysis (gt3a sequences)	Figure 6-1 Figure 6-11	AB691595, AB691596, AB792683, AF046866, AM423055, AM423056, AM423057, AM423058, AM423059, AM423060, AM423061, AM423062, AY956467, AY957985, AY958044, D17763, D28917, DQ430819, DQ430820, DQ437509, GQ275355, GQ356200, GQ356201, GQ356202, GQ356203, GQ356204, GQ356205, GQ356206, GQ356207, GQ356208, GQ356209, GQ356210, GQ356211, GQ356212, GQ356213, GQ356214, GQ356215, GQ356216, GQ356217, GQ379231, GU172375, GU570704, GU814263, HQ108092, HQ108096, HQ108100, HQ108104, HQ639941, HQ639942, HQ738645, HQ912953, JN714194, JQ664687, JQ717254, JQ717255, JQ717256, JQ717257, JQ717258, JQ717259, JQ717260, JQ802144, JQ802145, KC844041, KF035123, KF035124, KF035125, KF035126, KF035127

Table 8-2; Univariable and multivariable analysis of co-morbidities by age

Results of a univariable and a multivariable analysis of the co-morbidities affecting individuals infected with HCV analysed by age. P values ≤0.05 were considered statistically significant and have been highlighted. Values tending towards significance (between p>0.05 and p ≤0.1) have been underlined.

	Age (Univariable analysis)					Age (Multivariable analysis)				
Co-morbidity	**<45yrs (n=360)**	**%**	**>=45yrs (n=227)**	**%**	**p**	**<45yrs (n=360)**	**%**	**>= 45yrs (n=227)**	**%**	**p**
No recorded co-morbidities	76	21.11	37	16.30	0.36	76	21.11	37	16.30	0.21
High alcohol consumption	111	30.83	65	28.63	0.38	111	30.83	65	28.63	0.30
Obesity	10	2.78	6	2.64	0.24	10	2.78	6	2.64	0.80
Truama, stabbing or assault	62	17.22	28	12.33	0.21	62	17.22	28	12.33	0.26
Cancer	3	**0.83**	10	**4.41**	**0.02**	3	**0.83**	10	**4.41**	**<0.01**
Cardiovascular disease or stroke	18	5.00	25	11.01	0.06	18	**5.00**	25	**11.01**	**0.02**
Diabetes	7	1.94	10	4.41	0.09	16	4.44	8	3.52	0.90
Dermatology	16	4.44	8	3.52	0.85	7	1.94	10	4.41	0.10
Endocrine	5	1.39	6	2.64	0.93	5	1.39	6	2.64	0.83
Gastrointestinal disorders	37	**10.28**	43	**18.94**	**0.03**	37	**10.28**	43	**18.94**	**0.01**
Gynecological disorders	15	**4.17**	5	**2.20**	**0.04**	15	4.17	5	2.20	0.06
Haematological	4	1.11	7	3.08	0.06	4	1.11	2	0.88	0.30
Hemochromatosis	4	1.11	2	0.88	0.81	4	**1.11**	7	**3.08**	**0.01**
Liver transplant	0	0.00	1	0.44	0.62	0	0.00	1	0.44	0.19
Neurological disorders	38	10.56	22	9.69	0.61	38	10.56	22	9.69	0.45
Orthopaedic	15	4.17	17	7.49	0.06	15	4.17	17	7.49	0.12
Psychiatric disorders	160	44.44	98	43.17	1.00	160	44.44	98	43.17	0.63
Respiratory	36	10.00	20	8.81	0.24	36	10.00	20	8.81	0.60
Rheumatological	1	0.28	4	1.76	0.15	1	**0.28**	4	**1.76**	**0.04**
Urological or renal disorders	14	3.89	14	6.17	0.63	14	3.89	14	6.17	0.11
Vascular	23	6.39	12	5.29	0.48	23	6.39	12	5.29	0.82
HBV	4	1.11	0	0.00	0.31	4	1.11	0	0.00	0.13
HIV	8	2.22	1	0.44	0.06	8	2.22	1	0.44	0.15
TB	3	0.83	0	0.00	0.22	3	0.83	0	0.00	0.28

Table 8-3; Finney's table

Finney's table, used for the conversion of results into probit during Probit analysis

Table 3.2 Transformation of percentages to probits

%	0	1	2	3	4	5	6	7	8	9
0	—	2.67	2.95	3.12	3.25	3.36	3.45	3.52	3.59	3.66
10	3.72	3.77	3.82	3.87	3.92	3.96	4.01	4.05	4.08	4.12
20	4.16	4.19	4.23	4.26	4.29	4.33	4.36	4.39	4.42	4.45
30	4.48	4.50	4.53	4.56	4.59	4.61	4.64	4.67	4.69	4.72
40	4.75	4.77	4.80	4.82	4.85	4.87	4.90	4.92	4.95	4.97
50	5.00	5.03	5.05	5.08	5.10	5.13	5.15	5.18	5.20	5.23
60	5.25	5.28	5.31	5.33	5.36	5.39	5.41	5.44	5.47	5.50
70	5.52	5.55	5.58	5.61	5.64	5.67	5.71	5.74	5.77	5.81
80	5.84	5.88	5.92	5.95	5.99	6.04	6.08	6.13	6.18	6.23
90	6.28	6.34	6.41	6.48	6.55	6.64	6.75	6.88	7.05	7.33
—	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
99	7.33	7.37	7.41	7.46	7.51	7.58	7.65	7.75	7.88	8.09

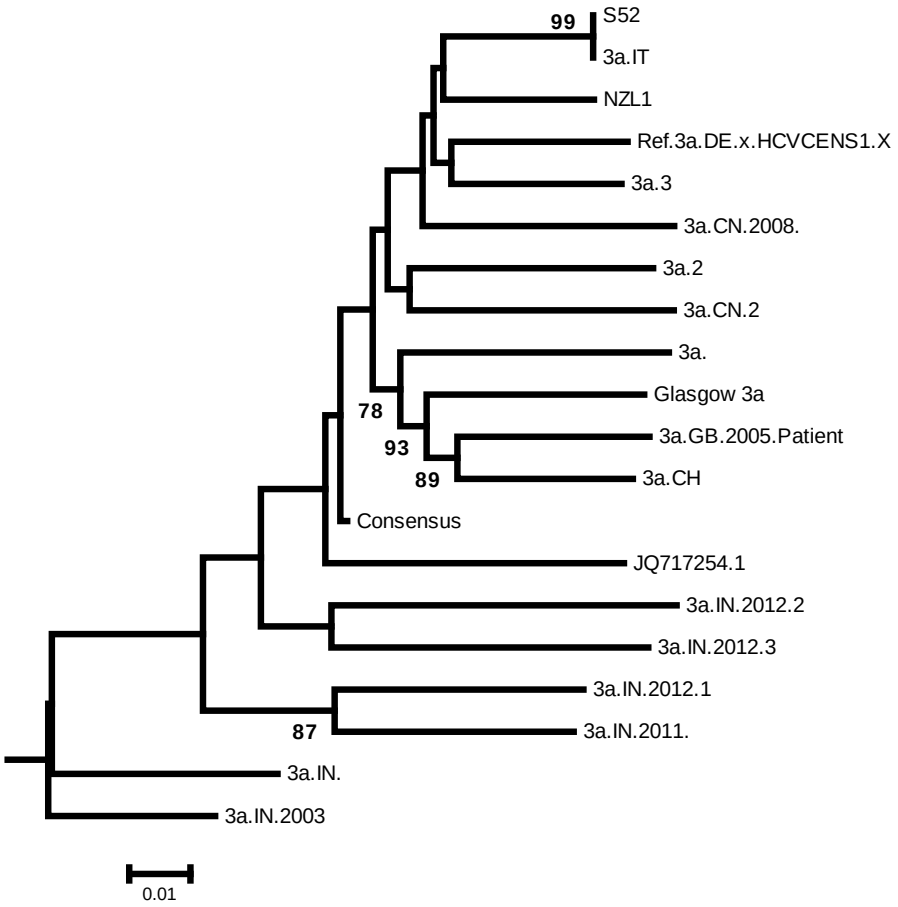


Figure 8-1; Selection of the gt3a control strain

Maximum likelihood phylogenetic tree of complete gt3a sequences with gt1 sequence used as an outgroup (not shown). The control selected (3a.GB.2005.Patient) can be seen to cluster closely to the Glasgow 3a sequence in the tree. Bootstrap support of >70% after 1000 replicates is indicated on the tree.

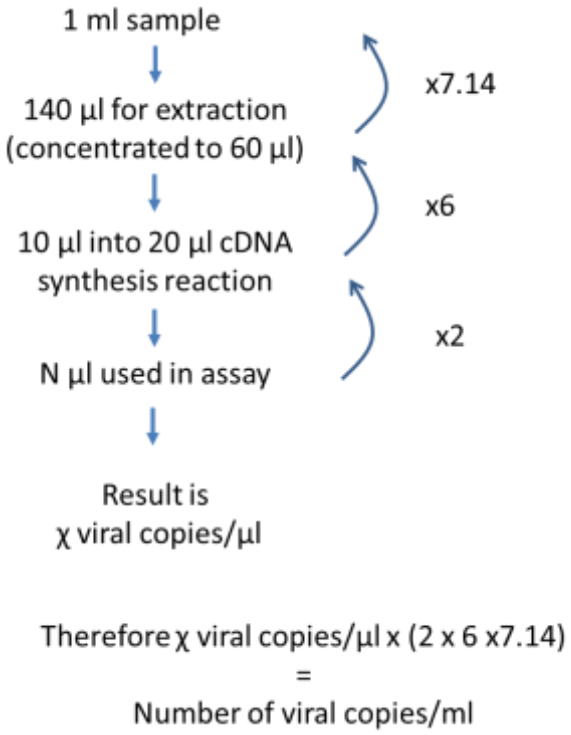


Figure 8-2; Calculation of viral load

Explanation of how the calculation to convert qPCR

North Glasgow Microbiology	VS BBV-SOP-109 v2
	Authorised by: Jane McOwan & Rory Gunson
Issued: 17/08/12	Author: Alasdair MacLean

- Internal control failures on control specimens should be discussed with a senior member of staff before validating the run.

10 RECORDING AND REPORTING OF RESULTS

- HCV genotyping reporting
  HCV genotypes will be reported by the following algorithm.

		Test result			
HCV genotype	Genotype interpretation	HCV Quant	HCV FAM	HCV VIC	IC
1	1	+	+	-	+
2	2	+	+/- (flat)	+	+
3	3	+	-	-	+
4	4 or 5	+	+	+	+
5	4 or 5	+	+	+	+

The criteria for HCV genotype sequencing in the new multiplex is a quantitation value less than 1000 IU/ml; genotyping result and **quantitation result having a Ct difference > 6;** or not fitting the genotyping interpretation algorithm in the Table above. All putative genotypes 4 or 5 should be sequenced as these cannot be distinguished by the above algorithm.

Results should be compared with previous results and if they do not match, the sample should be re-extracted and also coded for HCV genotype sequencing.

All genotypes 1 should not be reported but tested for subtype 1a or 1b by coding for H72 (worksheet HCV1GT).

The genotype results are assigned to the patients' records on TelePath and their paper cards as their HCV genotype determined by 5'NCR real time PCR.

Figure 8-3; WWSVC HCV genotyping SOP

Extract from North Glasgow Microbiology BBV-SOP-109 v2, used within the WSSVC clinical testing laboratory for HCV genotyping at the time of sample collection. Data provided by Dr Tamer Abdelrahman, Clinical Lecturer at the University of Glasgow.

Table 8-4; Raw data from CD-Hit analysis of Run 1

Raw data from CD-Hit analysis of Run 1 (using 0.9 similarity threshold) with gt2 and H77 contamination excluded.*denotes the number of distinct clusters identified with >500 or >1000 reads within the cluster.

	RUN 1		E1E2		NS5B	
	Sample	HCV genotype	>500*	>1000*	>500*	>1000*
Mixed gt1a/gt3 samples	g3-36	3a	9	4	17	10
	g3-41	3a	14	8	5	3
	g3-44	3a	16	8	11	7
	g3-51	3a	11	5	17	5
	g3-67	3a	13	3	16	5
	g3-73	3a	27	14	16	10
		1a	0	0	3	3
	g3-80	3a	4	4	3	2
		1a	0	0	1	0
	g3-125	3a	11	6	12	8
		1a	4	2	0	0
	g3-128	3a	16	9	13	7
	g3-155	3a	26	17	10	6
	g3-166	3a	15	11	9	4
	g3-176	3a	21	11	4	4
	g-27	3a	9	5	11	2
	g-30	3a	12	6	13	7
	g-45	3a	10	7	16	11
	e-15	3a	0	0	1	0
		1a	6	5	9	8
	e49	3a	9	5	13	5
Gt3 samples	3-1	3a	13	9	7	4
	3-2	3a	9	5	4	3
	3-3	3a	29	14	31	19
	3-4	3a	8	4	7	4
	3-5	3a	7	5	4	3
	3-6	3a	14	9	6	3
	3-7	ND	n/a	n/a	n/a	n/a
	3-8	3a	26	15	5	3
	3-9	3a	17	10	2	2
	3-10	3a	12	8	4	2
	3-11	3a	7	7	4	2
	3-12	3a	7	5	4	2
	3-13	3a	5	4	20	7
	3-14	3a	9	4	2	2
	3-15	3a	20	13	2	0
	3-16	3a	14	9	2	2
	3-17	3a	10	6	0	0
	3-18	3a	19	11	2	2
	3-19	3a	18	11	2	2
	3-20	3a	8	7	0	0
Controls	3a-transcript1	3a	9	5	9	4
	3a-transcript2	3a	5	4	7	3
	3a-transcript3	3a	9	7	10	4
	0.1% 1a-major3a	3a	6	6	2	2
		1a	0	0	0	0
	1% 1a-major3a	3a	7	4	4	2
		1a	2	2	2	0
	10% 1a-major3a	3a	62	33	0	0
		1a	49	24	3	3
	neg-serum-2	ND	n/a	n/a	n/a	n/a
	neg-h2o	ND	n/a	n/a	n/a	n/a

Table 8-5; Raw data from CD-Hit analysis of Run 2

Raw data from CD-Hit analysis of Run 2 (using 0.9 similarity threshold) with gt2 and H77 contamination excluded.*denotes the number of distinct clusters identified with >500 or >1000 reads within the cluster.

	RUN 2		E1E2		NS5B	
	Sample	HCV genotype	>500*	>1000*	>500*	>1000*
Mixed	g3-85	1a	38	22	18	13
	g3-92	1a	20	11	42	21
	1a-128	1a	14	6	0	0
Gt1a samples	1a-1	1a	1	0	22	14
	1a-2	1a	ND	ND	10	5
	1a-3	1a	ND	ND	12	5
	1a-4	1a	ND	ND	25	15
	1a-5	1a	12	8	7	4
	1a-6	1a	2	1	46	25
	1a-7	1a	2	0	36	16
	1a-8	1a	ND	ND	31	14
	1a-9	1a	15	10	27	14
	1a-10	1a	2	2	2	0
	1a-11	1a	ND	ND	27	13
	1a-12	1a	ND	ND	16	9
	1a-13	1a	1	1	32	17
	1a-14	1a	ND	ND	17	8
	1a-15	1a	ND	ND	43	22
	1a-16	1a	16	11	45	19
	1a-17	1a	ND	ND	13	7
	1a-18	1a	ND	ND	16	10
	1a-19	1a	3	3	17	5
	1a-20	1a	ND	ND	43	22
Gt1a samples (E1-E2 repeats)	E1a-1	1a	21	13	n/a	n/a
	E1a-2	1a	21	12	n/a	n/a
	E1a-3	1a	16	12	n/a	n/a
	E1a-4	1a	29	14	n/a	n/a
	E1a-5	1a	45	32	n/a	n/a
	E1a-6	1a	21	13	n/a	n/a
	E1a-7	1a	26	19	n/a	n/a
	E1a-8	1a	21	9	n/a	n/a
	E1a-9	1a	43	28	n/a	n/a
	E1a-10	1a	30	10	n/a	n/a
	E1a-11	1a	25	12	n/a	n/a
	E1a-12	1a	9	7	n/a	n/a
	E1a-13	1a	20	14	n/a	n/a
	E1a-14	1a	19	13	n/a	n/a
	E1a-15	1a	26	19	n/a	n/a
	E1a-16	1a	45	28	n/a	n/a
	E1a-17	1a	35	18	n/a	n/a
	E1a-18	1a	70	31	n/a	n/a
	E1a-20	1a	20	10	n/a	n/a
Controls	1a-transcript1	1a	30	13	26	16
	1a-transcript2	1a	24	11	30	13
	1a-transcript3	1a	21	8	30	13
	0.1% 1a-major3a	1a	5	5	7	5
		3a	0	0	2	2
	1% 1a-major3a	1a	3	2	7	5
		3a	0	0	3	3
	10% 1a-major3a	1a	4	3	8	4
		3a	0	0	4	2
	neg-serum-2	ND	n/a	n/a	n/a	n/a
	neg-h2o	ND	n/a	n/a	n/a	n/a

Chapter 9: List of References

Aach, R. D., Szmuness, W., Mosley, J. W., Hollinger, F. B., Kahn, R. A., Stevens, C. E., Edwards, V. M. and Werch, J. (1981) 'Serum alanine aminotransferase of donors in relation to the risk of non-A,non-B hepatitis in recipients: the transfusion-transmitted viruses study.', *The New England journal of medicine*. UNITED STATES, 304(17), pp. 989–994. doi: 10.1056/NEJM198104233041701.

AASLD-IDSA Guidance Panel (2016) *Initial Treatment of HCV Infection, HCV Guidance: Recommendations for Testing, Managing, and Treating Hepatitis C*.

Abdelrahman, T., Hughes, J., Main, J., McLauchlan, J., Thurz, M. and Thomson, E. (2015) 'Next generation sequencing sheds light on the natural history of hepatitis C infection in patients that fail treatment', *Hepatology*, 61(1), pp. 88–97.

Abid, K., Pazienza, V., De Gottardi, A., Rubbia-Brandt, L., Conne, B., Pugnale, P., Rossi, C., Mangia, A. and Negro, F. (2005) 'An in vitro model of hepatitis C virus genotype 3a-associated triglycerides accumulation', *Journal of Hepatology*, 42, pp. 744–751. doi: 10.1016/j.jhep.2004.12.034.

Aceijas, C. and Rhodes, T. (2007) 'Global estimates of prevalence of HCV infection among injecting drug users.', *The International journal on drug policy*. Netherlands, 18(5), pp. 352–358. doi: 10.1016/j.drugpo.2007.04.004.

Aird, D., Chen, W.-S., Ross, M., Connolly, K., Meldrim, J., Russ, C., Fisher, S., Jaffe, D., Nusbaum, C. and Gnirke, A. (2010) 'Analyzing and minimizing bias in Illumina sequencing libraries', *Genome Biology*. BioMed Central Ltd, 11(2), p. P3. doi: 10.1186/gb-2010-11-s1-p3.

Alberti, A., Morsica, G., Chemello, L., Cavalletto, D., Noventa, F., Pontisso, P. and Ruol, A. (1992) 'Hepatitis C viraemia and liver disease in symptom-free individuals with anti-HCV.', *Lancet*. ENGLAND, 340(8821), pp. 697–698.

Alizon, S. and van Baalen, M. (2008) ‘Multiple Infections, Immune Dynamics, and the Evolution of Virulence’, *The American Naturalist*. United States, 172(4), pp. E150–E168. doi: 10.1086/590958.

Alter, H. J., Holland, P. V, Morrow, a G., Purcell, R. H., Feinstone, S. M. and Moritsugu, Y. (1975) ‘Clinical and serological analysis of transfusion-associated hepatitis.’, *Lancet*, 2(7940), pp. 838–841. doi: 10.1016/S0140-6736(75)90234-2.

Alter, H. J., Purcell, R. H., Holland, P. V and Popper, H. (1978) ‘Transmissible agent in non-A, non-B hepatitis.’, *Lancet (London, England)*. ENGLAND, 1(8062), pp. 459–463.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. and Lipman, D. J. (1990) ‘Basic local alignment search tool.’, *Journal of molecular biology*. ENGLAND, 215(3), pp. 403–410. doi: 10.1016/S0022-2836(05)80360-2.

Anderson, N. (2008) ‘Molecular Mechanisms and Therapeutic Targets in Steatosis and Steatohepatitis’, *Pharmacological reviews*, 60(3), pp. 311–357. doi: 10.1124/pr.108.00001.311.

André, P., Komurian-Pradel, F., Deforges, S., Perret, M., Berland, J. L., Sodoyer, M., Pol, S., Bréchot, C., Paranhos-Baccalà, G. and Lotteau, V. (2002) ‘Characterization of low- and very-low-density hepatitis C virus RNA-containing particles.’, *Journal of virology*, 76(14), pp. 6919–28. doi: 10.1128/JVI.76.14.6919.

Andriulli, A., Mangia, A., Iacobellis, A., Ippolito, A., Leandro, G. and Zeuzem, S. (2008) ‘Meta-analysis: the outcome of anti-viral therapy in HCV genotype 2 and genotype 3 infected patients with chronic hepatitis.’, *Alimentary pharmacology & therapeutics*. England, 28(4), pp. 397–404. doi: 10.1111/j.1365-2036.2008.03763.x.

De Angelis, D., Hickman, M. and Yang, S. (2004) ‘Estimating long-term trends in the incidence and prevalence of opiate use/injecting drug use and the number of former users: Back-calculation methods and opiate overdose deaths’, *American Journal of Epidemiology*, 160(10), pp. 994–1004. doi: 10.1093/aje/kwh306.

Al Ashgar, H. I., Khan, M. Q., Al-Ahdal, M., Al Thawadi, S., Helmy, A. S., Al Qahtani, A.

and Sanai, F. M. (2013) ‘Hepatitis C genotype 4: genotypic diversity, epidemiological profile, and clinical relevance of subtypes in Saudi Arabia.’, *Saudi journal of gastroenterology*. India, 19(1), pp. 28–33. doi: 10.4103/1319-3767.105920.

Ashrafi, M., Modabbernia, A., Dalir, M., Taslimi, S., Karami, M., Ostovaneh, M. R., Malekzadeh, R. and Poustchi, H. (2012) ‘Predictors of mental and physical health in non-cirrhotic patients with viral hepatitis: a case control study.’, *Journal of psychosomatic research*. England, 73(3), pp. 218–224. doi: 10.1016/j.jpsychores.2012.06.006.

van Asten, L. and Prins, M. (2004) ‘Infection with concurrent multiple hepatitis C virus genotypes is associated with faster HIV disease progression.’, *AIDS*, 18(17), pp. 2319–2324. doi: 10.1097/00002030-200411190-00013.

Atkinson, N. J., Witteveldt, J., Evans, D. J. and Simmonds, P. (2014) ‘The influence of CpG and UpA dinucleotide frequencies on RNA virus replication and characterization of the innate cellular pathways underlying virus attenuation and enhanced replication.’, *Nucleic acids research*, 42(7), pp. 4527–45. doi: 10.1093/nar/gku075.

Ayesh, B. M., Zourob, S. S., Abu-Jadallah, S. Y. and Shemer-Avni, Y. (2009) ‘Most common genotypes and risk factors for HCV in Gaza strip: a cross sectional study’, *Virology journal*, 6, pp. 105–112. doi: 10.1186/1743-422X-6-105.

Babu, C. K., Suwansrinon, K., Bren, G. D., Badley, A. D. and Rizza, S. A. (2009) ‘HIV induces TRAIL sensitivity in hepatocytes.’, *PloS one*. United States, 4(2), p. e4623. doi: 10.1371/journal.pone.0004623.

Backmund, M., Reimer, J., Meyer, K., Gerlach, J. and Zachoval, R. (2005) ‘Hepatitis C virus infection and injection drug users: Prevention, risk factors, and treatment’, *Clinical Infectious Diseases*, 40(Supplement 5), pp. S330–S335.

Backus, L. I., Boothroyd, D. B., Phillips, B. R., Belperio, P., Halloran, J. and Mole, L. A. (2011) ‘A Sustained Virologic Response Reduces Risk of All-Cause Mortality in Patients With Hepatitis C’, *Clinical Gastroenterology and Hepatology*. Elsevier Inc., 9(6), p. 509–516.e1. doi: 10.1016/j.cgh.2011.03.004.

Baden, R., Rockstroh, J. K. J. K. and Buti, M. (2014) 'Natural history and management of hepatitis C: Does sex play a role?', *Journal of Infectious Diseases*, 209(SUPPL. 3), pp. 81–85. doi: 10.1093/infdis/jiu057.

Bader el-Din, N. G., Abd el-Meguid, M., Tabll, A. a, Anany, M. a, Esmat, G., Zayed, N., Helmy, A., el-Zayady, A. R., Barakat, A. and el-Awady, M. K. (2011) 'Human cytomegalovirus infection inhibits response of chronic hepatitis-C-virus-infected patients to interferon-based therapy.', *Journal of gastroenterology and hepatology*, 26(1), pp. 55–62. doi: 10.1111/j.1440-1746.2010.06319.x.

Bagaglio, S., Uberti-Foppa, C., Di Serio, C., Trentini, F., Andolina, A., Hasson, H., Messina, E., Merli, M., Porrino, L., Lazzarin, A. and Morsica, G. (2015) 'Dynamic of Mixed HCV Infection in Plasma and PBMC of HIV/HCV Patients Under Treatment With Peg-IFN/Ribavirin.', *Medicine*, 94(43), p. e1876. doi: 10.1097/MD.0000000000001876.

Baker, S. F., Nogales, A. and Martinez-Sobrido, L. (2015) 'Downregulating viral gene expression: codon usage bias manipulation for the generation of novel influenza A virus vaccines.', *Future virology*, 10(6), pp. 715–730. doi: 10.2217/fvl.15.31.

Baltimore, D. (1970) 'Viral RNA-dependent DNA Polymerase: RNA-dependent DNA Polymerase in Virions of RNA Tumour Viruses', *Nature*, 226(5252), pp. 1209–1211.

Baltimore, D. (1971) 'Expression of Animal Virus Genomes', *Bacteriological Reviews*, 35(3), pp. 235–241.

Barreiro, P., Labarga, P., Fernandez-Montero, J. V, Poveda, E., de Mendoza, C., Sanchez, C., Vispo, E. and Soriano, V. (2013) 'Longitudinal changes in viral RNA concentration in patients with chronic hepatitis C and/or HIV infection in the absence of antiviral therapy.', *Journal of clinical virology : the official publication of the Pan American Society for Clinical Virology*. Netherlands, 58(2), pp. 391–395. doi: 10.1016/j.jcv.2013.06.031.

Barrera, J. M., Bruguera, M., Ercilla, M. G., Gil, C., Celis, R., Gil, M. P., del Valle Onorato, M., Rodes, J. and Ordinas, A. (1995) 'Persistent hepatitis C viremia after acute self-limiting posttransfusion hepatitis C.', *Hepatology*. UNITED STATES, 21(3), pp. 639–644.

