Alternative tools to investigate HBV entry pathway and intra-host evolution

Chun Goddard MPharm (Hons)

July 2020

Principal supervisor: Professor William Irving

Second supervisor: Dr Alexander Tarr

Virus Research Group, School of Life Sciences, Queen's Medical Centre, University of Nottingham, UK

Student Declaration:

I declare that the data presented is entirely my work, except where otherwise stated.

Abstract

Hepatitis B virus (HBV) infects liver cells resulting in hepatitis B infection. In approximately 90% of childhood infections and 10% of adult cases the virus establishes a chronic, lifelong infection. In 2015, it was estimated that approximately 3.5% of the global population were living with chronic HBV infection. Long-term complications of chronic HBV infection such as cirrhosis and hepatocellular carcinoma cause a significant disease burden. The World Health Organisation has set a goal to eliminate the infection by 2030. However, current treatment for HBV does not cure the infection, and most people require the treatment for life. Therefore, a better understanding of the virus lifecycle and viral intra-host evolution is crucial in the development of antiviral drugs.

The error-prone nature of the HBV polymerase and combined with high replication rate result in genetic diversity manifesting as a quasispecies in a chronically HBV-infected host. However, the impact of genetic diversity in HBV infections is still poorly understood; while antigenic differences and drug resistant forms of HBV have been assigned to specific viral polymorphisms, the consequence of the majority of genetic variants have not been interrogated. In this study, genetic diversity of HBV infections was investigated using a nanopore sequencing platform to determine longitudinal HBV evolution within hosts. The data extends previous knowledge that greater viral diversity is observed in HBV e-antigen (HBeAg)-negative infection and the rate of evolution in the nonoverlapping region is much higher than that of the overlapping region.

The HBV surface proteins (HBsAg) comprise the Large (L), Middle (M) and Small (S). The HBsAg gene is completely overlapped by the polymerase gene in a -1 frameshift. Mutations occurring in one open reading frame (ORF) can also impact on the other;

therefore, it is challenging to investigate the phenotype of HBsAg mutations without altering phenotypic behaviour of the polymerase. Moreover, HBV research requires Biosafety Level (BSL)-3 containment, which hinders research progress. To address these limitations, retroviral pseudotyped viruses (PVs) were utilised in this study to investigate HBV entry and the effect of viral polymorphisms. Retroviruses pseudotyped with the HBV glycoproteins, possess the structural proteins of retrovirus and HBV surface glycoproteins and only require BSL-2 laboratory facilities, which offers an alternative to overcome the drawbacks of handling HBV *in vitro*. This study developed novel pseudotyping techniques to investigate the HBV entry pathway. A sodium taurocholate co-transporting polypeptide (NTCP) overexpressing hepatocellular carcinoma cell line was developed to determine the impact of receptor expression on HBV entry. A new cell line (293T.gtD4.S) constitutively expressing the S protein of a genotype D isolate has also been developed. Using an in-house engineered lentiviral vector pNL4.3.Nluc.R-E- and a plasmid expressing L co-transfecting 293T.gtD4.S cells, the infectivity of HBV PVs was quantified and monitored. The results revealed the importance of L/S ratio in HBV infectivity.

Acknowledgements

I wish to express my special gratitude to my principal supervisor Professor William Irving who gave me this opportunity in HBV research. This research work has taken me to a level that I would have never dreamed of. I am very grateful for his immense patience that has helped me throughout this research and writing of this thesis.

A huge thanks goes to my second supervisor Dr Alexander Tarr who is the fountain of knowledge, for his insightful comments and his challenging questions which inspired me to broaden my views on this study.

My big thanks also go to Dr Patrick McClure, the expert in construction of viral molecular cloning, for his ongoing guidance on gene cloning and site-directed mutagenesis. A big thanks goes to Dr Barnabas King, Dr Richard Urbanowicz and Dr Stuart Astbury, who guided me on many techniques required in this research. My gratitude also goes to my fellow lab mates within our virology team, such as Mosab, Jo, Emma, Fernando, Chidinma, Josh, who have also given me tremendous help, moral support and encouragement.

I would take this opportunity to show my gratitude to Dr Keith Spriggs. He is my Master's supervisor, but he has given me tremendous support and encouragement even during my PhD studies.

I would like to thank my sponsors Medical Research Council (MRC), Nottingham Molecular Pathology Node (NMPN) and Nottingham School of Life Sciences for giving me the opportunity to undertake a fully funded 3-year PhD.

Finally, I would like to thank my husband, David Goddard, who always shows faith in me and supports me financially and emotionally. Without all his help and encouragement, I would not have accomplished what I have done thus far.

Table of Contents

Table of figures

List of tables

Abbreviations

aa	Amino acid
BSA	Bovine serum albumin
C-terminus	Carboxyl-terminus
cccDNA	Covalently closed circular DNA
cDNA	Complementary DNA
DMEM	Dulbecco's Modified Eagle's Medium
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
dsL	Double stranded linear
DTT	Dithiothreitol
E.coli	*Escherichia coli*
EDTA	Ethylenediamine tetraacetic acid
eGFP	Enhanced green fluorescence protein
ELISA	Enzyme-linked immuno-sorbent assay
ER	Endoplasmic reticulum
FITC	Fluorescein isothiocyanate
HSPGs	Heparin sulphate proteoglycans
GAPDH	Glyceraldehyde-3-phosphate dehydrogenase
H77	Hepatitis C virus subtype 1a, Hutchinson strain isolate 77
HBV	Hepatitis B virus
HBV-PVs	Hepatitis B virus pseudotyped viruses
HCC	Hepatocellular carcinoma
HDV	Hepatitis D virus
HEK	Human embryonic kidney
HepG2	Hepatocellular carcinoma cell line G2
HIV	Human immunodeficiency virus

HRP	Horse radish peroxidase
HuH7	Human hepatocellular carcinoma cell line 7
IF	Immunofluorescence
IFN-α	Interferon-α
IP	Immunoprecipitation
IRES	Internal ribosomal entry site
LB	Luria-Bertani
mRNA	Messenger RNA
N-terminus	Amino-terminus
NEAA	Non-essential amino acids
NTCP	Sodium Taurocholate Cotransporting Polypeptide
ORF	Open reading frame
PAGE	Polyacrylamide gel electrophoresis
PBS	Phosphate buffered saline
PBST	PBS containing 0.1% v/v Tween 20
PCR	Polymerase chain reaction
PEI	Polyethylenimine
Pen-Strep	Penicillin-streptomycin
PVDF	Polyvinylidene fluoride
RC	Relax circular
RLU	Relative luminescence units
RNA	Ribonucleic acid
RNase	Ribonuclease
Rpm	Revolutions per minute
SDS	Sodium dodecyl sulphate
TBS	Tris buffered saline
UK	United Kingdom

VSV	Vesicular stomatitis virus
w/v	Weight per volume
WB	Western blot
WHO	World Health Organisation
Wt	Wild type
X-Gal	5-Bromo-4-Chloro-3-Indolyl β-D-Galactopyranoside

Chapter 1: Introduction

Hepatitis B virus (HBV) infects the liver, which can lead to chronic liver disease and in a small number of infections, cirrhosis and hepatocellular carcinoma (HCC). According to the World Health Organisation (WHO), in 2015, approximately 257 million individuals were living with chronic HBV infection and more than 887000 deaths occurred as a result of the disease (WHO, 2019a). However, the true burden of this disease may not be clear as accurate quantification is limited as many calculation models only include the adult population and are not adjusted by age and sex.

The WHO has set a goal aiming to eliminate hepatitis B as a public health concern by 2030 (WHO, 2016). In order to achieve this target, the WHO is focusing on transmission prevention, scaling up screening, care and treatment services to reduce new infections by 90% and HBV related death by 65% by 2030 (WHO, 2019a). Reducing new infections has been shown to be effective by high coverage of birth dose vaccine and vaccination among children. However, the main challenge is to reduce infection-associated mortality (Liu et al., 2019), and to develop new curative therapies. Joint effort worldwide is required to achieve this goal.

The elimination targets will require new approaches to therapy. In this study, the aims were to develop novel approaches to investigating HBV entry, with the future goal of screening entry inhibitors, a class of interventions that have been assessed in clinical trials and clinical practice for other viruses, including human immunodeficiency virus 1 (HIV-1), respiratory syncytial virus (RSV) and hepatitis C virus (HCV). In 2012, sodium taurocholate cotransporting polypeptide (NTCP) was identified as the functional receptor for HBV entry; exogenously expressing NTCP rendered non-

infectious hepatoma cells including HuH7 and HepG2 cells susceptible to HBV infection (Yan et al., 2012). To facilitate the study of investigation the contribution of the HBV surface antigen (HBsAg) to HBV entry, NTCP expressing cell lines were developed, which is described in Chapter 3; Chapter 4 reports the work involved in pseudotyping and HBV entry assays; Chapter 5 illustrates the longitudinal quasispecies analysis using MinION sequencing to determine HBsAg variation; and the Chapter 6 provides the final discussion and proposals for future research leading from this study.

1.1 General background of HBV

1.1.1 Discovery of HBV

A long history of research into what was termed “catarrhal” jaundice started in 1865. Outbreaks of “catarrhal” jaundice in 1885 following smallpox vaccination and serum-transmitted outbreaks in the 1930s as well as an epidemic of jaundice in soldiers following yellow fever vaccine in 1942, led to the discovery that distinct filterable agents caused the infectious hepatitis. In 1947, the term “hepatitis B” was introduced (Block et al., 2016). Despite the development of a variety of liver function tests leading to the recognition of chronic hepatitis, the causative agent(s) could not be isolated until 1961, when Blumberg discovered an antibody (termed “precipitin”) in patients who had received multiple blood transfusions. In 1965, Blumberg and his co-workers screened a panel of 24 samples from Blumberg’s global collection by utilising such antibodies and they discovered an antigen in a specimen from an Australian aboriginal person (known as “Au”) (Blumberg et al., 1965). “Au” was frequently found in sera from patients who received a blood transfusion and from patients diagnosed with viral hepatitis.

Evidence linking “Au” to hepatitis came from a report published by Okochi and colleagues in Japan, who found that patients who received Au-positive blood transfusion developed abnormal liver function tests and became antigen-positive (Okochi and Murakami, 1968, Block et al., 2016). Additional evidence was provided by electron microscopy (EM) of sucrose gradient-fractionated serum samples, showing 20 nm particles forming large aggregates with anti-Au antibodies (Bayer et al., 1968). In 1970, by examining Au antigen immune complexes under the EM, Dane and colleagues discovered that Au appeared on circular, tubular sub-viral particles and larger, complete enveloped virions, known as “Dane particles” (Dane et al., 1970, Gerlich, 2013).

1.1.2 Epidemiology and transmission of HBV

HBV is transmitted from human to human through exposure to infected blood and bodily fluids. There are two main routes involved in this process. One is vertical transmission, i.e. mother-to-child; the other is horizontal transmission, including sexual and parenteral/percutaneous transmissions.

1.1.2.1 Vertical transmission

Vertical transmission is defined when an infant born to an HBV infected mother is HBsAg or HBV DNA positive by the age of 6 to 12 months old. Mother-to-child transmission can occur intrauterine, during delivery or postpartum. The intrauterine route accounts for a very low incidence of HBV transmission. It is often associated with mothers with very high viral load and HBV e-antigen (HBeAg)-positivity (Xu et al., 2002). Transmission during delivery is the most common route of vertical transmission, arising from the newborn baby being in contact with the infectious blood or body fluids from the mother (Piratvisuth, 2013). While breast feeding does not appear to pose a risk of transmission, the spread of HBV through close contact has been demonstrated (Hill et al., 2002).

1.1.2.2 Horizontal transmission

Sexual transmission is a major form of HBV transmission in the world, especially in low endemicity countries such as North America. In the United States, 60% to 70% of homosexual men were estimated to have HBV infection (Alter and Margolis, 1990). However, the risk of HBV infection among heterosexuals has been increasing, especially in persons who have multiple sexual partners or with partners who are prostitutes or drug users (Alter and Mast, 1994, Alter and Margolis, 1990). HBV infection can also be spread through percutaneous/parenteral transmission, including blood or blood product transfusion, contaminated needles/sharps, dental work, acupuncture or tattooing. HBV infection can also occur by exposure to HBV-contaminated inanimate objects, as the virus can survive on environmental surfaces for more than seven days (Bond et al., 1981).

1.1.3 Symptoms

It is observed that, during acute HBV infection, the rate of patients with apparent symptoms increases with age. About 10% of patients under the age of 4 are symptomatic while approximately 30% of adults (age of 30 years or older) show symptoms (McMahon et al., 1985). Symptoms vary according to the severity of the disease. Fatigue, nausea and vomiting are mild symptoms, while jaundice can occur in more marked situations. Around 1% of patients experience a sudden onset of fever, abdominal pain, jaundice, and may soon progress to confusion, coma and even death (Trey and Davidson, 1970), (Berk and Popper, 1978). This is known as fulminant hepatitis.

Patients with chronic infection may be asymptomatic or just have mild symptoms such as fatigue or discomfort in the right upper quadrant area. Patients in more severe stages such as cirrhosis or HCC may have more marked symptoms including

jaundice, spider angiomata, splenomegaly, foetor hepaticus, ascites, gastrointestinal bleeding or encephalopathy (Liang, 2009).

1.1.4 Diagnosis of HBV

Diagnosis of HBV relies on laboratory blood tests for the detection of serological markers. These markers are complex and include hepatitis B surface antigen (HBsAg), hepatitis B e antigen (HBeAg), antibody to HBeAg (Anti-HBe), antibody to core antigen (Anti-HBc), immunoglobulin M (IgM) antibody to core antigen (Anti-HBc IgM), IgG antibody to core antigen (Anti-HBc IgG), antibody to HBsAg (Anti-HBs) (NICE, 2019). Acute hepatitis B is characterised by the detection of HBsAg and anti-HBc IgM, supported by raised alanine aminotransferase (ALT) in blood plasma. During the initial onset of the infection, HBeAg is also positive. Chronic hepatitis B is confirmed when HBsAg is persistent for more than six months with or without detection of HBeAg and anti-HBc IgG while anti-HBc IgM is negative (NICE, 2019, WHO, 2019b).

1.1.5 Disease prognosis

The disease outcome is influenced by the age at infection. In many adult cases, HBsAg spontaneously disappears after 4-6 months of infection. The proportion of acute infection which progresses to HBV chronicity decreases with age ranging from approximately 90% in infants to 30%-50% in children between the age of 1 to 6 years, and 5%-10% in children above the age of 6 years and adults (Beasley et al., 1981, Xu et al., 1985, Borgia et al., 2012). Compared to CHB, the number of fulminant hepatitis cases from acute HBV is less than 1% in adults. However, the outlook for these infections is bleak. Without liver transplantation, the mortality rate of fulminant hepatitis is approximately 80% (Trepo et al., 2014). During the course of CHB

infection, approximately 30-50% of CHB patients will enter a sustained clinical remission state with normal liver tests and normal platelet counts and no symptoms or signs of chronic liver disease (Tong et al., 2006, Hoofnagle et al., 1987). Among these patients, 0.7% will spontaneously achieve HBsAg seroconversion each year; the seroconversion significantly reduces the risk of developing HCC (Simonetti et al., 2010). It is estimated that approximately 15-40% of CHB patients will progress to cirrhosis, liver failure or HCC (Lok, 2002).

1.1.6 Immune response to HBV

HBV has been described as a “stealth virus”. Studies using animal and human liver specimens have shown that there is a lack of robust innate immunity during acute HBV infection in adults (Tan et al., 2015). *In vivo* studies of infected chimpanzees showed that HBV infection does not result in induction of interferons (IFNs) or an IFN-stimulated gene (ISG) response in infected hepatocytes. Consistent with this, evidence of type I/III IFN response towards HBV is lacking in sera of humans with acute HBV infection (Suslov et al., 2018). Furthermore, liver tissue from chronically HBV-infected patients demonstrated that HBV does not induce an innate immune response, but this response is activated when the HBV-infected tissue is challenged with other factors such as toll-like receptor (TLR) ligands or Sendai virus infection, suggesting that HBV evades innate immune recognition, rather than inactivating the innate recognition by pattern recognition receptors (PRRs) (Suslov et al., 2018).

A study examining longitudinal blood samples from nine patients who presented with acute HBV infection suggested that HBV does not go completely unchallenged by the innate immune system. The results revealed that CD56 (both bright and dim) natural killer (NK) cells were elevated in early infection and an increasing proportion of these cells expressed activation markers such as CD38, NKp46 and HLA-DR. The level of

interferon-gamma inducible protein 10 (IP-10), an ISG, was also significantly increased in early infection, indicating a broad innate immune response via the interferon pathway (Stelma et al., 2017). The discrepancy observed between these studies and previous studies in animals may be due to the timing of sample collection. Samples were collected in this study from patients who were symptomatic, suggesting local signals such as interferons produced by lymphocytes might have activated the ISG pathway (Stelma et al., 2017).

Patients who cleared acute HBV infection had significant higher levels of functional HBV-specific CD8+ and CD4+ response compared to those who progressed to chronic infection (Gehring and Protzer, 2019, Stelma et al., 2017). Clinical observations revealed that HBV is not completely eliminated in individuals who achieved spontaneous and sustained resolution of infection as HBV can be reactivated if these individuals are immunosuppressed (Maini and Pallett, 2018). It is important that after the natural resolution of infection, ongoing immune responses keep residual viruses at bay. Liver CD8+ T cells from a donor who had cleared HBV infection could mount an efficient response when re-challenged with peptides from all the major HBV proteins (Maini and Pallett, 2018). The majority of these T cells are hepatic memory T cells (T_{RM} cells), which were able to coordinate rapid antiviral responses. In contrast, a patient with CHB had very low frequency of liver CD8+ T_{RM} cells (< 0.2% of circulating CD8+ T cells) (Gehring and Protzer, 2019, Maini and Pallett, 2018). Furthermore, the remaining CD8+ T_{RM} cells in individuals with CHB also exhibit defective function on antigen recognition, leading to reduced production of antiviral cytokines interferon-gamma (Maini and Pallett, 2018).

Humoral immunity also plays an important role in controlling the residual infection. Clinical observations revealed that depletion of CD20-expressing cells by rituximab could reactivate HBV infection in some individuals whose infection has resolved as

serological evidence showed detectable antibodies of anti-HBc ± anti-HBs (Burton et al., 2018). A study using a fluorescent bait reagent to quantify HBsAg-specific B cells revealed individuals with CHB maintained comparable amounts of HBsAg-specific B cells compared with vaccinated or natural resolvers; however, the majority of these cells in patients with CHB have a phenotype of atypical memory B cells (atMBC) that lack CD21 and CD27 expression. These atMBCs were capable of differentiating into plasma cells but failed to produce detectable levels of anti-HBs (Burton et al., 2018). However, it is unclear to what extent these *in vitro* data are representative of *in vivo* situations.

1.1.7 Treatment

In order to achieve a complete virologic cure, all virologic markers should be eliminated, which include covalently closed circular DNA (cccDNA), the viral minichromosome that serves as a template for all viral RNA transcripts (Revill et al., 2016). The biggest challenge is eliminating persistent cccDNA from the nucleus of the infected hepatocytes, which can be reactivated even after a long period of quiescence. Frequent reports of HBV reactivation during the use of B-cell depleting agents have led to a warning issued by the Food and Drug Administration (FDA) indicating that HBV screening is strongly recommended prior to B-cell depletion therapy (Perrillo et al., 2015). Complete elimination of cccDNA is hard to achieve, therefore, sustained virologic response at 24 weeks (SVR24) after the completion of a treatment course where plasma HBV DNA is suppressed, and plasma HBsAg is undetectable, with or without seroconversion, was proposed as a primary efficacy endpoint, or 'functional cure' (Anderson et al., 2019).

Current antiviral treatment regimens aim to delay disease progression to liver cirrhosis and HCC, therefore improving long-term survival of the host. The medical therapies

include parenteral injections of pegylated (PEG) interferon (IFN) and nucleos(t)ide HBV polymerase inhibitors, known as nucleos(t)ide analogues (NUCs). Historically, interferon therapy achieved higher rates of functional cure, as defined by total loss of HBsAg in the blood. A study showed the rate of HBsAg clearance was 5% in patients with HBV DNA ≤ 2000 IU/mL at year 1 post-treatment, and this rate increased to 61.5% in patients with HBV DNA < 70 IU/mL (Marcellin et al., 2013). However, it is rarely used in HBV treatment as a result of tolerability, convenience and limited efficacy in the majority of patients (Anderson et al., 2019). Tenofovir Disoproxil Fumarate (TDF) is preferred in all adults, adolescents or children at the age of 12 or order who require NUC treatment. Entecavir is recommended for children at age 2-12 years. Although NUCs have been shown to be effective in inhibiting HBV DNA replication, rarely have they achieved HBsAg seroconversion and a cure. Hence, long-term NUCs treatment is required in the vast majority (WHO, 2015). The expectation of an individual to adhere to noncurative, potentially lifelong treatment is not always possible. Therefore, more effective and finite treatment is an unmet medical need, which requires novel drugs and new therapies to meet such need (Anderson et al., 2019).

1.2 Molecular virology of HBV

1.2.1 HBV genome

HBV is the prototype virus of the family of *Hepadnaviridae*. This family contains three genera, including *Orthohepadnavirus* (possessing viruses infecting primates and bats), *Avihepadnavirus* (whose members onlyinfect birds) (Schaefer, 2007) and the most recently discovered HBV group which infect fish and amphibians (Dill et al., 2016).

HBV contains a relaxed, circular, partially double-stranded DNA (rcDNA) molecule of approximately 3200 nucleotides in length. (Delius et al., 1983, Guo et al., 2007). The HBV genome is organised in a very compact manner; approximately two-thirds of the genome encode more than one protein. There are four overlapping open reading frames (ORFs) (Will et al., 1987, Minor and Slagle, 2014) including surface (HBs), preCore/Core, polymerase (Pol) and the X ORF. The surface ORF (nucleotides (nt) 2854-3221, 1-835), encodes three proteins, including small (HBS), middle (M) and large (L). They all share the same C-terminus (S) but have discrete N-terminal sequences. The M has an additional preS2 domain compared to the N-terminus of S and the L protein has an additional preS1 to the N-terminus of M (Figure 1.1). The preCore/Core ORF (nt.1814 to 2455) encodes HBeAg and HBcAg. The X gene (nt.1374 to nt.1835) is the smallest ORF and encodes the HBx antigen (HBxAg) (Bouchard and Schneider, 2004, Zhang and Cao, 2011). The Pol gene encodes the HBV polymerase (P) (nt. 2307 to 3221; 1 to 1632), which has four domains including the terminal domain acting as a primer for reverse transcriptase (RT), followed by a spacer region, with unknown function, the RT domain and a ribonuclease H domain (RNase H) (Januszkiewicz-Lewandowska et al., 2014, HBVdb, 2020) (Figure 1.1).

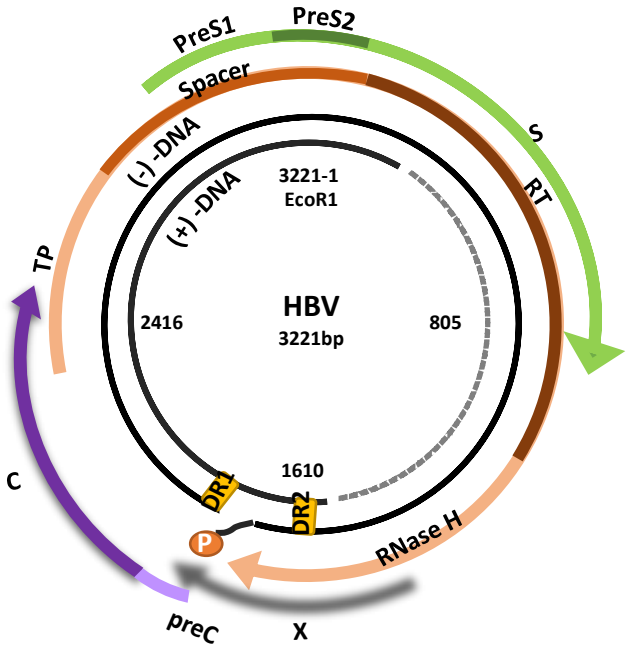

Figure 1.1 HBV genome structure

HBV genome is a partially double-stranded DNA molecule including a complete minus-strand DNA ((-)-DNA) linked to polymerase (P) protein at its 5' end. The plus-strand DNA ((+)-DNA) is incomplete (the incomplete part is symbolised with dashed lines). DR1 and DR2 are the direct repeat sequences. The genome consists of four overlapping open reading frames (ORFs), polymerase (Pol), surface, X and PreCore/Core. (Graph is adapted from: (HBVdb, 2020, Beck and Nassal, 2007)). PreC, preCore; C, core; TP, terminal protein; RT, reverse transcriptase; RNase H, ribonuclease H; DR, direct repeat.

1.2.2 The life cycle of HBV

A typical HBV virion is approximately 42 nm in size, consisting of an icosahedral nucleocapsid of approximately 30 nm in diameter surrounded by an outer lipid envelope containing the three surface proteins. The nucleocapsid encloses the DNA polymerase and viral DNA genome.

1.2.2.1 Viral entry

The current model of HBV entry proposes that initial interactions between virion and host cell are mediated by low specificity interactions with heparan sulphate proteoglycans (HSPG) on the cell surface (Schulze et al., 2007, Leistner et al., 2008). This is determined by the observation that HBV and hepatitis delta virus (HDV) infection of primary human hepatocytes and HepaRG cells can be completely blocked by treatment with highly sulphated compounds such as heparin and dextran sulphate (Zhang et al., 2016).

It was proposed that the preS1 domain of L protein initiated the binding to HSPG (Schulze et al., 2007). This claim was based on the observation that only S protein was detected in the washout of heparin sepharose columns after application of trypsinised HBV virions. However, in that experiment, only immune dot blot was performed. The protein integrity of both domains including the preS1 and S was undetermined. Therefore, it was unclear whether the fact that preS1 was undetectable was due to binding to heparin or because it was undetectable by anti-preS1 antibody in the washout. Similarly, it is unclear whether the detection of S in the flow-through was because the integrity of S, particularly the antigenic loop (AGL) structure, was disrupted that prevents its binding to heparin. More recent evidence showed that the AGL of the HBsAg participates in this interaction (Sureau and Salisse, 2013). This was determined by demonstrating that removal of the preS1 domain by treatment with

trypsin did not abrogate the heparin-binding capacity of the HBV virions. In that experiment, western blot was performed to determine the size of S post-trypsin-treatment to ensure the integrity of the AGL was not disrupted. Moreover, binding of S protein-coated sub-viral particles (SVPs) or HDV virions to immobilised heparin or target HepaRG cells was equivalent to the heparin binding capacity of HBsAg coated SVPs or HDV virions (Sureau and Salisse, 2013).

Following initial low specificity binding to cell surface HSPG, the preS1 domain of HBsAg associated with virions is thought to specifically bind to sodium taurocholate co-transporting protein (NTCP), a recently identified hepatocyte-specific HBV receptor (Yan et al., 2012) (more detailed discussion in 1.5.6). Upon binding, epidermal growth factor receptor (EGFR) plays a crucial role in HBV virion internalisation (Iwamoto et al., 2019). Evidence demonstrated that EGFR knockdown attenuated HBV virus internalisation, and conversely EGFR relocalisation triggered internalisation of the HBV/preS1-NTCP complex. It was shown that the interaction between EGFR and NTCP triggered endocytosis of the virus-receptor complex. Thereafter, in a poorly understood process, including un-coating and translocation, the rcDNA is transferred into the nucleus (Fourati and Pawlotsky, 2016). The rcDNA is then repaired by cellular enzymes and converted into cccDNA (Figure 1.2)

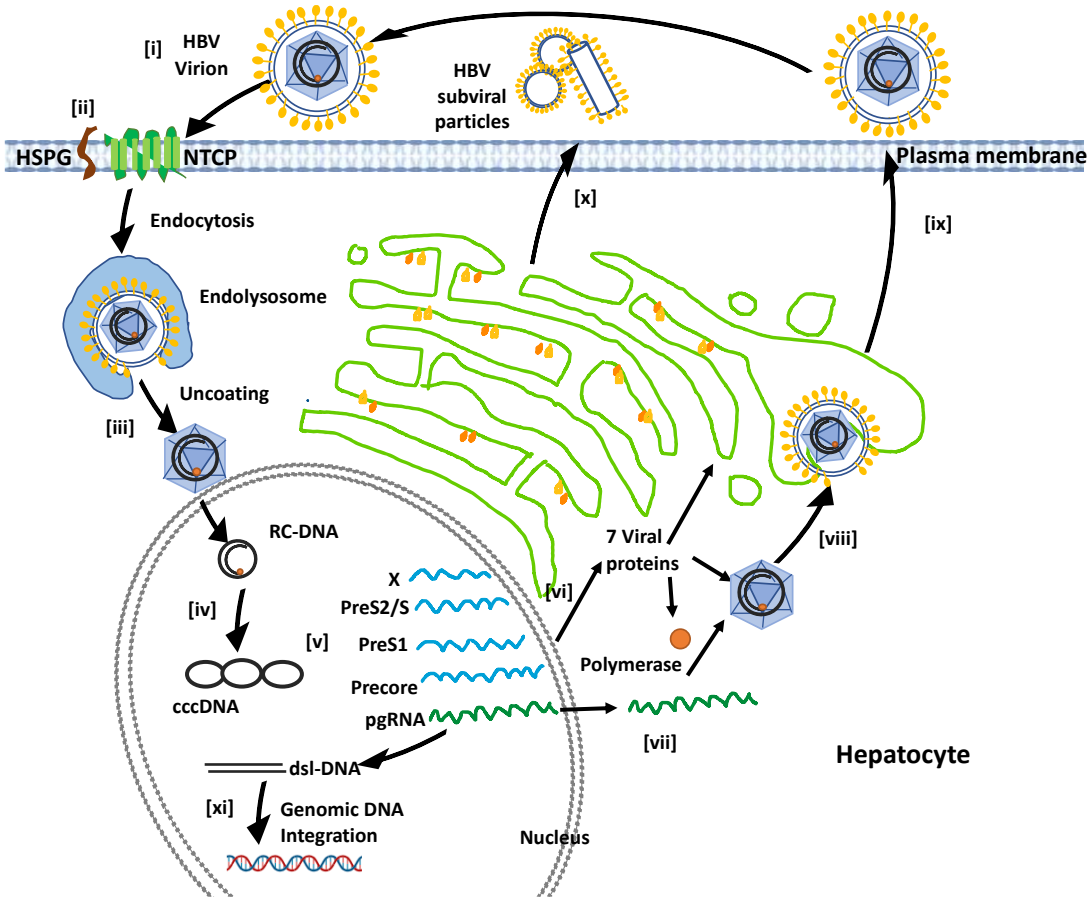

Figure 1.2 Schematic graph of the HBV life cycle

The HBV life cycle (adapted from (McNaughton et al., 2019a)) [i] Infectious HBV particles, also known as Dane particles, are approximately 42 nm in diameter; [ii] HBV virions enter hepatocytes via initial low-affinity binding with heparan sulphate proteoglycan (HSPG) and then binding to sodium taurocholate cotransporting polypeptide (NTCP) with high-affinity binding; [iii] upon endocytosis, with unclear mechanisms in uncoating and translocation, viral DNA enters the nucleus in the form of rc-DNA; [iv] the rc-DNA is repaired by cellular enzymes and converted into cccDNA; [v] the cccDNA is a template for all viral RNA transcripts; [vi] seven viral proteins are translated from five mRNAs: X mRNA (0.7-kb) encoding the X protein; preS2/S mRNA (2.1-kb) encoding Medium (M) and Small (S) proteins of the HBV surface proteins (HBsAg); preS1 mRNA (2.4-kb) encoding the Large (L) of HBsAg ; preCore mRNA (3.5-kb) encoding e antigen (HBeAg); and pgRNA (3.5-kb) encoding core and

polymerase proteins; [vii] In the icosahedral symmetrical capsid, the polymerase transcribes pgRNA into minus-strand DNA and mediates the synthesis of an incomplete plus-strand DNA; [viii] In the post-ER compartment, capsid is enveloped by HBsAg, which marks the completion of infectious virions; [ix] Infectious HBV particles are released from the hepatocytes; [x] without capsids, HBsAg can exit the hepatocyte as non-infectious empty spherical or filamentous subviral particles; [xi] double-stranded linear DNA (dsl-DNA) is integrated into the host genome.

1.2.2.2 Gene transcription

The cccDNA is transcribed into the major RNA transcripts, including the 3.5 kilobases (kb) preCore RNA, or pregenomic RNA (pgRNA) and subgenomic RNAs those that encode the surface antigens and X antigens. These RNAs have lengths of 2.4 kb, 2.1kb and 0.7kb. All RNAs have a 5' cap, and 3' poly-adenylated (A) tail; the RNAs start at different sites within the genome, but end at a common site and all RNAs serve as messenger RNAs (mRNAs) for protein production. The preCore mRNA (pcRNA) is 3.5-kb in length and contains the start codon of the preCore, which is translated to e antigen (HBeAg) (Beck and Nassal, 2007). The 2.4-kb mRNA is translated to L. The 2.1kb mRNA is translated predominantly to S and small amounts of M. The smallest 0.7kb mRNA is translated to X protein (Tong and Revill, 2016) (Figure 1.2). The pgRNA encompasses a unit of the entire genome length plus terminal redundancy of approximately 120 nt containing a second copy of the direct repeat 1 (DR1) and the epsilon (ε) signal (Figure 1.3a). The pgRNA encodes the Core protein and the reverse transcriptase. Furthermore, pgRNA is also the template for the generation of a new HBV genome (Beck and Nassal, 2007).

Gene expression is regulated by two enhancers and four promoters. Enhancer I (EnhI) is located upstream of the X promoter, and Enhancer II is located upstream of the preCore promoter. Both enhancers are crucial in HBV gene transcription by regulating all promoters (Doitsh and Shaul, 2004). Core promoter drives the transcription of the pgRNA and pcRNA. X promoter regulates the transcription of the 0.7kb mRNA. There are two transcriptional promoters that participate in the regulation of gene expression. One distal (SPI) is a TATA box, is upstream of the start codon of preS1, directs the transcription of the 2.4kb mRNA. SPII is a proximal simian virus 40-like promoter that promotes the transcription of 2.1kb mRNA (Yuh and Ting, 1990, Gallina et al., 1992, Chang et al., 1989, Chang and Ting, 1989).

1.2.2.3 RC-DNA formation

Initiation of rcDNA formation starts at the 3' end of DR1*. It is thought that pgRNA is organised in a close-loop formation with the 5'- ε and the DR1* in juxtaposition to each other (Shin et al., 2004). The 5'-UUCA at the apical loop of the 5'- ε serves as the priming site, which gives rise to its complementary motif 3'-AAGT bound to TP of the P protein (Figure 1.3a). This TP-bound motif is translocated to its complementary sequence 5'-UUCA of 3'-DR1* and the synthesis continues till the 5' end of pgRNA. The fate of the poly-adenylated 3' end is yet to be determined (Figure 1.3c). The end product of this synthesis is the minus-strand rcDNA containing a unit length of HBV DNA plus terminal redundancy "r" (5' and 3'r), approximately 10 nt (Beck and Nassal, 2007). The RNase H domain simultaneously catalyses the degradation of pgRNA up to the 5' terminus including 5'-DR1 and the cap. The remaining 5' terminal region of pgRNA serves as the primer for the plus-strand rcDNA. The synthesis of the plus-strand rcDNA involves primer translocation and circularisation (Liu et al., 2004, Shin et al., 2004). First, the 5'-terminus of the pgRNA is translocated to the 3'-DR2 motif of the minus-strand rcDNA and extended towards the P-bound 5'-terminus of the minus-strand rcDNA (Beck and Nassal, 2007). Further elongation requires circularisation where the growing DNA translocates to 3'r of minus-strand rcDNA. The final product of the plus-strand DNA is incomplete yielding heterogeneous lengths, which may arise as a result of the differential supply of deoxyribonucleotide triphosphates (dNTPs) inside individual nucleocapsids or due to space restrictions in the enclosed nucleocapsids (Beck and Nassal, 2007) (Figure 1.3).

The final product of rcDNA is characterised by: (1), the minus-strand DNA is longer than a unit genome length containing terminal redundancy sequence; (2), the P protein is covalently linked to the 5'-terminus of the minus-strand DNA; (3), the plus-strand DNA is incompletely synthesised and contains a capped RNA oligonucleotide at its 5'-terminus (Guo et al., 2007) (Figure 1.3).

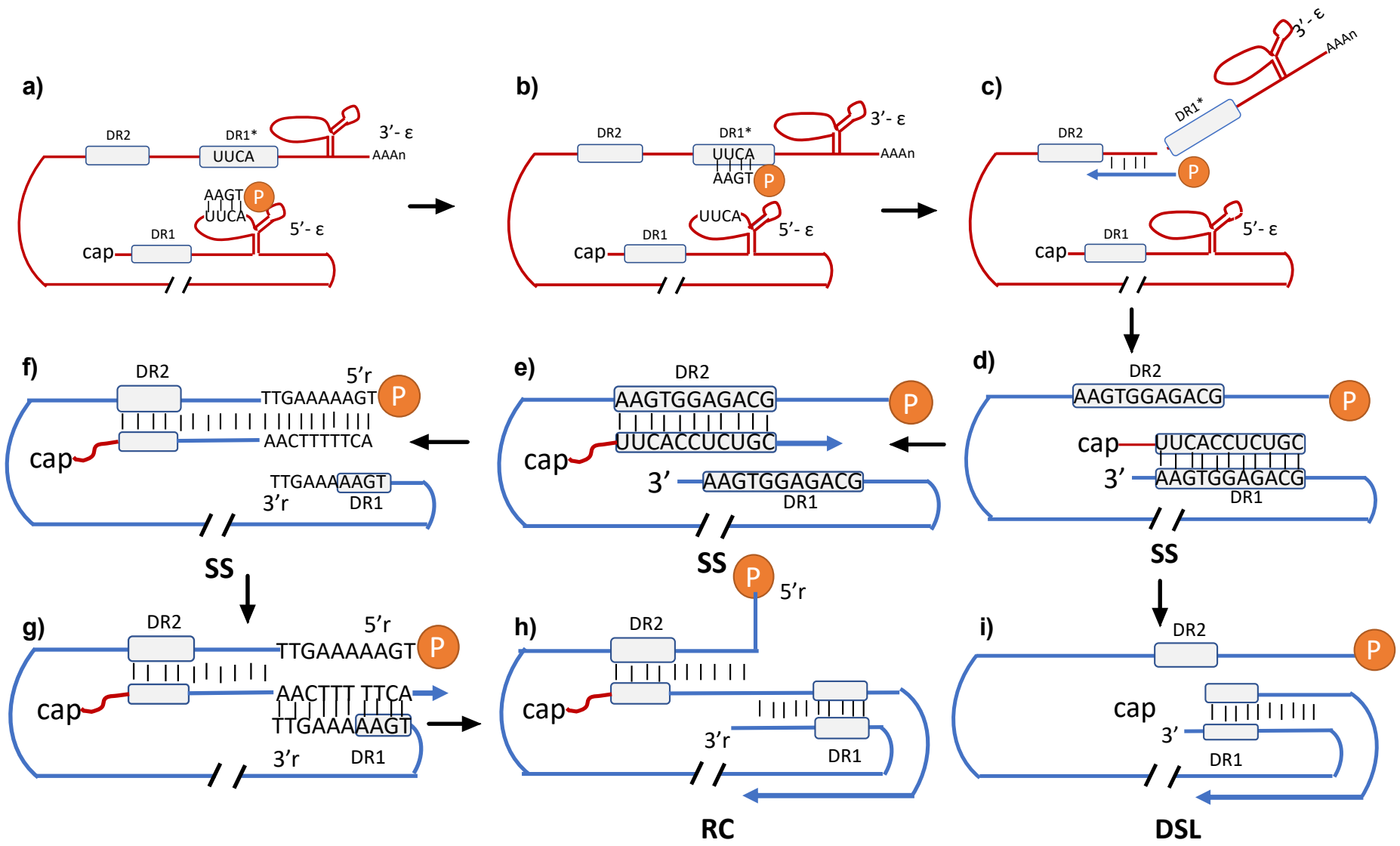
a)
DR2
DR1*
UUCA
3'- ε
AAAn
AAGT
P
UUCA
DR1
cap
5'- ε
b)
DR2
DR1*
UUCA
3'- ε
AAAn
AAGT
P
UUCA
DR1
cap
5'- ε
c)
3'- ε
AAAn
DR1*
DR2
P
DR1
cap
5'- ε
d)
DR2
AAGTGGAGACG
P
cap
UUCACCUCUGC
3'
AAGTGGAGACG
DR1
SS
e)
DR2
AAGTGGAGACG
P
cap
UUCACCUCUGC
3'
AAGTGGAGACG
DR1
SS
f)
DR2
5'r
TTGAAAAAGT
P
cap
AACTTTTTCA
TTGAAAAAGT
3'r
DR1
SS
g)
DR2
5'r
TTGAAAAAGT
P
cap
AACTTT TTCA
TTGAAAAAGT
3'r
DR1
h)
P
5'r
DR2
cap
3'r
DR1
RC
i)
DR2
P
cap
3'
DR1
DSL

Figure 1.3 Schematic diagram of HBV reverse transcription

a)*. Initiation of the minus-strand HBV DNA replication. The synthesis is initiated at the apical loop of the 5'- ε of the pgRNA. P protein serves as the primer for reverse transcriptase;* ***b)****. The first primer translocation. P protein-bound motif is translocated to its complementary sequence 5'-UUCA of 3'-DR1*;* ***c) & d)****. Completion of the minus-strand DNA synthesis. The synthesis continues till the 5' end of pgRNA. Meanwhile, the RNase H degrades the pgRNA except for the capped 5' end region including the DR1;* ***e).*** *The second primer translocation. The remaining pgRNA is translocated to DR2 and serves as the primer for the synthesis of the plus-strand DNA;* ***f).*** *Plus-strand DNA elongation resumes up to the 5' redundancy (5'r);* ***g).*** *DNA circularisation. The 3' end of the plus-strand DNA anneals to its complementary 3'r of the minus-strand DNA;* ***h).*** *Completion of the relaxed circular (RC)-DNA;* ***i).*** *Formation of double-strand linear DNA (DSL). DSL is synthesised when RNA primer fails to translocate to DR2, known as "in-situ priming". The red lines represent pgRNA and the blue lines represent DNA (Adapted from (Liu et al., 2004)).*

1.2.2.4 Viral assembly and secretion

An infectious Dane particle consists of an outer lipid envelope and an icosahedral nucleocapsid that encloses one copy of P protein-linked rcDNA. The viral capsid is composed of full-length core proteins. A full-length core protein contains 183 to 185 residues (depending on genotype) including an assembly domain (1-149) and a carboxy-terminal domain (CTD) essential for nucleic acid-binding (Selzer and Zlotnick, 2015). *In vitro*, the N-terminal 149 residues have been extensively used to study the HBV capsid assembly. A molecular structure of the HBV capsid has been determined by cryo-electron microscopy (EM) and X-ray diffraction. Approximately 95% of capsids have T=4 icosahedral symmetry, composed of 120 dimers. The assembly domain has 5 α-helices, of which helices three and four from each half dimer form a four-helix bundle into a shape of an upturned “T”, while helices one, two and five make up a hydrophobic core surrounding the “T” stalk that modulates the dimer-dimer interaction(Selzer and Zlotnick, 2015, Wynne et al., 1999). The dimers are the basic unit for capsid assembly; however, whether the assembly is via trimer of dimers, tetrameric or pentametric intermediates is still unknown (Holmes et al., 2015).

The CTD is a highly positively charged domain, including 16 arginines and 7 serines that provide interactions between core protein and pgRNAs (Selzer and Zlotnick, 2015). Experimental evidence showed that core variants ending at residue 164 were sufficient for pgRNA encapsidation; however, residues 165-173 are required for the synthesis of the positive-strand DNA and therefore, the formation of rcDNA (Nassal, 1992). The arginine-rich CTD has multiple effects on genome replication, including the first primer switch during the minus-strand DNA synthesis, second primer translocation and circularisation during synthesis of the positive-strand DNA (Lewellyn and Loeb, 2011).

Not all mature nucleocapsids are enveloped to infectious particles; some are recycled back to the nucleus, replenishing the pool of cccDNA (Selzer and Zlotnick, 2015). Some are enveloped via direct interactions between the L and Core proteins. Several core amino acids were found to participate in L and Core direct interactions including residues L60, L95, K96, I126 and Y132. The L proteins act as a molecular platform for the recruitment of S and Core into the ER compartment for envelopment (Pastor et al., 2019). L proteins facilitate the budding of capsid at the membrane of multivesicular endosomes and viral secretion (Selzer and Zlotnick, 2015). However, it remains unclear how surface proteins are transported from the pre-Golgi/ER membrane to the late endosomes for assembly.

1.2.3 HBV integration

Under certain conditions, a fragment of HBV viral DNA can be integrated into the host genome. This occurs when double-stranded linear (dsl)DNA is generated resulting from a failure of RNA primer translocation from the direct repeat (DR) 1 site to DR2 site during the process of plus-stranded DNA synthesis (Yang and Summers, 1995, Liu et al., 2004) (Figure 1.3). This dslDNA is then integrated into the host genome at the site of double-strand DNA breaks (Bill and Summers, 2004). However, the mechanism of integration is still unclear. A recent high-throughput viral integration detection (HIVID) analysis suggested that a microhomology-mediated end joining (MMEJ) pathway may account for integration events (Zhao et al., 2016).

HBV DNA integration is not essential for successful completion of the HBV life cycle. The majority of breakpoints of the integrated HBV genome occur at around the 3' end of X and 5' end of preCore/Core genes (Zhao et al., 2016). This structural arrangement interrupts the expression of all viral ORFs apart from the HBsAg ORF (Tu et al., 2017). Moreover, the dslDNA is a unit length of HBV genome plus terminal

redundancy of ~16 nts, unable to produce pgRNA, which means the integrated HBV DNA is incapable of replication (Tu et al., 2017). However, integration into the host genome is detected in approximately 90% HBV related HCC tumours (Saigo et al., 2008) indicating HBV genome integration plays a role in disease progression.

1.2.4 Genetic heterogeneity

HBV has been categorised into nine genotypes (A-I), with at least 8% divergence at the nucleotide level between each genotype. A tenth putative genotype J discovered in one Japanese individual is phylogenetically positioned in between human and ape genotypes of HBV. (Okamoto et al., 1988, Lin and Kao, 2015, Tatematsu et al., 2009). Genetic divergence between 4-8% is determined as subtypes and at least 35 subtypes have been identified (Kramvis et al., 2008, McNaughton et al., 2019a).

HBV infection in a host exists as a quasispecies containing closely related, yet genetically distinct variants, a phenomenon known as mutant spectrum. Mutations frequently occur due to the error-prone nature of the RNA-dependent DNA polymerase, combined with the high replication rate and viral turnover, and in some cases by recombination, such as genotype G has recombination events with genotype A in the S domain between nt. 250-350; genotype E and genotype D also have evidence of recombination in the core encoding region between at nt. 1950-2500 (Simmonds and Midgley, 2005).

Viral diversity is often assessed to monitor viral evolution. The estimated rate of evolution is 1.4-3.2 x 10^{-5} per substitution per site per year (Chotiyaputta and Lok, 2009, Mina et al., 2015), resulting in a cloud of genotypic and phenotypic variants. This rapid evolution would lead to loss of biological information were it not for constant elimination of unfit mutants (Domingo et al., 2012). Moreover, viral population and

genomic variants are also influenced by bottleneck events leading to severe reduction of population size, which occurs whenever there is a new event of infection, not only during host-to-host transmission, but also in the course of cell-to-cell spread within a host. The inter-host evolution appears to be relatively static compared with the intra-host evolution (Domingo and Perales, 2019). Evidence from two sets of 400-year-old mummified remains from Korea and Italy reveals that sequences of HBV isolates have surprisingly minimal genetic divergence from modern HBV genomes (Patterson Ross et al., 2018, Kahila Bar-Gal et al., 2012). However, the relative static inter-host evolution could be due to different approaches taken in the analysis. In general, the intra-host analysis examines viral population with low and high frequencies in a host within a short period of time, whereas, inter-host studies often use the consensus sequence, i.e. the high-frequency variant, therefore, inter-host comparison ignores the whole mutant spectrum (Renzette et al., 2017). However, a mutant spectrum can contribute to biologically significant perturbations without any modification in the consensus sequence (Domingo et al., 2012).

1.3 HBsAg proteins and their roles in HBV entry

While HBV entry is intensively studied, there is still more to discover. The 2.4kb mRNA and 2.1kb mRNA are transcribed from one single ORF surface, which are then translated into L, M and S by initiation at three different start codons: specifically AUG1, AUG2 and AUG3 (Gallina et al., 1992). Initiation at AUG3 yields S, which contains 226 amino acids; initiation at AUG2 adds a further 55 amino acids, known as the preS2 domain, to the N terminus of S that yields M; initiation at AUG1 adds 108 to 119 amino acids (depending on genotype), known as the preS1 domain to the N terminus of M, which generates L (Figure 1.4). The surface proteins not only are incorporated with nucleocapsids which are then released as 42 nm virions, but are also secreted as spherical and filamentous sub-viral particles (20 nm) in large

quantity, approximately 10,000 to 1,000,000 fold higher concentration than that of virions (Bruss and Vieluf, 1995, Ganem and Prince, 2004).

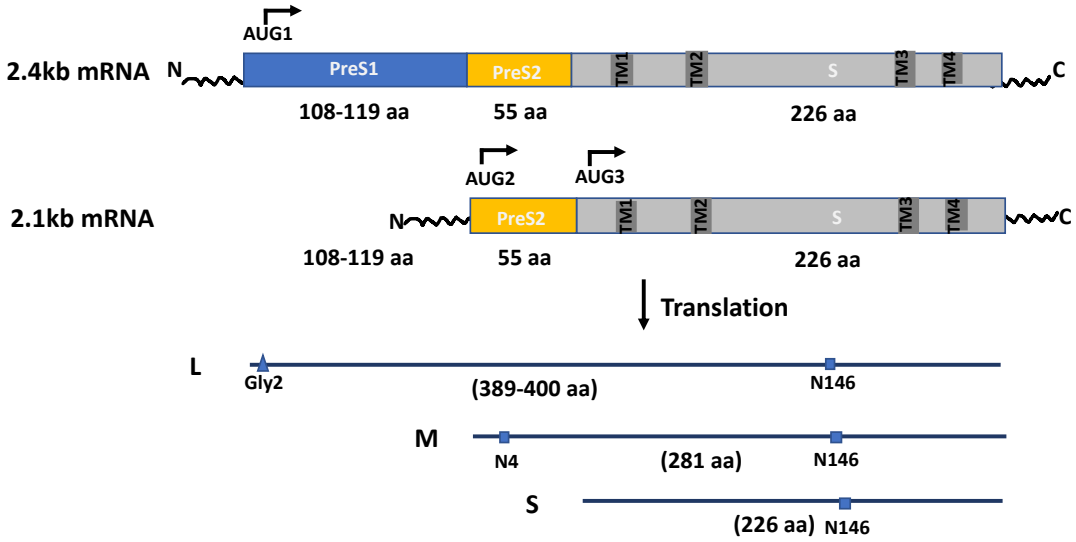

Figure 1.4 Schematic diagram of HBV surface antigen expression (HBsAg)

Two messenger RNAs including the 2.4kb mRNA and 2.1kb mRNA are translated into the HBsAg by initiation at different start codons including AUG1, AUG2 and AUG3. Initiation at AUG1 gives rise to L (389-400 aa depending on the specific genotype) containing preS1, preS2 and S domains. Initiation at AUG2 generates M (281 aa) including preS2 and the S domain. Initiation at AUG3 generates S (226 aa). Therefore, all proteins share the same C-terminus (226aa) with a common glycosylation site (blue square) at 146 aa. M adds an extra glycosylation site at 4 aa (blue square). L is myristoylated at 2 aa (blue triangle). Transmembrane (TM) represents the areas where the protein traverses the endoplasmic reticulum (ER).

1.3.1 PreS1 domain

A synthetic peptide analogue of the preS1 adw2 (aa21-47), corresponding to preS1 (aa10-36) in genotypes D, E and G, was reported to block the binding of HBsAg to HepG2 cells (a hepatocellular carcinoma cell line) (Neurath et al., 1986). In a consistent manner, a study demonstrated that, in the absence of the short peptide, the amount of HBV binding to HepG2 was similar to primary tree shrew hepatocytes (PTHs). (Leistner et al., 2008). However, the HepG2 cell line is not a reliable cell model for studying viral-receptor binding as these cells are not permissive for HBV infection. Since then, a number of studies mapping the preS1 domain have been reported using terminally differentiated HepaRG, PTH and primary human hepatocytes (PHH), all susceptible cell models for HBV infection.

Experimental evidence showed that myristoylation of preS1 was essential for HBV infectivity as well as the inhibitory activity of preS1 peptides (Glebe and Urban, 2007, Glebe et al., 2005). It was shown that the length of the acyl side chain was also important. Reducing the length by replacement of the C14-myristic acid chain with pentanoyl (C5) or octanoyl (C8) reduced inhibitory activity by more than 10-fold. Whereas, preS1 peptides acylated with palmitic acid (C16) or stearic acid (C18) increased the inhibitory effect by more than 10-fold (Gripon et al., 2005).

The function of the preS1 domain in HBV infectivity has long been recognised. PreS1 peptide blocking assay has been widely used to understand the early infection events of HBV gaining an initial insight into the amino acid sequence requirements within the preS1 domain. One early experiment demonstrated that the N-terminal 77 amino acids of preS1 was essential for infectivity (Le Seyec et al., 1999). Consistent with this finding, a synthetic peptide consisting of myristoylated 2-78 amino acids of preS1 ($2\text{-}78^{myr}$) could effectively block HBV infection. However, the peptide $2\text{-}48^{myr}$ increased

the inhibitory activity suggesting amino acids 49-78 act as a shield masking the activity of amino acids 2-48, which may play a role in viral secretion (Gripon et al., 2005). In a PTH cell model, infection inhibition assay using lipopeptides containing various length of myristoylated preS1 observed that: (1) residues containing aa 2-8 (genotype D) showed a weak inhibitory effect on HBV infection; (2) addition of aa 9-18 enhanced the inhibitory effect drastically by over 30-fold; (3) aa 19-28 reduced the inhibitory effect slightly; (4) peptides containing aa 19-48 abolished the inhibitory effect; (5) peptide 2-48 showed the maximal inhibitory effect. These observations suggested that aa 9-18 were essential, aa 19-28 were dispensable, whereas residues of aa 2-8 and aa 29-48 enhanced the inhibitory effects (Glebe et al., 2005). Consistent with this finding, a separate study showed that the short peptide 2-48^{myr} displayed specific inhibition of HBV infection and further C-terminal shortened peptides 2-39^{myr} and 2-28^{myr} exhibited progressively reduced but detectable inhibitory activities. Further C-terminal shortened peptides 2-18^{myr} and 2-8^{myr} displayed loss of inhibitory function. This led to the hypothesis that residues 2-18 were not required for efficient activity. However, a subsequent experiment using peptide 19-48^{myr} showed no detectable inhibition, suggesting residues 2-18 were essential either participating in direct receptor recognition or serving as a structural feature (Gripon et al., 2005). The discrepancy regarding the importance of aa 9-18 in infection-inhibition could be due to the recipient cells used. Glebe *et al* used PTH whereas Gripon *et al* used terminally differentiated HepaRG cells. Possibly, the binding between preS1 aa 9-18 and the receptors on PTH was more specific compared to that with human hepatocytes.

A short linear sequence of amino acids between aa 9-17 (NPLGFFPDH) is highly conserved in primate and bat hepadnaviruses, but not in rodent, such as Woodchuck hepatitis virus (Drexler et al., 2013). Residues between aa 29-33 (NPDWD) and aa 37-50 (NKDHWPEANKVGVG) have been found only in the tent-making bat, but not the Old World bat hepatitis virus, displaying high levels of sequence identity to primate

HBV (Drexler et al., 2013). Hepatitis D virus enveloped with surface proteins of roundleaf bat hepatitis virus (an old-world bat species) were unable to infect primary human hepatocytes. Interestingly, the hepatitis virus of tent-making bat (a new-world bat species) was capable of infecting human hepatocytes (Drexler et al., 2013) suggesting there is potential for cross-species transmission and zooNotIc infection.

The importance of the preS1 domain in viral attachment leading to infection has also been shown by the use of neutralising antibodies against this domain. MA18/7, a monoclonal antibody binding to aa 20-23 of the preS1 (DPAF), completely blocked the infection of purified plasma-derived HBV to PTH (Glebe et al., 2003). Introducing internal deletions at aa 20-23 of the myristoylated preS1 2-48 (preS1-2-48^{myr}) led to 10 to 100 times reduction in HBV infection in PTH and HepaRG cells (Gripon et al., 2005, Glebe et al., 2005), which confirmed that these residues are also involved in viral-receptor interaction. Further fine mapping displayed that N-terminal aa 2-9 contribute to inhibitory activity in a sequence-specific manner. Amino acid substitution of asparagine (N) to lysine (K) at residue 9 completely abolished the inhibitory activity of peptide 2-48^{myr} (Yan et al., 2012). Meanwhile, deletion of aa 11-15 or substitution of aa 11, 13, 14, and 15 with respective D-amino acids resulted in complete loss of inhibitory activity suggesting these segments are involved in stereo-selective interaction with the receptor (Schulze et al., 2010).

1.3.2 PreS2 domain

The preS2 domain contains 55 amino acids corresponding to the N-terminal domain of the M protein. While the M protein has been proven dispensable for the production of complete viral particles and viral infectivity (Lepere-Douard et al., 2009, Ni et al., 2010), the preS2 domain in L protein serves an important role in viral assembly and infectivity (Le Seyec et al., 1998, Ni et al., 2010). The first five amino acids of preS2

directly participate in envelope-core interaction as viral assembly and secretion were abolished when these five amino acids were lacking (Le Seyec et al., 1998, Pastor et al., 2019). One early experiment showed that apart from the first five amino acids, the remaining preS2 sequence was dispensable for viral assembly, secretion and infectivity (Le Seyec et al., 1998). However, this experiment only examined viral assembly, secretion and infectivity using plasmid constructs that contained deletions of up to 10 amino acids in the preS2 domain. A more recent study showed that viral assembly was progressively impaired when introducing deletions exceeding 20 amino acids within this domain (Ni et al., 2010). Evidence has shown that the preS2 deletion mutant proteins accumulate in the endoplasmic reticulum (ER) and upregulate ER stress signals, subsequently, increase the incidence of HCC development (Su et al., 2014). Interestingly, apart from the first five amino acids, the introduction of randomised sequence in the remaining preS2 domain still support viral assembly, secretion and infectivity. This indicates that preS2 domain serves as a spacer to support conformational changes of L protein in viral assembly (Ni et al., 2010).

1.3.3 S domain

In addition to the essential preS1 region, the S domain also plays a crucial role in HBV entry. It is hypothesised that the first transmembrane region (TM1) of the S participates in membrane fusion during viral entry. Consistent with this hypothesis, multiple alanine substitutions of three hydrophobic clusters within TM1 including 12-LLVL-15, 19-FFLL-22 and 25-ILTI-28 in the L and M proteins drastically reduced the viral infectivity by up to 95% (Lepere-Douard et al., 2009). Further evidence showed that the deletion of TM1 resulted in complete abrogation of viral infectivity (Ni et al., 2010); TM1 deletion can completely alter the confirmation of the protein, which might account for the phenotype.

In addition to the TM1, the “a” determinant consisting of aa 124-147 has been identified as another determinant in viral entry, which was evident in both HDV and HBV entry models (Jaoude and Sureau, 2005, Salisse and Sureau, 2009). Antibodies against this region display a potent neutralising activity as a result of aggregating viral particles as well as by blocking a functional motif borne by the “a” determinant (Salisse and Sureau, 2009).

There are eight cysteines within the major hydrophilic region (MHR, aa 99-169), which could participate in forming disulfide bonds to stabilise the viral envelope proteins. However, experimental evidence showed that disulfide bonds in this region are not obligatory for morphogenesis or stability of viral particles, but rather essential in viral entry. Disturbing the disulfide bonds within the “a” determinant by replacement of cysteine at positions 121, 124, 137, 138, 139, 147, 149 resulted in marked inhibitory effects on infectivity (Abou-Jaoude and Sureau, 2007). Viral Infectivity and conformation of the “a” determinant were also impaired when virions were challenged with reducing agents such as tris(2-carboxyethyl)phosphine (TCEP) prior or during the viral-cell interaction period (Le Duff et al., 2009, Abou-Jaoude and Sureau, 2007).