Barrett, S., Goh, J., Coughlan, B., Ryan, E., Stewart, S., Cockram, A., O'Keane, J. C. and Crowe, J. (2001) 'The natural course of hepatitis C virus infection after 22 years in a unique homogenous cohort: spontaneous viral clearance and chronic HCV infection.', *Gut*, 49(3), pp. 423–30. doi: 10.1136/gut.49.3.423.

Bartels, D. J., Zhou, Y., Zhang, E. Z., Marcial, M., Byrn, R. A., Pfeiffer, T., Tigges, A. M., Adiwijaya, B. S., Lin, C., Kwong, A. D. and Kieffer, T. L. (2008) 'Natural prevalence of hepatitis C virus variants with decreased sensitivity to NS3.4A protease inhibitors in treatment-naive subjects.', *The Journal of infectious diseases*. United States, 198(6), pp. 800–807. doi: 10.1086/591141.

Bartenschlager, R. (1999) 'The NS3/4A proteinase of the hepatitis C virus: unravelling structure and function of an unusual enzyme and a prime target for antiviral therapy', *Journal of Viral Hepatitis*. Blackwell Science Ltd, 6(3), pp. 165–181. doi: 10.1046/j.1365-2893.1999.00152.x.

Barth, H., Schäfer, C., Adah, M. I., Zhang, F., Linhardt, R. J., Toyoda, H., Kinoshita-Toyoda, A., Toida, T., Van Kuppevelt, T. H., Depla, E., Von Weizsäcker, F., Blum, H. E. and Baumert, T. F. (2003) 'Cellular Binding of Hepatitis C Virus Envelope Glycoprotein E2 Requires Cell Surface Heparan Sulfate', *Journal of Biological Chemistry*, 278(42), pp. 41003–41012. doi: 10.1074/jbc.M302267200.

Bartosch, B., Dubuisson, J. and Cosset, F.-L. L. (2003) 'Infectious Hepatitis C Virus Pseudo-particles Containing Functional E1-E2 Envelope Protein Complexes', *Journal of Experimental Medicine*. United States, 197(5), pp. 633–642. doi: 10.1084/jem.20021756.

Bassett, S. E., Guerra, B., Brasky, K., Miskovsky, E., Houghton, M., Klimpel, G. R. and Lanford, R. E. (2001) 'Protective immune response to hepatitis C virus in chimpanzees rechallenged following clearance of primary infection.', *Hepatology*. United States, 33(6), pp. 1479–1487. doi: 10.1053/jhep.2001.24371.

Bayram, A., Eksi, F., Mehli, M. and Sozen, E. (2007) 'Prevalence of hepatitis E virus antibodies in patients with chronic hepatitis B and chronic hepatitis C.', *Intervirology*. Switzerland, 50(4), pp. 281–286. doi: 10.1159/000103916.

Bedard, P. L., Hansen, A. R., Ratain, M. J. and Siu, L. L. (2013) ‘Tumour heterogeneity in the clinic’, *Nature*. Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved., 501(7467), pp. 355–364.

Belalov, I. S. and Lukashev, A. N. (2013) ‘Causes and Implications of Codon Usage Bias in RNA Viruses’, *PLoS ONE*, 8(2). doi: 10.1371/journal.pone.0056642.

Bennett, S. T., Barnes, C., Cox, A., Davies, L. and Brown, C. (2005) ‘Toward the 1,000 dollars human genome.’, *Pharmacogenomics*. England, 6(4), pp. 373–382. doi: 10.1517/14622416.6.4.373.

Benson, D. a., Cavanaugh, M., Clark, K., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J. and Sayers, E. W. (2013) ‘GenBank’, *Nucleic Acids Research*. England, 41(Database issue), pp. D36–D42. doi: 10.1093/nar/gks1195.

Benvegnu, L., Fattovich, G., Noventa, F., Tremolada, F., Chemello, L., Cecchetto, A. and Alberti, A. (1994) ‘Concurrent hepatitis B and C virus infection and risk of hepatocellular carcinoma in cirrhosis. A prospective study.’, *Cancer*. UNITED STATES, 74(9), pp. 2442–2448.

Berger, A., von Depka Prondzinski, M., Doerr, H. W., Rabenau, H. and Weber, B. (1996) ‘Hepatitis C plasma viral load is associated with HCV genotype but not with HIV coinfection.’, *Journal of medical virology*. UNITED STATES, 48(4), pp. 339–343. doi: 10.1002/(SICI)1096-9071(199604)48:4<339::AID-JMV7>3.0.CO;2-8.

Beutler, E., Gelbart, T., Han, J. H., Koziol, J. a and Beutler, B. (1989) ‘Evolution of the genome and the genetic code: selection at the dinucleotide level by methylation and polyribonucleotide cleavage.’, *Proceedings of the National Academy of Sciences of the United States of America*, 86(1), pp. 192–196. doi: 10.1073/pnas.86.1.192.

Bilofsky, H. S. and Burks, C. (1988) ‘The GenBank genetic sequence data bank.’, *Nucleic acids research*. ENGLAND, 16(5), pp. 1861–1863.

Bird, A. P. (1980) ‘DNA methylation and the frequency of CpG in animal DNA’, *Nucleic Acids Res.*, 8(4), pp. 1499–1504. doi: 10.1093/nar/8.7.1499.

Bitzegeio, J., Bankwitz, D., Hueging, K., Haid, S., Brohm, C., Zeisel, M. B., Herrmann, E., Iken, M., Ott, M., Baumert, T. F. and Pietschmann, T. (2010) 'Adaptation of Hepatitis C Virus to mouse CD81 permits infection of mouse cells in the absence of human entry factors', *PLoS Pathogens*, 6(7), pp. 1–15. doi: 10.1371/journal.ppat.1000978.

Blackard, J. T., Kemmer, N. and Sherman, K. E. (2006) 'Extrahepatic replication of HCV: insights into clinical manifestations and biological consequences.', *Hepatology (Baltimore, Md.)*, 44(1), pp. 15–22. doi: 10.1002/hep.21283.

Blackard, J. T. and Sherman, K. E. (2007) 'Hepatitis C Virus Coinfection and Superinfection', *The Journal of Infectious Diseases*, 195, pp. 519–524.

Blatt, L. M., Mutchnick, M. G., Tong, M. J., Klion, F. M., Lebovics, E., Freilich, B., Bach, N., Smith, C., Herrera, J., Tobias, H., Conrad, a, Schmid, P. and McHutchison, J. G. (2000) 'Assessment of hepatitis C virus RNA and genotype from 6807 patients with chronic hepatitis C in the United States.', *Journal of viral hepatitis*, 7(3), pp. 196–202.

Blight, K. J., Kolykhalov, A. A. and Rice, C. M. (2000) 'Efficient initiation of HCV RNA replication in cell culture.', *Science*. UNITED STATES, 290(5498), pp. 1972–1974.

Bochud, P.-Y., Cai, T., Overbeck, K., Bochud, M., Dufour, J.-F., Mullhaupt, B., Borovicka, J., Heim, M., Moradpour, D., Cerny, A., Malinverni, R., Francioli, P. and Negro, F. (2009) 'Genotype 3 is associated with accelerated fibrosis progression in chronic hepatitis C.', *Journal of hepatology*. England, 51(4), pp. 655–666. doi: 10.1016/j.jhep.2009.05.016.

Bowden, S., McCaw, R., White, P. A., Crofts, N. and Aitken, C. K. (2005) 'Detection of multiple hepatitis C virus genotypes in a cohort of injecting drug users', *Journal of Viral Hepatitis*, 12(3), pp. 322–324. doi: 10.1111/j.1365-2893.2005.00592.x.

Brant, L. J., Ramsay, M. E., Tweed, E., Hale, A., Hurrelle, M., Klapper, P., Ngui, S. L. and the Sentinel Surveillance of Hepatitis Testing Group (2010) 'Planning for the healthcare burden of hepatitis C infection: Hepatitis C genotypes identified in England, 2002–2007', *Journal of Clinical Virology*, 48(2), pp. 115–119. doi: 10.1016/j.jcv.2010.03.018.

Breslauer, K. J., Frank, R., Blöcker, H. and Marky, L. a (1986) ‘Predicting DNA duplex stability from the base sequence.’, *Proceedings of the National Academy of Sciences of the United States of America*, 83(11), pp. 3746–3750. doi: 10.1073/pnas.83.11.3746.

Brooner, R. K., King, V. L., Kidorf, M., Schmidt, C. W. and Bigelow, G. E. (1997) ‘Psychiatric and Substance Use Comorbidity Among Treatment-Seeking Opioid Abusers’, *Archives of General Psychiatry*, 54, pp. 71–80.

Brown, J. R., Morfopoulou, S., Hubb, J., Emmett, W. a., Ip, W., Shah, D., Brooks, T., Paine, S. M. L., Anderson, G., Virasami, a., Tong, C. Y. W., Clark, D. a., Plagnol, V., Jacques, T. S., Qasim, W., Hubank, M. and Breuer, J. (2015) ‘Astrovirus VA1/HMO-C: An Increasingly Recognized Neurotropic Pathogen in Immunocompromised Patients’, *Clinical Infectious Diseases*, pp. 1–8. doi: 10.1093/cid/ciu940.

Buckton, A. J., Ngui, S. L., Arnold, C., Boast, K., Kovacs, J., Klapper, P. E., Patel, B., Ibrahim, I., Rangarajan, S., Ramsay, M. E. and Teo, C. G. (2006) ‘Multitypic Hepatitis C Virus Infection Identified by Real-Time Nucleotide Sequencing of Minority Genotypes’, *Journal of Clinical Microbiology*, 44(8), pp. 2779–2784. doi: 10.1128/jcm.01638-05.

Bukh, J., Miller, R. H. and Purcell, R. H. (1995) ‘Genetic heterogeneity of hepatitis C virus: quasispecies and genotypes.’, *Seminars in liver disease*. UNITED STATES, 15(1), pp. 41–63. doi: 10.1055/s-2007-1007262.

Bukh, J., Purcell, R. H. and Miller, R. H. (1992) ‘Sequence analysis of the 5’ noncoding region of hepatitis C virus.’, *Proceedings of the National Academy of Sciences of the United States of America*, pp. 4942–4946.

Bulteel, N., Partha Sarathy, P., Forrest, E., Stanley, A. J., Innes, H., Mills, P. R., Valerio, H., Gunson, R. N., Aitken, C., Morris, J., Fox, R. and Barclay, S. T. (2016) ‘Factors associated with spontaneous clearance of chronic hepatitis C virus infection.’, *Journal of hepatology*. doi: 10.1016/j.jhep.2016.04.030.

Burns, C. C., Campagnoli, R., Shaw, J., Vincent, A., Jorba, J. and Kew, O. (2009) ‘Genetic inactivation of poliovirus infectivity by increasing the frequencies of CpG and UpA dinucleotides within and across synonymous capsid region codons.’, *Journal of virology*,

83(19), pp. 9957–69. doi: 10.1128/JVI.00508-09.

Busch, D., McBride, A. B. and Benaventura, L. M. (1986) ‘Chemical Dependency in Women: The Link to OB/GYN Problems’, *Journal of Psychosocial Nursing and Mental Health Services*, 24(4), pp. 26–30.

Butt, S., Idrees, M., Rehman, I. U., Akbar, H., Shahid, M., Afzal, S., Younas, S. and Amin, I. (2011) ‘Mixed Genotype Infections with Hepatitis C Virus, Pakistan’, *Emerging Infectious Diseases*, 17(8). doi: 10.1128/JCM.00899-09.

Cacopardo, B., Nunnari, G. and Nigro, L. (2009) ‘Clearance of HCV RNA following acute hepatitis A superinfection.’, *Digestive and liver disease : official journal of the Italian Society of Gastroenterology and the Italian Association for the Study of the Liver*. Netherlands, 41(5), pp. 371–374. doi: 10.1016/j.dld.2007.11.015.

Cacoub, P., Gragnani, L., Comarmond, C. and Zignego, A. L. (2014) ‘Extrahepatic manifestations of chronic hepatitis C virus infection’, *Digestive and Liver Disease*. Editrice Gastroenterologica Italiana, 46(S5), pp. S165–S173. doi: 10.1016/j.dld.2014.10.005.

Campo, D. S., Dimitrova, Z., Yokosawa, J., Hoang, D., Perez, N. O., Ramachandran, S. and Khudyakov, Y. (2012) ‘Hepatitis C Virus Antigenic Convergence’, *Scientific Reports*. England, 2, p. 267. doi: 10.1038/srep00267.

Campo, D. S., Xia, G.-L., Dimitrova, Z., Lin, Y., Forbi, J. C., Ganova-Raeva, L., Punkova, L., Ramachandran, S., Thai, H., Skums, P., Sims, S., Rytsareva, I., Vaughan, G., Roh, H.-J., Purdy, M. A., Sue, A. and Khudyakov, Y. (2015) ‘Accurate Genetic Detection of Hepatitis C Virus Transmissions in Outbreak Settings’, *Journal of Infectious Diseases*, p. jiv542. doi: 10.1093/infdis/jiv542.

Carroll, M. W., Matthews, D. a., Hiscox, J. a., Elmore, M. J., Pollakis, G., Rambaut, A., Hewson, R., García-Dorival, I., Bore, J. A., Koundouno, R., Abdellati, S., Afrough, B., Aiyepada, J., Akhilomen, P., Asogun, D., Atkinson, B., Badusche, M., Bah, A., Bate, S., Baumann, J., Becker, D., Becker-Ziaja, B., Bocquin, A., Borremans, B., Bosworth, A., Boettcher, J. P., Cannas, A., Carletti, F., Castilletti, C., Clark, S., Colavita, F., Diederich, S., Donatus, A., Duraffour, S., Ehichioya, D., Ellerbrok, H., Fenandez-Garcia, M. D.,

Fizet, A., Fleischmann, E., Gryseels, S., Hermelink, A., Hinzmann, J., Hopf-Guevara, U., Ighodalo, Y., Jameson, L., Kelterbaum, A., Kis, Z., Kloth, S., Kohl, C., Korva, M., Kraus, A., Kuisma, E., Kurth, A., Liedigk, B., Logue, C. H., Lüdtke, A., Maes, P., McCowen, J., Mély, S., Mertens, M., Meschi, S., Meyer, B., Michel, J., Molkenthin, P., Muñoz-Fontela, C., Muth, D., Newman, E. N. C., Ngabo, D., Oestereich, L., Okosun, J., Olokor, T., Omiunu, R., Omomoh, E., Pallasch, E., Pályi, B., Portmann, J., Pottage, T., Pratt, C., Priesnitz, S., Quartu, S., Rappe, J., Repits, J., Richter, M., Rudolf, M., Sachse, A., Schmidt, K. M., Schudt, G., Strecker, T., Thom, R., Thomas, S., Tobin, E., Tolley, H., Trautner, J., Vermoesen, T., Vitoriano, I., Wagner, M., Wolff, S., Yue, C., Capobianchi, M. R., Kretschmer, B., Hall, Y., Kenny, J. G., Rickett, N. Y., Dudas, G., Coltart, C. E. M., Kerber, R., Steer, D., Wright, C., Senyah, F., Keita, S., Drury, P., Diallo, B., de Clerck, H., Van Herp, M., Sprecher, A., Traore, A., Diakite, M., Konde, M. K., Koivogui, L., Magassouba, N., Avšič-Županc, T., Nitsche, A., Strasser, M., Ippolito, G., Becker, S., Stoecker, K., Gabriel, M., Raoul, H., Di Caro, A., Wölfel, R., Formenty, P. and Günther, S. (2015) ‘Temporal and spatial analysis of the 2014–2015 Ebola virus outbreak in West Africa’, *Nature*. doi: 10.1038/nature14594.

de Castellarnau, M., Aparicio, E., Parera, M., Franco, S., Tural, C., Clotet, B. and Martínez, M. A. (2012) ‘Deciphering the interleukin 28B variants that better predict response to pegylated interferon-α and ribavirin therapy in HCV/HIV-1 coinfected patients.’, *PloS one*, 7(2), p. e31016. doi: 10.1371/journal.pone.0031016.

Cento, V., Mirabelli, C., Salpini, R., Dimonte, S., Artese, A., Costa, G., Mercurio, F., Svicher, V., Parrotta, L., Bertoli, A., Ciotti, M., Di Paolo, D., Sarrecchia, C., Andreoni, M., Alcaro, S., Angelico, M., Perno, C. F. and Ceccherini-Silberstein, F. (2012) ‘HCV genotypes are differently prone to the development of resistance to linear and macrocyclic protease inhibitors’, *PLoS ONE*, 7(7). doi: 10.1371/journal.pone.0039652.

Chan, M. C. W., Cheung, C. Y., Chui, W. H., Tsao, S. W., Nicholls, J. M., Chan, Y. O., Chan, R. W. Y., Long, H. T., Poon, L. L. M., Guan, Y. and Peiris, J. S. M. (2005) ‘Proinflammatory cytokine responses induced by influenza A (H5N1) viruses in primary human alveolar and bronchial epithelial cells.’, *Respiratory research*. England, 6, p. 135. doi: 10.1186/1465-9921-6-135.

Chander, G. and McCaul, M. (2003) ‘Co-occurring psychiatric disorders in women with addicitions.’, *Obstetric Obstet Gynecol Clin North Am*, 30(3), pp. 469–81.

Chandler, D. E., Penin, F., Schulten, K. and Chipot, C. (2012) ‘The p7 protein of hepatitis C virus forms structurally plastic, minimalist ion channels’, *PLoS Comput Biol*, 8(9), p. e1002702. doi: 10.1371/journal.pcbi.1002702.

Chayama, K., Notsumata, K., Kurosaki, M., Sato, K., Rodrigues, L., Setze, C., Badri, P., Pilot-Matias, T., Vilchez, R. A. and Kumada, H. (2015) ‘Randomized trial of interferon- and ribavirin-free ombitasvir/paritaprevir/ritonavir in treatment-experienced hepatitis C virus-infected patients’, *Hepatology*, 61(5), pp. 1523–1532. doi: 10.1002/hep.27705.

Chayama, K., Takahashi, S., Toyota, J., Karino, Y., Ikeda, K., Ishikawa, H., Watanabe, H., Mcphee, F., Hughes, E. and Kumada, H. (2012) ‘Dual therapy with the nonstructural protein 5A inhibitor, daclatasvir, and the nonstructural protein 3 protease inhibitor, asunaprevir, in hepatitis C virus genotype 1b-infected null responders’, *Hepatology*, 55(3), pp. 742–748. doi: 10.1002/hep.24724.

Chen, S. and Wang, Y.-M. (2005) ‘Multigene tracking of quasispecies in viral persistence and clearance of hepatitis C virus.’, *World journal of gastroenterology*. China, 11(19), pp. 2874–2884.

Chen, Z.-W., Li, H., Ren, H. and Hu, P. (2016) ‘Global prevalence of pre-existing HCV variants resistant to direct-acting antiviral agents (DAAs): mining the GenBank HCV genome data.’, *Scientific reports*, 6(February), p. 20310. doi: 10.1038/srep20310.

Cheng, B. Y. H., Ortiz-Riano, E., Nogales, A., de la Torre, J. C. and Martinez-Sobrido, L. (2015) ‘Development of live-attenuated arenavirus vaccines based on codon deoptimization.’, *Journal of virology*. United States, 89(7), pp. 3523–3533. doi: 10.1128/JVI.03401-14.

Cheng, X., Virk, N., Chen, W., Ji, S., Ji, S., Sun, Y. and Wu, X. (2013) ‘CpG usage in RNA viruses: data and hypotheses.’, *PloS one*, 8(9), p. e74109. doi: 10.1371/journal.pone.0074109.

Chevaliez, S. (2011) ‘Virological tools to diagnose and monitor hepatitis C virus infection’, *Clinical Microbiology and Infection*. European Society of Clinical Infectious Diseases, 17(2), pp. 116–121. doi: 10.1111/j.1469-0691.2010.03418.x.

Chevaliez, S., Bouvier-Alias, M., Brillet, R. and Pawlotsky, J. M. (2009) ‘Hepatitis C virus (HCV) genotype 1 subtype identification in new HCV drug development and future clinical practice’, *PLoS ONE*, 4(12). doi: 10.1371/journal.pone.0008209.

Chiaramonte, M., Stroffolini, T., Vian, A., Stazi, M. A., Floreani, A., Lorenzoni, U., Lobello, S., Farinati, F. and Naccarato, R. (1999) ‘Rate of incidence of hepatocellular carcinoma in patients with compensated viral cirrhosis.’, *Cancer*. UNITED STATES, 85(10), pp. 2132–2137.

Chien, A., Edgar, D. B. and Trela, J. M. (1976) ‘Deoxyribonucleic acid polymerase from the extreme thermophile Thermus aquaticus.’, *Journal of bacteriology*. UNITED STATES, 127(3), pp. 1550–1557.

Cho, I. and Blaser, M. J. (2012) ‘The human microbiome: at the interface of health and disease.’, *Nature Reviews Genetics*, 13(4), pp. 260–270. doi: 10.1038/nrg3182.

Choo, Q. L., Kuo, G., Weiner, A. J., Overby, L. R., Bradley, D. W. and Houghton, M. (1989) ‘Isolation of a cDNA clone derived from a blood-borne non-A, non-B viral hepatitis genome.’, *Science*, 244(4902), pp. 359–362. doi: 10.1126/science.2523562.

Choo, Q. L., Richman, K. H., Han, J. H., Berger, K., Lee, C., Dong, C., Gallegos, C., Coit, D., Medina-Selby, R., Barr, P. J., Weiner, A. J., Bradley, D. W., Kuo, G. and Houghton, M. (1991) ‘Genetic organization and diversity of the hepatitis C virus.’, *Proceedings of the National Academy of Sciences of the United States of America*, 88(6), pp. 2451–5. doi: 10.1073/pnas.88.6.2451.

Choo, Q. L., Weiner, A. J., Overby, L. R., Kuo, G., Houghton, M. and Bradley, D. W. (1990) ‘Hepatitis C virus: the major causative agent of viral non-A, non-B hepatitis.’, *British Medical Bulletin*. ENGLAND, 46(2), pp. 423–441.

Chung, E., Ferns, R. B., He, M., Rigatti, R., Grant, P., McCormick, A., Bhagani, S.,

Webster, D. P., Nastouli, E. and Waters, L. J. (2015) ‘Ultra-deep sequencing provides insights into the virology of hepatitis C super-infections in a case of three sequential infections with different genotypes’, *Journal of Clinical Virology*. Elsevier B.V., 70(March), pp. 63–66. doi: 10.1016/j.jcv.2015.06.105.

Clark, V. C., Peter, J. A. and Nelson, D. R. (2013) ‘New therapeutic strategies in HCV: Second-generation protease inhibitors’, *Liver International*, 33(SUPPL. 1), pp. 80–84. doi: 10.1111/liv.12061.

Cobb, B., Heilek, G. and Vilchez, R. A. (2014) ‘Molecular diagnostics in the management of chronic hepatitis C: key considerations in the era of new antiviral therapies.’, *BMC infectious diseases*. BioMed Central Ltd, 14 Suppl 5(Suppl 5), p. S8. doi: 10.1186/1471-2334-14-S5-S8.

Coco, B., Oliveri, F., Maina, A. M., Ciccorossi, P., Sacco, R., Colombatto, P., Bonino, F. and Brunetto, M. R. (2007) ‘Transient elastography: a new surrogate marker of liver fibrosis influenced by major changes of transaminases.’, *Journal of viral hepatitis*. England, 14(5), pp. 360–369. doi: 10.1111/j.1365-2893.2006.00811.x.

Cocquerel, L., Quinn, E. R., Flint, M., Hadlock, K. G., Foung, S. K. and Levy, S. (2003) ‘Recognition of native hepatitis C virus E1E2 heterodimers by a human monoclonal antibody’, *J Virol*, 77(2), pp. 1604–1609. doi: 10.1128/JVI.77.2.1604.

Codere, G., Weir, A., Hutchinson, S. and Goldberg, D. (2015) ‘Surveillance of known hepatitis C antibody positive cases: in Scotland: results to 31 December 2014’, *Health Protection Scotland Weekly Report*, 49(28), pp. 264–268.

Colombo, M., Kuo, G., Choo, Q. L., Donato, M. F., Del Ninno, E., Tommasini, M. A., Dioguardi, N. and Houghton, M. (1989) ‘Prevalence of antibodies to hepatitis C virus in Italian patients with hepatocellular carcinoma.’, *Lancet*. ENGLAND, 2(8670), pp. 1006–1008.

Corman, V. M., Grundhoff, A., Baechlein, C., Fischer, N., Gmyl, A., Wollny, R., Dei, D., Ritz, D., Binger, T., Adankwah, E., Marfo, K. S., Annison, L., Annan, A., Adu-Sarkodie, Y., Oppong, S., Becher, P., Drosten, C. and Drexler, J. F. (2015) ‘Highly divergent

hepaciviruses from African cattle', *Journal of Virology*, 89(March), p. JVI.00393-15. doi: 10.1128/JVI.00393-15.

Cornberg, M., Razavi, H. A., Alberti, A., Bernasconi, E., Buti, M., Cooper, C., Dalgard, O., Dillion, J. F., Flisiak, R., Forns, X., Frankova, S., Goldis, A., Goulis, I., Halota, W., Hunyady, B., Lagging, M., Largen, A., Makara, M., Manolakopoulos, S., Marcellin, P., Marinho, R. T., Pol, S., Poynard, T., Puoti, M., Sagalova, O., Sibbel, S., Simon, K., Wallace, C., Young, K., Yurdaydin, C., Zuckerman, E., Negro, F. and Zeuzem, S. (2011) 'A systematic review of hepatitis C virus epidemiology in Europe, Canada and Israel', *Liver International*, 31(SUPPL. 2), pp. 30–60. doi: 10.1111/j.1478-3231.2011.02539.x.

Cox, A. L., Netski, D. M., Mosbruger, T., Sherman, S. G., Strathdee, S., Ompad, D., Vlahov, D., Chien, D., Shyamala, V., Ray, S. C. and Thomas, D. L. (2005) 'Prospective evaluation of community-acquired acute-phase hepatitis C virus infection.', *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America*. United States, 40(7), pp. 951–958. doi: 10.1086/428578.

Craggs, J. K., Ball, J. K., Thomson, B. J., Irving, W. L. and Grabowska, A. M. (2001) 'Development of a strand-specific RT-PCR based assay to detect the replicative form of hepatitis C virus RNA.', *Journal of virological methods*. Netherlands, 94(1–2), pp. 111–120.