Most of the non-cysteine residues identified as essential for viral entry cluster together with cysteines in the disulfide bond network constituting conformation-dependent motifs (Salisse and Sureau, 2009). In particular, residues 119-124 (GPCRTC) appeared to be the most important in viral infectivity (Jaoude and Sureau, 2005). This subdomain contains a CXXC motif, a known catalytic site for enzymes such as protein disulfide isomerase (PDI) to perform disulfide isomerisation or redox reaction. This motif has been described in murine leukaemia virus peripheral subunit (SU), where one cysteine participates in disulfide bond formation with the envelope transmembrane (TM) forming SU-TM subunit and the other is free. Upon receptor binding mediated by the SU, the free thiol attacks the disulfide bond between SU-TM

and forms a disulfide isomer within the CXXC motif, which frees up TM for membrane fusion to facilitate viral entry (Wallin et al., 2005). However, in the HDV model, disrupting this CXXC motif by inserting or deleting one amino acid did not affect viral infectivity with the exception of arginine (R) at residue 122. Deletion or alanine substitution of R122 led to reduced infectivity, whereas replacing R122 with lysine (K), R122K, corresponding to *Ad* subtypes of HBsAg, did not alter the viral infectivity. This suggests a preference for a basic group in this position (Abou-Jaoude and Sureau, 2007). It would be interesting to know the viral infectivity in the presence of histidine (H) at residue 122, another basic side chain. K141 is another basic group determined to be important in viral infectivity. R122 and K141 are positioned in a three-dimensional antigenic loop structure that is stabilised by disulfide bonds. These two positively-charged groups initiate electrostatic interaction with negatively charged heparan sulphate proteoglycans (HSPGs) on the cell surface. This weak interaction is then followed by specific interaction between preS1 and NTCP, ultimately facilitating viral entry (Sureau and Salisse, 2013).

1.3.4 HBsAg morphogenesis

HBsAg translation and post-translational modification occur at the rough endoplasmic reticulum (ER). Upon translation at membrane bound ribosomes, an N-terminal type I signal and an internal type II signal sequence in the S protein lead the peptide to traverse the ER membrane twice forming a cytosolic loop as well as a luminal loop (Figure 1.5Figure 1.5a) (Eble et al., 1987, Bruss and Vieluf, 1995). In a similar way to S, the M protein also traverses the ER membrane twice with its pre-S2 domain facing the ER (Figure 1.5b) (Eble et al., 1990). Unlike S and M, the L protein adopts a dual topology where half of the L chains have their pre-S1/S2 domain exposed to the ER lumen (e-pre-S) and half of the L chains have their pre-S1/S2 domain exposed to the cytosolic side (i-pre-S) (Figure 1.5c). Experimental evidence shows that the N-

terminal 80 amino acids contain the retention signal that mediates the intracellular retention of HBsAg, while the C-terminal 11 amino acids participate in envelope-core interaction that is essential in virion formation (Bruss and Vieluf, 1995, Bruss and Thomssen, 1994, Pastor et al., 2019). The e-pre-S domain appears on the surface of the secreted viral particles and is critical in cell receptor binding (Bruss and Vieluf, 1995).

During L translation, the N-terminus of the L chain remains exposed to the cytosolic side of the ER membrane. The mechanism of how half of i-pre-S domains change their orientation during maturation is still unclear. Bruss & Vieluf (1995) observed that the S protein was needed to override the N-terminal retention signal of L (Bruss and Vieluf, 1995). It was hypothesised that oligomerised S domains formed a channel in the ER membrane that enabled the translocation of L (Berting et al., 1995). If this was true, S would be absolutely necessary in viral secretion and entry. However, Lambert & Prange (2001) demonstrated that dual topology of L was independent of either M or S. They observed that the hydrophobic transmembrane (TM) 2 (aa 80 to 90) segment in the S domain of L was essential and sufficient for pre-S translocation during maturation (Lambert and Prange, 2001). However, mutation with rtA181T/sW172* in the at reverse transcriptase introduces a stop codon at its overlapping surface antigen aa 172 that leads to a truncated S protein with a loss of 55 amino acids at the C-terminus (Yatsuji et al., 2006, Ahn et al., 2014, Yoo et al., 2007). This mutation not only causes retention of surface proteins but also affects the secretion of virions. It even has a negative effect on virion secretion where wild type is co-expressed (Warner and Locarnini, 2008). Similar effects were seen in vitro. Garcia *et al* demonstrated that virion secretion was blocked either when L was the only surface protein expressed or when the ratio L/S was too high (Garcia et al., 2009, Bruss and Ganem, 1991). Thus, the importance of S and the L/S ratio in viral secretion is still debated.

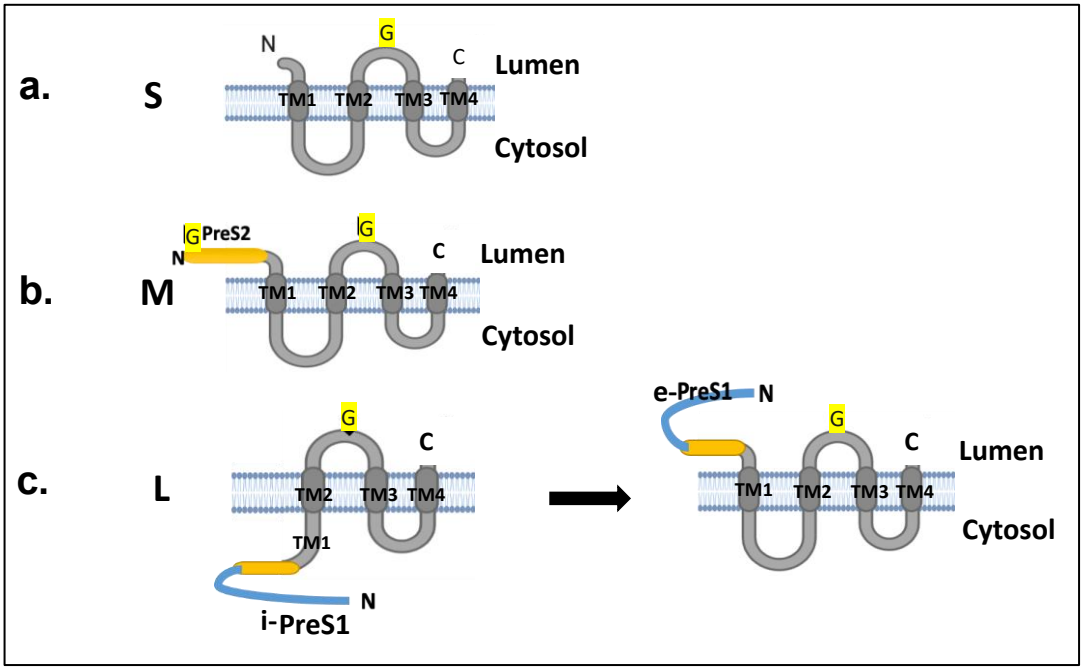

Figure 1.5 Transmembrane topology of HBV surface antigens (HBsAg) in the ER membrane.

a. *The Small surface protein (S) traverses the ER membrane twice led by the N-terminal type I signal (TM1) and an internal type II signal sequences (TM2) and twice with its hydrophobic C-terminal subdomain (TM3 and TM4);* ***b.*** *The middle surface protein (M) adopts the same topology as the S with the N-glycosylated N-terminal preS2 domain (55aa) (the orange bold line) traversing to the luminal side of the ER;* ***c.*** *The large surface protein (L) adopts a dual topology: (1) the N-terminal preS1 (blue line) and preS2 domains locate at the cytosolic side of the ER (i-preS1); (2) the N-terminal preS1 and preS2 domains traverse to the luminal side of the ER (e-preS1). TM: transmembrane; G: glycosylation. (the diagram is modified from (Lambert and Prange, 2001))*

1.4 Hepatitis delta virus-the satellite virus of HBV

Hepatitis Delta Virus (HDV) is a satellite RNA virus that requires HBV for its assembly and infection. It is estimated that at least 5% of people chronically infected with HBV are co-infected with HDV, leading to approximately 15-20 million people infected with HDV worldwide. HDV-HBV co-infection leads to increased chances of liver disease, HCC and death, therefore, is considered as the most severe form of chronic viral hepatitis (WHO, 2019c).

The HDV genome is a single-stranded negative sense RNA, consisting of approximately 1700 nucleotides. It encodes a single reading frame that translates into two isoforms of Delta antigens (HDAg) including a small S-HDAg (195 amino acids) and a large L-HDAg (214 amino acids). The two delta antigens share the identical sequence apart from that the L-HDAg has an additional 19 amino acids at its C-terminus (Hwang and Lai, 1993). The S-HDAg is responsible for genome replication, whereas the L-HDAg inhibits genome replication and facilitates viral assembly (Chao et al., 1990, Hwang and Lai, 1993). In the absence of HBV surface proteins (HBsAg), HDAg is localised in the nucleus; translocation of HDAg to the cytoplasm is mediated via the direct interaction between C211 of L-HDAg and the cytosolic loop of S HBsAg (Hwang and Lai, 1993). The HDV virion is composed of a circular RNA genome, HDAg and enveloped by HBsAg. Similar to HBV, HDV enters hepatocytes via the interaction of the preS1 domain of the L and the receptor NTCP (Lempp et al., 2019).

1.5 Target cell systems for studying HBV entry pathway

Since the discovery of HBV in 1965 and the development of a vaccine in the 1980s (Meireles et al., 2015), research into viral-host interaction has been limited. One of

the reasons for this is the lack of a reliable cell culture model (reviewed in (Allweiss and Dandri, 2016, Witt-Kehati et al., 2016)). The pros and cons of each cell culture models are discussed herein and summarised in Table 1.1.

Table 1.1 Advantages and disadvantages of in vitro cell culture systems

CELL SYSTEM	ADVANTAGES	DISADVANTAGES
PHH[1]	• Maintained functional integrity in vitro; • Support full HBV life cycle; • Can be infected by serum-derived or cell cultured derived HBV	• Infectivity varied between donor to donor; • Function drops rapidly after plating out; • Limited resources and technical difficulty • Absence of viral spreading
PTH[2]	• Permissive to serum-derived HBV; • Support full HBV life cycle; • Reproducible infections	• Low infection efficacy; • Infection inhibited by human serum • Absence of viral spreading
HEPARG	• Susceptible to HBV infection; • Support full HBV life cycle;	• Require DMSO[4] differentiation; • Infection efficacy depends on differentiation status • Absence of viral spreading
IPSC-HEPS[3]	• Can be differentiated to hepatocytes-like cells; • Support full HBV life cycle; • Support long-term infection; • Support viral spreading in vitro	• Require differentiation; • Delicate culture conditions; • Low yield of HBV viral biomarker output
NTCP-OVEREXPRESSING HEPATOMA CELLS	• Allow more reproducible infectivity; • Easy to maintain in cell culture; • Easy access	• Require high MOI[5] and PEG[6] for infection; • Absence of viral spreading

1. PHH: primary human hepatocytes; 2. PTH: primary tree shrew hepatocytes; 3. IPSC-heps: Induced pluripotent stem cell-derived hepatocyte-like cells; 4. DMSO: dimethylsulfoxide

5. MOI: multiplicity of infection; 6. PEG: polyethylene glycol

1.5.1 Primary human hepatocytes

Primary human hepatocytes (PHH) have been successfully maintained in *in vitro* cell culture. Gripon *et al.* showed that adult human primary hepatocytes could be infected by either HBV viruses from infectious sera or from the supernatants of transfected HepG2 cells. HBeAg, HBV DNA, major HBV transcripts, and complete HBV particles could be detected in culture medium. This indicates primary human hepatocytes maintain their functional integrity *in vitro* and support the full life cycle of HBV (Gripon et al., 1988). However, the infectivity for primary human hepatocytes varies from donor to donor. It was reported that inoculating freshly-plated primary hepatocytes with HBV, only 38% of all experimental cases yielded HBsAg production (Gripon et al., 1988, Galle et al., 1994). Moreover, the cell phenotype rapidly changes after plating *in vitro.* Culturing these cells is technically more difficult compared to transformed cell lines (Sai et al., 2016, Galle et al., 1994), making this cell model less tractable for complex experiments.

1.5.2 Hybrid cells

Recently, a new hybrid cell model has been developed to overcome the drawbacks of PHH culture. In this hybrid model, primary hepatocytes were fused with hypoxanthine-guanine phosphoribosyltransferase HGPRT(-) HepG2 cells, and the fused cells were able to survive in hypoxanthine-aminopterin-thymidine (HAT) medium. After multiple rounds of selection, a single colony of hybrid cells was established (Sai et al., 2016). After exposure to HBV positive serum, extracellular HBV DNA, HBsAg and HBeAg could be detected from the culture medium, as well as detection of intracellular HBcAg and HBV DNA. It was also shown that the infection could be blocked by pre-incubating HBV positive serum with various antibodies

against pre-S1 or pre-S2 (Sai et al., 2016). More data are required to fully appreciate the role this cell line plays in HBV research.

1.5.3 Primary tree shrew hepatocytes

Primary tree shrew hepatocytes (PTH), aka Tupaia hepatocytes (isolated from *Tupaia belangeri)*, have been used as target cells to study HBV infection. It was shown that incubation with HBV positive serum resulted in HBV RNA and DNA production in Tupaia hepatocytes as well as HBsAg, HBeAg being secreted into the culture medium (Walter et al., 1996, Kock et al., 2001). The use of Tupaia hepatocytes also led to the initial discovery of NTCP as a bona fide HBV receptor in 2012 (Yan et al., 2012). The infection was reproducible; however, its efficiency was low and its infectivity was inhibited by human serum, which requires HBV viral purification prior to the viral entry (Kock et al., 2001).

1.5.4 HepaRG cells

HepaRG cells were originally isolated from hepatocellular carcinoma tissues of a patient who suffered from chronic hepatitis C infection. These cells were able to differentiate to small clusters and displayed a hepatocyte-like morphology upon exposure to 2% DMSO and $5x10^{-5}$ M hydrocortisone (Gripon et al., 2002). It was demonstrated that these cells are susceptible to HBV infection and support the entire HBV replication cycle as viral cccDNA, main HBV RNA species and HBsAg were detected in terminally differentiated HepaRG cells following HBV infection. They also showed that HBV infection could be neutralised by monoclonal anti-HBsAg or blocked by myristoylated pre-S1 peptide (aa2-78) in a dose-dependent manner (Gripon et al., 2002). By using this cell culture, Schulze *et al* proposed HBV entry required preS-dependent binding to cell associated heparan sulphate proteoglycan (HSPG) as an

initial step and followed by another yet to be determined specific step for complete entry (Schulze et al., 2007). It was later discovered that the HBV attachment to HSPG is mediated by the antigenic loop (AGL) in the S domain (Sureau and Salisse, 2013). However, the downside of this system is that the infectivity is strongly reliant on the cells' differentiation status and requires DMSO and polyethylene glycol (PEG) to ensure infection efficiency (Gripon et al., 2002).

1.5.5 Induced pluripotent stem cell-derived hepatocyte-like cells (iPSC-heps)

Human induced pluripotent stem cells (iPSCs) are somatic cells genetically modified to yield pluripotency by ectopically expressing four transcription factors, including Oct4, Sox2, Klf4 and Myc (Kaneko et al., 2016). Human iPSCs can be differentiated into hepatocyte-like cells (iPSC-heps). Similar to the cell-culture models described above, iPSC-heps also support HBV viral infection and replication cycle as determined by detectable viral cccDNA, HBV RNAs, HBsAg, and HBeAg. Xia *et al* claimed that this model is superior to other models in several aspects. First of all, it supports long-term infection. Compared with 5-10 days infection maintained in PHH, iPSC-heps supported HBV infection and replication for at least four weeks (Xia et al., 2017). Moreover, it was shown that iPSC-heps support HBV virus spread, which has not been reported in other *in vitro* cell-culture models. This was achieved by co-culturing HBV infected iPSC-heps (donor cells) with GFP-expressing HBV naïve cells (target cells) and performing HBsAg staining or RNAscope on acceptor cells. It was further demonstrated that HBV spread could be blocked by treatment with Myrcludex B, a known NTCP inhibitor (Xia et al., 2017). However, Sakurai *et al* showed that the expression of HBsAg and HBcAg were approximately one $\log_{10}$ lower in iPSC-heps than in NTCP-overexpressing HepG2 cells (Sakurai et al., 2017).

1.5.6 Sodium Taurocholate Cotransporting Polypeptide expressing cells

Based on the discovery of NTCP as the bona fide receptor for HBV entry and the fact that hepatoma cell lines such as HepG2 and HuH7 lack NTCP expression, cell lines constitutively overexpressing NTCP have been developed and utilised to investigate the HBV/HDV life cycle as well to screen antiviral drugs targeting HBV/HDV entry. Recently, by using HDV as a surrogate of HBV and NTCP-overexpressing HuH7 (HuH7-NTCP) as the target cell, glypican 5 (GPC5) has been identified as a common factor that mediates the early attachment of virions to the host cell surface (Verrier et al., 2016). By using NTCP-overexpressing HepG2 (HepG2-NTCP) cells, DDX3, a member of the DEAD-box RNA helicases, has been demonstrated to inhibit HBV replication at a post-transcriptional level (Ko et al., 2014).

NTCP has become an attractive target for antiviral treatment. Myrcludex-B (MyrB), a synthetic lipopeptide corresponding to the myristoylated 2-48 aa region of preS1, is an entry inhibitor blocking the NTCP. A multicentre, open-label phase 2b clinical trial assessed the safety and efficacy of MyrB in combination with Tenofovir (TDF) in the treatment of patients with HBV/HDV co-infection. Treatment with MyrB in combination with TDF reduced HDV RNA replication over TDF alone. The inhibition of replication/spread was shown to be dose-dependent (Wedemeyer et al., 2018). In addition, cyclosporin A has also been shown to restrict HBV/HDV infection by blocking their binding to NTCP (Nkongolo et al., 2014).

NTCP-overexpressing hepatoma cell lines have their limitations. Very high multiplicity of infection (MOI) is required to infection 50% or more of the cells; a typical MOI ranges from 500 to 10000 copies of HBV genome equivalent per cell (Cheng et al., 2015). In addition, efficient infection also requires the presence of 2% dimethyl

sulfoxide (DMSO) and 4% polyethylene glycol (PEG) (Ni et al., 2014, Evripioti et al., 2019). Interestingly, HuH7-NTCP and HepG2-NTCP cell lines display very different phenotypes. HepG2-NTCP cells appear to be more susceptible to HBV infection, whereas, HuH7-NTCP cells support HDV infection but are restricted for HBV, suggesting additional factors might have contributed to these phenotypes (Ni et al., 2014). A study found that L protein coated nanocapsules (bio-nanocapsules, BNCs), myristoylated BNCs (Myr-BNCs) and HBsAg particles were able to enter HepG2 cells in an NTCP-independent fashion. Conversely, the entry of HBsAg particles was inhibited by myristoylated pre-S1 residues aa.2-48 irrespective of NTCP expression. Enriched NTCP was also observed in endosome/lysosomes in HepG2-NTCP cells (Somiya et al., 2016, Sarkar et al., 2006). This leads to a postulation that the entry process of HBV is independent of surface NTCPs but dependent on intracellular NTCPs present in the membrane of endosomes/lysosomes (Somiya et al., 2016).

1.6 Surrogate systems for studying HBV entry

Previous investigations into HBV entry predominantly utilised serum and/or cell-culture derived live HBV and cells that are known to be susceptible to HBV entry such as PHH, PTH or terminally differentiated HepaRG cells (Li et al., 2016, Yan et al., 2012). However, HBV is classified as a Hazard Group 3 hazardous biological agent requiring Containment Level 3 (CL-3) handling as a minimum, i.e. biosafety level 3 (BSL-3) laboratory equipment is required (Pastorino et al., 2017). Furthermore, numerous variations in the HBV S gene provide a challenge to the interrogation of HBV entry, as culture models for rapidly assessing the phenotype of naturally occurring variants are lacking (Gencay et al., 2017, King and Tarr, 2017). In addition, the gene encoding HBsAg completely overlaps with the polymerase gene within the HBV genome. Mutations occurring in the HBsAg can simultaneously affect the

polymerase ORF. It is difficult to interrogate the phenotype of HBsAg mutation without altering the behaviour of another viral component (King and Tarr, 2017).

1.6.1 HDV as a surrogate for studying HBV entry

In laboratory settings, HDV genome supplemented with HBsAg encoding genetic material *in trans* is known as HDV pseudotyping. HDV pseudetyped viruses (PVs) have been widely used as surrogates in the study of HBV entry (Yan et al., 2013, Drexler et al., 2013). Initially, HDVpv was regarded as a safer alternative to HBV live virus as HDV is a satellite virus that requires HBV for its assembly and cellular transmission. However, a recent *in vitro* and *in vivo* study demonstrated that HDV potentially can infect hosts without co-infection with HBV. *In vitro,* HDV genome could be enveloped by glycoproteins from six other non-HBV viruses, including vesiculovirus (VSV), hepatitis C (HCV), lymphocytic choriomeningitis virus (LCMV), human metapneumovirus (hMPV), dengue virus (DENV) and West Nile virus (WNV), and could produce infectious HDV particles. Furthermore, HCV could propagate HDV infection in mice with humanised liver (Perez-Vargas et al., 2019).

1.6.2 Pseudotyped viruses

Alternatively, HBV pseudotyping provides an attractive alternative to overcome the drawbacks of handling live HBV *in vitro*. Pseudotyped HBV is defined as a particle consisting of the structural proteins of a retrovirus, enveloped by HBV surface glycoproteins and carrying reporter genes. This pseudotyped HBV only requires CL-2 containment.

Pseudotyped viruses (PVs) is a chimeric viral particle harbouring the structural proteins of one virus and glycoproteins of another. The structural proteins are the

backbone of the pseudotyping system. Retroviruses such as lentiviral human immunodeficiency virus (HIV) and murine leukaemia virus (MuLV) are the most commonly used backbones for pseudotyping. Vesicular stomatitis virus (VSV) has also been developed as a pseudotyping vector (Cheresiz et al., 2014). However, the establishment of this system is complex. First, producer cells must be transfected with plasmids encoding the VSV nucleoprotein (N), phosphoprotein (P), and polymerase protein (L) as well as the pVSV- ΔG* genome plasmid (VSV-G replaced with reporter gene). Further addition of a VSV-G expression plasmid leads to the production of VSV-PVs encapsidating the VSV-ΔG* genome and coated with VSV-G. Furthermore, a source of bacteriophage T7 polymerase must be provided by *trans* in order to initiate the transcription of VSV- ΔG* genome. Finally, VSV-PVs coated with heterologous glycoproteins can be produced by the transfection of a heterologous glycoprotein expression plasmid followed by infection with VSV-PVs pseudotyped with VSV-G (Whitt, 2010, King et al., 2016). However, due to the high efficiency of VSV-G pseudotyping, it is important to incubate the resulting heterologous glycoprotein pseudotyped VSV-PVs with anti-VSV-G hybridoma cell culture supernatants to minimise background infection mediated by residual VSV-PVs pseudotyped with VSV-G (Steeds et al., 2020).

1.6.2.1 Retroviral backbones for pseudotyping

The primary genes provided by retrovirus are *gag-pol* that encode the structural proteins for pseudotyping. The *gag* gene encodes a polyprotein, which is cleaved into multiple domains including matrix (MA), capsid (CA) and nucleocapsid (NC). The retroviral *pol* gene encodes enzymes, including protease, reverse transcriptase (RT) and integrase that are responsible for processing Gag polyproteins, reverse transcribing genomic RNA to DNA and integration of the genome into the host chromosome (Dalba et al., 2007, Johnson, 2011)).

In addition to the structural proteins encoded by the *gag* and *pol* genes, the HIV has additional proteins encoded by the regulatory genes (*tat* and *rev*) and the accessory genes (*vpu*, *vpr*, *vif* and *nef*) that are essential for viral life cycle and pathogenesis. The tat protein enhances the promoter activity of the 5' long-terminal repeat (LTR), and as such it is necessary for the transcription from the 5' LTR. The rev protein is necessary for the transport of un-spliced RNA out of the nucleus by binding to the rev response element (RRE) within the viral RNA. It is also necessary for efficient expression of *gag* and *pol*. Although the accessory proteins are essential for the HIV life cycle, they are dispensable for the production of PVs ((reviewed in (Blesch, 2004)). To reduce the likelihood of producing replication-competent retroviruses that could originate in the vector producer or target cells, and to prevent recombination with wild-type HIV in the host, the HIV backbone has undergone several genetic modifications including (i) the deletion of accessory proteins; (ii) replacement of the tat protein for a safer promoter, such as the human cytomegalovirus (CMV) promoter; (iii) encoding rev in a separate plasmid. Plasmid pNL4.3.luc.R-E- is a first-generation lentiviral packaging vector (Chen et al., 1994) (Figure 1.6). It has non-functional *env* and *vpr* due to frameshifts in the *env* and *vpr* genes. The plasmid contains a firefly luciferase gene cloned into the *nef* position (NIH, 2017). The second-generation lentiviral packaging plasmids, such as plasmids psPAX2 (Addgene plasmid #12260) and pCMVR8.74 (Addgene plasmid #22036), gifts from Didier Trono, only contain the *gag*, *pol*, *tat* and *rev* genes. These plasmids lack all the accessory genes as well as the LTRs and a packaging signal, psi (Ψ); both the LTRs and psi are supplied in a separate transfer plasmid bearing a transgene (Zufferey et al., 1997). The transfer plasmid is designed such that the transgene is flanked by 5'- and 3'- LTRs and the packaging signal Ψ, which allows encapsulation of the transgene into the PVs and integration into the host genome. The U3 region of the 3' LTR acts as a promoter allowing the expression of the transgene after integration. To improve safety, this U3 region from the 3' LTR is removed resulting in a self-inactivating vector (Zufferey et

al., 1998). The commonly used transfer plasmids pCFLW (expressing firefly luciferase) and pCSGW (expressing green fluorescent protein) have this safety property in place ((reviewed in (Carnell et al., 2015)).

To further improve safety, a third-generation packaging system has also been developed. In this design, the system splits into two plasmids: one encoding *gag* and *pol*, the other encoding the *rev*. The gene *tat* is deleted from this system and its function is rescued by the addition of a constitutive promoter upstream of the reporter gene on the transfer plasmid (Dull et al., 1998). In this system, four plasmids are required to generate the PVs, which minimises the risk of producing HIV-1-like virus.

Similar to the HIV packaging vectors, MuLV provides *gag* and *pol* as the backbone for pseudotyping. Plasmid phCMV-5349, a commonly used MuLV packaging vector, and a luciferase encoding plasmid pTG126 (kind gifts from François-Loïc Cosset) have been used for the production of PVs (Urbanowicz et al., 2016) (Figure 1.6). In contrast to HIV-based PVs that are able to infect and integrate their genome into terminally differentiated cells, the MuLV-based PVs are only able to deliver their genome into actively dividing cells (Roe et al., 1993). Despite this difference, the choice of backbone for pseudotyping is often down to preference, availability. Not all envelop glycoproteins can be incorporated efficiently into the HIV or MuLV backbones, in the case of pseudotyping for a novel virus, the choice of backbones should be optimised by empirical studies ((reviewed in (King et al., 2016, Carnell et al., 2015)).

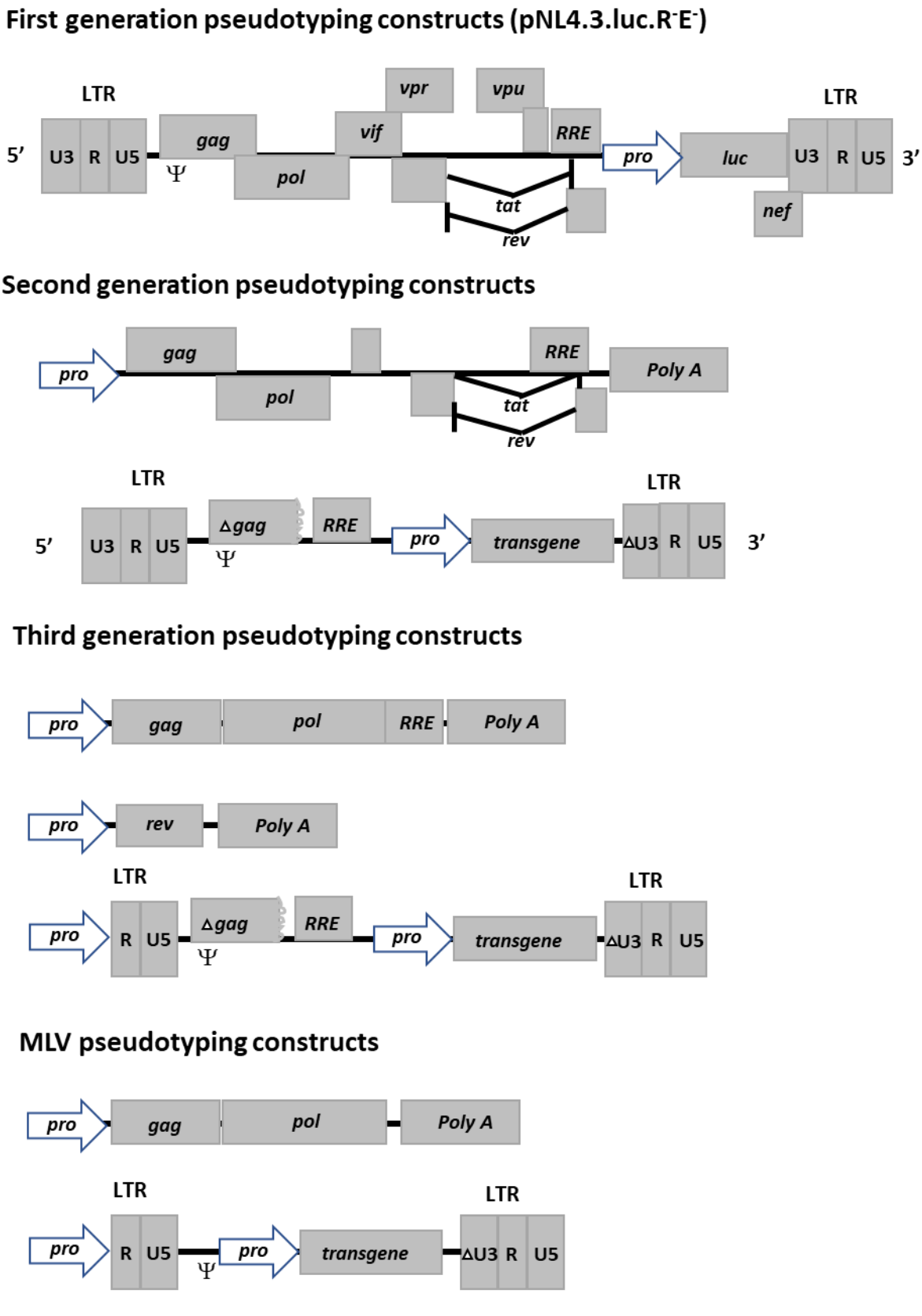

Figure 1.6 Schematic diagram of HIV and MLV pseudotyping systems. (This graph is adapted from (Carnell et al., 2015)).

1.6.2.2 Generation of pseudotyped viruses

To generate PVs, plasmids encoding the retroviral *gag-pol* genes, the envelope gene of choice and the reporter gene are introduced into producer cells such as HEK293T cells (Mather et al., 2013). By hijacking relevant cellular machinery, the imported genes driven by upstream promoters, such as the cytomegalovirus (CMV) promoter, are transcribed and translated. Two copies of RNA harbouring the reporter gene are packaged into the core driven by a packaging signal Ψ. In order to prevent progeny viruses and limit the PVs to a single round infection, Ψ is omitted in the plasmid carrying the *gag-pol* gene and therefore, this gene is not packaged into the core. Thus, pseudotyping is characterised as a safe alternative tool to study highly pathogenic viruses. In a poorly understood mechanism, pseudotype capsid is enveloped by the heterologous glycoproteins in the plasma membrane before budding extracellularly. As the reporter gene is flanked by LTRs, it integrates into the host genome after successful transduction of PVs into the target cells (Figure 1.7). This integration is catalysed by the polymerase/integrase that are co-packaged with the transgene in the PVs. Expression of the reporter gene is then measured, which can be used to quantify the titer or infectivity of the PVs (Johnson, 2011, Mather et al., 2013, Cooray et al., 2012).

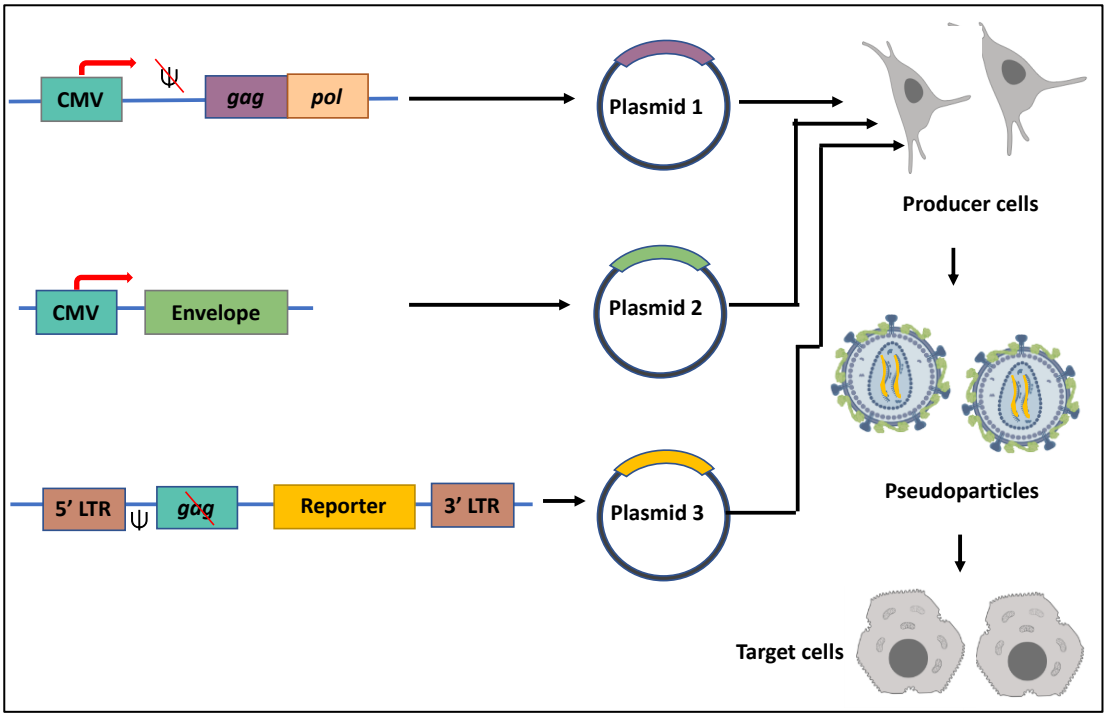

Figure 1.7. Schematic diagram of the production of pseudotyped viruses.

Typically, a three-plasmid system is used to obtain the pseudotyped viruses (PVs). Plasmid 1 bears the retroviral gag-pol, while plasmid 2 bears the envelope protein and plasmid 3 bears a reporter gene that is flanked by 5'- and 3'- long-terminal repeats (LTRs). Three plasmids are introduced into the HEK293T cells, 72 hours later, PVs are harvested and used to transduce target cells. CMV: cytomegalovirus immediate-early promoter.

1.6.2.3 Applications of pseudotyped viruses

The flexibility of construction with heterologous viral envelope proteins supports the use of PVs in a wide range of applications. The technique has been used to investigate antibody responses against viruses, to identify novel viral receptors, to develop vaccines and to deliver therapeutic genes (Tani et al., 2011, King et al., 2016). Particularly, PVs are used as a surrogate for wild type viruses in the study of several high-risk infectious viruses that require high-level containment laboratories. For example, Ebola is a Hazard Group 4 virus and the study of its replication has been severely hampered by safety concerns, due to its extreme pathogenicity. Analysis of infection by using PVs carrying Ebola virus envelope proteins revealed a broad range of host cells susceptible to Ebola infection, whereas, all tested lymphoid cell lines were completely resistant (Wool-Lewis and Bates, 1998). Severe Acute Respiratory Syndrome Coronavirus (SARS-CoV) is another extremely pathogenic virus requiring Containment Level-3 facilities to perform viral cell culture (WHO, 2003). The use of PVs carrying the SARS-CoV envelope protein showed that porcine cells were susceptible to SARS infection suggesting domestic pigs may be a source for interspecies of transmission of SARS (Giroglou et al., 2004). The above examples indicate that pseudotyping can be a powerful tool for characterisation of viral tropism and entry.

Lentiviral HBV pseudotyped viruses (HBV-PVs) have been used in the investigation of the HBV entry pathway (Chai et al., 2007, Meredith et al., 2016). This approach is particularly useful when studying the phenotypic behaviour of mutations in HBsAg ORF. As the HBsAg gene is completely overlapped by the *pol* gene, it is challenging to investigate the phenotype of HBsAg mutations without altering the other in a conventional live HBV system. By only expressing HBsAg in the HBV-PV system, it is more straightforward to investigate mutations, variants or differences between genotypes in this region without having any impact on *pol*.

1.7 Aims

The principal aims of this PhD were to take advantage of pseudotyping techniques to investigate the contribution of different proteins, including L, M and S surface proteins, to HBV entry. To achieve this goal, an in-house genetically engineered cell line over-expressing NTCP was developed. Furthermore, to further understand HBV intra-host evolution and to interrogate genetic variation in the HBsAg-coding region, alongside the main project, the MinION sequencing platform was employed to analyse longitudinal evolution of quasispecies of HBV using clinical samples.

Chapter 2: General Materials & Methods

This chapter describes the general materials and methods that were used throughout this PhD study. Parts of materials and methods were unique to a study are described under the experimental chapters.

2.1 Materials

2.1.1 Mammalian Cell lines and culture conditions

Human Embryonic Kidney 293T (HEK293T) cells expressing Simian Virus 40 (SV40) large T antigen, facilitate efficient transfection with lentiviral/retroviral packaging vectors and helper plasmids. HEK293T derived cells were stably transduced with a plasmid containing an HBsAg expression construct and blasticidin resistance (detailed in Chapter 4).

Huh7 cells ware derived from a human liver hepatocellular carcinoma cell line (Nakabayashi et al., 1984). This cell line is not susceptible to hepatitis B virus; but it is highly susceptible to hepatitis C virus. Huh7-NTCP-eGFP cells were stably transduced with NTCP-eGFP transgene that expressed NTCP and eGFP proteins. Huh7-NTCP-BSD cells were stably transduced with NTCP-BSD gene that expressed NTCP and blasticidin deaminase (BSD).

The HepG2 cell line was derived from a human liver hepatocellular carcinoma. This cell line is also referred as hepatoma cell line (Knowles et al., 1980) or hepatoblastoma cell line (Lopez-Terrada et al., 2009). This cell line is not susceptible to either hepatitis B or C viruses. HepG2-NTCP-BSD cells were stably transduced with NTCP-BSD gene that expressed NTCP and BSD.

All HEK293T, Huh7 cells were cultured in Dulbecco's Modified Eagle's medium (DMEM; Gibco) supplemented with 10% v/v foetal bovine serum (FBS) (Gibco) and 1% v/v Non-essential amino acids (NEAA) (Gibco). HepG2 cells were cultured in Minimum Essential Medium Eagle (MEM; Sigma-Aldrich) supplemented with 10% v/v FBS, 1% v/v NEAA and 1 mM sodium pyruvate (Gibco, ThermoFisher Scientific). All cells were incubated at 37°C in a humidified incubator with 5% CO_2.

For simplicity, DMEM and MEM containing the required concentration of FBS and NEAA are termed "complete medium" in this thesis.

2.1.2 Bacterial cell lines and culture conditions

NEB® 5-alpha competent *E.coli* cells (C2987; New England Biolabs) eliminate nonspecific endonuclease I (endoA1), supporting high quality of plasmid production. C2987 cells were used throughout the Q5 site directed mutagenesis and transformations in this study.

NEB® stable competent *E.coli* cells (C3040; New England Biolabs) have several features such as endoA I deficient; recombination deficient (recA1); methyl restriction deficient and restriction deficient, that are ideal for transformation and isolation for plasmid containing direct repeats and inverted repeats sequences, for example, any clones in retroviral or lentiviral vectors.

Culture medium for bacteria propagation was Luria-Bertani (LB) either in liquid form (1% w/v bacto-tryptone, 0.5% w/v Bacto-yeast extract, 171 mM NaCl, pH 7.0) or semi-solid form for single colony selection. Bacteria were incubated at 37°C (or 30°C for C3040 cells during colony selection stage). Cultures were supplemented with 100

µg.mL^{-1} ampicillin (or 50 µg.mL^{-1} kanamycin where specifically indicated in the method section).

2.1.3 Antibiotics

Ampicillin (100 mg.mL^{-1}) (Sigma) was used to select colonies that bear the ampicillin resistant gene. The working concentration was 100 µg.mL^{-1}.

Blasticidin hydrochloride (10 mg.mL^{-1}) in HEPES buffer (Invivogen) was used to select BSD gene transduced cells. The working concentrations were 20 µg.mL^{-1} for Huh7-NTCP-BSD selection, 30 µg.mL^{-1} for HepG2-NTCP-BSD selection and 20 µg.mL^{-1} for HEK293T derived BSD expressing cells.

The 100x Penicillin-Streptomycin solution (Pen-Strep) (ThermoFisher Scientific) contains 10,000 units/mL penicillin and 10,000 µg.mL^{-1} streptomycin. It was used in infection assay to prevent bacterial contamination of cell culture due to purification of HBV-PVs after ultracentrifugation.

2.1.4 Chemicals

Sodium dodecyl sulphate (SDS), Tween 20, Bovine Saline Albumin (BSA), Ethylenediaminetetraacetic acid disodium salt dihydrate (EDTA), Igepal-CA-630, Trizma base, Methanol, EDTA-free protease inhibitor cocktail tablets, dimethyl sulfoxide (DMSO), Triton X-100, Tris acetate-EDTA (TAE) buffer and Dithiothreitol (DTT) were obtained from Sigma.

Phosphate buffered saline (PBS) tablet (Oxoid). Ethanol, 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-gal), glycine, sucrose were products of Thermo fisher scientific. Polyethylenimine (PEI) was a product of Polysciences.

2.2 Methods

2.2.1 Passage of mammalian Cell lines

Growing cells were passaged at a density of ~1.5 x 10^6 in 15 mL of complete medium every 72-96 hours until 90% confluence was reached. This process involved aspiration of media followed by washing with 6-10 mL of Dulbecco's phosphate buffered saline (PBS) modified without calcium chloride or magnesium chloride (Sigma). Cells were dissociated from the flask by incubating with 3 mL (for T75 flask) or 7 mL (for T175 flask) of trypsin-EDTA (0.05%) (with phenol red; ThermoFisher Scientific) at 37°C (5% CO2); this usually took 2 minutes for HEK293T cells, 7 minutes for HuH7 and 10 minutes for HepG2 cells as well as their derived cell lines, As soon as cells detached from flasks, trypsin-EDTA was inactivated by addition of a double volume of the complete medium. Cells were then diluted in 30-50 mL complete medium. A 10 µL aliquot of the diluted cell suspension was applied onto a glass haemocytometer and cells were counted under a microscope. Total cells were centrifuged and resuspended in 10 mL of complete medium, and then seeded at desired density.

2.2.2 Freezing and Thawing of Human Cells Stores in liquid Nitrogen

Post passaging, cells were resuspended at 4 - 6 x 10^6 cells per mL of freezing media containing 10% dimethyl sulfoxide (DMSO; Hybri-Max ^{TM} sterile-filtered; Sigma)

diluted in FBS in cryovial tubes and placed in a Nalgene™ Freezing Container (Sigma). To achieve a cooling rate at -1°C per minute, the freezing container was stored at -80°C freezer overnight. 24 hours later, cells were then transferred to liquid nitrogen storage at -196°C.

On thawing, cryovial tubes were placed in a 37°C water bath until the last remaining drop of crystal and followed by slow addition of 1 mL of completed medium in dropwise to prevent osmotic damage. Cells were then transferred to a T25 flask with 7 mL of completed medium and incubated at 37°C (5% CO2). Twenty-four hours later, cells were transferred into T75 flask in 15 mL of completed medium.

2.2.3 Transformation of competent bacterial cells

Transformation of competent bacterial cells was carried out according to the protocols set up by the manufacturer. Throughout this PhD work, all plasmids with repeat elements and unstable inserts were transformed into NEB stable competent *E. coli*; and other plasmids were introduced into NEB 5-alpha Competent *E. coli.* In brief, approximately 50 ng of plasmid DNA or ligation reaction mix was added to 25 µL of competent bacterial cells and incubated on ice for 30 minutes. Cells were then heat-shocked at 42°C for 30 seconds followed by incubation on ice for 5 minutes. Competent cells were allowed to recover from the heat shock by incubation in 650 µL SOC medium (provided by the manufacturer) for one hour at a suitable temperature (30°C for NEB stable cells and 37°C for NEB 5-alpha cells) with constant agitation at 225 revolutions per minute (rpm). Finally, approximately 150 µL of the transformed bacterial mixture was spread onto LB-agar plates supplemented with the appropriate antibiotics and incubated overnight at an appropriate temperature depending on the cell type, i.e. 30°C for NEB stable cells and 37°C for NEB 5-alpha cells.

2.2.4 Colony screening

Well-isolated colonies were picked and directly transferred into a PCR mix containing 1x standard Taq buffer (Qiagen), 0.2 mM of each deoxynucleotide triphosphate (dNTPs), 2.5 units HotStarTaq DNA polymerase (Qiagen), 0.2 µM of forward primer and 0.2 µM of reverse primers. Meanwhile, corresponding screened bacterial colonies were further expanded on an LB agar plate supplemented with 100 $\mu g.mL^{-1}$ ampicillin and incubated inverted at 37°C overnight.

2.2.5 The Polymerase Chain Reaction (PCR)

The enzymes used for DNA amplification were either HotStarTaq DNA polymerase (Qiagen) or Q5 Hot Start High-Fidelity DNA polymerase (New England Biolabs). The thermocycling conditions were slightly different between these two enzymes. The profile for HotStarTaq DNA polymerase started an initial denaturation step at 95°C for 15 minutes; proceeding with 25-30 cycles of denaturation at 95°C for 20 seconds, annealing at 55-68°C (depending on primers) for 20 seconds, extension at 72°C for 1 minute per kilobase (kb); and a final extension at 72°C for 5 minutes. The thermocycling condition for Q5 Hot Start High-Fidelity DNA polymerase started with an initial denaturation at 98°C for 30 seconds; proceeding with 25-30 cycles of denaturation at 98°C for 20 seconds, annealing at 55-72°C (depending on primers) for 20 seconds, extension at 72°C for 30 seconds per kb; and a final extension at 72°C for 2 minutes.

2.2.6 DNA gel electrophoresis

A 2% w/v agarose gel was prepared by dissolving 2 g agarose powder in 10 mL of Tris acetate-EDTA (TAE) buffer (10x concentrate, pH 8.3, non-sterile, 0.2 µM filtered;

Sigma-Aldrich) and 90 mL of dH2O and boiling. The solution was then cooled to approximately 60°C and 5 µL of ethidium bromide (EtBr) (Electran®, 10 $mg.mL^{-1}$; VWR International) was added to a final concentration of 0.5 $\mu g.mL^{-1}$ prior to pouring into a sealed casting tray with well-forming combs. Approximately 30 minutes later, solidified upon cooling, the gel was immersed in 1000 mL of 1x TAE buffer with 0.5 $\mu g.mL^{-1}$ EtBr in a gel tank. A 15 µL of sample was mixed with 1.5 µL of 10x FastDigest green buffer (Thermo Scientific), and a 5 µL aliquot was used. In order to determine the molecular weight of PCR amplified DNA samples, a 5 µL aliquot of GeneRuler DNA ladder mix (0.1 µg/µl; Thermo Scientific) was also used . Gels were run at 90 volts (V) for 36-42 minutes, the samples migrating to the "+" electrode. Finally, by using an ultraviolet (UV) transilluminator, DNA bands were visualised at 312 nm wavelength.

2.2.7 DNA sequencing

DNA samples with bands at expected size were commercially sequenced by Source BioSciences (Nottingham, UK). Appropriate sequencing primers were sent along with the samples. A DNA sequence trace file as a result of using PeakTrace™ Basecaller was viewed using FinchTV (Geospiza Inc) and was aligned against known reference based on their DNA sequences using the ClustalW algorithm of the Molecular Evolutionary Genetics Analysis (MEGA) 7 software.

2.2.8 Plasmid purification

For plasmid DNA analysis, a small-scale culture was set up. Bacteria from a single colony were inoculated into 5 mL of LB broth containing appropriate antibiotic and incubated overnight at 37°C. The resulting bacterial stock was processed using GenElute™ Plasmid Miniprep kit (Sigma-Aldrich). As per the manufacturer's

instructions, bacteria were pelleted by centrifugation at 12000 x g and resuspended in 200 µL resuspension solution followed by addition of 200 µL lysis solution and gently mixed. After 5 minutes incubation, lysates were mixed with 350 µL neutralisation solution, and cellular debris was removed by centrifugation followed by transferral of the clear lysate into a binding column. Contaminants were removed by addition of 500 µL optional washing solution (for *EndA*$^+$ strains only) and the flow-through was discarded after centrifugation. The washing process repeated by the addition of 650 µL wash solution. Finally, the binding column was transferred to a fresh collection tube and followed by the addition of 65 µL elution solution. The final clear flow-through was the purified plasmid DNA, which could be stored at 4°C for experimental procedures or stored at -20°C for long term storage.

For larger scale purifications of 50 mL cultures, the GenElute™ Plasmid Miniprep kit was used with a broadly similar protocol.

2.2.9 DNA quantification

DNA was quantified using a NanoDrop® spectrophotometer (Thermo Scientific) and ND-1000 v3.7.1 software. Since RNA, DNA including dsDNA and ssDNA are all absorbed at a wavelength of 260 nm, a 260 / 280 ratio was used to determine the purity of the samples. A 260/280 ratio of approximately 1.8 was generally considered pure for DNA. Moreover, carbohydrates and chemicals used during the DNA purification such as phenol, EDTA were absorbed at 230, therefore, a ratio of 260 / 230 was used to determine the purity of the sample. A ration in the range of “2.0 – 2.2” was considered pure; whereas if the figure fell below, the sample was considered impure.

2.2.10 RNA extraction

7.0×10^6 to 9.0×10^6 cells were pelleted at 300 x g for 5 minutes and the total RNA was extracted using the GenElute Mammalian Total RNA Miniprep Kit (Sigma-Aldrich). As per the manufacturer's instructions, pelleted cells were thoroughly mixed with 500 µL lysis solution/2-mercaptoethanol followed by transferral of the lysate to a filtration column. After centrifugation, 500 µL 70% ethanol was added into the filtered lysate and mixed well. The mixture was transferred to a binding column followed by centrifugation and the resulting flow-through was discarded. Contaminants were removed by the addition of 500 µL wash solution and the binding column was transferred to a new collection tube after centrifugation. The washing process was then repeated twice with the addition of 500 µL wash solution. Finally, the binding column was transferred to another new collection tube followed by addition of 30 µL elution solution. After final centrifugation, the clear flow-through was the extracted RNA, which was collected and stored at -80°C for long term storage. RNA concentration was analysed by Nanodrop spectrophotometer.

2.2.11 One-Step qPCR

One-Step qPCR was performed using Luna® Universal One-Step RT-qPCR kit (New England Biolabs). Following the manufacturer's instruction, a reaction mix was set up containing 10 µL Luna Universal One-Step reaction 2x mix, 1 µL Luna WarmStart RT enzyme 10x mix, 0.4 µM of each forward and reverse primers, 1 µL RNA template and nuclease-free water to a total volume of 20 µL. Each reaction was prepared in triplets. The qPCR was performed using the BioRad CFX96 (Bio-Rad) real-time instrument. The thermocycling started with reverse transcription at 55°C for 10 minutes followed by initial denaturation at 95°C for 1 minute and proceeded with 45 cycles of denaturation at 95°C for 10 seconds, extension at 60°C for 30 seconds and

finally melt curve recorded between 60-95°C. The plate was read every 30 seconds during the extension period. The results were analysed using CFX Maestro™.

2.2.12 In-Fusion® HD cloning (TaKaRa) with Cloning Enhancer Treatment

Generation of linearised vector is described in Chapter 4 and generation of specific genes of interest to be ligated into the vector are described in Chapter 3, 4 and 5. The cloning was processed using In-Fusion HD Cloning Kit. As per manufacturer's instruction, 2.5 µL of the PCR reaction was purified by addition of 1 µL of Cloning Enhancer. The mixture was incubated at 37°C for 15 minutes then at 80°C for a further 15 minutes in a PCR thermal cycler. The ligation process was carried out by mixing 0.5 µL 5x In-Fusion HD enzyme premix, 0.5 µL linearised vector, 0.5 µL PCR reaction (the gene of interest), 1 µL dH_2O and incubating at 50 °C for 15 minutes followed by cooling on ice. The resulting product (1 µL) was introduced into competent bacterial cells as described in section 2.2.3 Transformation of competent bacterial cells.

2.2.13 Lysis of mammalian cells

Cells were washed three times in PBS before addition of Lysis buffer (50 mM Tris.HCL pH 7.4, 1mM EDTA, 1% Igepal-CA-630, 150 mM NaCl) and cOmplete™ Mini, EDTA-free protease inhibitor cocktail (Sigma). After incubation with agitation on ice for 15 minutes, lysates were centrifuged at 12,000 x g for 10 minutes and the resulting protein-containing supernatants were harvested and stored at -20°C.

2.2.14 Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE)

Protein samples in equal volume of 2x Laemmli buffer (10% v/v glycerol, 4% w/v sodium dodecyl sulphate (SDS), 0.01% w/v bromophenol blue, 125 mM Tris HCL pH 6.8 and 20 mM dithiothreitol (DTT)) were denatured at 95°C for 5 minutes prior to separation by a 12% Mini-Protean TGX™ precast gel (Bio-Rad). Multicolour broad range protein ladder (Thermo Fisher Scientific) was run alongside the samples to allow determination of protein size. The gel was run in 1x running buffer (25 mM Tris base, 192 mM glycine, 0.1% (w/v) SDS) at 90 volt (v) for 10 minutes then 180 v for 40 minutes.

2.2.15 Western blot

Gels were transferred to methanol (100% v/v) activated polyvinylidene difluoride (PVDF) membrane (Amersham, GE Healthcare) using a Transblot® SD semi-dry transfer cell (Bio-Rad) in transfer buffer (25 mM Tris-HCL pH 8.3, 192 mM glycine, 20% v/v methanol). The transfer was run at 100 mA for 1 hour. Transferred membrane was blocked in 10% w/v dried, skimmed milk powder (Marvel) in 1x PBS 0.1% v/v Tween 20 (PBST) with constant agitation for 1 hour at room temperature. The membrane was then incubated in primary antibody (specific primary antibodies were detailed in experimental chapters) in 5% milk / PBST at 4°C overnight. The unbound primary antibodies were removed by washing three times for 5 minutes in PBST before incubation in HRP-conjugated secondary antibody (specific secondary antibodies were detailed in experimental chapters) for 1 hour. The unbound secondary antibodies were then removed by washing three times for 5 minutes each in PBST. Finally, the membrane was immersed in 600 µL of detection substrate (1:1 ratio of radiance peroxide and radiance substrate; Azure Biosystems Radiane HRP)

for 1 minute prior to being placed into the dark chamber for varying exposures in a G box (SYNGENE).

Chapter 3: Construction of sodium taurocholate co-transporting polypeptide expressing cell lines

3.1 Introduction

Sodium taurocholate cotransporting polypeptide (NTCP) is encoded by a gene known as solute carrier family 10A1 (SLC10A1). It belongs to the SLC10A transporter family that consists of 7 members (SLC10A1-7). Among them, NTCP and SLC10A2, known as apical sodium-dependent bile salt transporter (ASBT), are both transporters for bile acids in a sodium-dependent fashion (Watashi et al., 2014). Moreover, NTCP also transports other molecules such as steroid hormones, thyroid hormones and drugs (Doring et al., 2012).

3.1.1 NTCP topology

NTCP is a multiple transmembrane protein consisting of 349 amino acids and mainly expressed at the basolateral membrane of hepatocytes (Hagenbuch and Meier, 1994) (Figure 3.1). It comprises an extracellular N-terminal domain and a cytoplasmic C-terminus resulting in an odd number of transmembrane (TM) domains (Hagenbuch and Meier, 1994, Hallen et al., 2002). Yet, whether it contains seven or nine membrane spanning regions is still yet to be determined. The nine transmembrane segment model was supported by using an alanine insertion scanning method. It was hypothesised that the insertion of alanine residues into the middle of a transmembrane α helix would disrupt structural associations with neighbouring membrane segments, in turn changing the position of the adjacent α helix.

Consequently, the function of the transporter would be inhibited. Mutants with alanine insertion into the different hydrophobic regions of NTCP were expressed in HEK293T cells. As predicted, the transporter activity was inhibited as a result of alanine insertion into any of the nine transmembrane segments (Hallen et al., 2002). This interpretation was refuted by a more recent study where a 7 TM model was proposed. In that experiment, a full-length C-terminal yellow fluorescent protein (YFP) was fused to NTCP and expressed in HEK-293T cells. By limited tryptic digestion, exoplasmic-accessible lysines or arginines were determined (Figure 3.1). This analysis obtained a model where the first six and the ninth hydrophobic regions were membrane-inserted, while the 7^{th} and 8^{th} segments were exoplasmic as indicated in the cartoon models in Figure 3.1 (Mareninova et al., 2005).

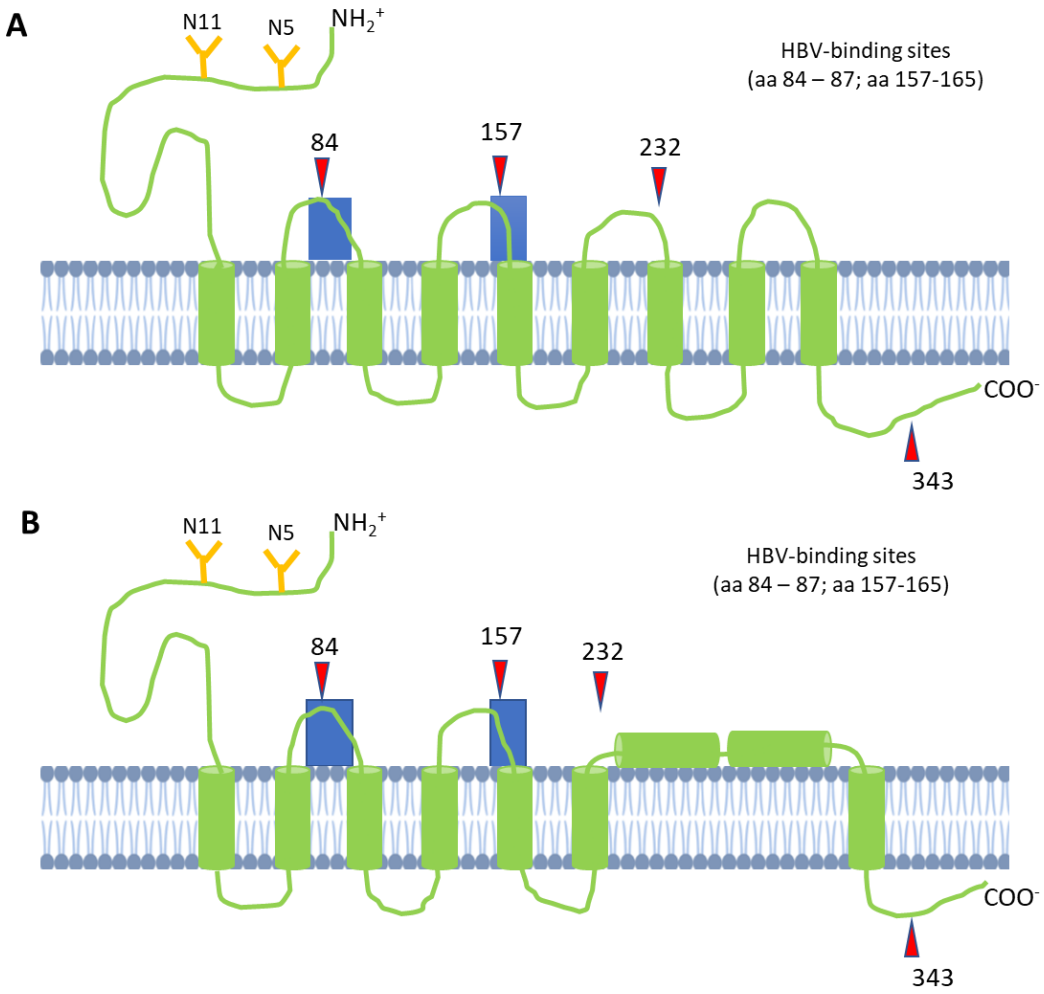

Figure 3.1 Cartoon representation of a sodium taurocholate cotransporting polypeptide (NTCP)

*The protein has an odd number of transmembrane segments with the N-terminus on the extracellular surface and C-terminus on the intracellular side. N5 and N11 are two glycosylation sites. Two HBV-binding sites are highlighted in blue. (**A**) A predicted nine transmembrane domain topology with nine hydrophobic segments membrane-inserted; **(B).** A predicted seven transmembrane segment model with the first six and the ninth hydrophobic segments membrane-inserted; the 7th and the 8th hydrophobic segments exoplasmic (Jacquet et al., 2019, Mareninova et al., 2005).*

3.1.2 NTCP N-glycosylation

NTCP contains two N-glycosylated sites at residues 5 and 11. It was reported that replacing asparagine (N) with glutamine (Q) at either residue 5 or 11 resulted in modest reduction of NTCP localisation at the plasma membrane, bile acid uptake and subsequent HBV infection compared to NTCP-wild type (WT); however, a double mutant (N5,11Q) significantly reduced NTCP plasma membrane localisation and failed to support bile acid uptake as well as HBV infection (Appelman et al., 2017). Glycosylation is essential for NTCP plasma membrane localisation and for HBV docking and subsequent entry (Appelman et al., 2017). Treatment with MG132, a proteasome inhibitor, did not enhance NTCP expression or localisation at the plasma membrane in HepG2 cells expressing NTCP-N5,11Q. However, treatment with bafilomycin, a potent inhibitor of autophagy, is able to block the degradation of NTCP-N5,11Q in the lysosomes. This experiment provided evidence that non-glycosylated NTCP is internalised and degraded in lysosomes (Appelman et al., 2017). However, a more recent study found that the introduction of N to A or N to Q mutations at residues N5 and/or N11 did not alter cell surface availability of NTCP or its plasma membrane localisation (Lee et al., 2018). Five different combination NTCP mutants, as well as the NTCP-WT were all able to express NTCP on the cell membrane; and all supported HBV infection; though the double mutants, especially N5/11Q, expressed lower levels of NTCP than the NTCP-WT. The endogenous NTCP expressed by differentiated HepaRG cells appeared non-glycosylated and supported both serum-derived HBV as well as cell culture-derived HBV; whereas, exogenously overexpressed NTCP in HepG2 cells supported cell culture-derived HBV but the cells were poorly infected by serum-derived HBV, which implies host factors other than NTCP play a role in infection efficiency by serum-derived HBV (Lee et al., 2018, Li et al., 2016).