Craig, D. W., O'Shaughnessy, J. A., Kiefer, J. A., Aldrich, J., Sinari, S., Moses, T. M., Wong, S., Dinh, J., Christoforides, A., Blum, J. L., Aitelli, C. L., Osborne, C. R., Izatt, T., Kurdoglu, A., Baker, A., Koeman, J., Barbacioru, C., Sakarya, O., Vega, F. M. D. La, Siddiqui, A., Hoang, L., Billings, P. R., Salhia, B., Tolcher, A. W., Trent, J. M., Mousses, S., Hoff, D. Von and Carpten, J. D. (2013) 'Genome and Transcriptome Sequencing in Prospective Metastatic Triple-Negative Breast Cancer Uncovers Therapeutic Vulnerabilities', *Molecular Cancer Therapeutics*, 12(1), pp. 104–116. doi: 10.1158/1535-7163.MCT-12-0781.

Crawford, V., Crome, I. B. and Clancy, C. (2003) 'Co-existing Problems of Mental Health and Substance Misuse [Dual Diagnosis]: A Literature Review', *Drugs: Education, Prevention, and Policy*, 10(November), p. 1. doi: 10.1080/0968763031000072990.

Cristina, J. and Colina, R. (2006) ‘Evidence of structural genomic region recombination in Hepatitis C virus.’, *Virology journal*. England, 3, p. 53. doi: 10.1186/1743-422X-3-53.

Culasso, a C. a, Baré, P., Aloisi, N., Monzani, M. C., Corti, M. and Campos, R. H. (2014) ‘Intra-host evolution of multiple genotypes of hepatitis C virus in a chronically infected patient with HIV along a 13-year follow-up period.’, *Virology*. Elsevier, 449, pp. 317–27. doi: 10.1016/j.virol.2013.11.034.

Cunningham, E. B., Applegate, T. L., Lloyd, A. R., Dore, G. J. and Grebely, J. (2015) ‘Mixed HCV infection and reinfection in people who inject drugs — impact on therapy’, *Nature Publishing Group*. Nature Publishing Group. doi: 10.1038/nrgastro.2015.36.

Dean, F. B., Hosono, S., Fang, L., Wu, X., Faruqi, a F., Bray-Ward, P., Sun, Z., Zong, Q., Du, Y., Du, J., Driscoll, M., Song, W., Kingsmore, S. F., Egholm, M. and Lasken, R. S. (2002) ‘Comprehensive human genome amplification using multiple displacement amplification.’, *Proceedings of the National Academy of Sciences of the United States of America*, 99(8), pp. 5261–5266. doi: 10.1073/pnas.082089499.

Delanty, N., Vaughan, C. J. and French, J. A. (1998) ‘Medical causes of seizures’, *Lancet*, 352(9125), pp. 383–390. doi: 10.1016/S0140-6736(98)02158-8.

Delwart, E. (2013) ‘A roadmap to the human virome.’, *PLoS pathogens*, 9(2), p. e1003146. doi: 10.1371/journal.ppat.1003146.

Deterding, K., Tegtmeyer, B., Cornberg, M., Hadem, J., Potthoff, A., Boker, K. H. W., Tillmann, H. L., Manns, M. P. and Wedemeyer, H. (2006) ‘Hepatitis A virus infection suppresses hepatitis C virus replication and may lead to clearance of HCV.’, *Journal of hepatology*. England, 45(6), pp. 770–778. doi: 10.1016/j.jhep.2006.07.023.

Deuffic-Burban, S., Deltenre, P., Buti, M., Stroffolini, T., Parkes, J., Mühlberger, N., Siebert, U., Moreno, C., Hatzakis, A., Rosenberg, W., Zeuzem, S. and Mathurin, P. (2012) ‘Predicted effects of treatment for HCV infection vary among European countries’, *Gastroenterology*. Elsevier Inc., 143(4), p. 974–985.e14. doi: 10.1053/j.gastro.2012.05.054.

Devalle, S., de Paula, V. S., de Oliveira, J. M., Niel, C. and Gaspar, A. M. C. (2003) 'Hepatitis A virus infection in hepatitis C Brazilian patients.', *The Journal of infection*. England, 47(2), pp. 125–128.

Diaz-San Segundo, F., Medina, G. N., Ramirez-Medina, E., Velazquez-Salinas, L., Koster, M., Grubman, M. J. and de Los Santos, T. (2015) 'Synonymous Deoptimization of Foot-and-Mouth Disease Virus Causes Attenuation In Vivo while Inducing a Strong Neutralizing Antibody Response.', *Journal of virology*. United States, 90(3), pp. 1298–1310. doi: 10.1128/JVI.02167-15.

Dierssen, U., Rehren, F., Henke-Gendo, C., Harste, G. and Heim, A. (2008) 'Rapid routine detection of enterovirus RNA in cerebrospinal fluid by a one-step real-time RT-PCR assay.', *Journal of clinical virology : the official publication of the Pan American Society for Clinical Virology*. Netherlands, 42(1), pp. 58–64. doi: 10.1016/j.jcv.2007.11.016.

Domingo, E. (2007) 'Virus evolution', in *Fields Virology (5th ed.)*, pp. 389–421.

Domingo, E., Sheldon, J. and Perales, C. (2012) 'Viral quasispecies evolution.', *Microbiology and molecular biology reviews : MMBR*, 76(2), pp. 159–216. doi: 10.1128/MMBR.05023-11.

Dore, G. J. and Micallef, J. (2007) 'Low incidence of HCV reinfection: Exposure, testing frequency, or protective immunity? [1]', *Hepatology*, 45(5), p. 1330. doi: 10.1002/hep.21577.

Dou, X.-G., Talekar, G., Chang, J., Dai, X., Li, L., Bonafonte, M.-T., Holloway, B., Fields, H. A. and Khudyakov, Y. E. (2002) 'Antigenic heterogeneity of the hepatitis C virus NS5A protein.', *Journal of clinical microbiology*. United States, 40(1), pp. 61–67.

Drucker, E., Alcabes, P. G. and Marx, P. A. (2001) 'The injection century: massive unsterile injections and the emergence of human pathogens.', *Lancet*. England, 358(9297), pp. 1989–92. doi: 10.1016/S0140-6736(01)06967-7.

Du, H., Qi, Y., Hao, F., Huang, Y., Mao, L., Ji, S., Huang, M., Qin, C., Yan, R., Zhu, X. and Zhang, C. (2012) 'Complex patterns of HCV epidemic in Suzhou: evidence for dual

infection and HCV recombination in East China.', *Journal of clinical virology : the official publication of the Pan American Society for Clinical Virology*. Elsevier B.V., 54(3), pp. 207–12. doi: 10.1016/j.jcv.2012.04.017.

Duret, L. and Galtier, N. (2000) 'The covariation between TpA deficiency, CpG deficiency, and G+C content of human isochores is due to a mathematical artifact.', *Molecular biology and evolution*, 17(11), pp. 1620–5. doi: 10.1093/oxfordjournals.molbev.a026261.

Duvoux, C., Pawlotsky, J.-M., Bastie, A., Cherqui, D., Soussy, C.-J., Dhumeaux, D. and Departments (1999) 'Low HCV replication levels in end-stage hepatitis C virus-related liver disease', *Journal of hepatology*, 31, pp. 593–597.

EASL (2015) 'EASL Recommendations on Treatment of Hepatitis C 2015', *Journal of Hepatology*. European Association for the Study of the Liver, 63(1), pp. 199–236. doi: 10.1016/j.jhep.2015.03.025.

Edgar, R. C. (2004) 'MUSCLE: multiple sequence alignment with high accuracy and high throughput', *Nucleic Acid Research*, 32(5), pp. 1792–1797. doi: 10.1093/nar/gkh340.

Egger, D., Wölk, B., Gosert, R., Bianchi, L., Blum, H. E., Moradpour, D. and Bienz, K. (2002) 'Expression of hepatitis C virus proteins induces distinct membrane alterations including a candidate viral replication complex.', *Journal of virology*, 76(12), pp. 5974–84. doi: 10.1128/JVI.76.12.5974.

El-Serag, H. B. and Rudolph, K. L. (2007) 'Hepatocellular carcinoma: epidemiology and molecular carcinogenesis.', *Gastroenterology*. United States, 132(7), pp. 2557–2576. doi: 10.1053/j.gastro.2007.04.061.

Esmat, G., El-Bendary, M., Zakarya, S., Ela, M. A. and Zalata, K. (2012) 'Role of Helicobacter pylori in patients with HCV-related chronic hepatitis and cirrhosis with or without hepatocellular carcinoma: possible association with disease progression.', *Journal of viral hepatitis*. England, 19(7), pp. 473–479. doi: 10.1111/j.1365-2893.2011.01567.x.

Evans, M. J., Rice, C. M. and Goff, S. P. (2004) 'Genetic Interactions between Hepatitis C

Virus Replicons', *Journal of Virology*, 78(21), pp. 12085–12089. doi: 10.1128/jvi.78.21.12085-12089.2004.

Eyster, M. E., Fried, M. W., Di Bisceglie, A. M. and Goedert, J. J. (1994) 'Increasing hepatitis C virus RNA levels in hemophiliacs: relationship to human immunodeficiency virus infection and liver disease. Multicenter Hemophilia Cohort Study.', *Blood*. UNITED STATES, 84(4), pp. 1020–1023.

Fan, X., Lang, D. M., Xu, Y., Lyra, A. C., Yusim, K., Everhart, J. E., Korber, B. T. M., Perelson, A. S. and Di Bisceglie, A. M. (2003) 'Liver transplantation with hepatitis C virus-infected graft: interaction between donor and recipient viral strains.', *Hepatology*, 38(1), pp. 25–33. doi: 10.1053/jhep.2003.50264.

Farci, P. (2000) 'The Outcome of Acute Hepatitis C Predicted by the Evolution of the Viral Quasispecies', *Science*, 288(5464), pp. 339–344. doi: 10.1126/science.288.5464.339.

Farci, P., Alter, H. J., Govindarajan, S., Wong, D. C., Engle, R., Lesniewski, R. R., Mushahwar, I. K., Desai, S. M., Miller, R. H. and Ogata, N. (1992) 'Lack of protective immunity against reinfection with hepatitis C virus.', *Science*. UNITED STATES, 258(5079), pp. 135–140.

Fattovich, G., Stroffolini, T., Zagni, I. and Donato, F. (2004) 'Hepatocellular carcinoma in cirrhosis: incidence and risk factors.', *Gastroenterology*. United States, 127(5 Suppl 1), pp. S35-50.

Feinstone, S. M., Mihalik, K. B., Kamimura, T., Alter, H. J., London, W. T. and Purcell, R. H. (1983) 'Inactivation of hepatitis B virus and non-A, non-B hepatitis by chloroform.', *Infection and immunity*, 41(2), pp. 816–21.

Finney, D. J. (1952) *Probit analysis: a statistical treatmentof thesigmoid response curve*. Cambridge University Press.

Fonseca-Coronado, S., Escobar-Gutierrez, A., Ruiz-Tovar, K., Cruz-Rivera, M. Y., Rivera-Osorio, P., Vazquez-Pichardo, M., Carpio-Pedroza, J. C., Ruiz-Pacheco, J. A., Cazares, F. and Vaughan, G. (2012) 'Specific detection of naturally occurring hepatitis C virus

mutants with resistance to telaprevir and boceprevir (protease inhibitors) among treatment-naive infected individuals.', *Journal of clinical microbiology*. United States, 50(2), pp. 281–287. doi: 10.1128/JCM.05842-11.

Forton, D. M., Thomas, H. C., Murphy, C. a., Allsop, J. M., Foster, G. R., Main, J., Wesnes, K. a. and Taylor-Robinson, S. D. (2002) 'Hepatitis C and cognitive impairment in a cohort of patients with mild liver disease', *Hepatology*, 35(2), pp. 433–439. doi: 10.1053/jhep.2002.30688.

Foster, G. R., Hezode, C., Bronowicki, J.-P., Carosi, G., Weiland, O., Verlinden, L., van Heeswijk, R., van Baelen, B., Picchio, G. and Beumont, M. (2011) 'Telaprevir alone or with peginterferon and ribavirin reduces HCV RNA in patients with chronic genotype 2 but not genotype 3 infections.', *Gastroenterology*. United States, 141(3), p. 881–889.e1. doi: 10.1053/j.gastro.2011.05.046.

Foy, E., Li, K., Wang, C., Sumpter, R. J., Ikeda, M., Lemon, S. M. and Gale, M. J. (2003) 'Regulation of interferon regulatory factor-3 by the hepatitis C virus serine protease.', *Science*. United States, 300(5622), pp. 1145–1148. doi: 10.1126/science.1082604.

Frank, C., Mohamed, M. K., Strickland, G. T., Lavanchy, D., Arthur, R. R., Magder, L. S., El Khoby, T., Abdel-Wahab, Y., Aly Ohn, E. S., Anwar, W. and Sallam, I. (2000) 'The role of parenteral antischistosomal therapy in the spread of hepatitis C virus in Egypt.', *Lancet*. ENGLAND, 355(9207), pp. 887–891.

Freeman, A. J., Pan, Y., Harvey, C. E., Post, J. J., Law, M. G., White, P. A., Rawlinson, W. D., Lloyd, A. R., Marinos, G. and Ffrench, R. A. (2003) 'The presence of an intrahepatic cytotoxic T lymphocyte response is associated with low viral load in patients with chronic hepatitis C virus infection.', *Journal of hepatology*. England, 38(3), pp. 349–356.

Friebe, P. and Bartenschlager, R. (2002) 'Genetic analysis of sequences in the 3' nontranslated region of hepatitis C virus that are important for RNA replication.', *Journal of virology*, 76(11), pp. 5326–38. doi: 10.1128/JVI.76.11.5326.

Friebe, P., Lohmann, V., Krieger, N. and Bartenschlager, R. (2001) 'Sequences in the 5 ′

Nontranslated Region of Hepatitis C Virus Required for RNA Replication', *Journal of Virology*, 75(24), pp. 12047–12057. doi: 10.1128/JVI.75.24.12047.

Frischer, M., Leyland, A. H., Cormack, R., Goldberg, D. J., Bloor, M., Green, S. T., Taylor, A., Covell, R., McKeganey, N. P. and Platt, S. (1993) 'Estimating the population of injection drug use and infection with human immunodeficiency virus among injection drug users in Glasgow, Scotland', *American Journal of Epidemiology*, 138(3), pp. 170–181.

Fu, L., Niu, B., Zhu, Z., Wu, S. and Li, W. (2012) 'CD-HIT: Accelerated for clustering the next-generation sequencing data', *Bioinformatics*, 28(23), pp. 3150–3152. doi: 10.1093/bioinformatics/bts565.

Gaeta, G. B., Stornaiuolo, G., Precone, D. F., Lobello, S., Chiaramonte, M., Stroffolini, T., Colucci, G. and Rizzetto, M. (2003) 'Epidemiological and clinical burden of chronic hepatitis B virus/hepatitis C virus infection. A multicenter Italian study.', *Journal of hepatology*. England, 39(6), pp. 1036–1041.

Gale, M. J., Korth, M. J., Tang, N. M., Tan, S. L., Hopkins, D. a, Dever, T. E., Polyak, S. J., Gretch, D. R. and Katze, M. G. (1997) 'Evidence that hepatitis C virus resistance to interferon is mediated through repression of the PKR protein kinase by the nonstructural 5A protein.', *Virology*, 230(2), pp. 217–227. doi: 10.1006/viro.1997.8493.

Gao, M., Nettles, R. E., Belema, M., Snyder, L. B., Nguyen, V. N., Fridell, R. A., Serrano-Wu, M. H., Langley, D. R., Sun, J.-H., O'Boyle II, D. R., Lemm, J. A., Wang, C., Knipe, J. O., Chien, C., Colonno, R. J., Grasela, D. M., Meanwell, N. A. and Hamann, L. G. (2010) 'Chemical genetics strategy identifies an HCV NS5A inhibitor with a potent clinical effect', *Nature*. Macmillan Publishers Limited. All rights reserved, 465(7294), pp. 96–100.

Garber, K. (2011) 'Hepatitis C: move over interferon', *Nature Biotechnology*, 29(11), pp. 963–966. doi: 10.1038/nbt.2031.

Gardner, J. P., Durso, R. J., Arrigale, R. R., Donovan, G. P., Maddon, P. J., Dragic, T. and Olson, W. C. (2003) 'L-SIGN (CD 209L) is a liver-specific capture receptor for hepatitis C virus.', *Proceedings of the National Academy of Sciences of the United States of America*,

100(8), pp. 4498–4503. doi: 10.1073/pnas.0831128100.

Ge, D., Fellay, J., Thompson, A. J., Simon, J. S., Shianna, K. V, Urban, T. J., Heinzen, E. L., Qiu, P., Bertelsen, A. H., Muir, A. J., Sulkowski, M., McHutchison, J. G. and Goldstein, D. B. (2009) 'Genetic variation in IL28B predicts hepatitis C treatment-induced viral clearance.', *Nature*. Nature Publishing Group, 461(7262), pp. 399–401. doi: 10.1038/nature08309.

George, S. L., Bacon, B. R., Brunt, E. M., Mihindukulasuriya, K. L., Hoffmann, J. and Di Bisceglie, A. M. (2009) 'Clinical, virologic, histologic, and biochemical outcomes after successful HCV therapy: a 5-year follow-up of 150 patients.', *Hepatology (Baltimore, Md.)*. United States, 49(3), pp. 729–738. doi: 10.1002/hep.22694.

Gerlach, J. T., Diepolder, H. M., Jung, M. C., Gruener, N. H., Schraut, W. W., Zachoval, R., Hoffmann, R., Schirren, C. A., Santantonio, T. and Pape, G. R. (1999) 'Recurrence of hepatitis C virus after loss of virus-specific CD4(+) T-cell response in acute hepatitis C.', *Gastroenterology*. UNITED STATES, 117(4), pp. 933–941.

Gerlach, J. T., Ulsenheimer, a, Grüner, N. H., Jung, M.-C., Schraut, W., Schirren, C., Heeg, M., Scholz, S., Witter, K., Zahn, R., Vogler, a, Zachoval, R., Pape, G. R. and Diepolder, H. M. (2005) 'Minimal T-cell-stimulatory sequences and spectrum of HLA restriction of immunodominant CD4+ T-cell epitopes within hepatitis C virus NS3 and NS4 proteins.', *Journal of virology*, 79(19), pp. 12425–12433. doi: 10.1128/JVI.79.19.12425-12433.2005.

Giannini, C., Giannelli, F., Monti, M., Careccia, G., Marrocchi, M. E., Laffi, G., Gentilini, P. and Zignego, a L. (1999) 'Prevalence of mixed infection by different hepatitis C virus genotypes in patients with hepatitis C virus-related chronic liver disease.', *The Journal of laboratory and clinical medicine*, 134(1), pp. 68–73.

Gilbert, W. and Maxam, A. (1973) 'The nucleotide sequence of the lac operator.', *Proceedings of the National Academy of Sciences of the United States of America*. UNITED STATES, 70(12), pp. 3581–3584.

Glynn, S. A., Wright, D. J., Kleinman, S. H., Hirschkorn, D., Tu, Y., Heldebrant, C.,

Smith, R., Giachetti, C., Gallarda, J. and Busch, M. P. (2005) 'Dynamics of viremia in early hepatitis C virus infection.', *Transfusion*. United States, 45(6), pp. 994–1002. doi: 10.1111/j.1537-2995.2005.04390.x.

Goldberg, D. and Anderson, E. (2004) 'Hepatitis C: who is at risk and how do we identify them?.', *Journal of viral hepatitis*, 11 Suppl 1, pp. 12–18.

Goncalves Rossi, L. M., Escobar-Gutierrez, A. and Rahal, P. (2016) 'Multiregion deep sequencing of hepatitis C virus: An improved approach for genetic relatedness studies.', *Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases*. Netherlands, 38, pp. 138–145. doi: 10.1016/j.meegid.2015.12.020.

González, V., Gomes-Fernandes, M., Bascuñana, E., Casanovas, S., Saludes, V., Jordana-Lluch, E., Matas, L., Ausina, V. and Martró, E. (2013) 'Accuracy of a commercially available assay for HCV genotyping and subtyping in the clinical practice', *Journal of Clinical Virology*, 58, pp. 593–597. doi: 10.1016/j.jcv.2013.09.010.

González-Candelas, F., López-Labrador, F. X. and Bracho, M. A. (2011) 'Recombination in hepatitis C virus.', *Viruses*, 3(10), pp. 2006–24. doi: 10.3390/v3102006.

Goodwin, S., McPherson, J. D. and McCombie, W. R. (2016) 'Coming of age: ten years of next-generation sequencing technologies', *Nature Reviews Genetics*. Nature Publishing Group, 17(6), pp. 333–351. doi: 10.1038/nrg.2016.49.

Goossens, N. and Negro, F. (2014) 'Is genotype 3 of the hepatitis C virus the new villain?', *Hepatology*, 59(6), pp. 2403–2412. doi: 10.1002/hep.26905.

Gottwein, J. M., Scheel, T. K. H., Callendret, B., Li, Y.-P., Eccleston, H. B., Engle, R. E., Govindarajan, S., Satterfield, W., Purcell, R. H., Walker, C. M. and Bukh, J. (2010) 'Novel infectious cDNA clones of hepatitis C virus genotype 3a (strain S52) and 4a (strain ED43): genetic analyses and in vivo pathogenesis studies.', *Journal of virology*, 84(10), pp. 5277–93. doi: 10.1128/JVI.02667-09.

Gottwein, J. M., Scheel, T. K. H., Jensen, T. B., Ghanem, L. and Bukh, J. (2011)

‘Differential efficacy of protease inhibitors against HCV genotypes 2a, 3a, 5a, and 6a NS3/4A protease recombinant viruses.’, *Gastroenterology*. Elsevier Inc., 141(3), pp. 1067–79. doi: 10.1053/j.gastro.2011.06.004.

Gotz, H. M., van Doornum, G., Niesters, H. G., den Hollander, J. G., Thio, H. B. and de Zwart, O. (2005) ‘A cluster of acute hepatitis C virus infection among men who have sex with men--results from contact tracing and public health implications.’, *AIDS*. England, 19(9), pp. 969–974.

Gouttenoire, J., Castet, V., Montserret, R., Arora, N., Raussens, V., Ruysschaert, J.-M., Diesis, E., Blum, H. E., Penin, F. and Moradpour, D. (2009) ‘Identification of a novel determinant for membrane association in hepatitis C virus nonstructural protein 4B.’, *Journal of virology*, 83(12), pp. 6257–68. doi: 10.1128/JVI.02663-08.

Gower, E., Estes C, C., Hindman, S., Razavi-Shearer, K. and Razavi, H. (2014) ‘Global epidemiology and genotype distribution of the hepatitis C virus.’, *Journal of hepatology*. European Association for the Study of the Liver, 61(1), pp. S45–S57. doi: 10.1016/j.jhep.2014.07.027.

Gowin, E., Bereszyńska, I., Adamek, A., Kowala-Piaskowska, A., Mozer-Lisewska, I., Wysocki, J., Michalak, M. and Januszkiewicz-Lewandowska, D. (2016) ‘The prevalence of mixed genotype infections in Polish patients with hepatitis C’, *International Journal of Infectious Diseases*, 43, pp. 13–16. doi: 10.1016/j.ijid.2015.12.007.

Grakoui, a, McCourt, D. W., Wychowski, C., Feinstone, S. M. and Rice, C. M. (1993) ‘Characterization of the hepatitis C virus-encoded serine proteinase: determination of proteinase-dependent polyprotein cleavage sites.’, *Journal of virology*, 67(5), pp. 2832–2843.

Gramenzi, A., Conti, F., Felline, F., Cursaro, C., Riili, A., Salerno, M., Gitto, S., Micco, L., Scuteri, A., Andreone, P. and Bernardi, M. (2010) ‘Hepatitis C Virus-related chronic liver disease in elderly patients: an Italian cross-sectional study.’, *Journal of viral hepatitis*. England, 17(5), pp. 360–366. doi: 10.1111/j.1365-2893.2009.01189.x.

Grard, G., Fair, J. N., Lee, D., Slikas, E., Steffen, I., Muyembe, J. J., Sittler, T.,

Veeraraghavan, N., Ruby, J. G., Wang, C., Makuwa, M., Mulembakani, P., Tesh, R. B., Mazet, J., Rimoin, A. W., Taylor, T., Schneider, B. S., Simmons, G., Delwart, E., Wolfe, N. D., Chiu, C. Y. and Leroy, E. M. (2012) 'A Novel Rhabdovirus Associated with Acute Hemorrhagic Fever in Central Africa', *PLoS Pathogens*, 8(9). doi: 10.1371/journal.ppat.1002924.

Gray, R. R., Parker, J., Lemey, P., Salemi, M., Katzourakis, A. and Pybus, O. G. (2011) 'The mode and tempo of hepatitis C virus evolution within and among hosts.', *BMC evolutionary biology*, 11(1), p. 131. doi: 10.1186/1471-2148-11-131.

Gray, R. R., Salemi, M., Klenerman, P. and Pybus, O. G. (2012) 'A new evolutionary model for hepatitis C virus chronic infection.', *PLoS pathogens*, 8(5), p. e1002656. doi: 10.1371/journal.ppat.1002656.

Grebely, J., Conway, B., Raffa, J. D., Lai, C., Krajden, M. and Tyndall, M. W. (2006) 'Hepatitis C virus reinfection in injection drug users.', *Hepatology*. United States, 44(5), pp. 1139–1145. doi: 10.1002/hep.21376.

Grebely, J. and Dore, G. J. (2011) 'What is killing people with hepatitis C virus infection?', *Seminars in liver disease*. United States, 31(4), pp. 331–339. doi: 10.1055/s-0031-1297922.

Grebely, J., Pham, S. T., Matthews, G. V, Petoumenos, K., Bull, R. A., Yeung, B., Rawlinson, W., Kaldor, J., Lloyd, A., Hellard, M., Dore, G. J. and White, P. A. (2012) 'Hepatitis C virus reinfection and superinfection among treated and untreated participants with recent infection', *Hepatology*, 55(4), pp. 1058–1069. doi: 10.1002/hep.24754.

Grebely, J., Prins, M., Hellard, M., Cox, A. L., Osburn, W. O., Lauer, G., Page, K., Lloyd, A. R. and Dore, G. J. (2012) 'Hepatitis C virus clearance, reinfection, and persistence, with insights from studies of injecting drug users: towards a vaccine', *The Lancet Infectious Diseases*, 12(5), pp. 408–414. doi: 10.1016/s1473-3099(12)70010-5.

Greenbaum, B. D., Levine, A. J., Bhanot, G. and Rabadan, R. (2008) 'Patterns of evolution and host gene mimicry in influenza and other RNA viruses', *PLoS Pathogens*, 4(6), pp. 1–9. doi: 10.1371/journal.ppat.1000079.

Guadalupe, M., Sankaran, S., George, M. D., Reay, E., Verhoeven, D., Shacklett, B. L., Flamm, J., Wegelin, J., Prindiville, T. and Dandekar, S. (2006) 'Viral suppression and immune restoration in the gastrointestinal mucosa of human immunodeficiency virus type 1-infected patients initiating therapy during primary or chronic infection.', *Journal of virology*, 80(16), pp. 8236–8247. doi: 10.1128/JVI.00120-06.

Guedj, J., Rong, L., Dahari, H. and Perelson, A. S. (2010) 'A perspective on modelling hepatitis C virus infection.', *Journal of viral hepatitis*. England, 17(12), pp. 825–833. doi: 10.1111/j.1365-2893.2010.01348.x.

Gutierrez, J. A., Lawitz, E. J. and Poordad, F. (2015) 'Interferon-free, direct-acting antiviral therapy for chronic hepatitis C.', *Journal of viral hepatitis*. England, 22(11), pp. 861–870. doi: 10.1111/jvh.12422.

Hairul Aini, H., Mustafa, M. I. a, Seman, M. R. and Nasuruddin, B. a (2012) 'Mixed-genotypes infections with hepatitis C virus in hemodialysis subjects.', *The Medical journal of Malaysia*, 67(2), pp. 199–203.

Hajarizadeh, B., Grebely, J. and Dore, G. J. (2013) 'Epidemiology and natural history of HCV infection.', *Nature reviews. Gastroenterology & hepatology*. Nature Publishing Group, 10(9), pp. 553–62. doi: 10.1038/nrgastro.2013.107.

Halfon, P., Bourlière, M., Halimi, G., Khiri, H., Bertezene, P., Portal, I., Botta-Fridlund, D., Gauthier, a P., Jullien, M., Feryn, J. M., Gerolami, V. and Cartouzou, G. (1998) 'Assessment of spontaneous fluctuations of viral load in untreated patients with chronic hepatitis C by two standardized quantitation methods: branched DNA and Amplicor Monitor.', *Journal of clinical microbiology*, 36(7), pp. 2073–5.

Harris, H. E., Eldridge, K. P., Harbour, S., Alexander, G., Teo, C.-G. G. and Ramsay, M. E. (2007) 'Does the clinical outcome of hepatitis C infection vary with the infecting hepatitis C virus type?', *Journal of Viral Hepatitis*. England, 14(3), pp. 213–220. doi: 10.1111/j.1365-2893.2006.00795.x.

Harris, R. J., Ramsay, M., Hope, V. D., Brant, L., Hickman, M., Foster, G. R. and De Angelis, D. (2012) 'Hepatitis C prevalence in England remains low and varies by ethnicity:

an updated evidence synthesis.', *European journal of public health*, 22(2), pp. 187–92. doi: 10.1093/eurpub/ckr083.

Hashimshony, T., Wagner, F., Sher, N. and Yanai, I. (2012) 'CEL-Seq: single-cell RNA-Seq by multiplexed linear amplification.', *Cell reports*. United States, 2(3), pp. 666–673. doi: 10.1016/j.celrep.2012.08.003.

Hay, G., Gannon, M., Casey, J. and McKeganey, N. (2009) *Estimating the National and Local Prevalence of Problem Drug Misuse in Scotland*.

He, L. F., Alling, D., Popkin, T., Shapiro, M., Alter, H. J. and Purcell, R. H. (1987) 'Determining the size of non-A, non-B hepatitis virus by filtration.', *The Journal of infectious diseases*. UNITED STATES, 156(4), pp. 636–640.

Heckmann, J. G., Kayser, C., Heuss, D., Manger, B., Blum, H. E. and Neundörfer, B. (1999) 'Neurological manifestations of chronic hepatitis C', *Journal of Neurology*, 246(6), pp. 486–491. doi: 10.1007/s004150050388.

Hedskog, C., Doehle, B., Chodavarapu, K., Gontcharova, V., Crespo Garcia, J., De Knegt, R., Drenth, J. P. H., McHutchison, J. G., Brainard, D., Stamm, L. M., Miller, M. D., Svarovskaia, E. and Mo, H. (2014) 'Characterization of hepatitis C virus intergenotypic recombinant strains and associated virological response to sofosbuvir/ribavirin.', *Hepatology (Baltimore, Md.)*, pp. 1–10. doi: 10.1002/hep.27361.

Hellard, M. E., Aitken, C. K. and Hocking, J. S. (2007) 'Tattooing in prisons--not such a pretty picture.', *American journal of infection control*. United States, 35(7), pp. 477–480. doi: 10.1016/j.ajic.2006.08.002.

Hellard, M. E., Hocking, J. S. and Crofts, N. (2004) 'The prevalence and the risk behaviours associated with the transmission of hepatitis C virus in Australian correctional facilities.', *Epidemiology and infection*. England, 132(3), pp. 409–415.

Hellard, M., McBryde, E., Sacks Davis, R., Rolls, D. A., Higgs, P., Aitken, C., Thompson, A., Doyle, J., Pattison, P. and Robins, G. (2015) 'Hepatitis C transmission and treatment as prevention - The role of the injecting network', *International Journal of Drug Policy*.

Elsevier B.V., 26(10), pp. 958–962. doi: 10.1016/j.drugpo.2015.05.006.

Herring, B. L., Page-Shafer, K., Tobler, L. H. and Delwart, E. L. (2004) 'Frequent hepatitis C virus superinfection in injection drug users.', *The Journal of infectious diseases*, 190(8), pp. 1396–1403. doi: 10.1086/424491.

Hilsabeck, R. C., Hassanein, T. I., Carlson, M. D., Ziegler, E. A. and Perry, W. (2003) 'Cognitive functioning and psychiatric symptomatology in patients with chronic hepatitis C', *J Int Neuropsychol Soc*, 9(6), pp. 847–854. doi: 10.1017/s1355617703960048.

Hofmann, W. P. and Zeuzem, S. (2011) 'A new standard of care for the treatment of chronic HCV infection.', *Nature reviews. Gastroenterology & hepatology*. England, 8(5), pp. 257–264. doi: 10.1038/nrgastro.2011.49.

Holland, J. J., De La Torre, J. C. and Steinhauer, D. A. (1992) 'RNA virus populations as quasispecies.', *Current topics in microbiology and immunology*. GERMANY, 176, pp. 1–20.

Holland, P. V and Alter, H. (1981) 'Non-A , Non-B Viral Hepatitis', *Human Pathology*. W. B. Saunders Company, 12(12), pp. 1114–1122. doi: 10.1016/S0046-8177(81)80332-2.

Hollinger, F. B., Gitnick, G. L., Aach, R. D., Szmuness, W., Mosley, J. W., Stevens, C. E., Peters, R. L., Weiner, J. M., Werch, J. B. and Lander, J. J. (1978) 'Non-A, non-B hepatitis transmission in chimpanzees: a project of the transfusion-transmitted viruses study group.', *Intervirology*. SWITZERLAND, 10(1), pp. 60–68.

Honda, M., Beard, M. R., Ping, L. H. and Lemon, S. M. (1999) 'A phylogenetically conserved stem-loop structure at the 5' border of the internal ribosome entry site of hepatitis C virus is required for cap-independent viral translation.', *Journal of virology*, 73(2), pp. 1165–74.

Honda, M., Brown, E. A. and Lemon, S. M. (1996) 'Stability of a stem-loop involving the initiator AUG controls the efficiency of internal initiation of translation on hepatitis C virus RNA.', *RNA (New York, N.Y.)*. UNITED STATES, 2(10), pp. 955–968.

Hong, S., Bunge, J., Leslin, C., Jeon, S. and Epstein, S. S. (2009) 'Polymerase chain reaction primers miss half of rRNA microbial diversity.', *The ISME journal*. Nature Publishing Group, 3(12), pp. 1365–1373. doi: 10.1038/ismej.2009.89.

Hoofnagle, J. H., Gerety, R. J., Tabor, E., Feinstone, S. M., Barker, L. F. and Purcell, R. H. (1977) 'Transmission of Non-A , Non-B Hepatitis', *Annals of Internal Medicine*, 87, pp. 14–20.

Horner, S. M. and Gale, M. (2013) 'Regulation of hepatic innate immunity by hepatitis C virus.', *Nature medicine*, 19(7), pp. 879–88. doi: 10.1038/nm.3253.

Hosseini-Moghaddam, S. M., Iran-Pour, E., Rotstein, C., Husain, S., Lilly, L., Renner, E. and Mazzulli, T. (2012) 'Hepatitis C core Ag and its clinical applicability: potential advantages and disadvantages for diagnosis and follow-up?', *Reviews in medical virology*. England, 22(3), pp. 156–165. doi: 10.1002/rmv.717.

Hraber, P. T., Fischer, W., Bruno, W. J., Leitner, T. and Kuiken, C. (2006) 'Comparative analysis of hepatitis C virus phylogenies from coding and non-coding regions: the 5' untranslated region (UTR) fails to classify subtypes.', *Virology journal*, 3, p. 103. doi: 10.1186/1743-422X-3-103.

Hsu, L.-C., Mo Park, J., Zhang, K., Luo, J.-L., Maeda, S., Kaufman, R. J., Eckmann, L., Guiney, D. G. and Karin, M. (2004) 'The protein kinase PKR is required for macrophage apoptosis after activation of Toll-like receptor 4', *Nature*, 428(6980), pp. 341–345.

Hsu, Y.-C., Lin, J.-T., Ho, H. J., Kao, Y.-H., Huang, Y.-T., Hsiao, N.-W., Wu, M.-S., Liu, Y.-Y. and Wu, C.-Y. (2014) 'Antiviral treatment for hepatitis C virus infection is associated with improved renal and cardiovascular outcomes in diabetic patients', *Hepatology*, 59(4), pp. 1293–1302. doi: 10.1002/hep.26892.

Hua, L., Andersen, J. W., Daar, E. S., Glesby, M. J., Hollabaugh, K. and Tierney, C. (2013) 'Hepatitis C virus/HIV coinfection and responses to initial antiretroviral treatment.', *AIDS*. England, 27(17), pp. 2725–2734. doi: 10.1097/01.aids.0000432470.46379.dd.

Humphreys, I., Fleming, V., Fabris, P., Parker, J., Schulenberg, B., Brown, A., Demetriou,

C., Gaudieri, S., Pfafferott, K., Lucas, M., Collier, J., Huang, K.-H. G., Pybus, O. G., Klenerman, P. and Barnes, E. (2009) 'Full-length characterization of hepatitis C virus subtype 3a reveals novel hypervariable regions under positive selection during acute infection.', *Journal of virology*, 83(22), pp. 11456–66. doi: 10.1128/JVI.00884-09.

Hutchinson, S. J., Bird, S. M. and Goldberg, D. J. (2005) 'Influence of alcohol on the progression of hepatitis C virus infection: a meta-analysis.', *Clinical gastroenterology and hepatology : the official clinical practice journal of the American Gastroenterological Association*. United States, 3(11), pp. 1150–1159.

Hutchinson, S. J., McIntyre, P. G., Molyneaux, P., Cameron, S., Burns, S., Taylor, a and Goldberg, D. J. (2002) 'Prevalence of hepatitis C among injectors in Scotland 1989-2000: declining trends among young injectors halt in the late 1990s.', *Epidemiology and infection*, 128(3), pp. 473–477. doi: 10.1017/S0950268802006945.

Hutchinson, S., KM Roy, Wadd, S., Bird, S., A Taylor, Anderson, E., Shaw, L., Codere, G. and Goldberg, D. (2006) 'Hepatitis C virus infection in Scotland: Epidemiological review and public health challenges', *Scottish Medical Journal*, 51(2), pp. 8–15.

Imhof, I. and Simmonds, P. (2010) 'Development of an intergenotypic hepatitis C virus (HCV) cell culture method to assess antiviral susceptibilities and resistance development of HCV NS3 protease genes from HCV genotypes 1 to 6.', *Journal of virology*, 84(9), pp. 4597–610. doi: 10.1128/JVI.02698-09.

Innes, H. A., Hutchinson, S. J., Allen, S., Bhattacharyya, D., Bramley, P., Carman, B., Delahooke, T. E. S., Dillon, J. F., Goldberg, D. J., Kennedy, N., Mills, P. R., Morris, J., Morris, J., Robertson, C., Stanley, A. J. and Hayes, P. (2012) 'Ranking predictors of a sustained viral response for patients with chronic hepatitis C treated with pegylated interferon and ribavirin in Scotland.', *European journal of gastroenterology & hepatology*, 24(6), pp. 646–55. doi: 10.1097/MEG.0b013e32835201a4.

Ito, H., Yoshioka, K., Ukai, K., Watanabe, K., Yano, M., Ishigami, M., Mizutani, T., Sasaki, Y., Katano, Y. and Goto, H. (2004) 'The fluctuations of viral load and serum alanine aminotransferase levels in chronic hepatitis C', *Hepatology Research*, 30(1), pp. 11–17. doi: http://dx.doi.org/10.1016/j.hepres.2004.06.002.

di Iulio, J., Ciuffi, A., Fitzmaurice, K., Kelleher, D., Rotger, M., Fellay, J., Martinez, R., Pulit, S., Furrer, H., Günthard, H. F., Battegay, M., Bernasconi, E., Schmid, P., Hirschel, B., Barnes, E., Klenerman, P., Telenti, A. and Rauch, A. (2011) 'Estimating the net contribution of interleukin-28B variation to spontaneous hepatitis C virus clearance.', *Hepatology (Baltimore, Md.)*, 53(5), pp. 1446–54. doi: 10.1002/hep.24263.

Jabara, C. B., Jones, C. D., Roach, J., Anderson, J. A. and Swanstrom, R. (2011) 'Accurate sampling and deep sequencing of the HIV-1 protease gene using a Primer ID.', *Proceedings of the National Academy of Sciences of the United States of America*, 108(50), pp. 20166–71. doi: 10.1073/pnas.1110064108.

Jacka, B., Lamoury, F., Simmonds, P., Dore, G. J., Grebely, J. and Applegate, T. (2013) 'Sequencing of the Hepatitis C Virus: A Systematic Review.', *PloS one*, 8(6). doi: 10.1371/journal.pone.0067073.

Jackel-Cram, C., Babiuk, L. a. and Liu, Q. (2007) 'Up-regulation of fatty acid synthase promoter by hepatitis C virus core protein: Genotype-3a core has a stronger effect than genotype-1b core', *Journal of Hepatology*, 46, pp. 999–1008. doi: 10.1016/j.jhep.2006.10.019.

Jardi, R., Rodriguez, F., Buti, M., Costa, X., Cotrina, M., Galimany, R., Esteban, R. and Guardia, J. (2001) 'Role of hepatitis B, C, and D viruses in dual and triple infection: influence of viral genotypes and hepatitis B precore and basal core promoter mutations on viral replicative interference.', *Hepatology (Baltimore, Md.)*. United States, 34(2), pp. 404–410. doi: 10.1053/jhep.2001.26511.

Jeannel, D., Fretz, C., Traore, Y., Kohdjo, N., Bigot, A., Pe Gamy, E., Jourdan, G., Kourouma, K., Maertens, G., Fumoux, F., Fournel, J. J. and Stuyver, L. (1998) 'Evidence for high genetic diversity and long-term endemicity of hepatitis C virus genotypes 1 and 2 in West Africa.', *Journal of medical virology*. UNITED STATES, 55(2), pp. 92–97.

Jenkins, G. M. and Holmes, E. C. (2003) 'The extent of codon usage bias in human RNA viruses and its evolutionary origin', *Virus Research*, 92, pp. 1–7. doi: 10.1016/S0168-1702(02)00309-X.

Jiang, J. and Luo, G. (2009) ‘Apolipoprotein E but not B is required for the formation of infectious hepatitis C virus particles.’, *Journal of virology*, 83(24), pp. 12680–12691. doi: 10.1128/JVI.01476-09.

Jirasko, V., Montserret, R., Lee, J. Y., Gouttenoire, J., Moradpour, D., Penin, F. and Bartenschlager, R. (2010) ‘Structural and functional studies of nonstructural protein 2 of the hepatitis C virus reveal its key role as organizer of virion assembly’, *PLoS Pathog*, 6(12), p. e1001233. doi: 10.1371/journal.ppat.1001233.

Jones, D. M., Domingues, P., Targett-Adams, P. and McLauchlan, J. (2010) ‘Comparison of U2OS and Huh-7 cells for identifying host factors that affect hepatitis C virus RNA replication’, *Journal of General Virology*, 91(9), pp. 2238–2248. doi: 10.1099/vir.0.022210-0.

JOSSE, J., KAISER, A. D. and KORNBERG, A. (1961) ‘Frequencies of nearest neighbor base sequences in deoxyribonucleic acid.’, *The Journal of biological chemistry*. Not Available, 236, pp. 864–875.

Judd, A., Hutchinson, S., Wadd, S., Hickman, M., Taylor, A., Jones, S., Parry, J. V, Cameron, S., Rhodes, T., Ahmed, S., Bird, S., Fox, R., Renton, A., Stimson, G. V and Goldberg, D. (2005) ‘Prevalence of, and risk factors for, hepatitis C virus infection among recent initiates to injecting in London and Glasgow: cross sectional analysis’, *Journal of Viral Hepatitis*, 12(6), pp. 655–662. doi: 10.1111/j.1365-2893.2005.00643.x.

Kamal, S. M., Turner, B., He, Q., Rasenack, J., Bianchi, L., Al Tawil, A., Nooman, A., Massoud, M., Koziel, M. J. and Afdhal, N. H. (2006) ‘Progression of fibrosis in hepatitis C with and without schistosomiasis: Correlation with serum markers of fibrosis’, *Hepatology*, 43(4), pp. 771–779. doi: 10.1002/hep.21117.

Kamili, S., Drobeniuc, J., Araujo, A. C. and Hayden, T. M. (2012) ‘Laboratory diagnostics for hepatitis C virus infection.’, *Clinical Infectious diseases*, 55 Suppl 1(Suppl 1), pp. 43–48. doi: 10.1093/cid/cis368.

Kanwal, F., Kramer, J. R., Ilyas, J., Duan, Z. and El-Serag, H. B. (2014) ‘HCV genotype 3 is associated with an increased risk of cirrhosis and hepatocellular cancer in a national

sample of U.S. Veterans with HCV', *Hepatology*, 60(1), pp. 98–105. doi: 10.1002/hep.27095.

Kao, J. H., Chen, P. J., Lai, M. Y., Yang, P. M., Sheu, J. C., Wang, T. H. and Chen, D. S. (1994) 'Mixed infections of hepatitis C virus as a factor in acute exacerbations of chronic type C hepatitis.', *The Journal of infectious diseases*, 170(5), pp. 1128–33.

Kao, J. H., Chen, P. J., Wang, J. T., Yang, P. M., Lai, M. Y., Wang, T. H. and Chen, D. S. (1996) 'Superinfection by homotypic virus in hepatitis C virus carriers: studies on patients with post-transfusion hepatitis.', *Journal of medical virology*, 50(4), pp. 303–8. doi: 10.1002/(SICI)1096-9071(199612)50:4<303::AID-JMV4>3.0.CO;2-C.

Kao, J. H., Lai, M. Y., Chen, P. J. and Chen, D. S. (1997) 'Immunoglobulin M antibody response to hepatitis C virus core protein in patients with chronic hepatitis C.', *Journal of the Formosan Medical Association = Taiwan yi zhi*. China (Republic : 1949-), 96(10), pp. 825–828.

Kapoor, A., Kumar, A., Simmonds, P., Bhuva, N., Chauhan, L. S., Lee, B., Alpha Sall, A., Jin, Z., Morse, S. S., Shaz, B., Burbelo, P. D. and Lipkin, W. I. (2015) 'Virome Analysis of Transfusion Recipients Reveals a Novel Human Virus That Shares Genomic Features with Hepaciviruses and', *mBio*, 6(October), pp. 1–12. doi: 10.1128/mBio.01466-15.Editor.

Kapoor, A., Simmonds, P., Gerold, G., Qaisar, N., Jain, K., Henriquez, J. a, Firth, C., Hirschberg, D. L., Rice, C. M., Shields, S. and Lipkin, W. I. (2011) 'Characterization of a canine homolog of hepatitis C virus.', *Proceedings of the National Academy of Sciences of the United States of America*, 108(28), pp. 11608–13. doi: 10.1073/pnas.1101794108.

Kapoor, A., Simmonds, P., Scheel, T. K. H., Hjelle, B., Cullen, J. M., Burbelo, P. D., Chauhan, L. V., Duraisamy, R., Leon, M. S., Jain, K., Vandegrift, K. J., Calisher, C. H., Rice, C. M. and Lipkin, W. I. (2013) 'Identification of Rodent Homologs of Hepatitis C Virus and Pegiviruses', *mBio*, 4(2), pp. e00216-13. doi: 10.1128/mBio.00216-13.Editor.

Kapoor, a, Simmonds, P., Lipkin, W. I., Zaidi, S. and Delwart, E. (2010) 'Use of nucleotide composition analysis to infer hosts for three novel picorna-like viruses.', *Journal of virology*, 84(19), pp. 10322–10328. doi: 10.1128/JVI.00601-10.

Karlin, S., Blaisdell, B. E. and Schachtel, G. A. (1990) ‘Contrasts in codon usage of latent versus productive genes of Epstein-Barr virus: data and hypotheses.’, *Journal of virology*. UNITED STATES, 64(9), pp. 4264–4273.

Karlin, S., Burge, C., Kariin, S. and Burge, C. (1995) ‘Dinucleotide relative abundance extremes: a genomic signature.’, *Trends in genetics : TIG*. ENGLAND: Elsevier, 11(7), pp. 283–290. doi: 10.1016/S0168-9525(00)89076-9.

Karlin, S., Doerfler, W. and Cardon, L. R. (1994) ‘Why is CpG suppressed in the genomes of virtually all small eukaryotic viruses but not in those of large eukaryotic viruses?’, *Journal of Virology*, 68(5), pp. 2889–2897.

Karlin, S. and Ladunga, I. (1994) ‘Comparisons of eukaryotic genomic sequences.’, *Proceedings of the National Academy of Sciences of the United States of America*. UNITED STATES, 91(26), pp. 12832–12836.

Karlin, S. and Mrazek, J. (1997) ‘Compositional differences within and between eukaryotic genomes.’, *Proceedings of the National Academy of Sciences of the United States of America*. UNITED STATES, 94(19), pp. 10227–10232.

Kato, N., Ootsuyama, Y., Ohkoshi, S., Nakazawa, T., Sekiya, H., Hijikata, M. and Shimotohno, K. (1992) ‘Characterization of hypervariable regions in the putative envelope protein of hepatitis C virus.’, *Biochemical and biophysical research communications*. UNITED STATES, 189(1), pp. 119–127.

Khan, A. G., Whidby, J., Miller, M. T., Scarborough, H., Zatorski, A. V., Cygan, A., Price, A. A., Yost, S. A., Bohannon, C. D., Jacob, J., Grakoui, A. and Marcotrigiano, J. (2014) ‘Structure of the Core Ectodomain of the Hepatitis C Virus Envelope Glycoprotein 2’, *Nature*, 509(7500), pp. 381–384. doi: 10.1021/ja8019214.Optimization.

Kim, A. Y. and Chung, R. T. (2009) ‘Coinfection With HIV-1 and HCV-A One-Two Punch’, *Gastroenterology*. Elsevier Inc., 137(3), pp. 795–814. doi: 10.1053/j.gastro.2009.06.040.

Kim, A. Y., Lauer, G. M., Ouchi, K., Addo, M. M., Lucas, M., Schulze Zur Wiesch, J.,

Timm, J., Boczanowski, M., Duncan, J. E., Wurcel, A. G., Casson, D., Chung, R. T., Draenert, R., Klenerman, P. and Walker, B. D. (2005) 'The magnitude and breadth of hepatitis C virus-specific CD8+ T cells depend on absolute CD4+ T-cell count in individuals coinfected with HIV-1.', *Blood*. United States, 105(3), pp. 1170–1178. doi: 10.1182/blood-2004-06-2336.

Kobasa, D., Jones, S. M., Shinya, K., Kash, J. C., Copps, J., Ebihara, H., Hatta, Y., Kim, J. H., Halfmann, P., Hatta, M., Feldmann, F., Alimonti, J. B., Fernando, L., Li, Y., Katze, M. G., Feldmann, H. and Kawaoka, Y. (2007) 'Aberrant innate immune response in lethal infection of macaques with the 1918 influenza virus.', *Nature*. England, 445(7125), pp. 319–323. doi: 10.1038/nature05495.

Kolykhalov, a. a., Agapov, E. V., Blight, K. J., Mihalik, K., Feinstone, S. M. and Rice, C. M. (1997) 'Transmission of Hepatitis C by Intrahepatic Inoculation with Transcribed RNA', *Science*, 277(5325), pp. 570–574. doi: 10.1126/science.277.5325.570.

Kong, F., Pan, Y., Chi, X., Wang, X., Chen, L., Lv, J., Sun, H., Wu, R., Jin, J., Yu, G., Ma, Z., Wang, Y., Huang, X., Li, H., Bai, Y., Jia, J., Minuk, G. Y., Zhong, J., Sun, B., Jiang, J. and Niu, J. (2014) 'Factors associated with spontaneous clearance of hepatitis C virus in Chinese population.', *BioMed*. United States, 2014, p. 527030. doi: 10.1155/2014/527030.

Kong, L., Giang, E., Nieusma, T., Kadam, R. U., Cogburn, K. E., Hua, Y., Dai, X., Stanfield, R. L., Burton, D. R., Ward, A. B., Wilson, I. A. and Law, M. (2013) 'Hepatitis C virus E2 envelope glycoprotein core structure', *Science*, 342(6162), pp. 1090–1094. doi: 10.1021/ja8019214.Optimization.

Korlach, J., Bjornson, K. P., Chaudhuri, B. P., Cicero, R. L., Flusberg, B. A., Gray, J. J., Holden, D., Saxena, R., Wegener, J. and Turner, S. W. (2010) 'Real-time DNA sequencing from single polymerase molecules.', *Methods in enzymology*. United States, 472, pp. 431–455. doi: 10.1016/S0076-6879(10)72001-2.

Kow, Y. W. (2002) 'Repair of deaminated bases in DNA.', *Free radical biology & medicine*. United States, 33(7), pp. 886–893.

Kuiken, C., Combet, C., Bukh, J., Shin-I, T., Deleage, G., Mizokami, M., Richardson, R.,

Sablon, E., Yusim, K., Pawlotsky, J. M., Simmonds, P., Korber, B., Abfalterer, W., Calef, C., Foley, B., Funkhouser, R., Gaschen, B., Lang, D., Leitner, T., Szinger, J. and Zhang, M. (2006) 'A comprehensive system for consistent numbering of HCV sequences, proteins and epitopes', *Hepatology*, 44(5), pp. 1355–1361. doi: 10.1002/hep.21377.

Kuiken, C., Yusim, K., Boykin, L. and Richardson, R. (2004) 'The Los Alamos hepatitis C sequence database', *Bioinformatics*, 21(3), pp. 379–384. doi: 10.1093/bioinformatics/bth485.

Kuiken, C., Yusim, K., Boykin, L. and Richardson, R. (2005) 'The Los Alamos hepatitis C sequence database.', *Bioinformatics (Oxford, England)*, 21(3), pp. 379–84. doi: 10.1093/bioinformatics/bth485.