NTCP glycosylation is modified in an age-dependent manner. Western blotting quantification analysis showed that total quantity of NTCP especially 55kDa NTCP was significantly less in liver tissue from neonates less than one year old compared with liver tissue from a 3-year-old infant. Furthermore, immunofluorescence (IF) microscopy was used to demonstrate that NTCP expression was relatively low and predominantly located in the cytoplasm in liver tissues from individuals below 1 year old; but expression was increased with patient age and localised predominantly to the cell membrane (Sargiacomo et al., 2018).

3.1.3 NTCP variants

To date, 147 NTCP variants have been reported including 38 single nucleotide polymorphisms (SNPs), among which, 10 variants are located in the coding regions and eight of them are non-synonymous (Yang et al., 2016). The most studied SNP is rs2296651, a substitution of Serine (Ser) by phenylalanine (Phe) at residue 267 (S267F). This variant is found exclusively among Asians with the highest occurrence in East Asian people (Pan et al., 2011). The first study that investigated the association between the genotype and HBV infection in a Chinese Han cohort found that the S267F variant was linked to increased risk of HBV infection. However, further two separate studies with much larger sample sizes (1899 and 3801 CHB patients respectively) in similar ethnic backgrounds, found that S267F was associated with resistance to CHB and with decreased incidence of cirrhosis and HCC (Hu et al., 2016, Peng et al., 2015). Although the S267F variant gives rise to normal expression of NTCP, the resulting receptor has almost lost its function at transporting bile acid. Its ability in binding pre-S1 peptide was also diminished, resulting in a defect in supporting HBV infection (Ho et al., 2004, Yan et al., 2014). Note this amino acid is not located within the critical region of aa 157 to 165 of the NTCP, which suggests this amino acid could be also conformationally or functionally important to viral entry.

3.1.4 Aims

The aims of this study were to develop an *in vitro* cell system and to use pseudotyped viruses to study the entry mechanism of HBV as well as assessing the potency of novel entry inhibitors.

3.2 Materials and Methods

3.2.1 Primers used

The primers shown in Table 3.1 were designed in the lab through a joint effort by Dr Alexander Tarr, Dr Patrick McClure and myself for the purposes of construction and screening of NTCP expressing cell lines.

Table 3.1 Primers used in this experiment

Name	Sequence (5' - 3')	Use
hu.NTCP.f	CACCATGGAGGCCCACAACGCG	To amplify Human NTCP (full length)
hu.NTCP.r	CTAGGCTGTGCAAGGGGAGCAGT	
pWPI-NTCP-BSD.f	CAGGCGCGCCGGATCCACCATGGAGGCCCACAACGCG	To amplify full length human NTCP with 15bp extensions at 5' end that were complementary to The end of pWPI-BSD vector
pWPI-NTCP-BSD.r	GATCCCCCGGGGCGACTAGGCTGTGCAAGGGGAGCAGT	
pWPI.scr.f	GGTACAGTGCAGGGGAAAGA	To select colonies that expressed plasmid harbouring pWPI-BSD vector
pWPI.scr.r	CACACCGGCCTTATTCCAAG	
pHIV-NTCP-eGFPf	ACCCGGGCTAGGATCCACCATGG AGGCCCACAACGCG	To amplify full length human NTCP with 15bp extensions at 5' end that were complementary to The end of pHIV-eGFP vector
pHIV-NTCP-eGFPr	GGAGAGGGGCGGATCCTAGGCTG TGCAAGGGGAGCAGT	
pHIV.scr.f	TGGAATTTGCCCTTTTTGAG	To select colonies that expressed plasmid harbouring pHIV-eGFP vector
pHIV.scr.r	AGGAACTGCTTCCTTCACGA	
qNTCP.f	CCGGCTGAAGAACATTGAGG	To amplify a small amplicon for quantification of NTCP DNA expression
qNTCP.r	CGATCCCTATGGTGCAAGGA	
Luc.scr.f	TGTAGCCATCCATCCTTGTCA	To amplify a small amplicon of luciferase gene
Luc.scr.r	CGGGCGCGGTCGGTAAA	

3.2.2 Generation of stable NTCP cell lines

3.2.2.1 Development of HuH7-NTCP-BSD cell line

Liver RNA was a kind donation from the Centre for Human Virology of the University of Birmingham, UK (Hedegaard et al., 2017). EcoDry Premix with Random hexamer Primers (TakaRa Clontech) was used to reverse transcribe RNA to cDNA. The resulting cDNA was amplified by polymerase chain reaction (PCR) using full length NTCP amplification primers with pWPI-BSD tag. The reaction mixture contained 1 µL cDNA, 12.5 µL 2x CloneAmp HiFi PCR premix, 0.2 µM forward primer, 0.2 µM reverse primer and 12.5 µL nuclease free water. The PCR involved 45 cycles of denaturation at 98°C for 10 seconds, annealing at 60°C for 15 seconds, extension at 72°C for 2 minutes; and one cycle of a final extension at 72°C for 2 minutes. The PCR products were ligated into the pWPI-BSD vector (a gift of Richard Brown, Twincore, Hannover) using In-Fusion® HD cloning kit (TaKaRa) following the manufacturer's protocol (Figure 3.2). In brief, 1 µL of the PCR fragment was first purified by mixing with 0.5 µL cloning enhancer and incubated at 37°C for 15 minutes then at 80°C for 15 minutes. 0.5 µL of the purified PCR fragment was then incubated with 0.5 µL In-Fusion HD enzyme premix, 0.5 µL BamHI digested pWPI-BSD vector and 1 µL nuclease free water at 50°C for 15 minutes. The reaction was terminated by incubating on ice. A diagram of the plasmid is shown in Figure 3.3.

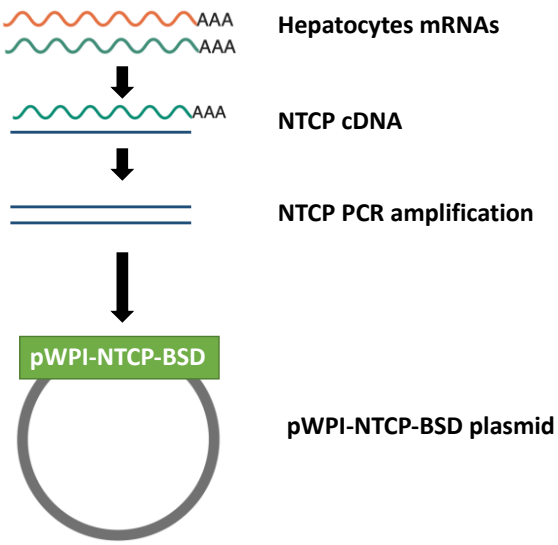

Figure 3.2 Flow chart of production of pWPI-NTCP-BSD plasmid.

NTCP mRNA was reverse-transcribed into complementary DNA (cDNA) using random hexamer primers (TakaRa Clontech). The liver RNA was a gift from the Centre for Human Virology of the University of Birmingham, UK (Hedegaard et al., 2017). PCR amplicons were cloned into vectors upstream of Blasticidin S Deaminase (BSD) gene.

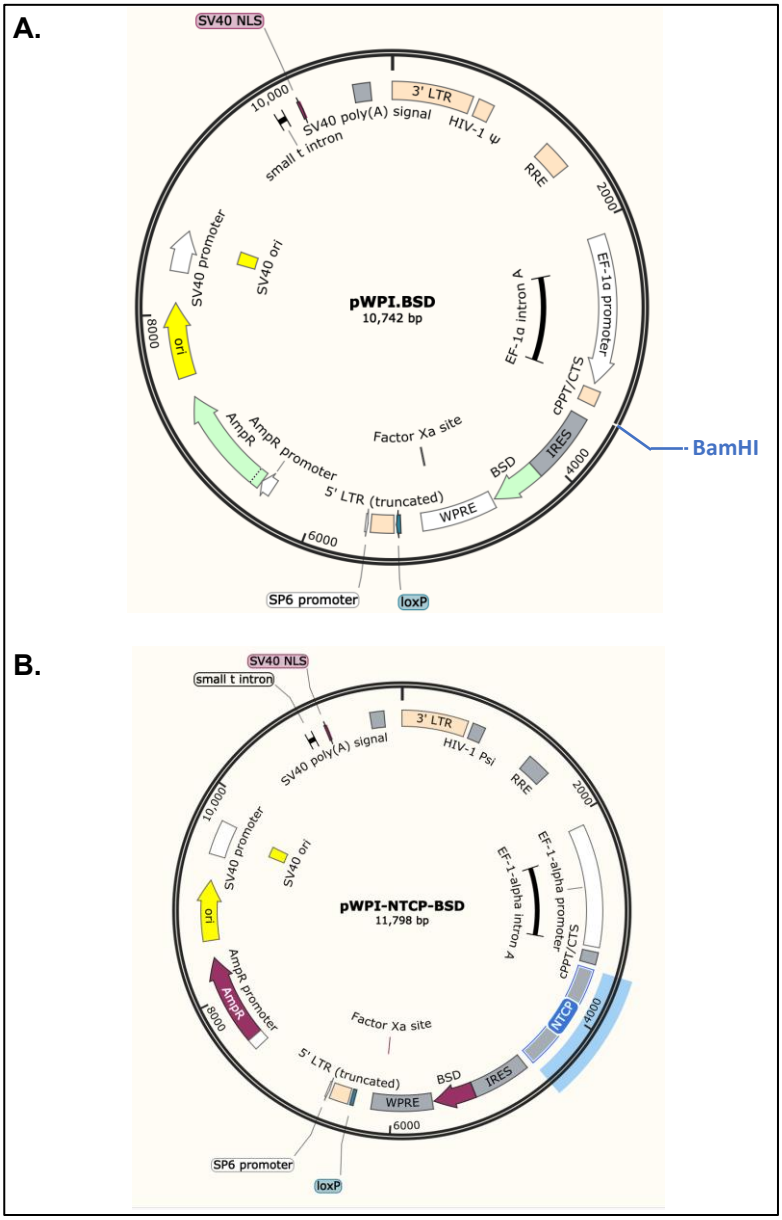

Figure 3.3 Plasmid map of pWPI-BSD and pWPI-NTCP-BSD.

(A). *The map shows the unique restriction site BamHI (coloured in blue);* ***(B).*** *the NTCP DNA was inserted into BamHI restriction site. The NTCP expression was under EF-1a promoter which gives consistent expression regardless of cell type or physiology.*

The resulting chimeric plasmid was transformed into competent cells. Harvested plasmid was sent to Source Bioscience for sequencing. Upon confirmation of NTCP sequence, a plasmid mix containing pWPI-NTCP-BSD, psPAX2 (coding for HIV packaging proteins) and pMD2-G (coding vesicular stomatitis virus (VSV)-glycoproteins) were delivered into HEK293T cells in a 10 cm^2 plate (cell concentration 1.2 x 10^6 / dish). Seventy-two hours later, the supernatant was removed from the plate and filtered through a 0.45 µm syringe filter. It was then used to transduce HuH7 and HepG2 cell lines, independently (Figure 3.4). Post two passages, cells were treated with blasticidin (20 $\mu g.mL^{-1}$ for HuH7 and 30 $\mu g.mL^{-1}$ for HepG2 cells). After selection and expansion, aliquots of the blasticidin resistant cells were stored in liquid nitrogen in 90% fetal bovine serum (FBS) in the presence of 10% dimethylsulfoxide (DMSO).

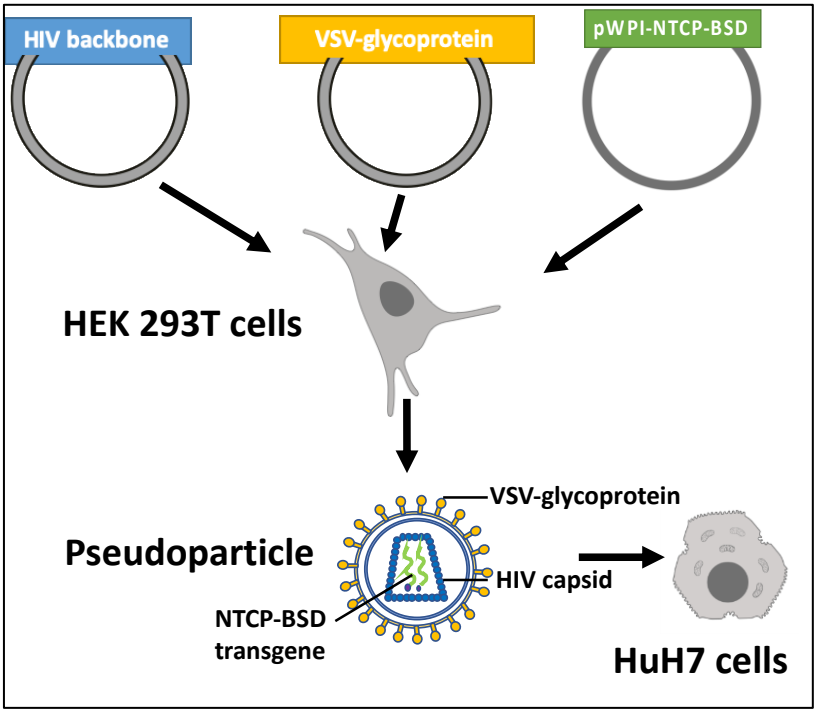

Figure 3.4 Development of HuH7-NTCP-BSD cell line

Pseudotyped viruses (PVs) bearing the vesicular stomatitis virus (VSV) glycoprotein were generated using a three-plasmid system including HIV backbone, VSV-glycoprotein and pWPI-NTCP-BSD. Seventy-two hours after transfecting these plasmids into HEK293T cells, supernatants containing PVs were used to infect target cells, e.g. HuH7 cells.

3.2.2.2 Development of pHIV-NTCP-eGFP cell line

NTCP cDNA was harvested as detailed in section 3.2.2.1 Development of HuH7-NTCP-BSD cell line. Resulting 1 µL of the cDNA template was mixed with 12.5 µL 2x CloneAmp HiFi PCR premix, 0.2 µM of each of the pHIV-NTCP.GFP forward and reverse primers and 12.5 µL nuclease free water to a total of 25 µL reaction mixture. The thermocycling started with 45 cycles of denaturation at 98°C for 10 seconds, annealing at 75°C for 15 seconds, extension at 72°C for 2 minutes; and a final extension at 72°C for 2 minutes. The resulting amplicon was ligated into BamHI digested pHIV-GFP vector using In-Fusion® HD cloning kit (TaKaRa). Finally, the recombinant pHIV-NTCP.EGFP plasmids (Figure 3.5) were taken up by competent NEB® stable competent *E.Coli* cells (NEB) as described above.

Harvested Plasmid was sent to Source Bioscience for sequencing. Upon confirmation of NTCP sequence, a plasmid mix containing pHIV-NTCP-EGFP, pCMVR8.74 (coding for HIV packaging proteins) and pMD2.G was delivered into HEK293T cells in a 10 cm^2 plate (cell concentration 1.2 x 10^6 per dish). Seventy-two hours later, the supernatant was removed from the plate and filtered through a 0.45 µm syringe filter. It was then used to transduce HuH7 and HEK293T cells. Post two passages, cells were subjected to flow cytometry live sorting (sorting was performed by Dr Richard Urbanowicz). After selection and expansion, aliquots of the eGFP positive cells were stored in liquid nitrogen in fetal bovine serum (FBS) in the presence of 10% DMSO.

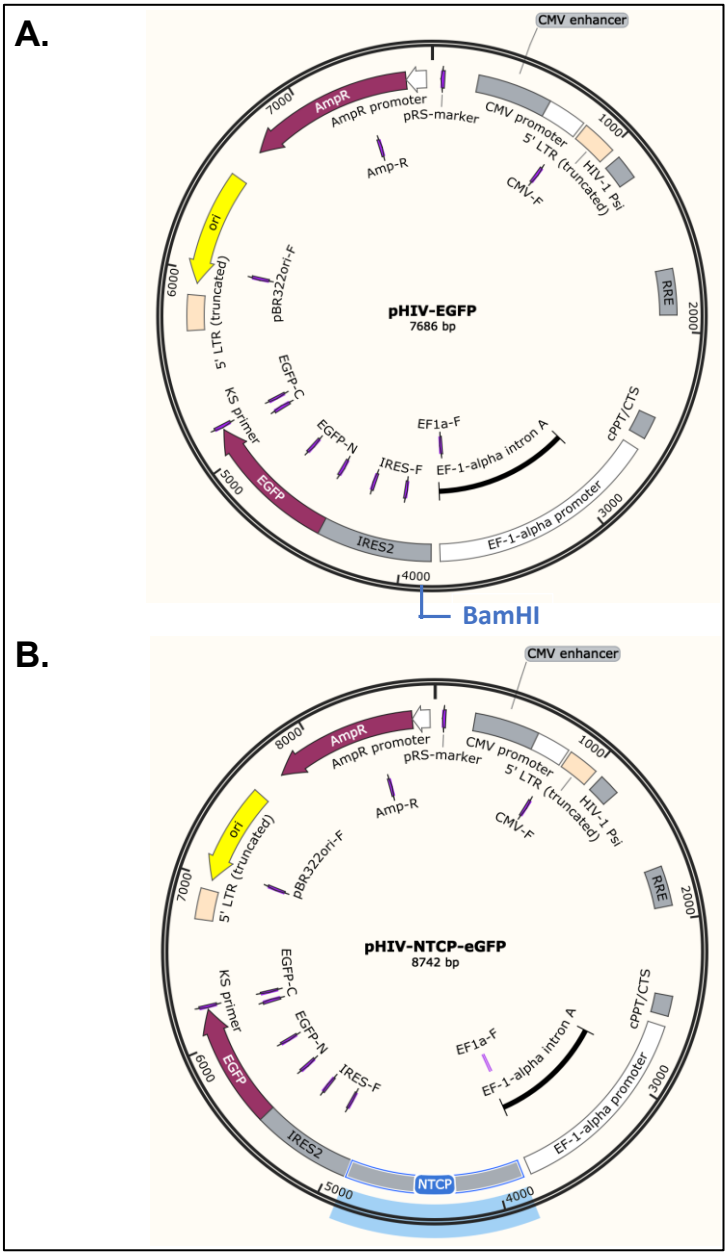

Figure 3.5 Plasmid map of pHIV-eGFP and pHIV-NTCP-GFP (A). The map *shows the unique restriction site of BamHI (coloured in blue) in plasmid pHIV-eGFP;* ***(B).*** *the NTCP DNA was inserted into the BamHI restriction site. The NTCP expression was under EF-1a promoter which gives consistent expression regardless of cell type or physiology.*

3.2.3 Indirect Immunofluorescence (IF) assay

Five hundred µL of Huh7 ± NTCP or HepG2 ± NTCP cells at a density of $1x10^5$ per well were seeded directly onto coverslips in a 12 well flat bottom cell culture plate (Costar). Twenty-four hours later (aiming for 80%-90% cell density), the cells were fixed with 100 µL 4% formaldehyde for 15 minutes. If permeabilisation was required, cells were incubated with 100 µL 0.1% Triton X-100 for 15 minutes, otherwise, this step was omitted and proceeded by blocking with 5% Bovine Serum Albumin (BSA) (Sigma) for one hour. For detecting NTCP, anti-SLC10A1 Rabbit polyclonal antibody either TA322912 (OriGene) at a concentration of $4\mu g.mL^{-1}$ or HPA042727 (Merck) at a concentration of $0.4\ \mu g.mL^{-1}$ diluted in 1% phosphate buffered saline containing 1% BSA (1% PBA/BSA) was added for 2 hours at room temperature (RT). For simplicity, TA322912 (OriGene) is termed anti-NTCP (Ori) and HPA042727 (Merck) is termed anti-NTCP (Mer). After washing, cells were incubated with Alexa Fluor® 594 goat anti-rabbit IgG (ThermoFisher Scientific) at a concentration of $4\ \mu g.mL^{-1}$ diluted in 1% PBA/BSA for one hour. Following washing, cells were stained with $0.04\mu g.mL^{-1}$ DAPI (4',6-Diamidino-2-Phenylindole Dihydrochloride; Thermo Fisher Scientific) for 5 minutes. Finally, a small drop of Dakocytomation Fluorescent Mounting Media (Agilent Dako) was added onto microscope slides and the coverslips were mounted on top of the medium with cell side down with gentle pressure and then dried at room temperature. Cells were visualised using a Zeiss LSM710 confocal microscope.

3.2.4 Fluorescence activated cell sorting (FACS) of live cells

3.2.4.1 HuH7-NTCP-BSD cells and HepG2 ± NTCP cells

HuH7 ± NTCP and HepG2 ± NTCP cells (2.0×10^5 cells in 5 mL media) were centrifuged at 300 x g for 7 minutes. After carefully removing the supernatant, 500 µL 1% BSA was added into each tube, followed by centrifugation at 300 x g for 7 minutes.

Resulting supernatant was removed and 5 µL of 4 µg.mL^{-1} anti-NTCP (Ori) was added for one hour followed by washing by addition of 1 mL 1% BSA and centrifugation at 300 x g for 5 minutes. This step was repeated twice to remove non-specific binding. A 5 µL of 4 µg.mL^{-1} Alexa Fluor® 594 goat anti-rabbit IgG (ThermoFisher Scientific) was added for one hour in the dark and the cells were then washed with 1% BSA twice. Finally, 500 µL 1% BSA was added into each tube and the cell were ready for FACS.

3.2.4.2 HuH7-NTCP-eGFP and HEK293T-NTCP-eGFP cells

HuH7 ± NTCP-eGFP and HEK293T ± NTCP-eGFP cells at density of 2.0x10^5 in 5 mL media was centrifuged at 300 x g for 7 minutes. The resulting supernatant was removed followed by addition of 500 µL 1% BSA. After centrifugation at 300 x g for 7 minutes, cells were resuspended in 500 µL1% BSA and were ready for FACS. For convenience, HuH7-NTCP-eGFP and HEK293T-NTCP-eGFP cells will be shortened as HuH7-NTCP and HEK293T-NTCP.

3.2.5 Western blot

Western blot analysis was detailed in Chapter 2 Materials and Methods. Total proteins were quantified using the Pierce™ BCA Protein Assay Kit (Thermo Fisher Scientific) as per the manufacturer's instruction. Ten to 20 µg of proteins each sample from cell lysates of either HuH7 ± NTCP or HepG2 ± NTCP were separated by running on SDS-PAGE and blotted with antibodies. Two primary antibodies were tested in this study including 4 µg.mL^{-1} anti-NTCP (Ori) and 0.4 µg.mL^{-1} anti-NTCP (Mer). The secondary antibody used was goat anti-rabbit HRP (Dako, 1:1000). All membranes were re-probed with a goat anti-GAPDH polyclonal antibody (AF5718, R&D system, 1:5000) and followed by rabbit anti-Goat IgG HRP (A8919, Sigma, 1:5000).

3.2.6 Endpoint reverse-transcription PCR (RT-PCR)

RNA extraction was performed as described in Chapter 2. EcoDry Premix with Random hexamer Primers (TakaRa Clontech) was used to reverse transcribe RNA to cDNA. The concentration of cDNA solution was titrated by 10-fold serial dilutions. These were then added into a PCR mixture including 1x standard Taq buffer (Qiagen), 0.2 mM of each deoxynucleotide triphosphate (dNTPs), 2.5 units HotStarTaq DNA polymerase (Qiagen), 0.2 µM qNTCP forward primer (5'-CCGGCTGAAGAACATTGAGG-3') and 0.2 µM qNTCP reverse primer (5'-CGATCCCTATGGTGCAAGGA-3'). The PCR reaction was initiated at 95°C for 15 minutes; 55 cycles of [denaturation at 95°C for 20 seconds, annealing at 55°C for 20 seconds, and extension at 72°C for 3 minutes]; and a final extension at 72°C for 1 minute. The PCR products were separated by electrophoresis in a 2% agarose gel at 90 v for 42 minutes.

3.2.7 NTCP quantification

As detailed in Chapter 2 section One-Step qPCR, NTCP was quantified using Luna® Universal One-Step RT-qPCR kit (NEB) and qNTCP forward and reverse primers. Each reaction was prepared in triplicates including HuH7, HuH7-NTCP non-DNase treated and HuH7-NTCP DNase treated samples. Meanwhile, 10-fold dilutions of known copies of pWPI-NTCP-BSD plasmid were reconstituted with nuclease-free water to construct a standard curve containing 10^2-10^8 copies of templates. The standards and samples were assayed in the same run.

3.2.8 Enzymatic de-glycosylation

Cell lysates of HuH7-NTCP, HepG2-NTCP and HEK293T-NTCP were treated with peptide-*N*-glycosidase F (PNGase F, NEB) according to the manufacturer's protocol. In brief, 20 µg of total protein from each lysates were first incubated with 1 µL of denaturing buffer (10x) at 100°C for 10 minutes and followed by incubation with 2 µL of NP40, 2 µL of GlycoBuffer 2 (10x), 5 µL of dH2O, 1 µL of PNGase at either room temperature overnight, 37°C for 4 hours or 37°C overnight. Negative controls were prepared in parallel without the enzyme. The resulting denatured products were ready for western blotting.

3.2.9 Bile acid uptake assay

Plasmid pCMV.NucleoBAS (Addgene plasmid # 62861) was a gift from Stan van de Graaf (van der Velden et al., 2013). The plasmid transiently transfected HEK293T and HEK293T-NTCP cells at a density of 1 µg plasmid per 200,000 cells. On the day of transfection, 1 µg of plasmid was incubated with 1 $mg.mL^{-1}$ polyethyleneimine (PEI) at a 1:4 ratio for 30 minutes before the mixture was added to the cells in a six-well clear bottom plate. Seventy-two hours later, fluorescent intensity was measured using BMG Labtech FLUOstar Omega. Excitation was set at 405nM and spectra range for emission detection was set at 495/10 nm and 540/25 nm for cerulean and citrine respectively. After baseline measurement, citrine/cerulean emission was measured every two minutes upon addition of 30 µM sodium taurocholate hydrate (Molecular Weight (Mr): 537.68; Sigma). Final measurements were taken after addition of 5 µM GW4064 (Mr: 542.84; Sigma).

3.2.10 Virus-receptor binding assay

Huh7-NTCP cells were seeded in 96 well clear flat bottom plate (Corning) at a concentration of 1.5 x 10^4 per well. Cells were cultivated in complete DMEM at 37°C in a humidified 5% CO_2 incubator overnight. Prior to the binding assay, cells were fixed with 4% formaldehyde for 10 minutes at room temperature followed by washing in 1% PBS. The binding assay was carried out by incubation of HBV containing serum diluted in DMEM (Sigma) at increasing concentration from 200 units to 1000 units of HBV DNA per cell for 1 hour at room temperature in a class I safety cabinet. Each concentration was performed in duplicate. Unbound virions were removed by three washes with 1% PBS. Next, 50uL of ice-cold methanol was added to cells for 10 minutes at -20°C. The methanol fixation was stopped by aspiration of the methanol followed by washing in 1% PBS. After blocking in 1% BSA (diluted in 1% PBS) for 1 hour, cells were incubated with Human Hepatitis B Immunoglobulin (HBIG, kindly donated by Samreen Liaz, Public Health England) (diluted in 0.5% BSA at 1:100) for 2 hours at room temperature. After washing in PBS, cells were incubated with goat anti-human IgG conjugated to alkaline phosphate (Sigma, diluted in 0.5% BSA at 1:2000) for 1 hour at room temperature. After a final wash in 1% Tris buffered saline (TBS; Sigma), cells were incubated in p-Nitrophenyl phosphate (pNPP; Sigma) substrate solution for 20 minutes and viral-cell binding was quantified by measuring the absorbance at a wavelength of 405 nm.

The HBV positive sera (viral load, 2.9 x 10^8 IU/mL) were obtained from Nottingham University Hospital NHS Trust, Nottingham, UK and held under ethical approval of the Nottingham Health Science Biobank. The negative control was known with undetectable HBV DNA but tested HBsAg positive. It was obtained from the same Biobank.

The choice of the negative control was based on the assumption that (i) HBsAg positive DNA negative sample only carried spherical or filamentous subviral particles (SVPs) and absence of full virus particles (known as Dane particles); (ii) the binding of SVPs to NTCP would be minimum. HBV DNA sera contain Dane particles as well as SVPs, therefore, HBsAg positive DNA negative sample would give a measurement of background binding to HuH7 cells through other unknown receptors that independent of NTCP, which would be the same in either DNA negative or positive samples, in which case, any additional binding would be a contribution of the virus-NTCP binding.

3.2.11 Statistical analysis

Statistical analyses were conducted using a One-way ANOVA in Prism 8. $P>0.01$ was considered statistically insignificant and indicated as 'ns'; while $P<0.01$ was deemed statistically significant and the adjusted P values are indicated in the individual figures.

3.3 Results

3.3.1 HuH7 cells lack of NTCP expression

Previous studies have demonstrated HuH7 cells to be lakcing expresson of NTCP. To determine whether NTCP mRNA expression was expressed in HuH7 cells before they were used to develop a NTCP overexpressing cell line, end-point RT-PCR was performed. During end-point RT-PCR, serial 10-fold dilutions of cDNA were made. The resulting PCR products were analysed by 2% gel electrophoresis. As shown in Figure 3.6, product of expected size (274bp) was undetectable after 1:10 dilution.

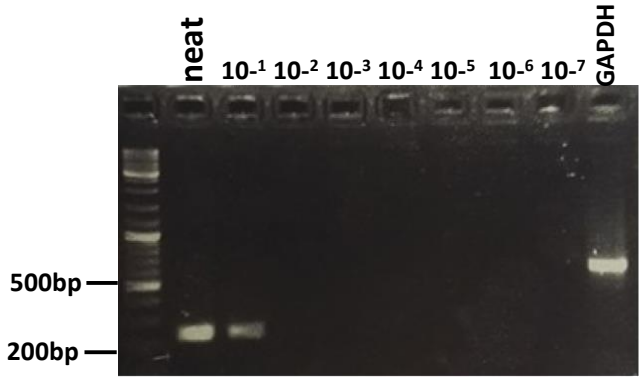

Figure 3.6 NTCP mRNA expression in HuH7 cells

RNAs were extracted using GenELute Mammalian Total RNA Miniprep Kit. Cell mRNAs were then reverse transcribed into complementary DNA (cDNA) using EcoDry™ Premix Oligo(dT) 18 Primers in a XP cycler for one hour. The cDNAs were prepared in a series of ten-fold dilution and amplified using qNTCP primers, and the products were analysed by 2% gel electrophoresis. GAPDH, glyceraldehyde 3-phosphate dehydrogenase.

3.3.2 HuH7-NTCP-BSD cell line

3.3.2.1 Development of a HuH7-NTCP-BSD cell line

To develop a cell line overexpressing NTCP, the first step was to clone the gene into a suitable expression vector. Upon successful reverse transcription of Liver RNA to cDNA, the resulting cDNA was further amplified using a pair of in house designed full length NTCP primers with pWPI-BSD tag. After amplification, the PCR products were cloned into the pWPI-BSD vector resulting in a chimeric plasmid harbouring pWPI-NTCP-BSD (Figure 3.3). This plasmid was designed to simultaneously express NTCP protein as well as Blasticidin S Deaminase (BSD). Next, this chimeric plasmid required propagation in a chemically competent cell, commonly *E.Coli.* Multiple attempts to transform into NEB 5-alpha Competent *E. coli* (NEB) yielded no colonies using standard transformation protocols. Subsequently, NEB stable competent *E.Coli* cells (NEB) were used and overnight incubation was lowered from 37°C to 30°C. Colony screening PCR was performed using pWPIscr.f and pWPIscr.r (Table 3.1). Products of the expected size (~1.2Kb) visualised by 2% agarose gel electrophoresis were confirmed by Sanger sequencing and alignment against human NTCP mRNA (GenBank: L21893.1). A 5 mL overnight bacterial culture was set up using a colony harbouring the correct sequence. A confirmed plasmid clone was designated 'pWPI-NTCP-BSD.2'. Following repeated attempts at optimising yield, the highest plasmid concentration achieved for this construct was 87.18 ng/ µL.

As shown in Figure 3.4, a three-plasmid system was used to generate PVs harbouring a gene expressing NTCP-BSD. By infecting the HuH7 and HepG2 cells with the PVs, the NTCP-BSD transgene was then transduced into target cells. Seventy-two hours later transduced cells were challenged with blasticidin (Invivogen) at a concentration of 20 $\mu g.mL^{-1}$ for HuH7 and 30 $\mu g.mL^{-1}$ for HepG2. Blasticidin-resistant cells were

further expanded under blasticidin. These cell lines were designated 'HuH7-NTCP-BSD' and HepG2-NTCP respectively.

3.3.2.2 NTCP mRNA expression in HuH7±NTCP-BSD cell lines

In order to determine whether NTCP mRNA was expressed in NTCP-BSD transduced cells, NTCP cDNA qPCR assay was performed. Total RNA was extracted from HuH7 ± NTCP cells. NTCP expression in both cell lines was quantified. A dilution series containing 10^1-10^8 copies of pWPI-NTCP-BSD plasmid was prepared and amplified in parallel in order to create a standard curve. The results were analysed using CFX Maestro™. Figure 3.7A shows the standard curve. The circles represent plasmid standards. Two points including the 4^{th} and 7^{th} $\log_{10}$ copies of pWPI-NTCP-BSD plasmid were removed from this standard curve in order to achieve the coefficient determination (r^2) at 0.998. Figure 3.7B illustrates single melt peaks indicating single products yielded for each sample. The inferred copies of NTCP cDNA are illustrated in Figure 3.7C. Near zero copies of NTCP cDNA per cell (10 copies in 1 million cells) were detected in HuH7 cells, which confirmed the low levels of NTCP cDNA detected in endpoint RT-PCR (Figure 3.6). The mean copies of NTCP cDNA per cell were 161 in DNase un-treated HuH7-NTCP-BSD cells and were 90.5 in DNase-treated HuH7-NTCP-BSD cells suggesting that genomic DNA might have been extracted along with mRNAs.

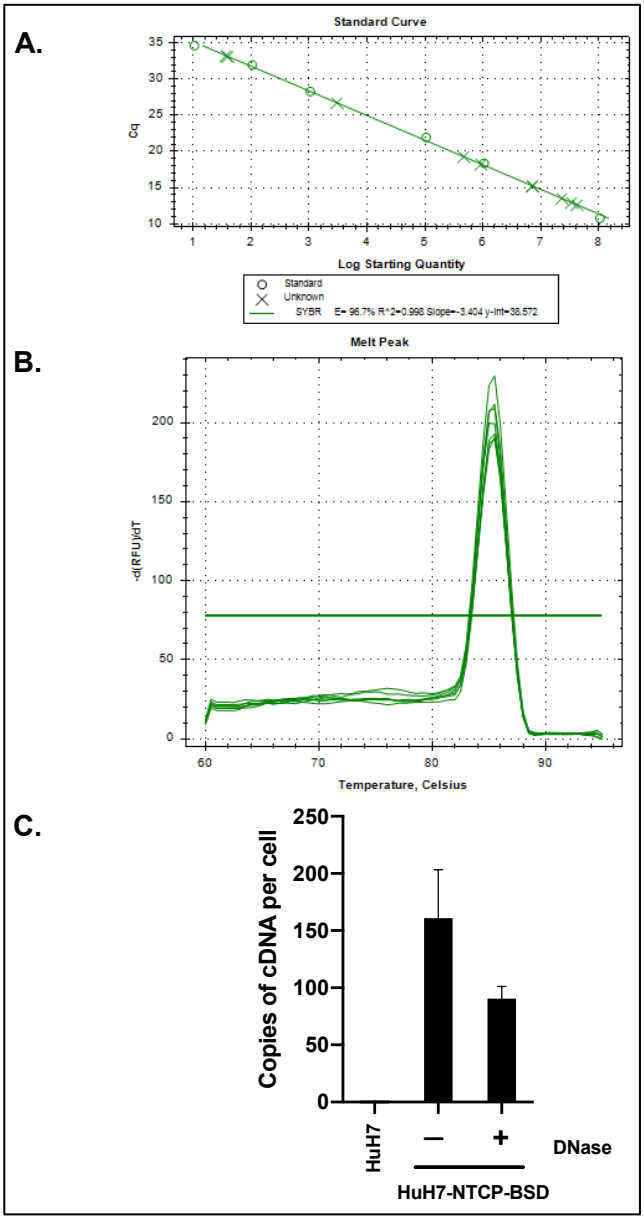

Figure 3.7 Comparison of NTCP mRNA in HuH7 ± NTCP-BSD cells

(A) *Messenger RNAs (mRNAs) were extracted from 6x10^6 cells. One µg of total RNAs were amplified and quantified using the One-step RT-qPCR kit (Biolab). Known copies of NTCP plasmid was used to set up a standard curve; (**B**). melting peaks of the PCR products were determined by raising temperature from 60-95°C. The peak is where 50% the double stranded DNA dissociated; (**C**) cDNA copy number per cell in HuH7 and HuH7-NTCP-BSD cells with or without DNase was interpolated from the standard curve.*

3.3.2.3 Detection of NTCP expression in HuH7 ± NTCP cell lines using anti-NTCP (Ori)

Upon confirmation of NTCP mRNA expression in HuH7-NTCP-BSD cells, the following step was to confirm whether this mRNA was translated to NTCP protein and if yes, where was the protein localised in relation to the cells. Indirect IF assay was employed to detect the expression of NTCP. All cell types including naïve and NTCP cloned cells were probed with anti-NTCP (Ori). It was expected that the cell membrane or/and the endoplasmic reticulum would be stained red. No signal was detected in either of the cell types (Figure 3.8A). This implied either a lack of NTCP expression or a problem with the antibody. The DAPI stain (in blue) showed that the nuclei of cells were evenly distributed (Figure 3.8A) indicating that cells were present on the coverslips.

To determine the lack of signal in IF was not due to the protocol used, cells were prepared for fluorescence activated cell sorting (FACS) using the primary and secondary antibodies used in IF. As shown in Figure 3.8B, the red line (representing the HuH7-NTCP-BSD or HepG2-NTCP cells) superimposed the black line (representing the HuH7 or HepG2 cells), which implied no fluorescence intensity was detected in either cell lines, suggesting this antibody was unsatisfactory for detecting NTCP.

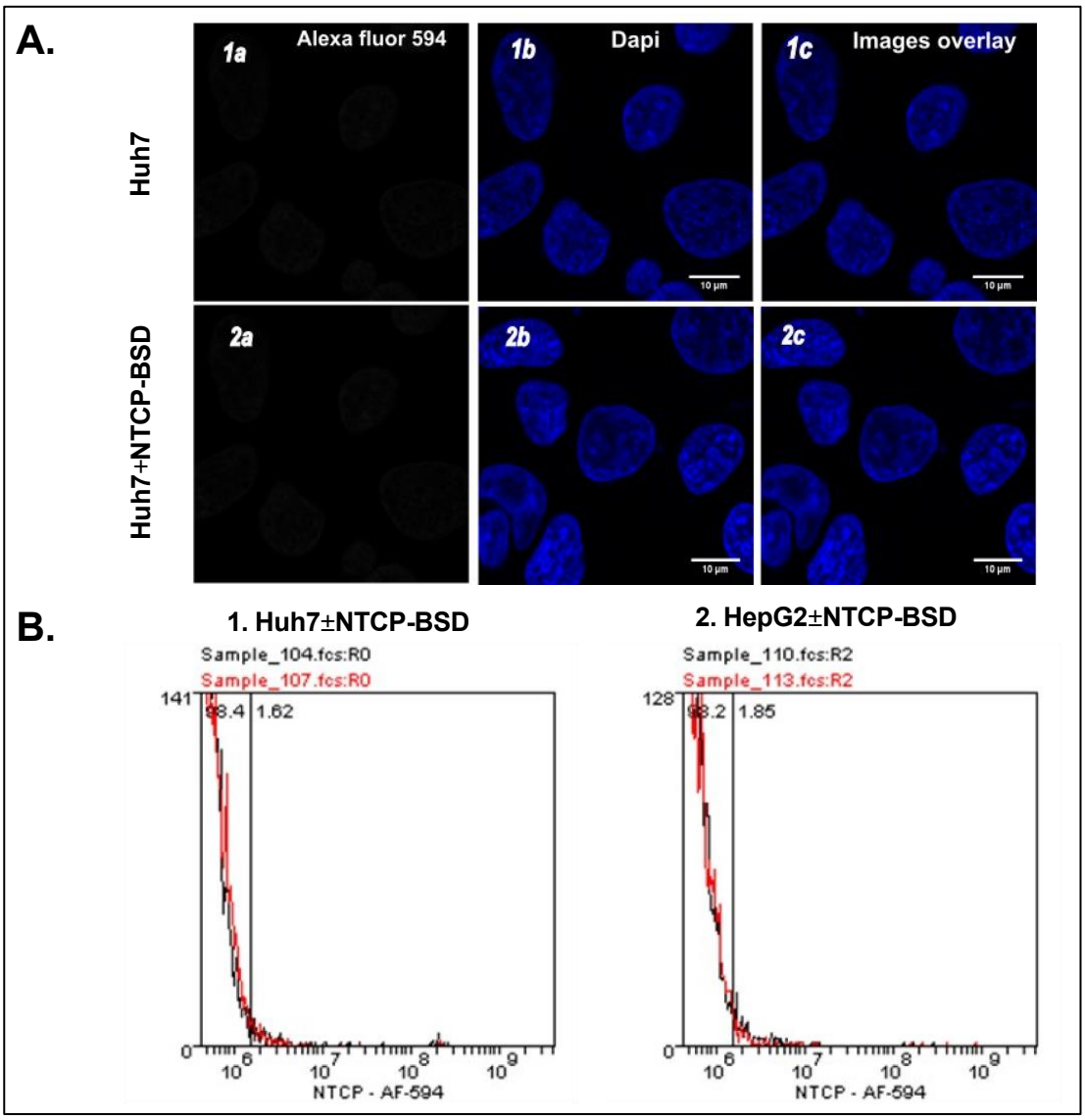

Figure 3.8 Detection of NTCP expression in HuH7 ± NTCP-BSD and HepG2 ± NTCP-BSD cell lines

***A**. HuH7 ± NTCP-BSD were seeded on coverslips in a 12-well flat bottom plate. Once cell density reached 90% confluence, cells were fixed, permeabilised, and stained with anti-SLC10A1 rabbit polyclonal antibody (anti-NTCP (Ori)) and Alexa Fluor® 594 goat anti-rabbit IgG* (ThermoFisher Scientific)*. The nuclei were stained with DAPI in blue. Cells were visualized by confocal imaging. **B.** HuH7 ± NTCP-BSD and HepG2 ± NTCP-BSD cells (~2.0 x 10^5 cells) were stained with 4µg.mL^{-1} anti-NTCP (Ori) and followed by staining with 4µg.mL^{-1} Alexa Fluor® 594 goat anti-rabbit IgG. The antibody binding was measured by fluorescence activated cell sorting (FACS).*

3.3.2.4 Application of NTCP-BSD overexpression cells

Following unsuccessful detection of NTCP using anti-NTCP (Ori), an alternative anti-NTCP antibody (Mer) was used. The epitope recognised by this antibody corresponds to a part of the C-terminal region of NTCP; it is not suitable for flow cytometry live cell sorting as the C-terminus of NTCP locates at the intracellular space. Nonetheless, western blot was performed to detect NTCP expression in the stably transduced cell lines. A clear band was resolved at an estimated molecular weight of 50 to 60 kDa in the NTCP over-expressing HuH7 cells, and two bands with a molecular weight at 50-60 kDa and 65-75 kDa in HepG2 derived NTCP cells. No bands were seen in either parental cell lines (Figure 3.9).

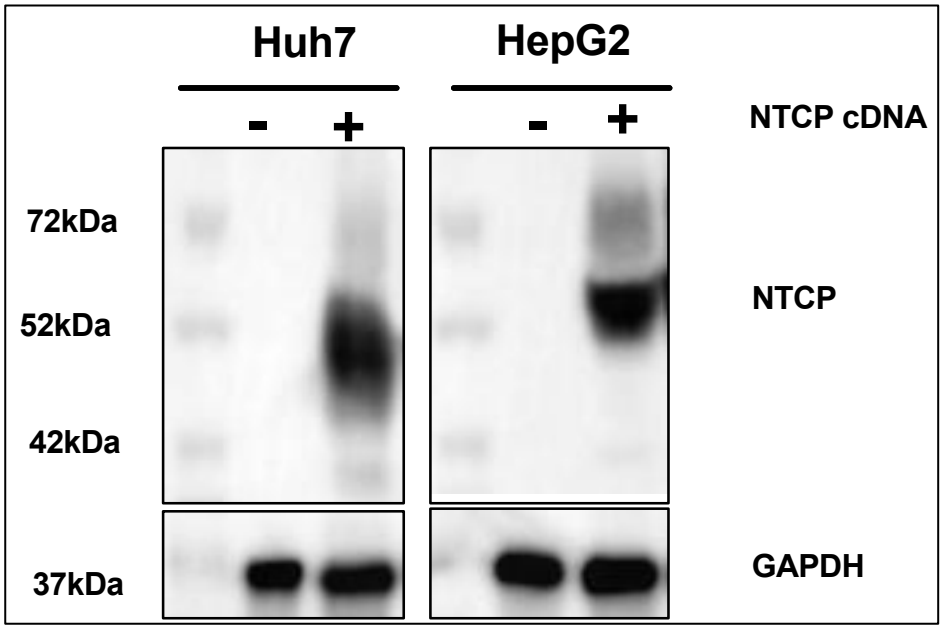

Figure 3.9 Detection of NTCP expression

Proteins were extracted from HuH7 ± NTCP-BSD and HepG2 ± NTCP-BSD cells for analysis by western blot. The primary antibody used was anti-SLC10A1 rabbit polyclonal antibody (HPA042727, Merck).

Upon confirmation of NTCP expression, these HuH7 ± NTCP-BSD cell lines were used in HBV infection assays to confirm whether NTCP supports HBV entry (detailed HBV infection assay is described in Chapter 4). In brief, this assay utilised HBV pseudotyped viruses (HBV-PVs) comprised of HBsAg envelope, HIV capsid and a copy of RNA encoding the luciferase gene, bounded by the HIV-1 LTRs. The hypothesis was that luminescence emission should be only detected in NTCP overexpressing cells that permitted the entry of HBV-PVs. In this assay, the positive control was PVs possessing the envelope glycoproteins of HCV strain J6, a widely used HCV strain for investigating HCV entry. Its E1E2 was known able to be pseudotyped and infect HuH7 cells independently of NTCP expression (Urbanowicz et al., 2015). Delta E (ΔE), the negative control, was a preparation of PVs with the HIV capsid and the luciferase gene, but the omission of an envelope glycoprotein expression construct. The test sample, gtB.LMS was a PV bearing the surface antigen derived from a genotype B hepatitis B virus (known as BR5B6, had been previously developed in-house).

As shown in Figure 3.10A, there was no significant difference in luminescence (shown as relative luminescence unit (RLU)) emitted from cells infected with PVs representing gtB.LMS and ΔE. Rather surprisingly, ~2.5 log units of luminescence were detected in HuH7-NTCP-BSD cells. This was repeatedly observed in following few attempts. Figure 3.10B shows relative infectivity of PVs prepared from different pseudotyping systems including psPAX2, pNL4.3.luc.R-E-. and the phCMV-5349. The data was not interpretable as the background noise was too high in the cell line transduced with NTCP-BSD gene. After verified with the provider who had modified the plasmid pWPI-BSD, it was transpired that luciferase gene was ligated into the vector.

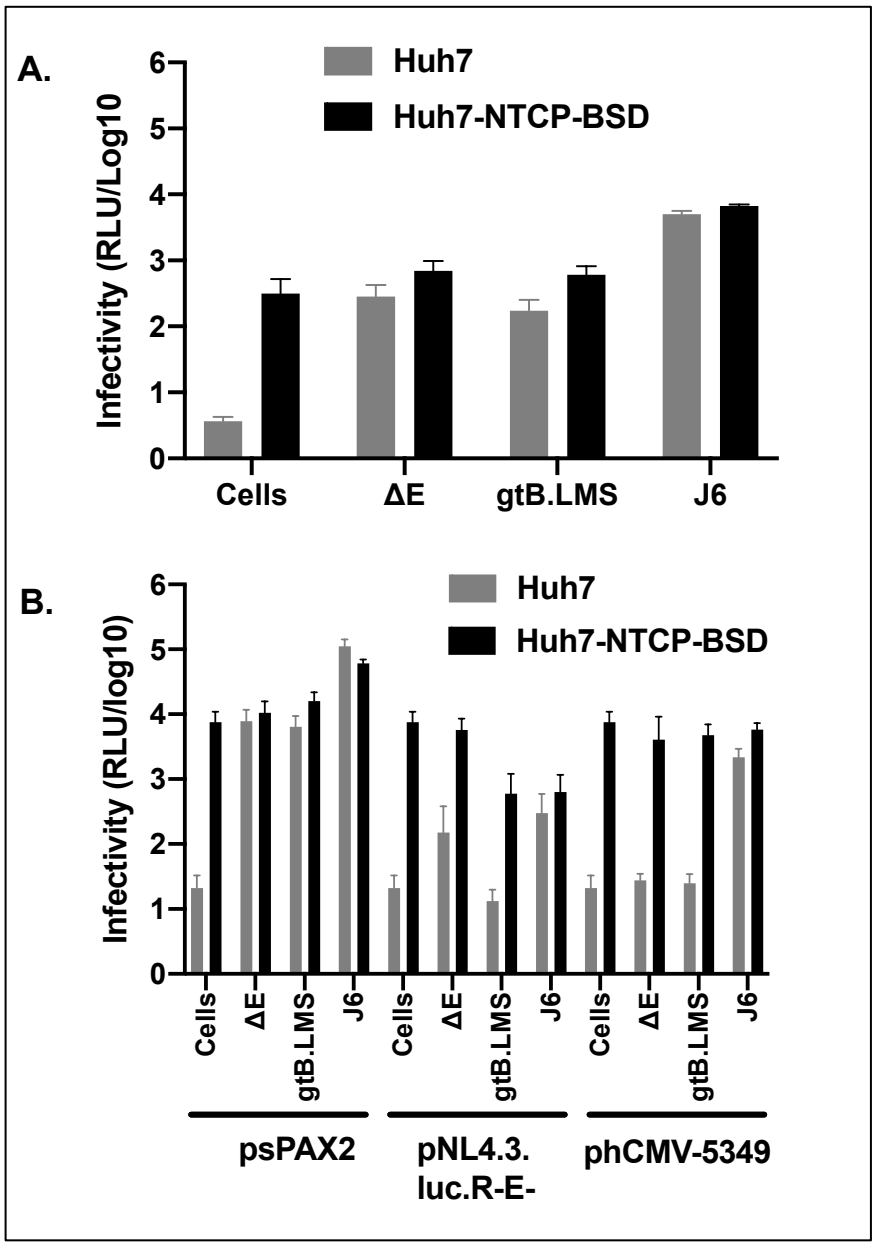

Figure 3.10 Luciferase activity was observed in HuH7-NTCP-BSD cells

*HEK293T cells were transfected with plasmids expressing lentiviral backbone, luciferase reporter gene and HBV surface antigen expression plasmid (J6 was a plasmid expressing E1E2 of HCV prepared as a positive control; ΔE was lentiviral packaging without surface antigens prepared as a negative control; Cell were used as another form of negative control). Plasmid psPAX2 expresses HIV packaging proteins only; it needs a luciferase expression plasmid being supplemented in trans (pcFLW was used in this study). MuLV (phCMV-5349***)** *is a plasmid expressing murine leukaemia virus packaging proteins; it requires a luciferase expression plasmid being*

supplemented in trans (pTG.126 was the reporter expression plasmid compatible for MuLV). Plasmid pNL4.3.Luc.R-E- expresses HIV packaging proteins and luciferase reporter in cis. Seventy-two hours later, pseudotyped viruses (PVs) were harvested and used to infect cells as indicated in the graphs. After a further 72 hours incubation, HBV-PVs infectivity was determined by reading luciferase activity shown as relative luminescence unit (RLU). ***A****. HBV-PVs were prepared using psPAX2 and pcFLW and target cells were HuH7 and HuH7-NTCP-BSD;* ***B****. HBV-PVs were prepared with different packaging plasmids including psPAX2, pNL4.3.Luc.R-E- and MuLV and target cells were the same as panel A.*

3.3.3 HuH7-NTCP-eGFP cell lines

As the HuH7-NTCP-BSD cell line expressed traces of firefly luciferase that affected the interpretation of HBV-PVs infection assays, this cell line was not used further. A new vector pHIV-eGFP (Welm et al., 2008) was used as an expression vector for stable expression of the receptor. As shown in Figure 3.5, the NTCP gene was cloned into the BamHI site enabling bicistronic expression with eGFP, subsequently, a chimeric plasmid pHIV-NTCP-eGFP was created.

As detailed above, the plasmid pHIV-NTCP-eGFP along with pMD2.G and pCMVR8.74 were used to produce PVs carrying the NTCP-eGFP genes. The resulting PVs were used to transduce HuH7 cells. Without the convenience of an antibiotic resistance marker with which to select transduced cells, alternative approaches were required.

3.3.3.1 NTCP-EGFP cell line sorting and verification

At passage three post transduction, cells were sent for live sorting by flow cytometry for evidence of GFP expression. This experiment was conducted by Dr Richard Urbanowicz. As shown in Figure 3.11, approximately 19.89% of cells were GFP positive. All GFP positive cells were collected in a separate tube and then placed back into a Corning cell culture flask in complete DMEM supplemented with 1x Pen-Strep (100 units penicillin and 100 µg.mL^{-1} streptomycin) for 4 passages to prevent bacterial contamination during the FACS.

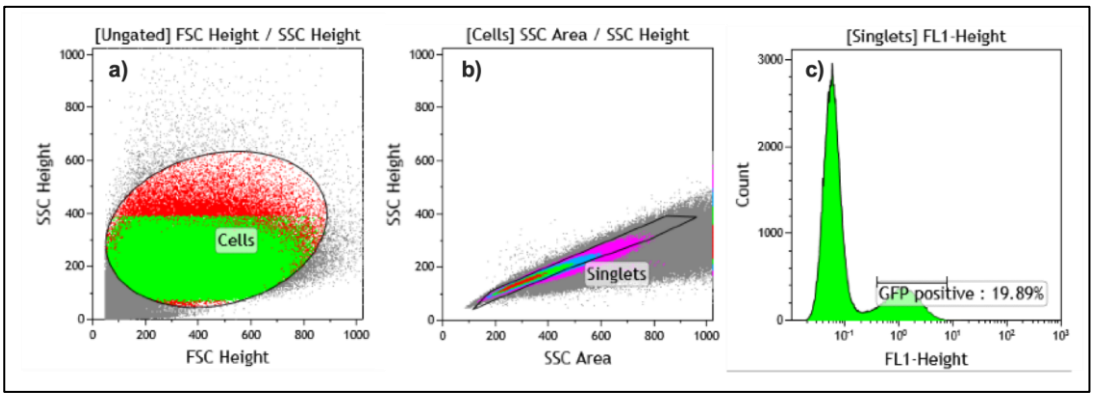

Figure 3.11 Flow cytometry cell sorting

RNA encoding the NTCP ORF was extracted from hepatocytes and reverse-transcribed to cDNA. The resulting cDNA was cloned into an expression vector. This was used to generate lentiviral delivery constructs. Transduction of the HuH7 hepatoma cell line generated a cell line over-expressing NTCP. Approximately 7x10^{6} NTCP-GFP transduced HuH7 cells were prepared for Flow Cytometry live cell sorting (This experiment was conducted by Dr Richard Urbanowicz). ***a).*** *HuH7 cells were used to set the gate in order to exclude cell debris by side scatter height versus forward scatter height;* ***b)*** *singlets selection by setting gate by the side scatter height versus side scatter area to exclude doublets;* ***c)*** *a single parameter histogram showed two different populations: GFP negative and GFP positive cells based on the intensity of the fluorescence.*

Following FACS cell sorting, indirect IF was conducted to determine whether GFP positive cells also expressed NTCP. Rabbit anti-NTCP (Mer) and Alexa Fluor™594 goat anti-rabbit IgG (H+L) were used. The results showed that there was no antibody-NTCP binding detected in HuH7 cells (Figure 3.12), whereas NTCP receptors were highly expressed in the plasma membrane of HuH7-NTCP-GFP cells (Figure 3.12). The results confirmed that the majority of GFP positive cells also expressed NTCP, however, as there was no GFP staining, a correlation between GFP and NTCP expression could not be determined.

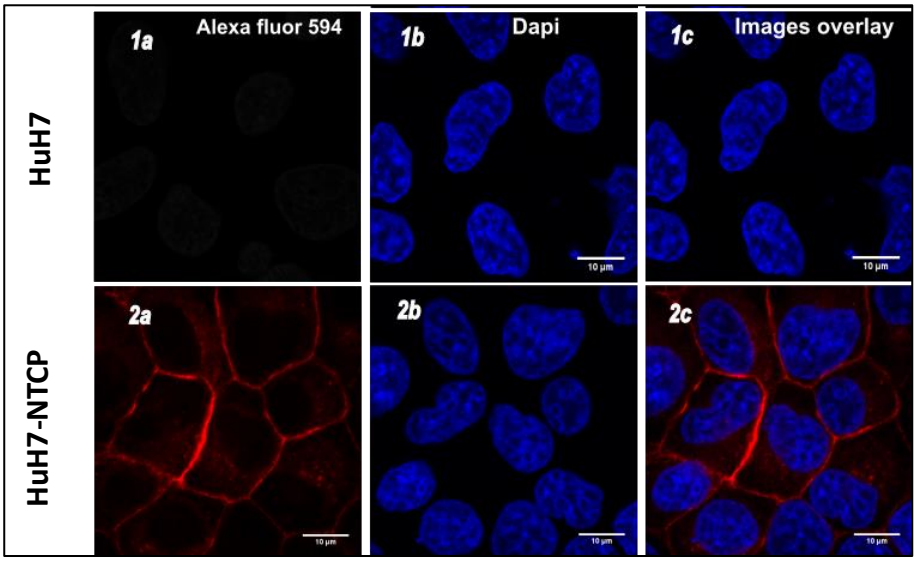

Figure 3.12 NTCP-eGFP cell validation

HuH7 ± NTCP-eGFP were seeded on coverslips in a 12-well flat bottom plate. Once cell density reached 90% confluence, cells were fixed, permeabilised, and stained with rabbit anti-NTCP (Mer) and Alexa Fluor® 594 goat anti-rabbit IgG. Cells were visualised using Zeiss LSM710 Confocal microscope. Panels 1a, 1b and 1c are images of HuH7 cells detected with filter sets for Alexa Fluor 594, Dapi and both respectively; panels 2a, 2b and 2c are images of HuH7-NTCP cells detected with filter sets for Alexa Fluor 594, Dapi and both respectively.