Kunec, D. and Osterrieder, N. (2016) 'Codon Pair Bias Is a Direct Consequence of Dinucleotide Bias', *Cell Reports*. The Authors, 14, pp. 55–67. doi: 10.1016/j.celrep.2015.12.011.

Kuntzen, T., Timm, J., Berical, A., Lennon, N., Berlin, A. M., Young, S. K., Lee, B., Heckerman, D., Carlson, J., Reyor, L. L., Kleyman, M., McMahon, C. M., Birch, C., Schulze Zur Wiesch, J., Ledlie, T., Koehrsen, M., Kodira, C., Roberts, A. D., Lauer, G. M., Rosen, H. R., Bihl, F., Cerny, A., Spengler, U., Liu, Z., Kim, A. Y., Xing, Y., Schneidewind, A., Madey, M. A., Fleckenstein, J. F., Park, V. M., Galagan, J. E., Nusbaum, C., Walker, B. D., Lake-Bakaar, G. V, Daar, E. S., Jacobson, I. M., Gomperts, E. D., Edlin, B. R., Donfield, S. M., Chung, R. T., Talal, A. H., Marion, T., Birren, B. W., Henn, M. R. and Allen, T. M. (2008) 'Naturally occurring dominant resistance mutations to hepatitis C virus protease and polymerase inhibitors in treatment-naive patients.', *Hepatology (Baltimore, Md.)*. United States, 48(6), pp. 1769–1778. doi: 10.1002/hep.22549.

Kuo, G., Choo, Q. L., Alter, H. J., Gitnick, G. L., Redeker, A. G., Purcell, R. H., Miyamura, T., Dienstag, J. L., Alter, M. J. and Stevens, C. E. (1989) 'An assay for circulating antibodies to a major etiologic virus of human non-A, non-B hepatitis.', *Science*. UNITED STATES, 244(4902), pp. 362–364.

van de Laar, T. J. W., van der Bij, A. K., Prins, M., Bruisten, S. M., Brinkman, K., Ruys,

T. A., van der Meer, J. T. M., de Vries, H. J. C., Mulder, J.-W., van Agtmael, M., Jurriaans, S., Wolthers, K. C. and Coutinho, R. A. (2007) 'Increase in HCV incidence among men who have sex with men in Amsterdam most likely caused by sexual transmission.', *The Journal of infectious diseases*. United States, 196(2), pp. 230–238. doi: 10.1086/518796.

van de Laar, T. J. W., Molenkamp, R., van den Berg, C., Schinkel, J., Beld, M. G. H. M., Prins, M., Coutinho, R. A. and Bruisten, S. M. (2009) 'Frequent HCV reinfection and superinfection in a cohort of injecting drug users in Amsterdam', *Journal of Hepatology*, 51(4), pp. 667–674. doi: 10.1016/j.jhep.2009.05.027.

Laar, T. van de, Pybus, O., Bruisten, S., Brown, D., Nelson, M., Bhagani, S., Vogel, M., Baumgarten, A., Chaix, M., Fisher, M., Gőtz, H., Matthews, G., Neifer, S., White, P., Rawlinson, W., Pol, S., Rockstroh, J., Coutinho1, R., Dore, G., Dusheiko, G. and Danta, M. (2013) 'Evidence of a large, international network of international hepatitis C virus transmission in HIV-positive men who have sex with men', 18(9), pp. 1199–1216. doi: 10.1016/j.micinf.2011.07.011.Innate.

Lagerkvist, U. (1978) '"Two out of three": an alternative method for codon reading.', *Proceedings of the National Academy of Sciences of the United States of America*, 75(4), pp. 1759–62. doi: 10.1073/pnas.75.4.1759.

Larrat, S., Poveda, J. D., Coudret, C., Fusillier, K., Magnat, N., Signori-Schmuck, A., Thibault, V. and Morand, P. (2013) 'Sequencing assays for failed genotyping with the versant hepatitis C virus genotype assay (LiPA), version 2.0', *Journal of Clinical Microbiology*, 51(9), pp. 2815–2821. doi: 10.1128/JCM.00586-13.

Larsen, C., Bousquet, V., Delarocque-Astagneau, E., Pioche, C., Roudot-Thoraval, F. and Desenclos, J.-C. (2010) '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*. United States, 82(10), pp. 1647–1654. doi: 10.1002/jmv.21850.

Laskus, T., Operskalski, E. A., Radkowski, M., Wilkinson, J., Mack, W. J., Degiacomo, M., Al-Harthi, L., Chen, Z., Xu, J. A. and Kovacs, A. (2007) 'Negative-strand hepatitis C virus (HCV) RNA in peripheral blood mononuclear cells from anti-HCV-positive/HIV-

infected women', *Journal of Infectious Diseases*, 195(1), pp. 124–133. doi: 10.1086/509897.

Laskus, T., Radkowski, M., Piasek, a, Nowicki, M., Horban, a, Cianciara, J. and Rakela, J. (2000) 'Hepatitis C virus in lymphoid cells of patients coinfected with human immunodeficiency virus type 1: evidence of active replication in monocytes/macrophages and lymphocytes.', *The Journal of infectious diseases*, 181(2), pp. 442–8. doi: 10.1086/315283.

Laskus, T., Wang, L. F., Radkowski, M., Vargas, H., Nowicki, M., Wilkinson, J. and Rakela, J. (2001) 'Exposure of Hepatitis C Virus (HCV) RNA-Positive Recipients to HCV RNA-Positive Blood Donors Results in Rapid Predominance of a Single Donor Strain and Exclusion and / or Suppression of the Recipient Strain', *Journal of Virology*, 75(5), pp. 2059–2066. doi: 10.1128/JVI.75.5.2059.

Lau, J. Y., Davis, G. L., Prescott, L. E., Maertens, G., Lindsay, K. L., Qian, K., Mizokami, M. and Simmonds, P. (1996) 'Distribution of hepatitis C virus genotypes determined by line probe assay in patients with chronic hepatitis C seen at tertiary referral centres in the United States. Hepatitis Interventional Therapy Group.', *Annals of Internal Medicine*, 124(10), pp. 868–876.

Lauring, A. S., Acevedo, A., Cooper, S. B. and Andino, R. (2012) 'Codon usage determines the mutational robustness, evolutionary capacity, and virulence of an RNA virus', *Cell Host and Microbe*. Elsevier Inc., 12(5), pp. 623–632. doi: 10.1016/j.chom.2012.10.008.

Lavanchy, D. (2009) 'The global burden of hepatitis C', *Liver International*, 29, pp. 74–81. doi: 10.1111/j.1478-3231.2008.01934.x.

Lavanchy, D. (2011) 'Evolving epidemiology of hepatitis C virus.', *Clinical microbiology and infection : the official publication of the European Society of Clinical Microbiology and Infectious Diseases*, 17(2), pp. 107–15. doi: 10.1111/j.1469-0691.2010.03432.x.

Leandro, G., Mangia, A., Hui, J., Fabris, P., Rubbia–Brandt, L., Colloredo, G., Adinolfi, L. E., Asselah, T., Jonsson, J. R., Smedile, A., Terrault, N., Pazienza, V., Giordani, M. T.,

Giostra, E., Sonzogni, A., Ruggiero, G., Marcellin, P., Powell, E. E., George, J. and Negro, F. (2006) 'Relationship Between Steatosis, Inflammation, and Fibrosis in Chronic Hepatitis C: A Meta-Analysis of Individual Patient Data', *Gastroenterology*, 130(6), pp. 1636–1642. doi: 10.1053/j.gastro.2006.03.014.

Lechner, F., Wong, D. K., Dunbar, P. R., Chapman, R., Chung, R. T., Dohrenwend, P., Robbins, G., Phillips, R., Klenerman, P. and Walker, B. D. (2000) 'Analysis of successful immune responses in persons infected with hepatitis C virus', *J Exp Med*, 191(9), pp. 1499–1512.

Lee, Y.-M., Lin, H.-J., Chen, Y.-J., Lee, C.-M., Wang, S.-F., Chang, K.-Y., Chen, T.-L., Liu, H.-F. and Chen, Y.-M. A. (2010) 'Molecular Epidemiology of HCV Genotypes Among Injection Drug Users in Taiwan: Full-Length Sequences of Two New Subtype 6w Strains and a Recombinant form_2b6w', *Journal of Medical Virology*, 82, pp. 57–68. doi: 10.1002/jmv.

Lefterova, M. I., Suarez, C. J., Banaei, N. and Pinsky, B. A. (2015) 'Next-Generation Sequencing for Infectious Disease Diagnosis and Management: A Report of the Association for Molecular Pathology.', *The Journal of molecular diagnostics : JMD*. American Society for Investigative Pathology and the Association for Molecular Pathology, 17(6), pp. 623–34. doi: 10.1016/j.jmoldx.2015.07.004.

Lehmann, J. and Libchaber, A. (2008) 'Degeneracy of the genetic code and stability of the base pair at the second position of the anticodon.', *RNA*, 14(7), pp. 1264–9. doi: 10.1261/rna.1029808.

Lemon, S. M. (2010) 'Induction and evasion of innate antiviral responses by hepatitis C virus.', *The Journal of biological chemistry*. United States, 285(30), pp. 22741–22747. doi: 10.1074/jbc.R109.099556.

Levin, M. K., Gurjar, M. and Patel, S. S. (2005) 'A Brownian motor mechanism of translocation and strand separation by hepatitis C virus helicase.', *Nature structural & molecular biology*. United States, 12(5), pp. 429–435. doi: 10.1038/nsmb920.

Li, H., Thomassen, L. V, Majid, A., McMahon, B. J., Bruden, D., McArdle, S., Bano, N.,

Chung, M., Carithers, R. L., Perkins, J. D., Sullivan, D. G. and Gretch, D. R. (2008) 'Investigation of putative multisubtype hepatitis C virus infections in vivo by heteroduplex mobility analysis of core/envelope subgenomes.', *Journal of virology*, 82(15), pp. 7524–32. doi: 10.1128/JVI.02220-07.

Li, K., Foy, E., Ferreon, J. C., Nakamura, M., Ferreon, A. C. M., Ikeda, M., Ray, S. C., Gale, M. and Lemon, S. M. (2005) 'Immune evasion by hepatitis C virus NS3/4A protease-mediated cleavage of the Toll-like receptor 3 adaptor protein TRIF.', *Proceedings of the National Academy of Sciences*, 102(8), pp. 2992–7. doi: 10.1073/pnas.0408824102.

Li, S., Hu, P., Zhang, Q.-Q., Liu, Y.-H., Hu, H.-D., Zhang, D.-Z. and Ren, H. (2011) 'Single nucleotide polymorphisms of the IL28B and sustained virologic response of patients with chronic hepatitis C to PEG-interferon/ribavirin therapy: A meta-analysis: Meta-analysis of IL28B.', *Hepatitis monthly*. Iran, 11(3), pp. 163–172.

Li, W. and Godzik, A. (2006) 'Cd-hit: A fast program for clustering and comparing large sets of protein or nucleotide sequences', *Bioinformatics*, 22(13), pp. 1658–1659. doi: 10.1093/bioinformatics/btl158.

Di Liberto, G., Roque-Afonso, A.-M., Kara, R., Ducoulombier, D., Fallot, G., Samuel, D. and Feray, C. (2006) 'Clinical and therapeutic implications of hepatitis C virus compartmentalization.', *Gastroenterology*, 131, pp. 76–84. doi: 10.1053/j.gastro.2006.04.016.

Lin, W., Weinberg, E. M., Tai, A. W., Peng, L. F., Brockman, M. A., Kim, K.-A., Kim, S. S., Borges, C. B., Shao, R.-X. and Chung, R. T. (2008) 'HIV increases HCV replication in a TGF-beta1-dependent manner.', *Gastroenterology*. United States, 134(3), pp. 803–811. doi: 10.1053/j.gastro.2008.01.005.

Lindenbach, B. D. and Rice, C. M. (2007) 'Flaviviridae: The Viruses and Their Replication', in *Fields Virology*, pp. 1101–1151. doi: 10.1016/0038-092X(88)90131-4.

Linnarsson, S. (2010) 'Recent advances in DNA sequencing methods - general principles of sample preparation.', *Experimental cell research*. Elsevier Inc., 316(8), pp. 1339–43. doi: 10.1016/j.yexcr.2010.02.036.

Liu, S., Yang, W., Shen, L., Turner, J. R., Coyne, C. B. and Wang, T. (2008) 'Tight Junction Proteins Claudin-1 and Occludin Control Hepatitis C Virus Entry and Are Downregulated during Infection To Prevent Superinfection', *Journal of Virology*, 83(4), pp. 2011–2014. doi: 10.1128/jvi.01888-08.

Lloyd, A. R., Jagger, E., Post, J. J., Crooks, L.-A., Rawlinson, W. D., Hahn, Y. S. and Ffrench, R. A. (2007) 'Host and viral factors in the immunopathogenesis of primary hepatitis C virus infection.', *Immunology and cell biology*. Australia, 85(1), pp. 24–32. doi: 10.1038/sj.icb.7100010.

Lobo, F. P., Mota, B. E. F., Pena, S. D. J., Azevedo, V., Macedo, A. M., Tauch, A., Machado, C. R. and Franco, G. R. (2009) 'Virus-host coevolution: Common patterns of nucleotide motif usage in Flaviviridae and their hosts', *PLoS ONE*, 4(7). doi: 10.1371/journal.pone.0006282.

Lohmann, V., Koch, J. O. and Bartenschlager, R. (1996) 'Processing pathways of the hepatitis C virus proteins.', *Journal of hepatology*. DENMARK, 24(2 Suppl), pp. 11–19.

Lohmann, V., Korner, F., Dobierzewska, A. and Bartenschlager, R. (2001) 'Mutations in hepatitis C virus RNAs conferring cell culture adaptation.', *Journal of virology*. UNITED STATES, 75(3), pp. 1437–1449. doi: 10.1128/JVI.75.3.1437-1449.2001.

Lohmann, V., Korner, F., Koch, J., Herian, U., Theilmann, L. and Bartenschlager, R. (1999) 'Replication of subgenomic hepatitis C virus RNAs in a hepatoma cell line.', *Science*. UNITED STATES, 285(5424), pp. 110–113.

Lontok, E., Harrington, P., Howe, A., Kieffer, T., Lennerstrand, J., Lenz, O., Mcphee, F., Mo, H., Parkin, N., Pilot-Matias, T. and Miller, V. (2015) 'Hepatitis C virus drug resistance-associated substitutions: State of the art summary', *Hepatology*, 62(5), pp. 1623–1632. doi: 10.1002/hep.27934.

Lopez-Labrador, F. X., Moya, A. and Gonzalez-Candelas, F. (2008) 'Mapping natural polymorphisms of hepatitis C virus NS3/4A protease and antiviral resistance to inhibitors in worldwide isolates.', *Antiviral therapy*. England, 13(4), pp. 481–494.

Louie, K. S., St Laurent, S., Forssen, U. M., Mundy, L. M. and Pimenta, J. M. (2012) 'The high comorbidity burden of the hepatitis C virus infected population in the United States', *BMC Infectious Diseases*. BioMed Central Ltd, 12(1), p. 86. doi: 10.1186/1471-2334-12-86.

Loulergue, P., Mir, O. and Sogni, P. (2012) 'Super-infection with genotype 4 hepatitis C virus patient treated for genotype 3 acute hepatitis C', *AIDS*, 5, pp. 655–56.

Lozach, P.-Y., Lortat-Jacob, H., De Lacroix De Lavalette, A., Staropoli, I., Foung, S., Amara, A., Houles, C., Fieschi, F., Schwartz, O., Virelizier, J.-L., Arenzana-Seisdedos, F. and Altmeyer, R. (2003) 'DC-SIGN and L-SIGN Are High Affinity Binding Receptors for Hepatitis C Virus Glycoprotein E2', *Journal of Biological Chemistry*, 278(22), pp. 20358–20366. doi: 10.1074/jbc.M301284200.

Luca, A. De, Giambenedetto, S. Di, Presti, A. Lo, Sierra, S., Prosperi, M., Cella, E., Giovanetti, M., Torti, C., Caudai, C., Vicenti, I., Saladini, F., Almi, P., Grima, P., Blanc, P., Fabbiani, M., Rossetti, B., Gagliardini, R., Kaiser, R., Ciccozzi, M. and Zazzi, M. (2015) 'Two Distinct Hepatitis C Virus Genotype 1a Clades Have Different Geographical Distribution and Association With Natural Resistance to NS3 Protease Inhibitors', *Open Forum Infectious Disease*, 2(2), pp. 1–8. doi: 10.1093/o.

Lutchman, G., Danehower, S., Song, B.-C., Liang, T. J., Hoofnagle, J. H., Thomson, M. and Ghany, M. G. (2007) 'Mutation rate of the hepatitis C virus NS5B in patients undergoing treatment with ribavirin monotherapy.', *Gastroenterology*. United States, 132(5), pp. 1757–1766. doi: 10.1053/j.gastro.2007.03.035.

Lyons, S., Kapoor, A., Sharp, C., Schneider, B. S., Wolfe, N. D., Culshaw, G., Corcoran, B., McGorum, B. C. and Simmonds, P. (2012) 'Nonprimate hepaciviruses in domestic horses, United Kingdom', *Emerging Infectious Diseases*, 18(12), pp. 1976–1982. doi: 10.3201/eid1812.120498.

Malboeuf, C. M., Yang, X., Charlebois, P., Qu, J., Berlin, A. M., Casali, M., Pesko, K. N., Boutwell, C. L., Devincenzo, J. P., Ebel, G. D., Allen, T. M., Zody, M. C., Henn, M. R. and Levin, J. Z. (2013) 'Complete viral RNA genome sequencing of ultra-low copy samples by sequence-independent amplification', *Nucleic Acids Research*, 41(1), pp. 1–10.

doi: 10.1093/nar/gks794.

Mancuso, M. E., Linari, S., Aghemo, A., Bartolozzi, D., Santagostino, E., Rumi, M. G., Fognani, E., Fasulo, M. R., Gragnani, L., Bruno, R., Morfini, M., Zignego, A. L. and Colombo, M. (2014) 'Interferon lambda 3 rs12979860 polymorphism in patients with haemophilia and HCV infection: a predictor of spontaneous viral clearance and sustained virological response.', *Thrombosis and haemostasis*. Germany, 111(6), pp. 1067–1076. doi: 10.1160/TH13-11-897.

Manns, M. P., McHutchison, J. G., Gordon, S. C., Rustgi, V. K., Shiffman, M., Reindollar, R., Goodman, Z. D., Koury, K., Ling, M. and Albrecht, J. K. (2001) 'Peginterferon alfa-2b plus ribavirin compared with interferon alfa-2b plus ribavirin for initial treatment of chronic hepatitis C: a randomised trial.', *Lancet*. England, 358(9286), pp. 958–965.

Manns, M. P., Wedemeyer, H. and Cornberg, M. (2006) 'Treating viral hepatitis C: efficacy, side effects, and complications.', *Gut*, 55(9), pp. 1350–9. doi: 10.1136/gut.2005.076646.

March, J. C., Oviedo-Joekes, E. and Romero, M. (2006) 'Drugs and social exclusion in ten European cities', *European Addiction Research*, 12(1), pp. 33–41. doi: EAR2006012001033 [pii]\r10.1159/000088581.

Margulies, M., Egholm, M., Altman, W. E., Attiya, S., Bader, J. S., Bemben, L. A., Berka, J., Braverman, M. S., Chen, Y.-J., Chen, Z., Dewell, S. B., Du, L., Fierro, J. M., Gomes, X. V, Godwin, B. C., He, W., Helgesen, S., Ho, C. H., Irzyk, G. P., Jando, S. C., Alenquer, M. L. I., Jarvie, T. P., Jirage, K. B., Kim, J.-B., Knight, J. R., Lanza, J. R., Leamon, J. H., Lefkowitz, S. M., Lei, M., Li, J., Lohman, K. L., Lu, H., Makhijani, V. B., McDade, K. E., McKenna, M. P., Myers, E. W., Nickerson, E., Nobile, J. R., Plant, R., Puc, B. P., Ronan, M. T., Roth, G. T., Sarkis, G. J., Simons, J. F., Simpson, J. W., Srinivasan, M., Tartaro, K. R., Tomasz, A., Vogt, K. A., Volkmer, G. A., Wang, S. H., Wang, Y., Weiner, M. P., Yu, P., Begley, R. F. and Rothberg, J. M. (2005) 'Genome sequencing in microfabricated high-density picolitre reactors.', *Nature*. England, 437(7057), pp. 376–380. doi: 10.1038/nature03959.

Marin, M. G., Bresciani, S., Puoti, M., Rodella, A., Gussago, A., Ravaggi, A., Pizzocolo,

G., Albertini, A. and Cariani, E. (1994) ‘Clinical significance of serum hepatitis C virus (HCV) RNA as marker of HCV infection.’, *Journal of clinical microbiology*. UNITED STATES, 32(12), pp. 3008–3012.

Markin, R. S. (1994) ‘Manifestations of Epstein-Barr virus-associated disorders in liver.’, *Liver*. DENMARK, 14(1), pp. 1–13.

Massard, J., Ratziu, V., Thabut, D., Moussalli, J., Lebray, P., Benhamou, Y. and Poynard, T. (2006) ‘Natural history and predictors of disease severity in chronic hepatitis C’, *Journal of Hepatology*, 44(SUPPL. 1), pp. 42–47. doi: 10.1016/j.jhep.2005.11.009.

Matzener, P., Magkouras, I., Rumenapf, T., Peterhans, E. and Schweizer, M. (2009) ‘The viral RNase E(rns) prevents IFN type-I triggering by pestiviral single- and double-stranded RNAs.’, *Virus research*. Netherlands, 140(1–2), pp. 15–23. doi: 10.1016/j.virusres.2008.10.015.

Mauger, D. M., Golden, M., Yamane, D., Williford, S., Lemon, S. M., Martin, D. P. and Weeks, K. M. (2015) ‘Functionally conserved architecture of hepatitis C virus RNA genomes’, *Proceedings of the National Academy of Sciences*, 112(12), pp. 3692–3697. doi: 10.1073/pnas.1416266112.

McCaffrey, K., Gouklani, H., Boo, I., Poumbourios, P. and Drummer, H. E. (2011) ‘The variable regions of hepatitis C virus glycoprotein E2 have an essential structural role in glycoprotein assembly and virion infectivity.’, *The Journal of general virology*. England, 92(Pt 1), pp. 112–121. doi: 10.1099/vir.0.026385-0.

McCarthy, M. and Ortega, M. (2012) ‘Neurological complications of hepatitis C infection’, *Current Neurology and Neuroscience Reports*, 12(6), pp. 642–654. doi: 10.1007/s11910-012-0311-6.

McDonald, S. A., Hutchinson, S. J., Bird, S. M., Mills, P. R., Hayes, P., Dillon, J. F. and Goldberg, D. J. (2011) ‘Excess morbidity in the hepatitis C-diagnosed population in Scotland, 1991-2006’, *Epidemiol Infect*, 139(3), pp. 344–353. doi: 10.1017/S0950268810001421.

McDonald, S. a, Hutchinson, S. J., Bird, S. M., Robertson, C., Mills, P. R., Graham, L., Dillon, J. F. and Goldberg, D. J. (2010) 'The growing contribution of hepatitis C virus infection to liver-related mortality in Scotland.', *Euro surveillance*, 15(18), pp. 1–6.

McElroy, K., Thomas, T. and Luciani, F. (2014) 'Deep sequencing of evolving pathogen populations: applications, errors, and bioinformatic solutions.', *Microbial informatics and experimentation*, 4(1), p. 1. doi: 10.1186/2042-5783-4-1.

McHutchison, J. G., Poynard, T., Pianko, S., Gordon, S. C., Reid, A. E., Dienstag, J., Morgan, T., Yao, R. and Albrecht, J. (2000) 'The impact of interferon plus ribavirin on response to therapy in black patients with chronic hepatitis C. The International Hepatitis Interventional Therapy Group.', *Gastroenterology*. UNITED STATES, 119(5), pp. 1317–1323.

McLauchlan, J. (2000) 'Properties of the hepatitis C virus core protein: a structural protein that modulates cellular processes.', *Journal of viral hepatitis*. ENGLAND, 7(1), pp. 2–14.

McLeish, N. J., Witteveldt, J., Clasper, L., McIntyre, C., McWilliam Leitch, E. C., Hardie, A., Bennett, S., Gunson, R., Carman, W. F., Feeney, S. a, Coyle, P. V, Vipond, B., Muir, P., Benschop, K., Wolthers, K., Waris, M., Osterback, R., Johannessen, I., Templeton, K., Harvala, H. and Simmonds, P. (2012) 'Development and assay of RNA transcripts of enterovirus species A to D, rhinovirus species a to C, and human parechovirus: assessment of assay sensitivity and specificity of real-time screening and typing methods.', *Journal of clinical microbiology*, 50(9), pp. 2910–7. doi: 10.1128/JCM.01172-12.

McNaughton, A. L. (2012) *Inhibition of hepatitis C virus entry into hepatocytes*.

Mcnaughton, A. L., Cameron, D., Wignall-fleming, E. B., Biek, R., Mclauchlan, J., Gunson, R. N., Templeton, K., Tan, H. M., Carol, E., Leitch, M., Cameron, I. D., Wignall-fleming, E. B., Biek, R., Mclauchlan, J., Gunson, R. N., Templeton, K., Tan, H. M. and McWilliam Leitch, E. C. (2015) 'Spatiotemporal Reconstruction of the Introduction of Hepatitis C Virus into Scotland and Its Subsequent Regional Transmission', *Journal of Virology*, 89(22), pp. 11223–11232. doi: 10.1128/JVI.02106-15.Editor.

McNaughton, A. L., Thomson, E. C., Templeton, K., Gunson, R. N. and Leitch, E. C. M.

(2014) ‘Mixed genotype hepatitis C infections and implications for treatment.’, *Hepatology*, 59(3), p. 1209. doi: 10.1002/hep.26544.

McWilliam Leitch, E. C. and McLauchlan, J. (2013) ‘Determining the cellular diversity of HCV quasispecies by single-cell viral sequencing.’, *Journal of virology*, (September). doi: 10.1128/JVI.01602-13.

Mehta, S. H., Cox, A., Hoover, D. R., Wang, X.-H., Mao, Q., Ray, S., Strathdee, S. A., Vlahov, D. and Thomas, D. L. (2002) ‘Protection against persistence of hepatitis C.’, *Lancet*. England, 359(9316), pp. 1478–1483. doi: 10.1016/S0140-6736(02)08435-0.

Merchante, N., Giron-Gonzalez, J. A., Gonzalez-Serrano, M., Torre-Cisneros, J., Garcia-Garcia, J. A., Arizcorreta, A., Ruiz-Morales, J., Cano-Lliteras, P., Lozano, F., Martinez-Sierra, C., Macias, J. and Pineda, J. A. (2006) ‘Survival and prognostic factors of HIV-infected patients with HCV-related end-stage liver disease.’, *AIDS*. England, 20(1), pp. 49–57.

Messina, J. P., Humphreys, I., Flaxman, A., Brown, A., Cooke, G. S., Pybus, O. G. and Barnes, E. (2015) ‘Global distribution and prevalence of hepatitis C virus genotypes’, *Hepatology*, 61(1), pp. 77–87. doi: 10.1002/hep.27259.

Micalessi, M. I., Gerard, C., Ameye, L., Plasschaert, S., Brochier, B. and Vranckx, R. (2008) ‘Distribution of hepatitis C virus genotypes among injecting drug users in contact with treatment centers in Belgium, 2004-2005.’, *Journal of medical virology*. United States, 80(4), pp. 640–645. doi: 10.1002/jmv.21145.

Micallef, J. M., Kaldor, J. M. and Dore, G. J. (2006) ‘Spontaneous viral clearance following acute hepatitis C infection: a systematic review of longitudinal studies’, *Journal of Viral Hepatitis*, 13(1), pp. 34–41. doi: 10.1111/j.1365-2893.2005.00651.x.

Miller, R. H. and Purcell, R. H. (1990) ‘Hepatitis C virus shares amino acid sequence similarity with pestiviruses and flaviviruses as well as members of two plant virus supergroups.’, *Proceedings of the National Academy of Sciences*, 87(6), pp. 2057–61. doi: 10.1073/pnas.87.6.2057.

Mindikoglu, A. L. and Miller, R. R. (2009) ‘Hepatitis C in the elderly: epidemiology, natural history, and treatment.’, *Clinical gastroenterology and hepatology : the official clinical practice journal of the American Gastroenterological Association*. United States, 7(2), p. 128–34; quiz 124. doi: 10.1016/j.cgh.2008.07.017.

Minola, E., Prati, D., Suter, F., Maggiolo, F., Caprioli, F., Sonzogni, A., Fraquelli, M., Paggi, S. and Conte, D. (2002) ‘Age at infection affects the long-term outcome of transfusion-associated chronic hepatitis C.’, *Blood*. United States, 99(12), pp. 4588–4591. doi: 10.1182/blood-2001-12-0192.

Missale, G., Bertoni, R., Lamonaca, V., Valli, A., Massari, M., Mori, C., Rumi, M. G., Houghton, M., Fiaccadori, F. and Ferrari, C. (1996) ‘Different clinical behaviors of acute hepatitis C virus infection are associated with different vigor of the anti-viral cell-mediated immune response.’, *The Journal of clinical investigation*, 98(3), pp. 706–14. doi: 10.1172/JCI118842.

Mohsen, A. H., Hcv, T. and Group, S. (2001) ‘The epidemiology of hepatitis C in a UK health regional population of 5.12 million’, *Gut*, 48, pp. 707–713.

Mokili, J. L., Dutilh, B. E., Lim, Y. W., Schneider, B. S., Taylor, T., Haynes, M. R., Metzgar, D., Myers, C. A., Blair, P. J., Nosrat, B., Wolfe, N. D. and Rohwer, F. (2013) ‘Identification of a novel human papillomavirus by metagenomic analysis of samples from patients with febrile respiratory illness.’, *PloS one*. United States, 8(3), p. e58404. doi: 10.1371/journal.pone.0058404.

Moore, K. and Dusheiko, G. (2005) ‘Opiate abuse and viral replication in hepatitis C.’, *The American journal of pathology*, 167(5), pp. 1189–91. doi: 10.1016/S0002-9440(10)61207-2.

Moqueet, N., Infante-Rivard, C., Platt, R. W., Young, J., Cooper, C., Hull, M., Walmsley, S. and Klein, M. B. (2015) ‘Favourable IFNL3 genotypes are associated with spontaneous clearance and are differentially distributed in Aboriginals in Canadian HIV-hepatitis C co-infected individuals.’, *International journal of molecular sciences*. Switzerland, 16(3), pp. 6496–6512. doi: 10.3390/ijms16036496.

Moreno, C., Berg, T., Tanwandee, T., Thongsawat, S., Van Vlierberghe, H., Zeuzem, S., Lenz, O., Peeters, M., Sekar, V. and De Smedt, G. (2012) 'Antiviral activity of TMC435 monotherapy in patients infected with HCV genotypes 2-6: TMC435-C202, a phase IIa, open-label study.', *Journal of hepatology*. England, 56(6), pp. 1247–1253. doi: 10.1016/j.jhep.2011.12.033.

Moureau, G., Cook, S., Lemey, P., Nougairede, A., Forrester, N. L., Khasnatinov, M., Charrel, R. N., Firth, A. E., Gould, E. a. and de Lamballerie, X. (2015) 'New Insights into Flavivirus Evolution, Taxonomy and Biogeographic History, Extended by Analysis of Canonical and Alternative Coding Sequences', *Plos One*, 10(2), p. e0117849. doi: 10.1371/journal.pone.0117849.

Mueller, S., Papamichail, D., Coleman, J. R., Skiena, S. and Wimmer, E. (2006) 'Reduction of the rate of poliovirus protein synthesis through large-scale codon deoptimization causes attenuation of viral virulence by lowering specific infectivity.', *Journal of virology*, 80(19), pp. 9687–96. doi: 10.1128/JVI.00738-06.

Mullis, K., Faloona, F., Scharf, S., Saiki, R., Horn, G. and Erlich, H. (1986) 'Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction.', *Cold Spring Harbor symposia on quantitative biology*. UNITED STATES, 51 Pt 1, pp. 263–273.

Naccache, S. N., Greninger, A. L., Lee, D., Coffey, L. L., Phan, T., Rein-Weston, A., Aronsohn, A., Hackett, J., Delwart, E. L. and Chiu, C. Y. (2013) 'The Perils of Pathogen Discovery: Origin of a Novel Parvovirus-Like Hybrid Genome Traced to Nucleic Acid Extraction Spin Columns', *Journal of Virology*, 87(22), pp. 11966–11977. doi: 10.1128/JVI.02323-13.

Nascimbeni, M., Mizukoshi, E., Bosmann, M., Major, M. E., Mihalik, K., Rice, C. M., Feinstone, S. M. and Rehermann, B. (2003) 'Kinetics of CD4+ and CD8+ memory T-cell responses during hepatitis C virus rechallenge of previously recovered chimpanzees.', *Journal of virology*, 77(8), pp. 4781–93. doi: 10.1128/JVI.77.8.4781.

National Treatment Agency for Substance Misuse (2010) *Injecting drug use in England: a declining trend*.

Natov, S. N., Lau, J. Y. N., Ruthazer, R., Schmid, C. H., Levey, A. S. and Pereira, B. J. G. (1999) 'Hepatitis C virus genotype does not affect patient survival among renal transplant candidates', *Kidney International*, 56, pp. 700–706.

Navas, S., Martín, J., Quiroga, J. a, Castillo, I. and Carreño, V. (1998) 'Genetic diversity and tissue compartmentalization of the hepatitis C virus genome in blood mononuclear cells, liver, and serum from chronic hepatitis C patients.', *Journal of virology*, 72(2), pp. 1640–1646.

Negash, A. A., Ramos, H. J., Crochet, N., Lau, D. T. Y., Doehle, B., Papic, N., Delker, D. A., Jo, J., Bertoletti, A., Hagedorn, C. H. and Gale, M. (2013) 'IL-1b Production through the NLRP3 Inflammasome by Hepatic Macrophages Links Hepatitis C Virus Infection with Liver Inflammation and Disease', *PLoS Pathogens*, 9(4), pp. 1–13. doi: 10.1371/journal.ppat.1003330.

Neumann, A. U., Lam, N. P., Dahari, H., Gretch, D. R., Wiley, T. E., Layden, T. J. and Perelson, A. S. (1998) 'Hepatitis C viral dynamics in vivo and the antiviral efficacy of interferon-alpha therapy.', *Science*. UNITED STATES, 282(5386), pp. 103–107.

Neumann-Haefelin, C., McKiernan, S., Ward, S., Viazov, S., Spangenberg, H. C., Killinger, T., Baumert, T. F., Nazarova, N., Sheridan, I., Pybus, O., von Weizsacker, F., Roggendorf, M., Kelleher, D., Klenerman, P., Blum, H. E. and Thimme, R. (2006) 'Dominant influence of an HLA-B27 restricted CD8+ T cell response in mediating HCV clearance and evolution.', *Hepatology*. United States, 43(3), pp. 563–572. doi: 10.1002/hep.21049.

Neumann-Haefelin, C., Oniangue-Ndza, C., Kuntzen, T., Schmidt, J., Nitschke, K., Sidney, J., Caillet-Saguy, C., Binder, M., Kersting, N., Kemper, M. W., Power, K. A., Ingber, S., Reyor, L. L., Hills-Evans, K., Kim, A. Y., Lauer, G. M., Lohmann, V., Sette, A., Henn, M. R., Bressanelli, S., Thimme, R. and Allen, T. M. (2011) 'Human leukocyte antigen B27 selects for rare escape mutations that significantly impair hepatitis C virus replication and require compensatory mutations.', *Hepatology*. United States, 54(4), pp. 1157–1166. doi: 10.1002/hep.24541.

NHS (2004) *Hepatitis C - Essential information for professionals and guidance on testing*.

NHS England (2014) *Interim Clinical Commissioning Policy Statement : Daclatasvir / Ledipasvir Sofosbuvir depending on for defined patients with Ribivirin Hepatitis, Policy Statement*.

NIH Consensus Development Panel (2002) 'National Institutes of Health Consensus Development Conference Statement: Management of hepatitis C: 2002--June 10-12, 2002.', *Hepatology*. United States, 36(5 Suppl 1), pp. S3-20. doi: 10.1053/jhep.2002.37117.

Nkontchou, G., Ziol, M., Aout, M., Lhabadie, M., Baazia, Y., Mahmoudi, A., Roulot, D., Ganne-Carrie, N., Grando-Lemaire, V., Trinchet, J.-C., Gordien, E., Vicaut, E., Baghad, I. and Beaugrand, M. (2011) 'HCV genotype 3 is associated with a higher hepatocellular carcinoma incidence in patients with ongoing viral C cirrhosis', *Journal of Viral Hepatitis*. Blackwell Publishing Ltd, 18(10), pp. e516–e522. doi: 10.1111/j.1365-2893.2011.01441.x.

Noguchi, T., Satoh, S., Noshi, T., Hatada, E., Fukuda, R., Kawai, A., Ikeda, S., Hijikata, M. and Shimotohno, K. (2001) 'Effects of mutation in hepatitis C virus nonstructural protein 5A on interferon resistance mediated by inhibition of PKR kinase activity in mammalian cells.', *Microbiology and immunology*. Japan, 45(12), pp. 829–840.

Ohto, U., Shibata, T., Tanji, H., Ishida, H., Krayukhina, E., Uchiyama, S., Miyake, K. and Shimizu, T. (2015) 'Structural basis of CpG and inhibitory DNA recognition by Toll-like receptor 9', *Nature*. Nature Publishing Group, 520(7549), pp. 4–6. doi: 10.1038/nature14138.

Okamoto, H., Okada, S., Sugiyama, Y., Kurai, K., Iizuka, H., Machida, A., Miyakawa, Y. and Mayumi, M. (1991) 'Nucleotide sequence of the genomic RNA of hepatitis C virus isolated from a human carrier: Comparison with reported isolates for conserved and divergent regions', *Journal of General Virology*. ENGLAND, 72(11), pp. 2697–2704. doi: 10.1099/0022-1317-72-11-2697.

Osburn, W. O., Fisher, B. E., Dowd, K. a, Urban, G., Liu, L., Ray, S. C., Thomas, D. L. and Cox, A. L. (2010) 'Spontaneous control of primary hepatitis C virus infection and immunity against persistent reinfection.', *Gastroenterology*. Elsevier Inc., 138(1), pp. 315–24. doi: 10.1053/j.gastro.2009.09.017.

Ou, J. J., Bull, R. A., Luciani, F., McElroy, K., Gaudieri, S., Pham, S. T., Chopra, A., Cameron, B., Maher, L., Dore, G. J., White, P. A. and Lloyd, A. R. (2011) 'Sequential Bottlenecks Drive Viral Evolution in Early Acute Hepatitis C Virus Infection', *PLoS Pathogens*, 7(9), p. e1002243. doi: 10.1371/journal.ppat.1002243.

P. Simmonds Zt T.-A. Cha, S.-W. Chan, F. McOmish, B. Irvine, E. neall, P. L. Yap, J. Kolberg and M. S. Urdea, E. C. H., Simmonds, P., Holmes, E. C., Cha, T. A., Chan, S. W., McOmish, F., Irvine, B., Beall, E., Yap, P. L., Kolberg, J. and Urdea, M. S. (1993) 'Classification of hepatitis C virus into six major genotypes and a series of subtypes by phylogenetic analysis of the NS-5 region', *Journal of General Virology*. ENGLAND, 74, pp. 2391–2399. doi: 10.1099/0022-1317-74-11-2391.

Page, K., Hahn, J. A. A., Evans, J., Shiboski, S., Lum, P., Delwart, E., Tobler, L., Andrews, W., Avanesyan, L., Cooper, S. and Busch, M. P. P. (2009) 'Acute Hepatitis C Virus Infection in Young Adult Injection Drug Users: A Prospective Study of Incident Infection, Resolution, and Reinfection', *The Journal of Infectious Diseases*. United States, 200(8), pp. 1216–1226. doi: 10.1086/605947.

Page-Shafer, K., Pappalardo, B. L., Tobler, L. H., Phelps, B. H., Edlin, B. R., Moss, A. R., Wright, T. L., Wright, D. J., O'Brien, T. R., Caglioti, S. and Busch, M. P. (2008) 'Testing strategy to identify cases of acute hepatitis C virus (HCV) infection and to project HCV incidence rates', *Journal of Clinical Microbiology*, 46(2), pp. 499–506. doi: 10.1128/JCM.01229-07.

Pang, P. S., Planet, P. J. and Glenn, J. S. (2009) 'The Evolution of the Major Hepatitis C Genotypes Correlates with Clinical Response to Interferon Therapy', *PLoS ONE*, 4(8). doi: 10.1371/journal.pone.0006579.g001.

Paolucci, S., Fiorina, L., Mariani, B., Gulminetti, R., Novati, S., Barbarini, G., Bruno, R. and Baldanti, F. (2013) 'Naturally occurring resistance mutations to inhibitors of HCV NS5A region and NS5B polymerase in DAA treatment-naïve patients.', *Virology journal*, 10, p. 355. doi: 10.1186/1743-422X-10-355.

Pariat, Y. F., Berka, J., Heiger, D. N., Schmitt, T., Vilenchik, M., Cohen, A. S., Foret, F. and Karger, B. L. (1993) 'Separation of DNA fragments by capillary electrophoresis using

replaceable linear polyacrylamide matrices.', *Journal of chromatography. A*. NETHERLANDS, 652(1), pp. 57–66. doi: 10.1016/0021-9673(93)80645-O.

Park, S.-H. and Rehermann, B. (2014) 'Immune responses to HCV and other hepatitis viruses.', *Immunity*. Elsevier Inc., 40(1), pp. 13–24. doi: 10.1016/j.immuni.2013.12.010.

Parola, P. and Raoult, D. (2001) 'Ticks and Tickborne Bacterial Diseases in Humans : An Emerging Infectious Threat', *Clinical Infectious Diseases*, 32, pp. 897–928.

Parsons, J., Hickman, M., Turnbull, P. J., McSweeney, T., Stimson, G. V, Judd, A. and Roberts, K. (2002) 'Over a decade of syringe exchange: results from 1997 UK survey.', *Addiction*, 97(7), pp. 845–50. doi: 10.1046/j.1360-0443.2002.00128.x.

Paulino, A. D., Ubhi, K., Rockenstein, E., Adame, A., Crews, L., Letendre, S., Ellis, R., Everall, I. P., Grant, I. and Masliah, E. (2011) 'Neurotoxic effects of the HCV core protein are mediated by sustained activation of ERK via TLR2 signaling.', *Journal of neurovirology*, 17(4), pp. 327–40. doi: 10.1007/s13365-011-0039-0.

Pauvolid-Correa, A., Solberg, O., Couto-Lima, D., Kenney, J., Serra-Freire, N., Brault, A., Nogueira, R., Langevin, S. and Komar, N. (2015) 'Nhumirim virus, a novel flavivirus isolated from mosquitoes from the Pantanal, Brazil.', *Archives of virology*. Austria, 160(1), pp. 21–27. doi: 10.1007/s00705-014-2219-8.

Pawlotsky, J. M. (2014) 'New hepatitis C therapies: The toolbox, strategies, and challenges', *Gastroenterology*. Elsevier, Inc, 146(5), pp. 1176–1192. doi: 10.1053/j.gastro.2014.03.003.

Pawlotsky, J.-M., Negro, F., Aghemo, A., Back, D., Dusheiko, G., Forns, X., Puoti, M. and Sarrazin, C. (2016) *EASL Recommendations on Treatment of Hepatitis C 2016*.

Pearlman, B. L. and Traub, N. (2011) 'Sustained virologic response to antiviral therapy for chronic hepatitis C virus infection: A cure and so much more', *Clinical Infectious Diseases*, 52(7), pp. 889–900. doi: 10.1093/cid/cir076.

Pearson, W. R. and Lipman, D. J. (1988) 'Improved tools for biological sequence

comparison.', *Proceedings of the National Academy of Sciences of the United States of America*. UNITED STATES, 85(8), pp. 2444–2448.

Perico, N., Cattaneo, D., Bikbov, B. and Remuzzi, G. (2009) 'Hepatitis C infection and chronic renal diseases', *Clinical Journal of the American Society of Nephrology*, 4(1), pp. 207–220. doi: 10.2215/CJN.03710708.

Peterson, P. K., Sharp, B., Gekker, G., Brummitt, C. and Keane, W. F. (1987) 'Opioid-mediated suppression of cultured peripheral blood mononuclear cell respiratory burst activity', *J.Immunol.*, 138(11), pp. 3907–3912.

Pfaender, S., Brown, R. J., Pietschmann, T. and Steinmann, E. (2014) 'Natural reservoirs for homologs of hepatitis C virus', *Emerging Microbes & Infections*, 3(3), p. e21. doi: 10.1038/emi.2014.19.

Pfeiffer, J. K. and Kirkegaard, K. (2005) 'Increased fidelity reduces poliovirus fitness and virulence under selective pressure in mice.', *PLoS pathogens*. United States, 1(2), p. e11. doi: 10.1371/journal.ppat.0010011.

Pham, S. T., Bull, R. A., Bennett, J. M., Rawlinson, W. D., Dore, G. J., Lloyd, A. R. and White, P. A. (2010) 'Frequent multiple hepatitis C virus infections among injection drug users in a prison setting', *Hepatology*, 52(5), pp. 1564–1572. doi: 10.1002/hep.23885.

Pickett, B. E., Striker, R. and Lefkowitz, E. J. (2011) 'Evidence for separation of HCV subtype 1a into two distinct clades', *Journal of Viral Hepatitis*, 18(9), pp. 608–618. doi: 10.1111/j.1365-2893.2010.01342.x.

Pietschmann, T., Kaul, A., Koutsoudakis, G., Shavinskaya, A., Kallis, S., Steinmann, E., Abid, K., Negro, F., Dreux, M., Cosset, F.-L. and Bartenschlager, R. (2006) 'Construction and characterization of infectious intragenotypic and intergenotypic hepatitis C virus chimeras.', *Proceedings of the National Academy of Sciences of the United States of America*. United States, 103(19), pp. 7408–7413. doi: 10.1073/pnas.0504877103.

Pileri, P., Uematsu, Y., Campagnoli, S., Galli, G., Falugi, F., Petracca, R., Weiner, A. J., Houghton, M., Rosa, D., Grandi, G. and Abrignani, S. (1998) 'Binding of hepatitis C virus

to CD81.', *Science*. UNITED STATES, 282(5390), pp. 938–941.

Ploss, A., Evans, M. J., Gaysinskaya, V. a, Panis, M., Jong, Y. P. De and Rice, C. M. (2009) 'Human occludin is a hepatitis C virus entry factor required for infection of mouse cells', *Nature*, 457(7231), pp. 882–886. doi: 10.1038/nature07684.Human.

Plotkin, J. B. and Kudla, G. (2011) 'Synonymous but not the same: the causes and consequences of codon bias.', *Nature reviews. Genetics*, 12(1), pp. 32–42. doi: 10.1038/nrg2899.

Polz, M. F. and Cavanaugh, C. M. (1998) 'Bias in template-to-product ratios in multitemplate PCR.', *Applied and environmental microbiology*. UNITED STATES, 64(10), pp. 3724–3730.

Poordad, F., Bronowicki, J.-P., Gordon, S. C., Zeuzem, S., Jacobson, I. M., Sulkowski, M. S., Poynard, T., Morgan, T. R., Molony, C., Pedicone, L. D., Sings, H. L., Burroughs, M. H., Sniukiene, V., Boparai, N., Goteti, V. S., Brass, C. a, Albrecht, J. K. and Bacon, B. R. (2012) 'Factors that predict response of patients with hepatitis C virus infection to boceprevir.', *Gastroenterology*. Elsevier Inc., 143(3), pp. 608–18. doi: 10.1053/j.gastro.2012.05.011.

Popescu, C.-I., Callens, N., Trinel, D., Roingeard, P., Moradpour, D., Descamps, V., Duverlie, G., Penin, F., Héliot, L., Rouillé, Y. and Dubuisson, J. (2011) 'NS2 Protein of Hepatitis C Virus Interacts with Structural and Non-Structural Proteins towards Virus Assembly', *PLoS Pathogens*, 7(2), p. e1001278. doi: 10.1371/journal.ppat.1001278.

Poynard, T. (2003) 'Effect of treatment with peginterferon or interferon alfa-2b and ribavirin on steatosis in patients infected with hepatitis C', *Hepatology*, 38(1), pp. 75–85. doi: 10.1053/jhep.2003.50267.

Poynard, T., Bedossa, P. and Opolon, P. (1997) 'Natural history of liver fibrosis progression in patients with chronic hepatitis C', *The Lancet*, 349(9055), pp. 825–832. doi: 10.1016/S0140-6736(96)07642-8.

Poynard, T., Bedossa, P. and Opolon, P. (1997) 'Natural history of liver fibrosis

progression in patients with chronic hepatitis C. The OBSVIRC, METAVIR, CLINIVIR, and DOSVIRC groups.', *Lancet*. ENGLAND, 349(9055), pp. 825–832.

Prachayangprecha, S., Schapendonk, C. M. E., Koopmans, M. P., Osterhaus, A. D. M. E., Schurch, A. C., Pas, S. D., van der Eijk, A. A., Poovorawan, Y., Haagmans, B. L. and Smits, S. L. (2014) 'Exploring the potential of next-generation sequencing in detection of respiratory viruses.', *Journal of clinical microbiology*. United States, 52(10), pp. 3722–3730. doi: 10.1128/JCM.01641-14.

Prince, A. M., Kuhns, W. J., Grady, G. F., Levine, R. W., Hazzi, C. and Millian, S. J. (1974) 'Hepatitis Without Serological Evidence of Exposure to Hepatitis-B', *The Lancet*, August, pp. 241–246.

Probst, a., Dang, T., Bochud, M., Egger, M., Negro, F. and Bochud, P. Y. (2011) 'Role of Hepatitis C virus genotype 3 in liver fibrosis progression - A systematic review and meta-analysis', *Journal of Viral Hepatitis*, 18(11), pp. 745–759. doi: 10.1111/j.1365-2893.2011.01481.x.

Public Health England (2015) *Hepatitis C in the UK, 2015 report*.

Puig-Basagoiti, F., Forns, X., Furcic, I., Ampurdanes, S., Gimenez-Barcons, M., Franco, S., Sanchez-Tapias, J. M. and Saiz, J.-C. (2005) 'Dynamics of hepatitis C virus NS5A quasispecies during interferon and ribavirin therapy in responder and non-responder patients with genotype 1b chronic hepatitis C.', *The Journal of general virology*. England, 86(Pt 4), pp. 1067–1075. doi: 10.1099/vir.0.80526-0.

Purcell, R. H., Alter, H. J. and Dienstag, J. L. (1976) 'Non-A, non-B hepatitis.', *The Yale Journal of Biology and Medicine*, 49, pp. 243–250.

Pybus, O. G., Cochrane, A., Holmes, E. C. and Simmonds, P. (2005) 'The hepatitis C virus epidemic among injecting drug users.', *Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases*, 5(2), pp. 131–9. doi: 10.1016/j.meegid.2004.08.001.

Pybus, O. G., Markov, P. V., Wu, A. and Tatem, A. J. (2007) 'Investigating the endemic

transmission of the hepatitis C virus', *International Journal for Parasitology*, 37(8–9), pp. 839–849. doi: 10.1016/j.ijpara.2007.04.009.

Quail, M. a, Otto, T. D., Gu, Y., Harris, S. R., Skelly, T. F., McQuillan, J. a, Swerdlow, H. P. and Oyola, S. O. (2011) 'Optimal enzymes for amplifying sequencing libraries', *Nature Methods*. Nature Publishing Group, 9(1), pp. 10–11. doi: 10.1038/nmeth.1814.

Radford, A. D., Chapman, D., Dixon, L., Chantrey, J., Darby, A. C. and Hall, N. (2012) 'Application of next-generation sequencing technologies in virology.', *The Journal of general virology*, 93(Pt 9), pp. 1853–68. doi: 10.1099/vir.0.043182-0.

Radkowski, M., Gallegos-Orozco, J. F., Jablonska, J., Colby, T. V., Walewska-Zielecka, B., Kubicka, J., Wilkinson, J., Adair, D., Rakela, J. and Laskus, T. (2005) 'Persistence of hepatitis C virus in patients successfully treated for chronic hepatitis C', *Hepatology*, 41(1), pp. 106–114. doi: 10.1002/hep.20518.

Radkowski, M., Wilkinson, J., Nowicki, M., Adair, D., Vargas, H., Ingui, C., Rakela, J., Laskus, T. and * (2002) 'Search for Hepatitis C Virus Negative-Strand RNA Sequences and Analysis of Viral Sequences in the Central Nervous System: Evidence of Replication', *Jounral of Virology*, 76(2), pp. 600–608.

Ramirez, S., Perez-del-Pulgar, S., Carrion, J. A., Coto-Llerena, M., Mensa, L., Dragun, J., Garcia-Valdecasas, J. C., Navasa, M. and Forns, X. (2010) 'Hepatitis C virus superinfection of liver grafts: a detailed analysis of early exclusion of non-dominant virus strains', *Journal of General Virology*, 91(5), pp. 1183–1188. doi: 10.1099/vir.0.018929-0.

Ramos Paesa, C., Arazo Garces, P., Pascual Catalan, A., Hermida Lazcano, I. and Aguirre Errasti, J. M. (1998) '[Characteristics of chronic hepatitis C in patients infected by human immunodeficiency virus].', *Revista clinica espanola*. SPAIN, 198(4), pp. 207–211.