The purity of GFP expression post FACS was 99.8%, which suggested that there were 0.2% non-transduced cells which might outgrow transduced cells. Hence over time, the purity of GFP expressing cells would reduce. Based on this hypothesis, GFP expression was monitored in sorted cells over a period of two months, the length of time that a batch of cells were used in cell culture after thawing from liquid nitrogen store. Four post sorting passages (PSP), 2, 4, 14 and 18 were analysed. The results showed that the purity of GFP overexpressing cells dropped from 99.8% at PSP2 to 95.6% at PSP18. The cells were passaged twice a week, and as such, after 8 to 9 weeks post sorting, 95.6% cells in a mixed population still expressed GFP (Figure 3.13A).

Western blot was performed on these passaged cells in parallel to monitor NTCP expression. As shown in Figure 3.13B, the NTCP expression was weak in PSP2 and 4, but improved overtime as bands were intensified in PSP14 and 18.

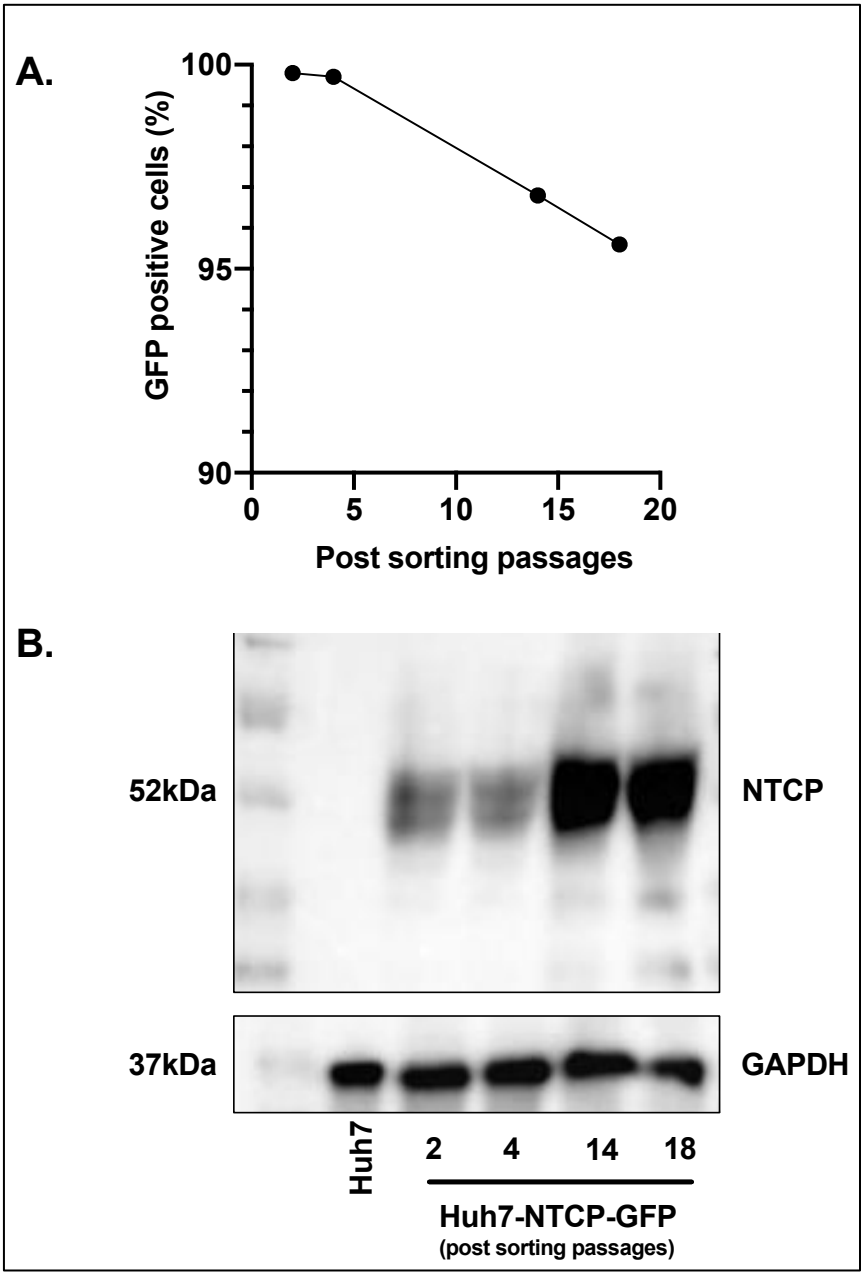

Figure 3.13 NTCP and GFP expression in HuH7-NTCP-GFP cells at different cell passages

GFP and NTCP expression monitored at various time points post sorting passage (psp) 2, 4, 14 and 18. ***A.*** *Approximately 5.0x10^{-4} cells were sent for flow cytometry at the specified time points.* ***B.*** *At same time points, approximately 5.0x10^{-6} cells were lysed and 20 µg of total protein from each lysate was separated by SDS-PAGE gel electrophoresis and then blotted with rabbit anti-NTCP (Mer) followed by goat anti-rabbit HRP.*

3.3.3.2 Luciferase activity was not detected in HuH7-NTCP-eGFP cells

Once the NTCP and GFP expression was confirmed, luciferase assay was performed to determine whether luciferase expression could be detected in HuH7-NTCP-eGFP cells. As the purpose of the assay was to determine the luciferase activity in HuH7±NTCP-eGFP cells, one each of the positive and negative control (Delta E) were included in this assay. The H77 was a plasmid expressing envelope glycoproteins E1E2 of HCV strain H77, another widely used HCV strain for investigating HCV entry in addition to J6. As demonstrated in Figure 3.14, light emission from the activity of luciferase was not detected in HuH7-NTCP-eGFP cells.

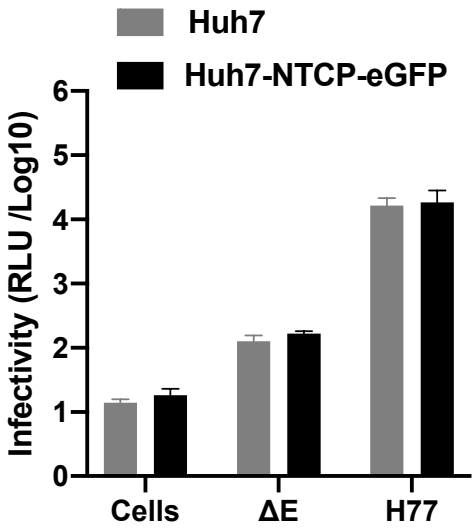

Figure 3.14 Luciferase activity was not detected in HuH7-NTCP-eGFP cells

HEK293T cells were transfected with plasmid pNL4.3.luc.R-E- expressing lentiviral backbone and luciferase reporter gene and plasmid expressing envelope glycoproteins. The H77 was a plasmid expressing E1E2 of HCV prepared as a positive control; ΔE was lentiviral packaging without surface antigens prepared as a negative control; Cell were used as another form of negative control. Seventy-two hours later, pseudotyped viruses (PVs) were harvested and used to infect cells as indicated in the graphs. After a further 72 hours incubation, infectivity was determined by reading luciferase activity shown as relative luminescence unit (RLU).

3.3.4 *N*-linked glycosylation of NTCP

The expected human NTCP molecular weight was 38 kDa. In order to determine whether the alternative band patterns observed were due to alternative glycosylation, enzymatic deglycosylation was performed. lysates of HuH7-NTCP and HepG2-NTCP cells were first denatured at 100°C for 10 minutes and followed by incubation with NP40, and PNGase F at 37°C for one hour. In parallel, non-PNGase treated samples were incubated with dH2O instead of PNGase. The proteins were then separated by SDS-PAGE gel electrophoresis. The gels were blotted with rabbit anti-NTCP (Mer) followed by goat anti-rabbit HRP. Interestingly, a defined band at ~38kDa was seen in the HepG2-NTCP sample post deglycosylation, but double bands (~36kDa and ~38kDa) were observed in the deglycosylated HuH7-NTCP sample (Figure 3.15A).

To exclude a potential cause that the *N*-linked glycosylation in HuH7-NTCP was resistant to PNGase F, incubation time with PNGase was increased to overnight. In addition, HEK293T-NTCP cells were analysed in parallel. After overnight incubation, one single sharp band at ~38kDa was yielded in both HepG2 and HEK293T derived NTCP cell samples. Once again, two bands were observed in HuH7-NTCP samples (Figure 3.15B).

To exclude whether the HuH7-NTCP cell line harboured mutations in the NTCP gene, total RNA was extracted and reverse transcribed to cDNA in one-step RT-PCR using NTCPf and NTCPr primers. An aliquot of PCR product was sequenced using the NTCPf and NTCPr primers, providing sequence data covering the entire coding sequence. No point mutation was noted aligned to the human NTCP mRNA from GenBank (L21893.1). Surprisingly, even though the loading quantities were the same for PNGase treated and none treated samples, the bands yielded in non-PNGase treated samples were much weaker than their counterparts (Figure 3.15).

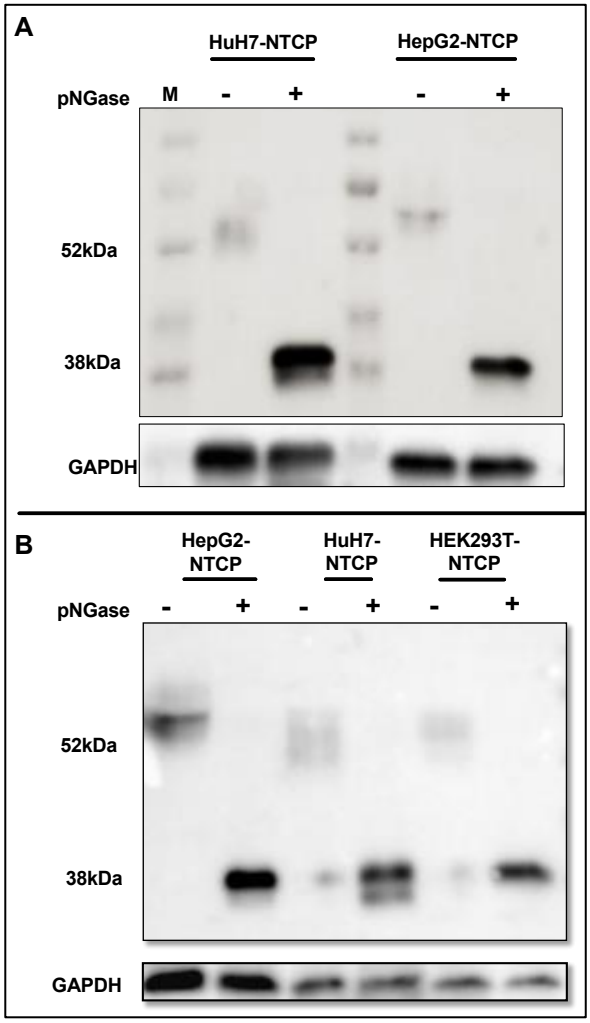

Figure 3.15 Comparison of NTCP expression in different cell lines

Twenty µg of total protein from cell lysates were first denatured at 100ºC for 10 minutes followed by incubation with NP40, and PNGase at 37ºC. Non-PNGase treated samples were treated similarly but dH2O replaced the PNGase. The proteins were then separated by SDS-PAGE gel electrophoresis. The gels were blotted with rabbit anti-NTCP (Mer) followed by goat anti-rabbit HRP. ***A****. Lysates of HuH7-NTCP and HepG2-NTCP cells were incubated at 37ºC for 4 hours;* ***B****. lysates of HuH7-NTCP, HepG2-NTCP and HEK293T-NTCP cells were incubated at 37ºC overnight.*

3.3.5 Bile acid uptake assay

To determine if the transduced cells possessed NTCP's biological function of relation to transporting bile acid, a plasmid (NucleoBAS) genetically encoded fluorescent bile acid sensor (BAS) was employed. The bile acid sensor consisted of two fluorophores including citrine and cerulean fused to the farnesoid X receptor ligand binding domain (FXR-LBD), which can be activated by a large influx of bile acids. Intensity of fluorescence can be quantified upon excitation by laser light in cells transfected with plasmid expressing the fluorescent bile acid sensor if the FXR-LBD is activated, using Forster Resonance Energy Transfer (FRET) (Van de Wiel et al., 2016, van der Velden et al., 2013) (Figure 3.16).

To minimise any potential bile acid uptake into hepatic cells, HEK293T-NTCP cells were initially used, and to determine whether the sensor was already saturated, HEK293T cells that did not express NTCP were used as a control. Moreover, to determine whether a false positive result might arise due to cell autofluorescence, cells without expression of the bile acid sensor were used as negative controls. GW4064 is the most potent FXR agonist that can maximise the ratio between citrine and cerulean. At the time of experiment, a simultaneous dual emission detector to measure FRET was unavailable, therefore, fluorescent signals of cerulean and citrine were measured in sequence. As demonstrated in Figure 3.16, emission ratios between citrine and cerulean in cells mimics that in cells transfected with NucleoBAS regardless of NTCP expression suggesting either the assay was not optimal or the NTCP was not functional.

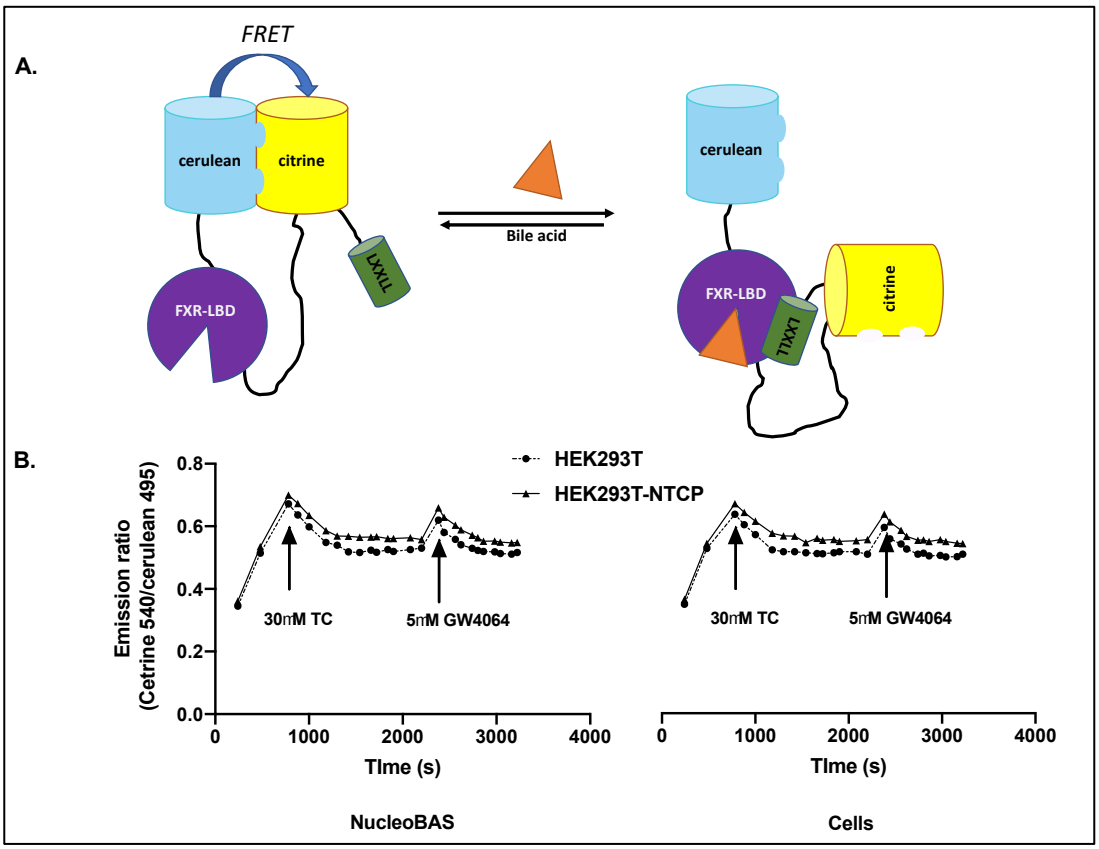

Figure 3.16 Bile Acid Sensor and Forster Resonance Energy Transfer

A. *A bile acid sensor (BAS) is composed of a donor fluorophore (cerulean) and an acceptor fluorophore (citrine) connected via FXR ligand bind domain (FXR-LBD). This molecule is fused to a FXR-cofactor (LXXLL motif). In the absence of bile acids, intramolecular interaction between the two fluorophores promotes maximal Forster Resonance Energy Transfer (FRET) signal. Binding to bile acids recruits LXXLL motif to FXR-LBD resulting in conformational change, which separates the two fluorophores and ultimately reduces FRET signal (van der Velden et al., 2013, Van de Wiel et al., 2016).* ***B.*** *HEK293T and HEK293T-NTCP were transfected with plasmid pCMV.NucleoBAS. Cells without plasmid transfection were prepared in parallel as negative controls. Fluorescent intensity of citrine and cerulean was measured using BMG Labtech FLUOstar Omega. FRET signal was recorded as fluorescent emission ratio between citrine and cerulean. A 30 µM taurocholate acid (TC) and a 5 µM GW4064 were added as indicated in the graph.*

3.3.6 Virus-receptor binding assay

To investigate virus-receptor interaction, HBV-NTCP binding assay was performed. Huh7 ± NTCP cells were used to capture serum derived HBV (Figure 3.17A). Furthermore, in order to determine whether the binding was dose-dependent, HBV containing serum was diluted with DMEM to obtain HBV DNA concentrations at 200, 400, 600, 800 and 1000 IU/mL. As a control, serum tested negative for HBV DNA but positive for HBsAg was used. As demonstrated in Figure 3.17B, except for the anomaly of the DNA concentration at 400 IU/mL, a clear trend showed that virus-cell binding enhanced as the amount of virus added was increased. There was an approximately 30% increase in viral binding observed in DNA concentration at 1000 IU/mL compared with the control sample; however, the difference was not statistically significant. Moreover, there was no statistically significant difference in virus-cell interaction between Huh7 and Huh7-NTCP cells suggesting either the assay was not optimal or NTCP was not the principle receptor for HBV (Figure 3.17B).

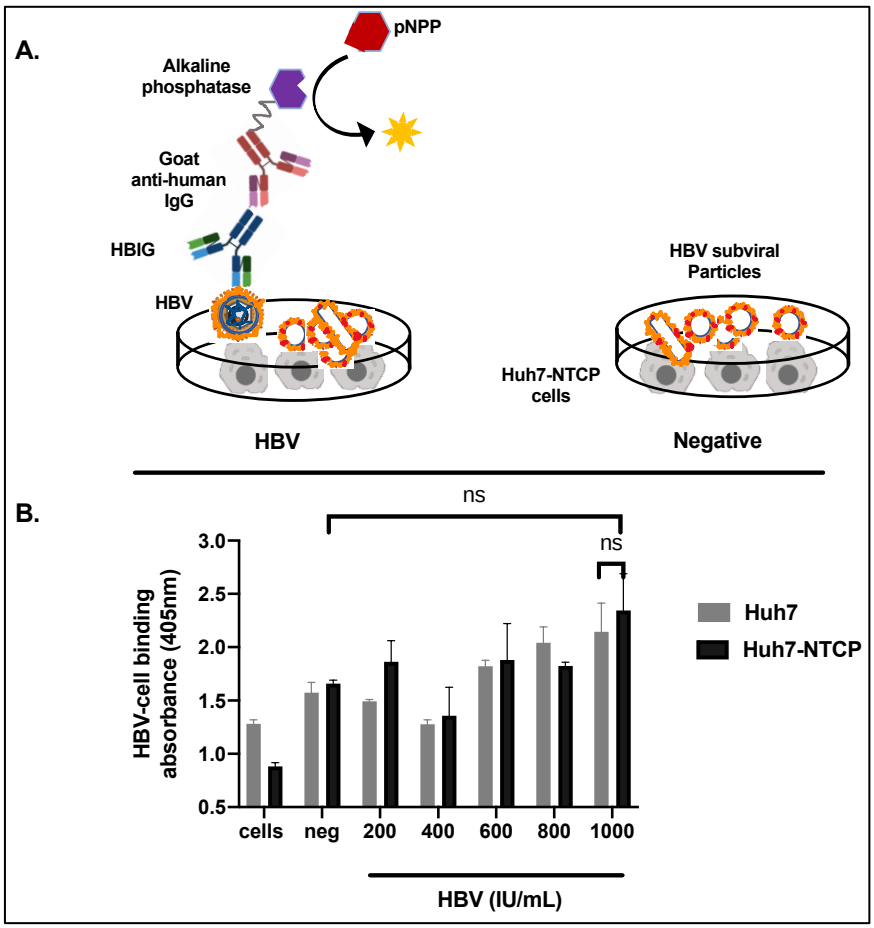

Figure 3.17 HBV-Cell binding

A. *Huh7 ± NTCP cells were fixed with 4% formaldehyde prior to incubation with HBV containing serum for 1 hour at room temperature. After methanol fixation for 10 minutes, cells were incubated with Human hepatitis B immunoglobulin (HBIG, 1:100) for 2 hours followed by incubation with goat anti-human IgG conjugated to alkaline phosphatase for 1 hour. A negative sample was tested in parallel that had been determined to be negative for HBV DNA, but positive for HBsAg.* ***B.*** *Virus-cell binding was quantified by measuring the absorbance at 405nm after incubation in p-Nitrophenyl phosphate (pNPP) substrate solution. HBV containing serum was diluted with Dulbecco's modified Eagle medium (DMEM) to final Viral DNA concentrations as indicated.* Statistical analyses were conducted using a One-way ANOVA in Prism 8. P>0.01 was considered statistically insignificant and indicated as 'ns'.

3.4 Discussion

The principle aim of this Chapter was to develop NTCP-overexpressing hepatoma cell lines using lentiviral transduction method. The expression plasmid pWPI-BSD was used initially with the convenience of an antibiotic resistance marker with which to select transduced cells.

3.4.1 Quantification of NTCP mRNA

In HuH7 cells, the end point PCR showed that NTCP mRNA was undetectable at > 1:10 dilution suggesting the expression of NTCP in these cells was extremely low, which is consistent with previous studies (Yan et al., 2012). The qPCR results further confirmed this claim. More than a 100-fold higher NTCP mRNA was present in HuH7-NTCP-BSD cells compared to HuH7 cells. A DNA-based standard curve is generally not a good calibrator for absolute RNA quantification because the efficiency of reverse transcription from mRNA to cDNA cannot be controlled. Not all mRNA is converted to cDNA in a RT reaction and not all transcripts are converted in a 1:1 ratio or proportionally to the input RNA templates. To limit variables during reverse transcription, a one-step qPCR approached was adopted providing an accurate, reproducible assay and minimising variations in variable such as input RNA concentration, RT enzyme and buffer components (Nolan et al., 2013).

There were approximately 70 copies more of NTCP cDNA per cell in non-DNase treated HuH7-NTCP-BSD cells compared to that calculated for DNase treated counterparts, indicating the presence of a genomic DNA contamination. One of the limitations of using DNase treatment is that any active enzyme residues could degrade newly synthesised DNA in qPCR and subsequently lower the true total quantity. The qPCR primers could have been designed to amplify across specific

NTCP exons only, such that, genomic DNA would not be amplified. However, the genomic NTCP DNA contamination in this assay could originate from the exogenously delivered NTCP gene, which had been reverse transcribed from NTCP mRNA, i.e. containing exons only. Nevertheless, precision measurement of the NTCP mRNA expression was not the primary aim of this assay, which was rather to provide evidence that HuH7-NTCP-BSD cells expressed 100-fold higher amounts of NTCP than the parental cells.

3.4.2 Propagation of plasmid bearing LTRs

The propagation of plasmid pWPI-NTCP-BSD in competent bacteria proved to be a challenging task. NEB 5-alpha Competent *E. coli* were used for the first attempt. Repeat attempts obtained either no colonies or colonies without the desired recombinant plasmid, suggesting that this plasmid was unstable. Instability of plasmids could be due to a number of factors such as repeat elements or unstable inserts present in the plasmid, toxic gene products, the metabolic burden of plasmid maintenance, plasmid copy number (Corchero and Villaverde, 1998). The inherited 5' and 3' long terminal repeats (LTRs) in pWPI-BSD as a lentiviral construct could be the main cause. LTRs are necessary for successful stable transduction because they contain the requisite signals for gene expression and DNA integration (Klaver and Berkhout, 1994). However, cloning of this recombinant DNA into standard *E.coli* such as NEB 5-alpha cells often leads to plasmid instability, often manifesting as deletions of the sequences between LTRs (Chakiath and Esposito, 2007). Deletion of the nucleotides between LTRs could happen. Such deletions remove the most important sequences leading to a much smaller recombinant plasmid often containing the replication origin and the antibiotic resistant marker. This smaller recombinant plasmid is strongly selected and outnumbers the desired clones (Chakiath and Esposito, 2007). Similar results have been observed by other researchers i.e. that a *DH5-*α

strain such as NEB 5-alpha is not a competent recipient for HIV-based vectors (Al-Allaf et al., 2013). It was reported that deletion arises due to homologous recombination events (DasGupta et al., 1987).

To improve plasmid stability, NEB® stable competent *E.coli* was used for the second round transformation. This strain is genetically modified for retroviral/lentiviral vector cloning as stated by NEB. It was claimed that the *recA*1 gene was silenced in the NEB® stable competent *E.coli* (Biolabs, 2020). As the recA protein mediates the rearrangement of homologous sequences such as LTRs, silencing recA leads to plasmid stability. It was attempted to draw a conclusion that *recA1* deficiency was responsible for the successful transformation into NEB® stable competent *E.coli* in the current experiment. However, NEB 5-alpha also has *recA1* mutation, and therefore the exact mechanism of stability generation is still unclear.

Plasmid stability was improved in the NEB® stable competent *E.coli*. However, although the PCR products from colony screening confirmed the desired NTCP sequences, bacterial growth in both minipreps and midipreps was far from satisfactory. After several attempts, the best plasmid yield was 87.17 ng/µL. At the time of conducting this experiment, the temperature and length of incubation were optimised. For example, bacterial culture incubation was changed from 37°C for 16-18 hours to 30°C for 24 hours. Previous evidence demonstrated that, at lower temperature, the plasmid was more stable and product toxicity was a less of an obstacle for the host propagation (Corchero and Villaverde, 1998, Al-Allaf et al., 2013). However, these modifications made no significant difference to plasmid yield.

An alternative method that could have been attempted is the use terrific broth (TB) for bacterial culture. Compared with Luria-Bertani broth (LB), a commonly used standard culture media, TB contains additional glycerol and potassium phosphate. Glycerol

offers an extra source of carbon for energy and potassium phosphate buffers the acid released during the growth of *E.coli,* which enhance the plasmid yield by extending the exponential phase of *E.Coli*. A 10-fold increase of the number of bacterial cells was observed in standard LB enriched with glycerol (Kram and Finkel, 2015).

3.4.3 Selection of antibodies for detection of NTCP expression

The seletion of a suitable antibody is crucial for the detection of expression of a protein. Depending on the immunogen (whether they are synthetic peptides or purified proteins), the application and performance of an antibody can vary. Although synthetic peptides provide known corresponding amino acid sequences, their tertiary structures and post-translational modifications may not recapitulate their native protein counterparts (Bordeaux et al., 2010). Antibodies generated against synthetic peptides may lose their ability to bind proteins in their native forms in applications such as FACS. On the contrary, they may be more useful for SDS-PAGE and western blot when binding to fully denatured proteins. Similarly, antibodies generated against purified proteins may bind better to proteins in their native conformation, but not the fully denatured ones (Bordeaux et al., 2010).

The rationale of choosing the anti-NTCP (Ori) antibody was because it was generated to against an N-terminal region of NTCP, which is located in the extracellular region of the protein. As such, it could be exploited for live cell sorting. However, this antibody did not perform well in either IF or FACS.

3.4.4 Co-expression transgenes from bicistronic vectors

In this study, two lentiviral vectors including pWPI-BSD and pHIV-GFP were used in the development of the NTCP cell line. Both vectors rely on human elongation factor-

1 alpha (EF-1 alpha) promoter-driven production of a single messenger RNA (mRNA) carrying two separate protein coding regions linked by the internal ribosomal entry site (IRES) sequences. NTCP was translated via a cap-dependent mechanism, while the translation of either Blasticidin S Deaminase or eGFP was mediated by IRESs.

It is claimed that protein translation via a cap-dependent mechanism is more efficient than protein expression mediated via an IRES (Thompson, 2012). However, the efficiency of co-expression of the two transgenes was unknown. Western blot and indirect IF were conducted to determine the expression of NTCP as the NTCP-BSD and NTCP-eGFP cell lines were selected by the presence of either BSD or eGFR respectively. As shown in Figure 3.9, an NTCP band was seen in the blasticidin resistant NTCP-BSD cell lines but not their parental cells, and indirect IF also provided evidence that NTCP was expressed in eGFP positive Hu7-NTCP-EGFP cells (Figure 3.12). However, both HuH7-NTCP-BSD and HuH7-NTCP-eGFP were in polyclonal cell populations, which means over time, transgene expression may decline. The growth rate may be much slower in cells that express high levels of transgene than in ones with lower levels of transgene expression or even absence of expression. In the case of HuH7-NTCP-eGFP, after initial cell sorting, the purity of eGFP expressing cells was 99.8%. Eventually, the minor population of eGFP-negative cells would outgrow the positive ones. To overcome this, aliquots of early post sorting passage cells were stored in liquid nitrogen. The purity of eGFP positive cells was monitored over a two month period. As shown in Figure 3.13, both NTCP and eGFP expression were maintained for more than two months. Therefore, a protocol was implemented, in which a batch of cells would be maintained in cell culture for two months only and then a new batch of cells would be thawed from liquid nitrogen storage.

Generating a monoclonal cell line is another way to overcome this issue. A typical NTCP-overexpressing hepatoma single cell clone could be generated by (1)

transfecting or transducing hepatoma cells with NTCP expressing gene; (2) splitting and growing the hepatoma cells in a 10cm^2 until single colonies are easily visible; (3) picking and transferring single colonies into a 48 well plate by digestion with trypsin in cloning cylinders; (4) growing and expanding the single clones; (5) performing HBV infection assay to select a cell line with high susceptibility to HBV infection (Sun et al., 2017). This procedure is laborious and time consuming. As such, a single clone cell line was not developed in the current study. Moreover, the polyclonal cell population could be more representative to the cell population *in vivo*. Hence, the polyclonal cell population was considered to be acceptable for further experimentation.