Rao, V. B., Johari, N., du Cros, P., Messina, J., Ford, N. and Cooke, G. S. (2016) 'Hepatitis C seroprevalence and HIV co-infection in sub-Saharan Africa: a systematic review and meta-analysis', *The Lancet Infectious Diseases*. Elsevier, 15(7), pp. 819–824. doi: 10.1016/S1473-3099(15)00006-7.

Ratziu, V., Munteanu, M., Charlotte, F., Bonyhay, L. and Poynard, T. (2003) 'Fibrogenic impact of high serum glucose in chronic hepatitis C.', *Journal of hepatology*. England, 39(6), pp. 1049–1055.

Razali, K., Thein, H. H., Bell, J., Cooper-Stanbury, M., Dolan, K., Dore, G., George, J., Kaldor, J., Karvelas, M., Li, J., Maher, L., McGregor, S., Hellard, M., Poeder, F., Quaine, J., Stewart, K., Tyrrell, H., Weltman, M., Westcott, O., Wodak, A. and Law, M. (2007) 'Modelling the hepatitis C virus epidemic in Australia.', *Drug and alcohol dependence*. Ireland, 91(2–3), pp. 228–235. doi: 10.1016/j.drugalcdep.2007.05.026.

Lo Re, V. 3rd, Kallan, M. J., Tate, J. P., Localio, A. R., Lim, J. K., Goetz, M. B., Klein, M. B., Rimland, D., Rodriguez-Barradas, M. C., Butt, A. A., Gibert, C. L., Brown, S. T., Park, L., Dubrow, R., Reddy, K. R., Kostman, J. R., Strom, B. L. and Justice, A. C. (2014) 'Hepatic decompensation in antiretroviral-treated patients co-infected with HIV and hepatitis C virus compared with hepatitis C virus-monoinfected patients: a cohort study.', *Annals of internal medicine*. United States, 160(6), pp. 369–379. doi: 10.7326/M13-1829.

Regier, D. a, Farmer, M. E., Rae, D. S., Locke, B. Z., Keith, S. J., Judd, L. L. and Goodwin, F. K. (1990) 'Cmorbidity of Mental Disorders With Alcohol and Other Drugs', *JAMA*, 264(19), pp. 2511–2518.

Reiberger, T., Ferlitsch, A., Sieghart, W., Kreil, A., Breitenecker, F., Rieger, A., Schmied, B., Gangl, A. and Peck-Radosavljevic, M. (2010) 'HIV-HCV co-infected patients with low CD4+ cell nadirs are at risk for faster fibrosis progression and portal hypertension.', *Journal of viral hepatitis*. England, 17(6), pp. 400–409. doi: 10.1111/j.1365-2893.2009.01197.x.

Richter, S. (2002) 'Laboratory Assays for Diagnosis and Management of Hepatitis C Virus Infection', *Journal of clinical microbiology*, 40(12), pp. 4407–4412. doi: 10.1128/JCM.40.12.4407.

Robinson, M., Swann, R., Sigruener, A., Barclay, S. T., Mills, P. R., McLauchlan, J. and Patel, A. H. (2015) 'Elevated interferon-stimulated gene transcription in peripheral blood mononuclear cells occurs in patients infected with genotype 1 but not genotype 3 hepatitis C virus.', *Journal of viral hepatitis*. England, 22(4), pp. 384–390. doi: 10.1111/jvh.12310.

Robinson, M. W., Aranday-Cortes, E., Gatherer, D., Swann, R., Liefhebber, J. M. P., Filipe, A. D. S., Sigruener, A., Barclay, S. T., Mills, P. R., Patel, A. H. and McLauchlan, J. (2015) 'Viral genotype correlates with distinct liver gene transcription signatures in chronic hepatitis C virus infection', *Liver International*, 35, pp. 2256–64. doi: 10.1111/liv.12830.

Rodriguez-Torres, M., Jeffers, L. J., Sheikh, M. Y., Rossaro, L., Ankoma-Sey, V., Hamzeh, F. M. and Martin, P. (2009) 'Peginterferon alfa-2a and ribavirin in Latino and non-Latino whites with hepatitis C.', *The New England journal of medicine*. United States, 360(3), pp. 257–267. doi: 10.1056/NEJMoa0805062.

Rong, X., Lu, L., Wang, J., Xiong, H., Huang, J., Chen, J., Huang, K., Xu, R., Wang, M., Zhang, X., Guo, T., Liu, Y., Gao, G., Fu, Y. and Nelson, K. E. (2012) 'Correlation of Viral Loads with HCV Genotypes: Higher Levels of Virus Were Revealed among Blood Donors Infected with 6a Strains', *PLoS ONE*, 7(12), pp. 1–7. doi: 10.1371/journal.pone.0052467.

Roque-Afonso, A.-M., Ducoulombier, D., Di Liberto, G., Kara, R., Gigou, M., Dussaix, E., Samuel, D. and Feray, C. (2005) 'Compartmentalization of Hepatitis C Virus Genotypes between Plasma and Peripheral Blood Mononuclear Cells', *Journal of Virology*, 79(10), pp. 6349–6357. doi: 10.1128/JVI.79.10.6349.

Rosseel, T., Scheuch, M., Hoper, D., De Regge, N., Caij, A. B., Vandenbussche, F. and Van Borm, S. (2012) 'DNase SISPA-next generation sequencing confirms Schmallenberg virus in Belgian field samples and identifies genetic variation in Europe.', *PloS one*. United States, 7(7), p. e41967. doi: 10.1371/journal.pone.0041967.

Roy, K. M., Hutchinson, S. J., Wadd, S., Taylor, A., Cameron, S. O., Burns, S., Molyneaux, P., McIntyre, P. G. and Goldberg, D. J. (2007) 'Hepatitis C virus infection among injecting drug users in Scotland: a review of prevalence and incidence data and the methods used to generate them', *Epidemiol Infect*, 135(3), pp. 433–442. doi: 10.1017/s0950268806007035.

Roychowdhury, S., Iyer, M. K., Robinson, D. R., Lonigro, R. J., Wu, Y.-M., Cao, X., Kalyana-Sundaram, S., Sam, L., Balbin, O. A., Quist, M. J., Barrette, T., Everett, J., Siddiqui, J., Kunju, L. P., Navone, N., Araujo, J. C., Troncoso, P., Logothetis, C. J., Innis,

J. W., Smith, D. C., Lao, C. D., Kim, S. Y., Roberts, J. S., Gruber, S. B., Pienta, K. J., Talpaz, M. and Chinnaiyan, A. M. (2012) 'Personalized Oncology Through Integrative High-Throughput Sequencing: A Pilot Study', *Science Translational Medicine*, 3(111). doi: 10.1126/scitranslmed.3003161.Personalized.

Russell, G. J., Walker, P. M., Elton, R. A. and Subak-Sharpe, J. H. (1976) 'Doublet frequency analysis of fractionated vertebrate nuclear DNA.', *Journal of molecular biology*. ENGLAND, 108(1), pp. 1–23.

Sacks-Davis, R., Grebely, J., Dore, G. J., Osburn, W., Cox, A. L., Rice, T. M., Spelman, T., Bruneau, J., Prins, M. and Kim, A. Y. (2013) 'Hepatis C virus reinfection and spontaneous clearance of reinfection - the InC3 study', 742(12). doi: 10.1016/S1001-0742(12)60252-6.

Sacks-Davis, R., Mcbryde, E., Grebely, J., Hellard, M., Vickerman, P. and Sacks-Davis, R. (2015) 'Many hepatitis C reinfections that spontaneously clear may be undetected: Markov-chain Monte Carlo analysis of observational study data.', *Journal of the Royal Society, Interface / the Royal Society*, 12(104). doi: 10.1098/rsif.2014.1197.

Saeed, M., Andreo, U., Chung, H.-Y., Espiritu, C., Branch, A. D., Silva, J. M. and Rice, C. M. (2015) 'SEC14L2 enables pan-genotype HCV replication in cell culture', *Nature*. doi: 10.1038/nature14899.

Sakamoto, M., Akahane, Y., Tsuda, F., Tanaka, T., Woodfield, D. G. and Okamoto, H. (1994) 'Entire nucleotide sequence and characterization of a hepatitis C virus of genotype V/3a.', *The Journal of general virology*. ENGLAND, 75 (Pt 7), pp. 1761–1768. doi: 10.1099/0022-1317-75-7-1761.

Salemi, M. and Vandamme, A.-M. (2002) 'Hepatitis C virus evolutionary patterns studied through analysis of full-genome sequences.', *Journal of molecular evolution*, 54(May), pp. 62–70. doi: 10.1007/s00239-001-0018-9.

Sang, R. C., Gichogo, A., Gachoya, J., Dunster, M. D., Ofula, V., Hunt, A. R., Crabtree, M. B., Miller, B. R. and Dunster, L. M. (2003) 'Isolation of a new flavivirus related to cell fusing agent virus (CFAV) from field-collected flood-water Aedes mosquitoes sampled

from a dambo in central Kenya.', *Archives of virology*. Austria, 148(6), pp. 1085–1093. doi: 10.1007/s00705-003-0018-8.

Sanger, F. and Coulson, A. R. (1975) 'A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase.', *Journal of molecular biology*. ENGLAND, 94(3), pp. 441–448.

Sanger, F., Coulson, A. R., Friedmann, T., Air, G. M., Barrell, B. G., Brown, N. L., Fiddes, J. C., Hutchison, C. A. 3rd, Slocombe, P. M. and Smith, M. (1978) 'The nucleotide sequence of bacteriophage phiX174.', *Journal of molecular biology*. ENGLAND, 125(2), pp. 225–246.

Sanger, F., Nicklen, S. and Coulson, A. R. (1977) 'DNA sequencing with chain-terminating inhibitors.', *Proceedings of the National Academy of Sciences of the United States of America*. UNITED STATES, 74(12), pp. 5463–5467.

Sarobe, P., Lasarte, J.-J., Garcia, N., Civeira, M. P., Borras-Cuesta, F. and Prieto, J. (2006) 'Characterization of T-cell responses against immunodominant epitopes from hepatitis C virus E2 and NS4a proteins.', *Journal of viral hepatitis*. England, 13(1), pp. 47–55. doi: 10.1111/j.1365-2893.2005.00653.x.

Sarrazin, C. (2016) 'The importance of resistance to direct antiviral drugs in HCV infection in clinical practice', *Journal of Hepatology*. European Association for the Study of the Liver, 64(2), pp. 486–504. doi: 10.1016/j.jhep.2015.09.011.

Sarrazin, C., Hézode, C., Zeuzem, S. and Pawlotsky, J.-M. (2012) 'Antiviral strategies in hepatitis C virus infection.', *Journal of hepatology*, 56 Suppl 1, pp. S88-100. doi: 10.1016/S0168-8278(12)60010-5.

Scarselli, E., Ansuini, H., Cerino, R., Roccasecca, R. M., Acali, S., Filocamo, G., Traboni, C., Nicosia, A., Cortese, R. and Vitelli, A. (2002) 'The human scavenger receptor class B type I is a novel candidate receptor for the hepatitis C virus.', *The EMBO journal*, 21(19), pp. 5017–5025. doi: 10.1093/emboj/cdf529.

Schaller, T., Appel, N., Koutsoudakis, G., Kallis, S., Lohmann, V., Pietschmann, T. and

Bartenschlager, R. (2007) ‘Analysis of Hepatitis C Virus Superinfection Exclusion by Using Novel Fluorochrome Gene-Tagged Viral Genomes’, *Journal of Virology*, 81(9), pp. 4591–4603. doi: 10.1128/jvi.02144-06.

Schmucker, D. L. (2005) ‘Age-related changes in liver structure and function: Implications for disease ?’, *Experimental gerontology*. England, 40(8–9), pp. 650–659. doi: 10.1016/j.exger.2005.06.009.

Schröter, M., Feucht, H.-H., Zöllner, B., Schäfer, P. and Laufs, R. (2003) ‘Multiple infections with different HCV genotypes: prevalence and clinical impact’, *Journal of Clinical Virology*, 27(2), pp. 200–204. doi: 10.1016/s1386-6532(02)00264-0.

Schwaiger, M. and Cassinotti, P. (2003) ‘Development of a quantitative real-time RT-PCR assay with internal control for the laboratory detection of tick borne encephalitis virus (TBEV) RNA.’, *Journal of clinical virology : the official publication of the Pan American Society for Clinical Virology*. Netherlands, 27(2), pp. 136–145.

Schweizer, M. and Peterhans, E. (2001) ‘Noncytopathic Bovine Viral Diarrhea Virus Inhibits Double-Stranded RNA-Induced Apoptosis and Interferon’, *Journal of Virology*, 75(10), pp. 4692–4698. doi: 10.1128/JVI.75.10.4692.

Selby, M. J., Choo, Q. L., Berger, K., Kuo, G., Glazer, E., Eckart, M., Lee, C., Chien, D., Kuo, C. and Houghton, M. (1993) ‘Expression, identification and subcellular localization of the proteins encoded by the hepatitis C viral genome.’, *The Journal of General Virology*. ENGLAND, 74 (Pt 6), pp. 1103–1113. doi: 10.1099/0022-1317-74-6-1103.

Sereno, S., Perinelli, P. and Laghi, V. (2009) ‘Changes in the prevalence of hepatitis C virus genotype among Italian injection drug users-Relation to period of injection started’, *Journal of Clinical Virology*, 45(4), pp. 354–357. doi: 10.1016/j.jcv.2009.04.022.

Serpaggi, J., Chaix, M.-L., Batisse, D., Dupont, C., Vallet-Pichard, A., Fontaine, H., Viard, J.-P., Piketty, C., Rouveix, E., Rouzioux, C., Weiss, L. and Pol, S. (2006) ‘Sexually transmitted acute infection with a clustered genotype 4 hepatitis C virus in HIV-1-infected men and inefficacy of early antiviral therapy.’, *AIDS*. England, 20(2), pp. 233–240. doi: 10.1097/01.aids.0000200541.40633.56.

Shao, W., Boltz, V. F., Spindler, J. E., Kearney, M. F., Maldarelli, F., Mellors, J. W., Stewart, C., Volfovsky, N., Levitsky, A., Stephens, R. M. and Coffin, J. M. (2013) 'Analysis of 454 sequencing error rate, error sources, and artifact recombination for detection of Low-frequency drug resistance mutations in HIV-1 DNA.', *Retrovirology*, 10(1), p. 18. doi: 10.1186/1742-4690-10-18.

Sharma, N. R., Mateu, G., Dreux, M., Grakoui, A., Cosset, F. L. and Melikyan, G. B. (2011) 'Hepatitis C virus is primed by CD81 protein for low pH-dependent fusion', *Journal of Biological Chemistry*, 286(35), pp. 30361–30376. doi: 10.1074/jbc.M111.263350.

Shimizu, I., Mizobuchi, Y., Yasuda, M., Shiba, M., Ma, Y. R., Horie, T., Liu, F. and Ito, S. (1999) 'Inhibitory effect of oestradiol on activation of rat hepatic stellate cells in vivo and in vitro.', *Gut*, 44(1), pp. 127–36. doi: 10.1136/gut.44.1.127.

Shirai, M., Arichi, T., Chen, M., Masaki, T., Nishioka, M., Ikeda, K., Takahashi, H., Enomoto, N., Saito, T., Major, M. E., Nakazawa, T., Akatsuka, T., Feinstone, S. M. and Berzofsky, J. A. (1999) 'T cell recognition of hypervariable region-1 from hepatitis C virus envelope protein with multiple class II MHC molecules in mice and humans: preferential help for induction of antibodies to the hypervariable region.', *Journal of immunology (Baltimore, Md. : 1950)*. UNITED STATES, 162(1), pp. 568–576.

Shuhart, M. C., Sullivan, D. G., Bekele, K., Harrington, R. D., Kitahata, M. M., Mathisen, T. L., Thomassen, L. V, Emerson, S. S. and Gretch, D. R. (2006) 'HIV infection and antiretroviral therapy: effect on hepatitis C virus quasispecies variability.', *The Journal of infectious diseases*. United States, 193(9), pp. 1211–1218. doi: 10.1086/502974.

Simen, B. B., Simons, J. F., Hullsiek, K. H., Novak, R. M., MacArthur, R. D., Baxter, J. D., Huang, C., Lubeski, C., Turenchalk, G. S., Braverman, M. S., Desany, B., Rothberg, J. M., Egholm, M. and Kozal, M. J. (2009) 'Low-Abundance Drug-Resistant Viral Variants in Chronically HIV-Infected, Antiretroviral Treatment–Naive Patients Significantly Impact Treatment Outcomes', *Journal of Infectious Diseases*, 199(5), pp. 693–701. doi: 10.1086/596736.

Simmonds, P. (2004) 'Genetic diversity and evolution of hepatitis C virus - 15 years on',

Journal of General Virology, 85, pp. 3173–3188. doi: 10.1099/vir.0.80401-0.

Simmonds, P. (2012) 'SSE: a nucleotide and amino acid sequence analysis platform.', *BMC research notes*. BioMed Central Ltd, 5(1), p. 50. doi: 10.1186/1756-0500-5-50.

Simmonds, P., Bukh, J., Combet, C., Deléage, G., Enomoto, N., Feinstone, S., Halfon, P., Inchauspé, G., Kuiken, C., Maertens, G., Mizokami, M., Murphy, D. G., Okamoto, H., Pawlotsky, J.-M., Penin, F., Sablon, E., Shin-I, T., Stuyver, L. J., Thiel, H.-J., Viazov, S., Weiner, A. J. and Widell, A. (2005) 'Consensus proposals for a unified system of nomenclature of hepatitis C virus genotypes', *Hepatology*, 42(4), pp. 962–973. doi: 10.1002/hep.20819.

Simmonds, P., Xia, W., Baillie, J. K. and McKinnon, K. (2013) 'Modelling mutational and selection pressures on dinucleotides in eukaryotic phyla--selection against CpG and UpA in cytoplasmically expressed RNA and in RNA viruses.', *BMC genomics*. BMC Genomics, 14(1), p. 610. doi: 10.1186/1471-2164-14-610.

Simmons, B., Saleem, J., Heath, K., Cooke, G. S. and Hill, A. (2015) 'Long-term treatment outcomes of patients infected with hepatitis C virus: A systematic review and meta-analysis of the survival benefit of achieving a sustained virological response', *Clinical Infectious Diseases*, 61(5), pp. 730–740. doi: 10.1093/cid/civ396.

Simmons, N. B., Seymour, K. L., Habersetzer, J. and Gunnell, G. F. (2008) 'Primitive Early Eocene bat from Wyoming and the evolution of flight and echolocation.', *Nature*. England, 451(7180), pp. 818–821. doi: 10.1038/nature06549.

Sims, D., Sudbery, I., Ilott, N. E., Heger, A. and Ponting, C. P. (2014) 'Sequencing depth and coverage: key considerations in genomic analyses.', *Nature reviews. Genetics*. Nature Publishing Group, 15(2), pp. 121–32. doi: 10.1038/nrg3642.

Singh, S., Malhotra, V. and Sarin, S. K. (2004) 'Distribution of hepatitis C virus genotypes in patients with chronic hepatitis C infection in India', *Indian Journal of Medical Research*, (April), pp. 145–148.

Siva, N. (2015) 'UK gears up to decode 100,000 genomes from NHS patients.', *Lancet*.

England, pp. 103–104. doi: 10.1016/S0140-6736(14)62453-3.

Smith, D. B., Bukh, J., Kuiken, C., Muerhoff, A. S., Rice, C. M., Stapleton, J. T. and Simmonds, P. (2014) 'Expanded classification of hepatitis C virus into 7 genotypes and 67 subtypes: updated criteria and genotype assignment Web resource.', *Hepatology*, 59(1), pp. 318–27. doi: 10.1002/hep.26744.

Smith, D. B., Pathirana, S., Davidson, F., Lawlor, E., Power, J., Yap, P. L. and Simmonds, P. (1997) 'The origin of hepatitis C virus genotypes.', *The Journal of general virology*, 78 (Pt 2), pp. 321–328.

Smith, J. a, Aberle, J. H., Fleming, V. M., Ferenci, P., Thomson, E. C., Karayiannis, P., McLean, A. R., Holzmann, H. and Klenerman, P. (2010) 'Dynamic coinfection with multiple viral subtypes in acute hepatitis C.', *The Journal of infectious diseases*, 202(12), pp. 1770–9. doi: 10.1086/657317.

Smyk-Pearson, S., Tester, I. a, Klarquist, J., Palmer, B. E., Pawlotsky, J.-M., Golden-Mason, L. and Rosen, H. R. (2008) 'Spontaneous recovery in acute human hepatitis C virus infection: functional T-cell thresholds and relative importance of CD4 help.', *Journal of virology*, 82(4), pp. 1827–1837. doi: 10.1128/JVI.01581-07.

Sobesky, R., Feray, C., Rimlinger, F., Derian, N., Dos Santos, A., Roque-Afonso, A. M., Samuel, D., Br??chot, C. and Thiers, V. (2007) 'Distinct hepatitis C virus core and F protein quasispecies in tumoral and nontumoral hepatocytes isolated via microdissection', *Hepatology*, 46(6), pp. 1704–1712. doi: 10.1002/hep.21898.

Soriano, V., Mocroft, A., Rockstroh, J., Ledergerber, B., Knysz, B., Chaplinskas, S., Peters, L., Karlsson, A., Katlama, C., Toro, C., Kupfer, B., Vogel, M. and Lundgren, J. (2008) 'Spontaneous viral clearance, viral load, and genotype distribution of hepatitis C virus (HCV) in HIV-infected patients with anti-HCV antibodies in Europe.', *The Journal of infectious diseases*. United States, 198(9), pp. 1337–1344. doi: 10.1086/592171.

Spiegel, B. M. R., Younossi, Z. M., Hays, R. D., Revicki, D., Robbins, S. and Kanwal, F. (2005) 'Impact of hepatitis C on health related quality of life: A systematic review and quantitative assessment', *Hepatology*, 41(4), pp. 790–800. doi: 10.1002/hep.20659.

Stamenkovic, G., Zerjav, S., Velickovic, Z. M., Krtolica, K., Samardzija, V. L., Jemuovic, L., Nozic, D. and Dimitrijevic, B. (2000) 'Distribution of HCV genotypes among risk groups in Serbia', *European Journal of Epidemiology*, 16(10), pp. 949–954. doi: 10.1023/A:1011060505152.

Steinhauer, D. A., Domingo, E. and Holland, J. J. (1992) 'Lack of evidence for proofreading mechanisms associated with an RNA virus polymerase.', *Gene*. NETHERLANDS, 122(2), pp. 281–288.

Steinmann, E. and Pietschmann, T. (2010) 'Hepatitis C virus P7-A viroporin crucial for virus assembly and an emerging target for antiviral therapy', *Viruses*, 2(9), pp. 2078–2095. doi: 10.3390/v2092078.

Stiffler, J. D., Nguyen, M., Sohn, J. A., Liu, C., Kaplan, D. and Seeger, C. (2009) 'Focal distribution of hepatitis C virus RNA in infected livers.', *PloS one*. United States, 4(8), p. e6661. doi: 10.1371/journal.pone.0006661.

Stiffler, J. D., Nguyen, M., Sohn, J. a, Liu, C., Kaplan, D. and Seeger, C. (2009) 'Focal distribution of hepatitis C virus RNA in infected livers.', *PloS one*. United States, 4(8), p. e6661. doi: 10.1371/journal.pone.0006661.

Sugawara, Y., Makuuchi, M., Kato, N., Shimotohno, K. and Takada, K. (1999) 'Enhancement of hepatitis C virus replication by Epstein-Barr virus- encoded nuclear antigen 1', *Embo J*, 18(0261–4189 SB–IM), pp. 5755–5760.

Sugawara, Y., Mizugaki, Y., Uchida, T., Torii, T., Imai, S., Makuuchi, M. and Takada, K. (1999) 'Detection of Epstein–Barr Virus (EBV) in Hepatocellular Carcinoma Tissue: A Novel {EBV} Latency Characterized by the Absence of EBV-Encoded Small {RNA} Expression', *Virology*, 256(2), pp. 196–202. doi: http://dx.doi.org/10.1006/viro.1999.9619.

Sulkowski, M. S. (2001) 'Hepatitis C Virus Infection in HIV-infected Patients.', *Current infectious disease reports*, 3(5), pp. 469–476.

Sutton, A. J., McDonald, S. A., Palmateer, N., Taylor, A. and Hutchinson, S. J. (2012) 'Estimating the variability in the risk of infection for hepatitis C in the Glasgow injecting

drug user population', *Epidemiology and Infection*, 140(12), pp. 2190–2198. doi: 10.1017/s0950268812000489.

Svarovskaia, E. S., Martin, R., McHutchison, J. G., Miller, M. D. and Mo, H. (2012) 'Abundant drug-resistant NS3 mutants detected by deep sequencing in hepatitis C virus-infected patients undergoing NS3 protease inhibitor monotherapy', *Journal of Clinical Microbiology*, 50(10), pp. 3267–3274. doi: 10.1128/JCM.00838-12.

Swerdlow, H. and Gesteland, R. (1990) 'Capillary gel electrophoresis for rapid, high resolution DNA sequencing.', *Nucleic acids research*. ENGLAND, 18(6), pp. 1415–1419.

Tabor, E., Gerety, R. J., Drucker, J. A., Seeff, L. B., Hoofnagle, J. H., Jackson, D. R., April, M., Barker, L. F. and Pineda-Tamondong, G. (1978) 'Transmission of non-A, non-B hepatitis from man to chimpanzee.', *Lancet*. ENGLAND, 1(8062), pp. 463–466.

Takahashi, K., Asabe, S., Wieland, S., Garaigorta, U., Gastaminza, P., Isogawa, M. and Chisari, F. V (2010) 'Plasmacytoid dendritic cells sense hepatitis C virus-infected cells, produce interferon, and inhibit infection.', *Proceedings of the National Academy of Sciences of the United States of America*, 107(16), pp. 7431–7436. doi: 10.1073/pnas.1002301107.

Takahashi, T. and Nakayama, T. (2006) 'Novel Technique of Quantitative Nested Real-Time PCR Assay for Mycobacterium tuberculosis DNA', *Society*, 44(3), pp. 1029–1039. doi: 10.1128/JCM.44.3.1029.

Takaki, A., Wiese, M., Maertens, G., Depla, E., Seifert, U., Liebetrau, A., Miller, J. L., Manns, M. P. and Rehermann, B. (2000) 'Cellular immune responses persist and humoral responses decrease two decades after recovery from a single-source outbreak of hepatitis C.', *Nature medicine*. UNITED STATES, 6(5), pp. 578–582. doi: 10.1038/75063.

Takhampunya, R., Kim, H. C., Tippayachai, B., Lee, D.-K., Lee, W.-J., Chong, S.-T., Kim, M. S., Lee, J. S. and Klein, T. A. (2014) 'Distribution and mosquito hosts of Chaoyang virus, a newly reported flavivirus from the Republic of Korea, 2008-2011.', *Journal of medical entomology*. United States, 51(2), pp. 464–474.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M. and Kumar, S. (2011) 'MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods.', *Molecular biology and evolution*, 28(10), pp. 2731–9. doi: 10.1093/molbev/msr121.

Tellinghuisen, T. L., Foss, K. L. and Treadaway, J. (2008) 'Regulation of hepatitis C virion production via phosphorylation of the NS5A protein.', *PLoS pathogens*, 4(3), p. e1000032. doi: 10.1371/journal.ppat.1000032.

Temin, H. M. and Mizutani, S. (1970) 'Viral RNA-dependent DNA Polymerase: RNA-dependent DNA Polymerase in Virions of Rous Sarcoma Virus', *Nature*, 226(5252), pp. 1211–1213.

Terrault, N. A., Dodge, J. L., Murphy, E. L., Tavis, J. E., Kiss, A., Levin, T. R., Gish, R. G., Busch, M. P., Reingold, A. L. and Alter, M. J. (2013) 'Sexual transmission of hepatitis C virus among monogamous heterosexual couples: the HCV partners study.', *Hepatology*. United States, 57(3), pp. 881–889. doi: 10.1002/hep.26164.

The Medical Letter (2016) 'Sofosbuvir/velpatasvir (Epclusa) for hepatitis C.', *The Medical letter on drugs and therapeutics*. United States, 58(1501), pp. 107–108.

Thein, H.-H., Yi, Q., Dore, G. J. and Krahn, M. D. (2008) 'Estimation of stage-specific fibrosis progression rates in chronic hepatitis C virus infection: A meta-analysis and meta-regression', *Hepatology*, 48(2), pp. 418–431. doi: 10.1002/hep.22375.

Thimme, R., Oldach, D., Chang, K. M., Steiger, C., Ray, S. C. and Chisari, F. V (2001) 'Determinants of viral clearance and persistence during acute hepatitis C virus infection', *J.Exp.Med.*, 194(0022–1007), pp. 1395–1406.

Thomas, D. L., Thio, C. L., Martin, M. P., Qi, Y., Ge, D., O'Huigin, C., Kidd, J., Kidd, K., Khakoo, S. I., Alexander, G., Goedert, J. J., Kirk, G. D., Donfield, S. M., Rosen, H. R., Tobler, L. H., Busch, M. P., McHutchison, J. G., Goldstein, D. B. and Carrington, M. (2009) 'Genetic variation in IL28B and spontaneous clearance of hepatitis C virus.', *Nature*. England: Nature Publishing Group, 461(7265), pp. 798–801. doi: 10.1038/nature08463.

Thompson, J. F. and Steinmann, K. E. (2010) ‘Single molecule sequencing with a HeliScope genetic analysis system.’, *Current protocols in molecular biology / edited by Frederick M. Ausubel ... [et al.]*. United States, Chapter 7, p. Unit7.10. doi: 10.1002/0471142727.mb0710s92.

Thomson, E. C., Fleming, V. M., Main, J., Klenerman, P., Weber, J., Eliahoo, J., Smith, J., McClure, M. O. and Karayiannis, P. (2011) ‘Predicting spontaneous clearance of acute hepatitis C virus in a large cohort of HIV-1-infected men’, *Gut*, 60(6), pp. 837–845. doi: 10.1136/gut.2010.217166.

Thomson, E. C., Smith, J. A. and Klenerman, P. (2011) ‘The natural history of early hepatitis C virus evolution; lessons from a global outbreak in human immunodeficiency virus-1-infected individuals’, *Journal of General Virology*, 92(10), pp. 2227–2236. doi: 10.1099/vir.0.033910-0.

Thomson, E., Ip, C. L. C., Badhan, A., Christiansen, M. T., Adamson, W., Ansari, M. A., Bibby, D., Breuer, J., Brown, A., Bowden, R., Bryant, J., Bonsall, D., Silva, D., Hinds, C., Hudson, E., Klenerman, P., Lythgow, K., Mbisa, J. L., Mclauchlan, J., Myers, R., Piazza, P., Roy, S. and Trebes, A. (2016) ‘Comparison of Next-Generation Sequencing Technologies for Comprehensive Assessment of Full-Length Hepatitis C Viral Genomes’, *Journal of clinical microbiology*, 54(10). doi: 10.1128/JCM.00330-16.Editor.

Thomson, E., Ip, C. L. C., Badhan, A., Christiansen, M. T., Ansari, M. A., Bibby, D., Breuer, J., Brown, A., Bowden, R., Bryant, J., Bonsall, D., Da, A., Filipe, S., Hinds, C., Hudson, E., Lythgow, K. K., Mbisa, J. L., Mclauchlan, J., Myers, R., Roy, S., Trebes, A., Vattipally, S. B. and Witteveldt, J. (2016) ‘Comparison of next generation sequencing technologies for the comprehensive assessment of full-length hepatitis C viral genomes’, (July), pp. 1–39. doi: 10.1128/JCM.00330-16.

Thorburn, F., Bennett, S., Modha, S., Murdoch, D., Gunson, R. and Murcia, P. R. (2015) ‘The use of next generation sequencing in the diagnosis and typing of respiratory infections’, *Journal of Clinical Virology*. Elsevier B.V., 69, pp. 96–100. doi: 10.1016/j.jcv.2015.06.082.

Tiffin, P. and Ross-Ibarra, J. (2014) ‘Swarm v2: highly-scalable and high-resolution

amplicon clustering', *PeerJ PrePints*, pp. 1–20. doi: http://dx.doi.org/10.7287/peerj.preprints.488v1.

Tillmann, H. L., Thompson, A. J., Patel, K., Wiese, M., Tenckhoff, H., Nischalke, H. D., Lokhnygina, Y., Kullig, U., Gobel, U., Capka, E., Wiegand, J., Schiefke, I., Guthoff, W., Grungreiff, K., Konig, I., Spengler, U., McCarthy, J., Shianna, K. V, Goldstein, D. B., McHutchison, J. G., Timm, J. and Nattermann, J. (2010) 'A polymorphism near IL28B is associated with spontaneous clearance of acute hepatitis C virus and jaundice.', *Gastroenterology*. United States, 139(5), p. 1586–92, 1592.e1. doi: 10.1053/j.gastro.2010.07.005.

Timpe, J. M., Stamataki, Z., Jennings, A., Hu, K., Farquhar, M. J., Harris, H. J., Schwarz, A., Desombere, I., Roels, G. L., Balfe, P. and McKeating, J. A. (2008) 'Hepatitis C virus cell-cell transmission in hepatoma cells in the presence of neutralizing antibodies.', *Hepatology (Baltimore, Md.)*. United States, 47(1), pp. 17–24. doi: 10.1002/hep.21959.

Tokita, H., Okamoto, H., Tsuda, F., Song, P., Nakata, S., Chosa, T., Iizuka, H., Mishiro, S., Miyakawa, Y. and Mayumi, M. (1994) 'Hepatitis C virus variants from Vietnam are classifiable into the seventh, eighth, and ninth major genetic groups.', *Proceedings of the National Academy of Sciences of the United States of America*. UNITED STATES, 91(23), pp. 11022–11026.

Toyoda, H., Fukuda, Y., Nakano, I., Katano, Y., Takayama, T., Kumada, T., Nakano, S., Takamatsu, J., Saito, H. and Hayakawa, T. (1998) 'Quasispecies Nature of Hepatitis C Virus (HCV) in HCV Subtypes', *Journal of Medical Virology*, 85(October 1997), pp. 80–85.

Tran, T. M., Aghili, A., Li, S., Ongoiba, A., Kayentao, K., Doumbo, S., Traore, B. and Crompton, P. D. (2014) 'A nested real-time PCR assay for the quantification of Plasmodium falciparum DNA extracted from dried blood spots', *Malar J*, 13, p. 393. doi: 10.1186/1475-2875-13-393.

Trepo, C. (2000) 'Genotype and viral load as prognostic indicators in the treatment of hepatitis C.', *Journal of viral hepatitis*, 7(4), pp. 250–257.

Trepo, C. and Pradat, P. (1999) ‘Hepatitis C virus infection in Western Europe.’, *Journal of hepatology*. DENMARK, 31 Suppl 1, pp. 80–83.

Tsatsralt-Od, B., Takahashi, M., Nishizawa, T., Endo, K., Inoue, J. and Okamoto, H. (2005) ‘High prevalence of dual or triple infection of hepatitis B, C, and delta viruses among patients with chronic liver disease in Mongolia.’, *Journal of medical virology*. United States, 77(4), pp. 491–499. doi: 10.1002/jmv.20482.

Tscherne, D. M., Evans, M. J., von Hahn, T., Jones, C. T., Stamataki, Z., McKeating, J. A., Lindenbach, B. D. and Rice, C. M. (2007) ‘Superinfection Exclusion in Cells Infected with Hepatitis C Virus’, *Journal of Virology*, 81(8), pp. 3693–3703. doi: 10.1128/jvi.01748-06.

Tulloch, F., Atkinson, N. J., Evans, D. J., Ryan, M. D. and Simmonds, P. (2014) ‘RNA virus attenuation by codon pair deoptimisation is an artefact of increases in CpG/UpA dinucleotide frequencies’, *eLife*, 3, pp. 1–15. doi: 10.7554/eLife.04531.

Tully, D. C., Hjerrild, S., Leutscher, P. D., Renvillard, S. G., Ogilvie, C. B., Bean, D. J., Videbech, P., Allen, T. M., McKeating, J. A. and Fletcher, N. F. (2016) ‘Deep sequencing of hepatitis C virus reveals genetic compartmentalization in cerebrospinal fluid from cognitively impaired patients.’, *Liver international : official journal of the International Association for the Study of the Liver*. doi: 10.1111/liv.13134.

Tuplin, A., Evans, D. J. and Simmonds, P. (2004) ‘Detailed mapping of RNA secondary structures in core and NS5B-encoding region sequences of hepatitis C virus by RNase cleavage and novel bioinformatic prediction methods’, *Journal of General Virology*, 85(10), pp. 3037–3047. doi: 10.1099/vir.0.80141-0.

Uden, J. and Raz, E. (2000) *Introduction to immunostimulatory DNA sequences, Springer Seminars in Immunopathology*. doi: 10.1007/s002810050010.

United Nations (1982) ‘Provisional Guidelines on Standard International Age Classifications’, *Department of International Economic and Social Affairs*, (74), p. 32.

Urbanus, A. T., van de Laar, T. J., Stolte, I. G., Schinkel, J., Heijman, T., Coutinho, R. A. and Prins, M. (2009) ‘Hepatitis C virus infections among HIV-infected men who have sex

with men: an expanding epidemic.', *AIDS*. England, 23(12), pp. F1-7. doi: 10.1097/QAD.0b013e32832e5631.

Vargas, H. E., Laskus, T., Radkowski, M., Wilkinson, J., Balan, V., Douglas, D. D., Harrison, M. E., Mulligan, D. C., Olden, K., Adair, D. and Rakela, J. (2002) 'Detection of hepatitis C virus sequences in brain tissue obtained in recurrent hepatitis C after liver transplantation.', *Liver transplantation : official publication of the American Association for the Study of Liver Diseases and the International Liver Transplantation Society*, 8(11), pp. 1014–9. doi: 10.1053/jlts.2002.36393.

Vattipally, S. (2014) 'TANOTI: a rapid BLAST-guided read mapper for small, divergent genomes'.

Veldt, B. J., Heathcote, E. J., Wedemeyer, H., Reichen, J., Hofmann, W. P., Zeuzem, S., Manns, M. P., Hansen, B. E., Schalm, S. W. and Janssen, H. L. (2007) 'Sustained virologic response and clinical outcomes in patients with chronic hepatitis C and advanced fibrosis', *Ann Intern Med*, 147(10), pp. 677–684.

Vento, S., Garofano, T., Renzini, C., Cainelli, F., Casali, F., Ghironzi, G., Ferraro, T. and Concia, E. (1998) 'Fulminant hepatitis associated with hepatitis A virus superinfection in patients with chronic hepatitis C.', *The New England journal of medicine*. UNITED STATES, 338(5), pp. 286–290. doi: 10.1056/NEJM199801293380503.

Vickerman, P., Grebely, J., Dore, G. J., Sacks-Davis, R., Page, K., L. Thomas, D., O. Osburn, W., L. Cox, A., Aitken, C. K., Hickman, M. and Hellard, M. (2012) 'The More You Look, the More You Find: Effects of Hepatitis C Virus Testing Interval on Reinfection Incidence and Clearance and Implications for Future Vaccine Study Design', *Journal of Infectious Diseases*, 205(9), pp. 1342–1350. doi: 10.1093/infdis/jis213.

Vignuzzi, M., Stone, J. K., Arnold, J. J., Cameron, C. E. and Andino, R. (2006) 'Quasispecies diversity determines pathogenesis through cooperative interactions in a viral population.', *Nature*. England, 439(7074), pp. 344–348. doi: 10.1038/nature04388.

Vogel, M., van de Laar, T., Kupfer, B., Stellbrink, H.-J., Kümmerle, T., Mauss, S., Knecht, G., Berger, A., Bruisten, S. and Rockstroh, J. K. (2010) 'Phylogenetic analysis of acute

hepatitis C virus genotype 4 infections among human immunodeficiency virus-positive men who have sex with men in Germany.', *Liver international : official journal of the International Association for the Study of the Liver*, 30(8), pp. 1169–1172. doi: 10.1111/j.1478-3231.2010.02305.x.

Wakita, T., Pietschmann, T., Kato, T., Date, T., Miyamoto, M., Zhao, Z., Murthy, K., Habermann, A., Krausslich, H.-G., Mizokami, M., Bartenschlager, R. and Liang, T. J. (2005) 'Production of infectious hepatitis C virus in tissue culture from a cloned viral genome.', *Nature medicine*. United States, 11(7), pp. 791–796. doi: 10.1038/nm1268.

Wang, C., Sun, J.-H., O'Boyle, D. R. 2nd, Nower, P., Valera, L., Roberts, S., Fridell, R. A. and Gao, M. (2013) 'Persistence of resistant variants in hepatitis C virus-infected patients treated with the NS5A replication complex inhibitor daclatasvir.', *Antimicrobial agents and chemotherapy*. United States, 57(5), pp. 2054–2065. doi: 10.1128/AAC.02494-12.

Wang, J., Charboneau, R., Balasubramanian, S., Barke, R. a, Loh, H. H. and Roy, S. (2002) 'The immunosuppressive effects of chronic morphine treatment are partially dependent on corticosterone and mediated by the mu-opioid receptor.', *Journal of leukocyte biology*, 71(5), pp. 782–90.

Washenberger, C. L., Han, J.-Q., Kechris, K. J., Jha, B. K., Silverman, R. H. and Barton, D. J. (2008) 'Hepatitis C Virus RNA: Dinucleotide Frequencies and Cleavage by RNase L', *Virus Research*, 15(10), pp. 1203–1214. doi: 10.1016/j.drugalcdep.2008.02.002.A.

Watson, J. and Crick, F. H. (1953) 'Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid.', *Nature*. Not Available, 171(4356), pp. 737–738.

Webster, B., Herker, E., Ott, M., Greenea, W. C., Wissing, S., Herker, E., Ott, M. and Greene, W. C. (2013) 'Rapid Intracellular Competition between Hepatitis C Viral Genomes as a Result of Mitosis', *Journal of Virology*, 87(1), pp. 581–596. doi: 10.1128/jvi.01047-12.

Webster, B., Ott, M. and Greene, W. C. (2013) 'Evasion of Superinfection Exclusion and Elimination of Primary Viral RNA by an Adapted Strain of HCV.', *Journal of virology*, (October). doi: 10.1128/JVI.02465-13.

Weeratna, R. D., McCluskie, M. J., Xu, Y. and Davis, H. L. (2000) 'CpG DNA induces stronger immune responses with less toxicity than other adjuvants.', *Vaccine*. ENGLAND, 18(17), pp. 1755–1762.

Wei, B., Kang, J., Kibukawa, M., Chen, L., Qiu, P., Lahser, F., Marton, M. and Levitan, D. (2016) 'Development and Validation of a Template- Independent Next-Generation Sequencing Assay for Detecting Low-Level Resistance-Associated Variants of Hepatitis C Virus', *The Journal of Molecular Diagnostics*. American Society for Investigative Pathology and the Association for Molecular Pathology, 18(5), pp. 643–656. doi: 10.1016/j.jmoldx.2016.04.001.

Westbrook, R. H. and Dusheiko, G. (2014) 'Natural history of hepatitis C', *Journal of Hepatology*. European Association for the Study of the Liver, 61(1), pp. S58–S68. doi: 10.1016/j.jhep.2014.07.012.

White, P. A., Li, Z., Zhai, X., Marinos, G. and Rawlinson, W. D. (2000) 'Mixed Viral Infection Identified Using Heteroduplex Mobility Analysis (HMA)', *Virology*, 271(2), pp. 382–389. doi: 10.1006/viro.2000.0323.

WHO and World Health Organization (2012) *Prevention & Control of Viral Hepatitis Infection : Framework for Global Action*, *World Health Organization*.

Widell, A., Mansson, S., Persson, N. H., Thysell, H., Hermodsson, S. and Blohme, I. (1995) 'Hepatitis C Superinfection in Hepatitis C Virus (HCV)-Infected Patients Transplanted with an HCV-Infected Kidney', *Transplantation*. UNITED STATES, 60(7), pp. 642–647.

Wiese, M., Berr, F., Lafrenz, M., Porst, H. and Oesen, U. (2000) 'Low frequency of cirrhosis in a hepatitis C (genotype 1b) single-source outbreak in germany: a 20-year multicenter study.', *Hepatology*, 32(1), pp. 91–6. doi: 10.1053/jhep.2000.8169.

Wilke, C. O., Li, H., Stoddard, M. B., Wang, S., Blair, L. M., Giorgi, E. E., Parrish, E. H., Learn, G. H., Hraber, P., Goepfert, P. A., Saag, M. S., Denny, T. N., Haynes, B. F., Hahn, B. H., Ribeiro, R. M., Perelson, A. S., Korber, B. T., Bhattacharya, T. and Shaw, G. M. (2012) 'Elucidation of Hepatitis C Virus Transmission and Early Diversification by Single

Genome Sequencing', *PLoS Pathogens*, 8(8), p. e1002880. doi: 10.1371/journal.ppat.1002880.

Wilke, C. O., Wang, J. L., Ofria, C., Lenski, R. E. and Adami, C. (2001) 'Evolution of digital organisms at high mutation rates leads to survival of the flattest.', *Nature*. England, 412(6844), pp. 331–333. doi: 10.1038/35085569.

Williams, I. T., Bell, B. P., Kuhnert, W. and Alter, M. J. (2011) 'Incidence and transmission patterns of acute hepatitis C in the United States, 1982-2006.', *Archives of internal medicine*. United States, 171(3), pp. 242–248. doi: 10.1001/archinternmed.2010.511.

Wilsnack, R. W., Wilsnack, S. C., Kristjanson, A. F., Vogeltanz-Holm, N. D. and Gmel, G. (2009) 'Gender and alcohol consumption: patterns from the multinational GENACIS project.', *Addiction*, 104(9), pp. 1487–500. doi: 10.1111/j.1360-0443.2009.02696.x.

Witteveldt, J., Martin-Gans, M. and Simmonds, P. (2016) '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(March), p. AAC.02932-15. doi: 10.1128/AAC.02932-15.

Wong, G. L.-H. (2013) 'Update of liver fibrosis and steatosis with transient elastography (Fibroscan).', *Gastroenterology report*, 1(1), pp. 19–26. doi: 10.1093/gastro/got007.

Wong, J. B., McQuillan, G. M., McHutchison, J. G. and Poynard, T. (2000) 'Estimating future hepatitis C morbidity, mortality, and costs in the United States', *American Journal of Public Health*, 90(10), pp. 1562–1569. doi: 10.2105/AJPH.90.10.1562.

World Health Organisation (2015) *Hepatitis C (Fact sheet N°164)*.

World Health Organization (2011) *Blood Safety: Global Database on Blood Safety reports*.

Wu, J. C., Chen, C. L., Hou, M. C., Chen, T. Z., Lee, S. D. and Lo, K. J. (1994) 'Multiple viral infection as the most common cause of fulminant and subfulminant viral hepatitis in

an area endemic for hepatitis B: application and limitations of the polymerase chain reaction.', *Hepatology*. UNITED STATES, 19(4), pp. 836–840.

Wünschmann, S., Medh, J. D., Klinzmann, D., Schmidt, W. N. and Stapleton, J. T. (2000) 'Characterization of hepatitis C virus (HCV) and HCV E2 interactions with CD81 and the low-density lipoprotein receptor.', *Journal of virology*, 74(21), pp. 10055–10062. doi: 10.1128/JVI.74.21.10055-10062.2000.

Yang, C., Skiena, S., Futcher, B., Mueller, S. and Wimmer, E. (2013) 'Deliberate reduction of hemagglutinin and neuraminidase expression of influenza virus leads to an ultraprotective live vaccine in mice.', *Proceedings of the National Academy of Sciences of the United States of America*. United States, 110(23), pp. 9481–9486. doi: 10.1073/pnas.1307473110.

Yu, M.-L., Dai, C.-Y., Huang, C.-F., Lee, J.-J., Yeh, M.-L., Yeh, S.-M., Kuo, H.-T., Huang, J.-F., Chang, J.-M., Chen, H.-C., Juo, S.-H. H., Hwang, S.-J. and Chuang, W.-L. (2014) 'High hepatitis B virus surface antigen levels and favorable interleukin 28B genotype predict spontaneous hepatitis C virus clearance in uremic patients.', *Journal of hepatology*. Netherlands, 60(2), pp. 253–259. doi: 10.1016/j.jhep.2013.09.023.

Yuan, H., Jain, M., Snow, K. K., Gale, M., Lee, W. M., the HALT-C Trail Group and HALT-C (2010) 'Evolution of Hepatitis C Virus NS5A Region in Breakthrough Patients During Pegylated Interferon and Ribavirin Therapy', *Journal of Viral Hepatitis*, 17(3), pp. 208–216. doi: 10.1016/j.surg.2006.10.010.Use.

Yun, H., Kim, D., Kim, S., Kang, S., Jeong, S., Cheon, Y., Joe, K., Gwon, D.-H., Cho, S.-N. and Jee, Y. (2008) 'High prevalence of HBV and HCV infection among intravenous drug users in Korea', *Journal of Medical Virology*, 80(9), pp. 1570–1575. doi: 10.1002/jmv.21255.

Yun, Z. B., Lindh, G., Weiland, O., Johansson, B. and Sonnerborg, A. (1993) 'Detection of hepatitis C virus (HCV) RNA by PCR related to HCV antibodies in serum and liver histology in Swedish blood donors.', *Journal of medical virology*. UNITED STATES, 39(1), pp. 57–61.

Zaltron, S., Spinetti, A., Biasi, L., Baiguera, C. and Castelli, F. (2012) 'Chronic HCV infection: epidemiological and clinical relevance.', *BMC infectious diseases*, 12 Suppl 2(Suppl 2), p. S2. doi: 10.1186/1471-2334-12-S2-S2.

Zarski, J. P., Bohn, B., Bastie, A., Pawlotsky, J. M., Baud, M., Bost-Bezeaux, F., Tran van Nhieu, J., Seigneurin, J. M., Buffet, C. and Dhumeaux, D. (1998) 'Characteristics of patients with dual infection by hepatitis B and C viruses.', *Journal of hepatology*. DENMARK, 28(1), pp. 27–33.

Zayas, M., Long, G., Madan, V. and Bartenschlager, R. (2016) 'Coordination of Hepatitis C Virus Assembly by Distinct Regulatory Regions in Nonstructural Protein 5A', *PLOS Pathogens*, 12(1), p. e1005376. doi: 10.1371/journal.ppat.1005376.

Zein, N. N. (2000) 'Clinical Significance of Hepatitis C Virus Genotypes Clinical Significance of Hepatitis C Virus Genotypes', *Cinical Microbiology Reviews*, 13(2), pp. 223–35. doi: 10.1128/CMR.13.2.223-235.2000.Updated.

Zeuzem, S., Berg, T., Moeller, B., Hinrichsen, H., Mauss, S., Wedemeyer, H., Sarrazin, C., Hueppe, D., Zehnter, E. and Manns, M. P. (2009) 'Expert opinion on the treatment of patients with chronic hepatitis C.', *Journal of viral hepatitis*, 16(2), pp. 75–90. doi: 10.1111/j.1365-2893.2008.01012.x.

Zeuzem, S., Buti, M., Ferenci, P., Sperl, J., Horsmans, Y., Cianciara, J., Ibranyi, E., Weiland, O., Noviello, S., Brass, C. and Albrecht, J. (2006) 'Efficacy of 24 weeks treatment with peginterferon alfa-2b plus ribavirin in patients with chronic hepatitis C infected with genotype 1 and low pretreatment viremia', *Journal of Hepatology*, 44(1), pp. 97–103. doi: 10.1016/j.jhep.2005.10.003.

Zhang, G., Cowled, C., Shi, Z., Huang, Z., Bishop-Lilly, K. A., Fang, X., Wynne, J. W., Xiong, Z., Baker, M. L., Zhao, W., Tachedjian, M., Zhu, Y., Zhou, P., Jiang, X., Ng, J., Yang, L., Wu, L., Xiao, J., Feng, Y., Chen, Y., Sun, X., Zhang, Y., Marsh, G. A., Crameri, G., Broder, C. C., Frey, K. G., Wang, L.-F. and Wang, J. (2013) 'Comparative analysis of bat genomes provides insight into the evolution of flight and immunity.', *Science*. United States, 339(6118), pp. 456–460. doi: 10.1126/science.1230835.

Zhang, Z. X., Chen, M., Sonnerborg, A. and Sallberg, M. (1994) ‘Antigenic structure of the complete nonstructural (NS) 2 and 5 proteins of hepatitis C virus (HCV): anti-HCV NS2 and NS5 antibody reactivities in relation to HCV serotype, presence of HCV RNA, and acute HCV infection.’, *Clinical and diagnostic laboratory immunology*. UNITED STATES, 1(3), pp. 290–294.

Zhou, J. H., Su, J. H., Chen, H. T., Zhang, J., Ma, L. N., Ding, Y. Z., Stipkovits, L., Szathmary, S., Pejsak, Z. and Liu, Y. S. (2013) ‘Clustering of low usage codons in the translation initiation region of hepatitis C virus’, *Infection, Genetics and Evolution*, 18, pp. 8–12. doi: 10.1016/j.meegid.2013.03.043.

Zhou, J., Liu, W. J., Peng, S. W., Sun, X. Y. and Frazer, I. (1999) ‘Papillomavirus capsid protein expression level depends on the match between codon usage and tRNA availability.’, *Journal of virology*, 73(6), pp. 4972–4982.