3.4.5 Luciferase activity was detected in HuH7-NTCP-BSD transduced cells

According to previous research, NTCP is the principal receptor for the entry of HBV (Yan et al., 2012, Ni et al., 2014). If this is true, HBV entry should be significantly enhanced in NTCP overexpressing cells. Therefore, an infection assay was performed. However, the results showed (Figure 3.10) that HBV-PV entry was not greater in HuH7-NTCP-BSD cells compared to the parental cells. Unexpectedly, luminescence was emitted from HuH7-NTCP-BSD-transduced cells indicating endogenous luciferase activity in these cells. PCR screening for the presence of a luciferase gene was performed for all pWPI-BSD containing plasmids as well as cDNAs extracted from the cell lines derived from this construct. As a result, the HuH7-NTCP-BSD cell line was no longer used in HBV-PV assay as traces of luciferase activity could interfere with the interpretation of the HBV-PV entry assay.

3.4.6 *N*-linked glycosylation of NTCP

Enzymatic deglycosylation of NTCP revealed two unexpected findings. The western blot results showed that after loading the same amounts of total proteins (as determined by protein assay), lower amounts of glycosylated NTCP (gNTCP) were detected compared with that of deglycosylated NTCP (dNTCP) (Figure 3.15). More surprisingly, there were two distinct bands of dNTCP in the sample from HuH7 derived NTCP, while one single sharp band was observed in samples from HepG2 and HEK293T derived NTCP.

The lower detection in gNTCP was probably because of reduced antibody-antigen interaction. Anti-NTCP (Mer) was raised against non-glycosylated synthetic peptides. The immunogen fragment is located at the C-terminus of NTCP between 306-aa to 342-aa, whereas, the two N-linked glycans N5-aa and N11-aa are at the N terminus. After SDS-PAGE separation, the two glycans might act as a shield, so that, the immunogen sequence was less accessible to the anti-NTCP antibody. In comparison, the dNTCP would have lost the glycan shield resulting in increased exposure to the antibody, hence, more antibody-antigen binding would be detected (Figure 3.18).

MEAHNASAPFNFTLPPNFGKRPTDLALSVILVFMLFFIMLSLGCTMEFS
KIKAHLWKPKGLAIALVAQYGIMPLTAFVLGKVFRLKNIEALAILVCGCSP
GGNLSNVFSLAMKGDMNLSIVMTTCSTFCALGMMPLLLYIYSRGIYDG
DLKDKVPYKGIVISLVLVLIPCTIGIVLKSKRPQYMRYVIKGGMIIILLCSVA
VTVLSAINVGKSIMFAMTPLLIATSSLMPFIGFLLGYVLSALFCLNGRCRR
TVSMETGCQNVQLCSTILNVAFPPEVIGPLFFFPLLYMIFQLGEGLLLIAI
FWCYEKFKTPKDKTKMIYTAATTEETIPGALGNGTYKGEDCSPCTA*

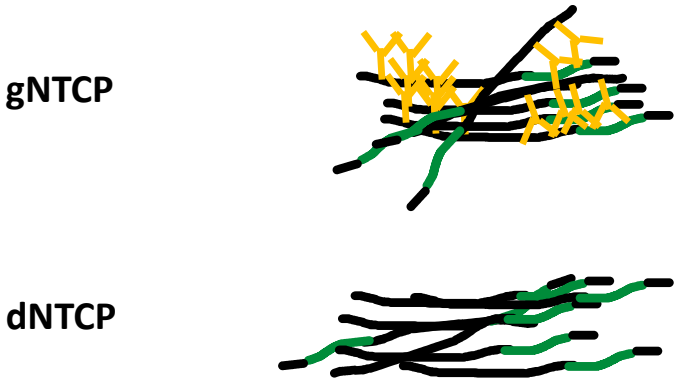

Figure 3.18 Proposed structure of NTCP after denaturing during western blotting

The two glycosylated amino acids (aa) N5-aa and N11-aa are highlighted in yellow and the immunogen of anti-NTCP (Mer) is highlighted in green. gNTCP, glycosylated NTCP; dNTCP, deglycosylated NTCP.

N-linked glycosylation is cell-type specific, depending on the features and availability of the glycosylation machinery in an individual cell (Rudd and Dwek, 1997). A previous study reported that the molecular weight of NTCP from human liver tissues was 56 kDa, which was reduced to 39 kDa post deglycosylation; when human NTCP was expressed in the HEK293 cell line, both the 50 kDa smear, as well as a 37 kDa band were detectable; when the NTCP was expressed in HeLa cells, two bands of approximately 40 kDa and 50 kDa were detected (Lee et al., 2018). Consistent with reported data, results reported herein showed two bands with molecular weight between 50-60 kDa and 65-75 kDa in HepG2 cells exogenously expressing NTCP and a broad smear band spanned 45 kDa to 60 kDa in HuH7 cells expressing NTCP (Figure 3.9), which could be reduced to ~38 kDa after enzymatic deglycosylation with the exception of HuH7 derived NTCP. The double bands observed in HuH7-NTCP lysates following deglycosylation have not been reported previously or been seen in other cell lines in this study. Even though all exogenous NTCP expressed in HuH7, HepG2 and HEK293T were from the same NTCP construct, to exclude the possibility of point mutation, NTCP cDNA extracted from HuH7-NTCP cells was sequenced. The results confirmed that there was no mutation in the HuH7 derived NTCP sequence. There could be unique post translational modification of NTCP in HuH7 cells. To confirm this, further research is required. For example, it would be helpful to separate the proteins via SDS-PAGE followed by silver staining. The resulting two bands could then be extracted from the gel and sent for mass spectrometric analysis. However, this procedure is beyond the scope of the present study.

3.4.7 Molecular function of NTCP

Although IF confirmed the expression of NTCP in the plasma membrane of all cell lines, the correct biological function of NTCP in this context was not verified. In the available literature, the classic methods to validate the biological function of NTCP has been the use of a [^{3}H] taurocholate uptake assay or a FITC-labelled pre-S1 peptide binding assay (Yan et al., 2014, Konig et al., 2014, Appelman et al., 2017). The sourcing of these compounds was challenging as [^{3}H] taurocholate is costly ($2414 per 250μCi) and its use is licensed. Furthermore, FITC-preS1peptide was not commercially available at the time of the experiment.

The purpose of using NucleoBAS was to determine the molecular function of NTCP without the need of fluorescent labelled peptides or radiolabelled compounds. However, there were no significant changes in emission ratios between citrine and cerulean across all assay setups, suggesting that this experiment was not successful. It was unclear whether it was due to no intracellular expression of BAS or due to failure in detection of FRET. In future work, BAS expression could be confirmed by confocal imaging, and once confirmed, FRET can be detected using FACS or confocal imaging.

A virus-receptor binding assay was used as an alternative approach to determine the correct conformation and expression of NTCP in the transduced cells. The main challenge experienced in this study was that the morphology of the cells was altered, with cell shrinkage and change to an oval shape shortly after inoculation with HBV containing sera. The cells were then easily detached from the surface during washing. Therefore, after optimisation, cells were fixed with 4% formaldehyde before treatment with HBV. Formaldehyde, similar to other aldehydes such as paraformaldehyde and glutaraldehyde, cross-links proteins, thereby preserving cell morphology. It is a widely

used fixation agent resulting in less cell shrinkage and good preservation of cellular structure as well as having less structural changes to proteins (Hobro and Smith, 2016). The following step of methanol fixation post treatment with HBV-containing serum was to deactivate live viruses and therefore render the sample safe to be handled in a Containment Level 2 laboratory. Previous evidence has shown that cell surface marker staining is not affected by sequential paraformaldehyde and methanol fixation (Pollice et al., 1992). However, fixation with both methods may alter antigenicity of proteins, which could affect virus-receptor binding. Moreover, formaldehyde contains ~1.5% of methanol as a stabiliser, which might have permeabilised the cell membranes and physically trapped viruses inside cells, resulting in false positive staining.

There was no statistically significant HBV-NTCP binding observed, a result which could be interpreted as (i) that the NTCP receptor may not in fact be the cell surface receptor for HBV, but instead it may support HBV infection through intracellular NTCP expression as postulated previously (Somiya et al., 2016); (ii) the concentration of HBV may have been too low for binding to be detectable. Interestingly, as the HBV DNA concentration increased, increasing HBV-cell binding was observed in both Huh7 and Huh7-NTCP cells suggesting other unidentified receptors might have played a role in this interaction. The choice of the negative control could be a contribution to the unsuccessful results. As the choice was based on the assumption of either no Dane particles present in the negative sample or the SVPs would not bind to NTCP. However, both assumptions could be untrue. Therefore, the assay should have been repeated with a sample negative in both HBV DNA and HBsAg. However, the experiment was not able to be repeated because, at the time of the experiment, HBV DNA positive sera with a high viral load was unavailable. Future work also could be improved by using a cell line known to be susceptible to HBV infection, such as terminally differentiated HepaRG cells to measure the viral-cell binding.

3.5 Conclusion and future work

In this study, two chimeric NTCP expression constructs including pWPI-NTCP-BSD and pHIV-NTCP-EGFP were generated. Subsequently, several NTCP overexpressing cell lines were obtained. Initially, the HuH7-NTCP-BSD cell line was utilised in an HBV-PV entry assay. Unexpectedly, these cells alone emitted luminescence, which interfered with the interpretation of the assay. Therefore, HuH7-NTCP-eGFP (simplified as HuH7-NTCP) was developed. NTCP deglycosylation was analysed using HEK293T and HepG2 derived NTCP cells along with HuH7-NTCP cells. The results revealed that NTCP post translational modification is cell dependent. NTCP expressed in HuH7 cells was unique compared with other cell lines, where two products were observed post deglycosylation. It would be interesting to compare the NTCP expression in HuH7 against native primary human hepatocytes to see if this is the natural case. After sequencing analysis, mutations were excluded. Further mass spectrometric analysis may be required to elucidate any other post translational modification involved in these two products. NTCP biological function assays, e.g. virus-NTCP binding assay and bile acid uptake assay using plasmid pCMV.NucleoBAS were not successful. In future work, FITC-labelled pre-S1 peptide binding assay will be a better choice to test the molecular function of NTCP.

Chapter 4: Pseudotyping as a tool to investigate HBV entry pathway

4.1 Introduction

4.1.1 Foreign glycoprotein recruitment to retroviral assembly sites

While pseudotyping can be a powerful tool to investigate viral entry pathways of a broad range of virus species, there are limitations that must be considered, particularly when it comes to developing novel pseudotyped viruses. For successful formation of infectious PVs, the foreign glycoproteins must translocate to the plasma membrane for successful incorporation with the retroviral capsid (Johnson, 2011). Some glycoproteins, such as Ebola glycoprotein 1/2 and VSVG are efficiently pseudotyped; whereas others are less efficient, such as Rous sarcoma virus (RSV) envelope (Env); or almost resistant to pseudotyping, such as gibbon ape leukemia virus Env (Jorgenson et al., 2009). There is still no adequate explanation why some virus envelope proteins work better than others in this system.

In some viruses, passive incorporation takes place when glycoproteins are present in the membrane where virion assembly occurs. Hepatitis C virus (HCV) pseudotyped viruses is an example of this incorporation. HCV E1E2 glycoproteins are retained in the ER, but incomplete ER retention leads to E1E2 'leaking' to the cell surface when high levels of E1E2 are expressed in the producer cells (Hsu et al., 2003). However, with other viruses, this incorporation appears to be an active process (Jorgenson et al., 2009). Among lentiviruses, evidence suggests incorporation of Env into virions is due to direct interaction between the cytoplasmic tail domain (CTD) of the Env and

the matrix domain of the Gag proteins (Gregory et al., 2014). However, for pseudotyped viruses, this direct interaction only appears to be true between very closely related native lentiviruses that can exchange structural components, such as HIV-1 and HIV-2 (Gregory et al., 2014). Mutations in Gag do not block the incorporation of foreign Env as experiments show deleting most of the matrix domain still allows efficient incorporation of MLV Env and VSV-G into PVs (Gregory et al., 2014, Jorgenson et al., 2009). Indirect interaction mediated through a cellular intermediate provides another mechanism of foreign Env incorporation. Tail-interacting protein (TIP) 47 is required for efficient HIV-1 Env incorporation into virions. However, this cellular protein was dispensable in foreign Env recruitment (Jorgenson et al., 2009). Another mechanism to explain the foreign Env recruitment is the lipid membrane microdomains, also termed lipid rafts. Confocal imaging showed that HIV Env and Ebola glycoproteins segregated to distinct lipid membrane microdomains within the same cell, where each associated with distinct Gag particles that gave rise to virions enveloped with one or the other foreign glycoprotein (Leung et al., 2008). However, a more recent finding disagreed with this claim. Co-packaging was found between glycoproteins that were actively recruited into virions, which included MLV glycoproteins, VSV-G and Ebola Env (Gregory et al., 2014).

4.1.2 HBV pseudotyping using HIV backbone

Although pseudotyping is a popular tool in a wide field, only a small handful of studies have utilised HBV-PVs to investigate virus entry (Meredith et al., 2016, Chai et al., 2007, Sun et al., 2018) suggesting there are underlying difficulties in HBV pseudotyping. Experimental evidence showed that primary Human Hepatocytes (PHHs) but not hepatoma cell lines permitted HBV-PV infection (Chai et al., 2007). In addition, NTCP overexpressing HuH7, HepG2 and Hep3B cells were all susceptible to HBV-PV infection (Meredith et al., 2016). Chai and co-workers used *lacZ* as the

reporter gene and counted blue-stained cells to determine transducing units (TU) as a measure of infectivity. The observation that 7,000 genome equivalents (GE) of lentiviral particles were required to yield 1TU suggesting these particles have either no or insufficient envelope proteins to be infective. This low HBsAg incorporation also explained the inability to detect HBsAg, even after carrying out immunoprecipitation using anti-HBsAg followed by immunoblotting (Chai et al., 2007). They also investigated the contributions of L and S in HBV-PV formation and entry. They observed that L only did not produce infectious particles (Chai et al., 2007). This observation was in disagreement with a finding reported by Meredith and co-workers, who found that infectivity of particles generated by L alone was comparable to the particles expressing L, M and S (Meredith et al., 2016). One difference between these two groups was the producer cells used. Chai used HEK293T cell while Meredith and colleagues used HuH7-NTCP cells. It is tempting to speculate that the packaging cells contributed to this discrepancy. Some post-translational modifications may differ in HEK293T cells compared to HuH7-NTCP cells. Furthermore, some cellular proteins can also be actively recruited to the site of budding virions and co-packaged with other glycoproteins (Gregory et al., 2014, King and Tarr, 2017), which may play a role in the discrepancy observed in above two studies.

4.1.3 Pseudotyping using HBV backbone

Recently, a two-plasmid HBV trans-complementation system has been developed. One plasmid (pHBV/NL) carries a reporter gene expressing Nanoluciferase (Nluc). In this construct, 562 nt spanning from the first codon of preCore/Core was replaced with the Nluc gene (513 nt). The missing proteins are provided by a helper plasmid (pHBV/D) that lacks a packaging signal. Therefore, only the Nluc-bearing pre-genomic RNA is packaged and transduced into target cells (Nishitsuji et al., 2015). Nluc is the smallest luciferase but approximately 100-fold brighter than other

luciferases, such as Renilla luciferase. The strength of this system is that it does not alter the general properties or pathogenicity of the virus. In addition it provides a tool to quantify HBV infectivity with high sensitivity and greater efficiency than the conventional method (Nishitsuji et al., 2015).

4.1.4 Aims

Having established cell lines expressing the NTCP receptor, this chapter aimed to explore the roles of L, M & S surface antigens in an *in vitro* model of hepatitis B virus (HBV) entry, and thereby to achieve an optimal *in vitro* cell system using pseudotyping system to investigate HBV entry mechanisms.

4.2 Materials & Methods

All primers (Table 4.1) were synthesised by Eurofins Genomics and posted in lyophilised powder form. Upon receipt, by adding nuclease free H2O, primers were reconstituted to 100 pmol/µL and further diluted to 10 pmol/µL ready for PCR mix. Once reconstituted, the primers were stored at -20°C.

A study found that hepatitis D virus enveloped with surface proteins of roundleaf bat hepatitis virus (an old-world bat species) were unable to infect primary human hepatocytes (Drexler et al., 2013), therefore, an open reading frame including 675bp of synthetic Roundleaf bat surface glycoprotein was ligated into the pcDNA3.1 vector, resulting in a construct expressing roundleaf bat HBsAg (RLBS).

Table 4.1 Primers used in this study

NAME	PRIMER SEQUENCE 5' → 3'	USE
gtB.LMf	GTACCGAACAcGGAGAACATCGC	Silencing the start codon of S protein to express L, M only
gtB.LMr	AGGGTCCCCAGTCTTCGA	
gtB.LSf	CCTCAGACCAcGCAGTGGAAC	Silencing the start codon of M protein to express L and S only
gtB.LSr	ATGAGTGTCCCTTAGAGG	
gtB.MSf	CCCTTCACCAcGGGAGGTTGG	Silencing the start codon of L protein to express M and S only
gtB.MSr	TACTGGATCCGAGCTCGG	
RLBSf	CAGTGTGGTGGAATTCCACCATG GGGAACATCTCATCAGAACTC	To amplify roundleaf bat Hepatitis virus S gene with 15bp extensions at 5' end that were complementary to the ends of pcDNA3.1 vector for In-fusion cloning
RLBSr	GATATCTGCAGAATTCTTAAATGTATC CCCAGAGAAGAAAAAAGAG	
pI18.gtB.LMSf	CAGTGTGGTGGAATTCCACCA TGGGAGGTTGGTCATCCAAACCT	To amplify L, M and S of genotype B HBsAg. 15bp extensions at 5' end that were complementary to the ends of pI.18 vector
pI18.gtB.LMSr	GATATCTGCAGAATTCTTAAATGTATA CCCAAAGACAAAAGAAAATTGG	
pI18.Sf	CAGTGTGGTGGAATTCCACCA ACATGGAGAACATCGCATC	To amplify S only of genotype B HBsAg. 15bp extensions at 5' end that were complementary to the ends of pI.18 vector
pI.18.Sr	GATATCTGCAGAATTCTTAAAT GTATACCCAAAGACAAAAGAA	
qHBVf	GATGTGTCTGCGGCGTTTTA	To amplify a small amplicon for quantification of HBsAg expression
qHBVr	CTGAGGCCCACTCCCATAGG	

Table 4.2 HBsAg expression plasmids

NAME	PROPERTIES	NOTES
gtB.LMS	L. M & S of Hepatitis B geneotype B cloned into pCDNA3.1	This plasmid had been previously developed in-house
gtB.L	A mutant of gtB.LMS with silenced start codon for M & S	
gtB.S	A mutant of gtB.LMS with silenced start codon for L & M	
gtB.LS	A mutant of gtB.LMS with silenced start codon for M	
gtB.MS	A mutant of gtB.LMS with silenced start codon for L	
gtD4.LMS	L. M & S of Hepatitis B geneotype D cloned into pCDNA3.1	This plasmid had been previously developed in-house
gtD4.LM	A mutant of gtD4.LMS with silenced start codon for S	
gtD4.S	A mutant of gtD4.LMS with silenced start codon for L & M	
pI18.gtB.LMS	gtB.LMS cloned into pI18 vector	
RLBS	Plasmid expressing surface protein of Round leaf bat hepatitis B	Used with packaging vector and reporter plasmid to produce round leaf bat Hepatitis B pseudotyped viruses
PWPI.LMS.BSD	Lentiviral shuttle plasmid facilites ligation of LMS (genotype D) ORF into the host genome	
PWPI.S.BSD	Lentiviral shuttle plasmid facilites ligation of S ORF (genotype D) into the host genome	

To silence start codons of L, M, S, ATG was substituted by AcG at corresponding start codons. These substitutions were incorporated into the parental plasmid via the use of primers carrying the mutations.

Table 4.3 Other plasmids used in this study

NAME	PROPERTIES	APPICATION
PNL4.3.LUC.R$^-$E$^-$	1st generation lentiviral packaging plasmid	Used with Envelope expressing plasmid to generate PVs[2]
PNL4.3.NLUC.R$^-$E$^-$	A mutant of pNL4.3.luc.R-E- by replacing luciferase gene with nano luciferase gene	Used with Envelope expressing plasmid to generate PVs
PSPAX2	2nd generation lentiviral packaging plasmid	Used with Envelope expressing plasmid & reporter plasmid to generate PVs
PCSFLW	Firefly luciferase reporter plasmid	Used with packaging vector & Envelope expressing plasmid to generate PVs that can be detected by luminescence reader
MULV.GAG/POL.ΔENV	Retroviral vector packaging plasmid	Used with Envelope expressing plasmid to generate PVs
PCSGW	Green fluorescence reporter plasmid	Used with packaging vector to generate PVs that can produce green fluorescence under light excitation
PCMVR8.74	2nd generation lentiviral packaging plasmid	Used with Envelope expressing plasmid & reporter plasmid to generate PVs
PMD2.G (AKA[1].VSVG)	Plasmid expressing vesicular stomatitis virus glycoprotein	Used with packaging vector to generate PVs
PSIN4MRFP	mCherry fluorescence reporter plasmid	Used with packaging vector & Envelope expressing plasmid to generate PVs that can produce red fluorescence under light excitation
PCDNA3.1D/V5-HIS-TOPO®H77 (AKA*. H77)	Hepatitis C E1E2 glycoprotein expressing plasmid	Used with packaging vector and reporter plasmid to produce Hepatitis C PVs
PCNDA3.1J6 (AKA*. J6)	Hepatitis C E1E2 glycoprotein expressing plasmid	Used with packaging vector and reporter plasmid to produce Hepatitis C PVs

1. aka—also known as; 2. PVs- pseudotyped viruses

4.2.1 Q5 Site-directed mutagenesis

Q5 Site-directed Mutagenesis (SDM) kit was used. Primers used for the SDM were designed using NEBaseChanger (New England Biolabs Inc.). Start codons at the beginning of preS1, beginning of preS2 and beginning of S were silenced by the substitution of ATG to ACG.

The DNA template for all SDM reactions was gtB.LMS (HBV genotype B wild type surface sequence inserted in pcDNA 3.1D/V5-His-TOPO). This plasmid has been previously developed in-house. All SDM reactions were conducted following Q5 SDM kit protocol. One to 25 ng template DNA were amplified by Polymerase Chain Reaction (PCR) using the appropriate primers with initial denaturation at 98°C for 30 seconds, followed by 25 cycles of denaturation at 98°C for 10 seconds, annealing at 60°C for 15 seconds, extension at 72°C for 3 minutes, final further extension at 72°C for 2 minutes. The resulting PCR products were confirmed by gel electrophoresis running on 2% agarose gel. A 5 - 6 kb band was assumed to contain the mutated sequences. Upon confirmation by Sanger sequencing, 1 µL PCR products were then incubated in *kinase-Ligase-Dpn1* (KLD; NEB) mixture for 5 minutes at room temperature. Two µL resulting *KLD* mix was transformed into 25 µL competent *E.coli* cells (C2987; NEB®). The bacterial culture was then plated on a *Luria Bertani* (LB, sigma) agar plate containing 100 $\mu g.mL^{-1}$ ampicillin and incubated at 37°C overnight.

4.2.2 Generation of pNL4.3Nluc.R^-.E^-

To enhance detection of entry of HBV-PVs, a derivative of pNL4.3Luc.R^-E^- was generated. The firefly luciferase gene was substituted by a gene encoding Nanoluc luciferase. First, the luciferase gene was removed from the vector by a double restriction enzyme digestion method using *NotI* and *XhoI* cleavage. Upon confirmation

of successful double digestion via gel electrophoresis, the gel with the band representing the digested pNL4.3R$^-$E$^-$ was harvested and further purified. Meanwhile, the NanoLuc^{TM} Luciferase (NLuc) gene from pNL1.1 (a gift from Dr. Lucy Wheatley, Promega) was PCR amplified. The primers used were 5'-ATGGGTGGCGCGGCCGCCACCATGGTCTTCACA-3' (forward) and 5'-GTTTTTCTAGGTCTCGATTACGCCAGAATGCGTTCG-3' (reverse). Using an In-Fusion cloning method according to the manufacturer's protocol, the Nluc gene was inserted into the linearised pNL4.3R$^-$E$^-$ vector (Figure 4.1).

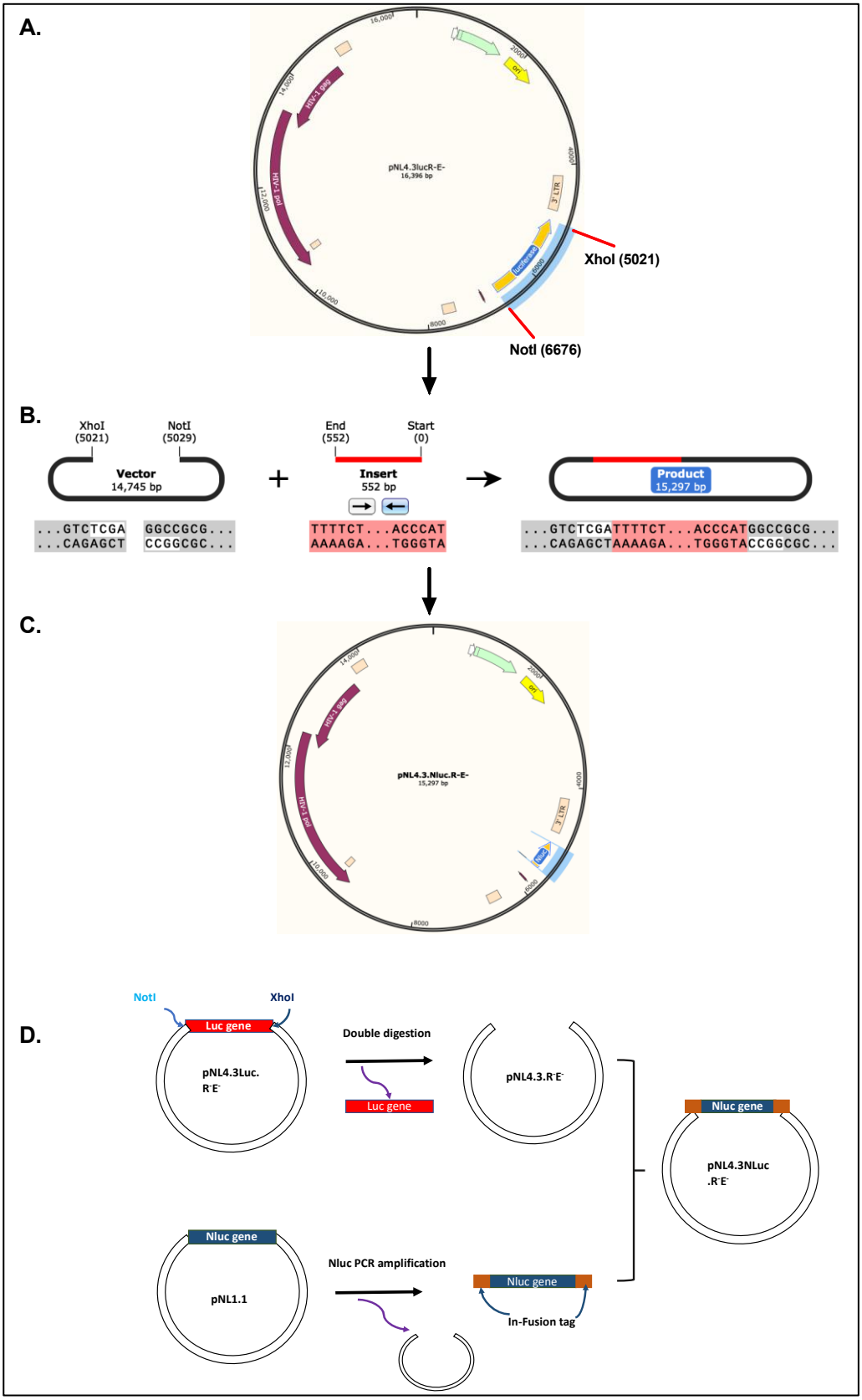
A.
pNL4.3lucR-E-
16,396 bp
XhoI (5021)
NotI (6676)
B.
XhoI
(5021)
NotI
(5029)
Vector
14,745 bp
...GTCTCGA
...CAGAGCT
GGCCGCG...
CCGGCGC...
End
(552)
Start
(0)
Insert
552 bp
TTTTCT...ACCCAT
AAAAGA...TGGGTA
Product
15,297 bp
...GTCTCGATTTTCT...ACCCATGGCCGCG...
...CAGAGCTAAAAGA...TGGGTACCGGCGC...
C.
pNL4.3.Nluc.R-E-
15,297 bp
D.
NotI
XhoI
Luc gene
pNL4.3Luc.
R-E-
Double digestion
Luc gene
pNL4.3.R-E-
Nluc gene
pNL4.3NLuc
.R-E-
Nluc gene
pNL1.1
Nluc PCR amplification
Nluc gene
In-Fusion tag

Figure 4.1 Replacement of luciferase (luc) gene with nano-luciferase (Nluc) gene in pNL4.3.luc.R-E- vector

***A**. With the aid of SnapGene software, NotI and XhoI were identified as the unique enzymes to remove the luciferase gene from the packaging vector; **B.** The linearised vector has sticky ends post double digestion and the Nluc gene with the vector tags is ligated into the vector in the same orientation as the luc gene; **C.** The completed pNL4.3.Nluc.R-E- (Nluc is highlighted in blue). **D.** The process of the construction of plasmid pNL4.3Nluc.R$^-$.E$^-$. First, luciferase gene was removed from pNL4.3luc.R$^-$.E$^-$ using NotI and XhoI double digestion; and then the Nluc gene amplified from pNL1.1 (gift from Dr. Lucy Wheatley, Promega) was ligated to the double digested pNL4.3.R-E- by In-Fusion cloning. (Luc, Luciferase; Nluc, NanoLuc Luciferase)*

4.2.3 Generation of Stable cell lines expressing HBsAg

HEK293T cells were transfected with three plasmids, 2 µg of each, including pCMVR.87.4 (encoding HIV packaging proteins), pMD2.G (encoding VSV glycoproteins) and pWPI.BSD bearing either LMS or S (genotype D) ORFs (containing the transgene expressing either LMS or S). Seventy-two hours later, supernatants were collected and filtered through a 0.45 µM pore size membrane, 100 µL of the filtered supernatants were used to infect HEK293T cells in a 6 well plate seeded 24 hours prior to the infection. After incubation for a further 72 hours, cells were challenged with Blasticidin (20 $\mu g.mL^{-1}$). Blasticidin-resistant cells were further expanded and aliquots were stored in liquid nitrogen.

4.2.4 Restriction enzyme digestion

During the site directed mutagenesis (SDM), the *DpnI* restriction enzyme (20000 units/mL; NEB) was used to remove any template plasmid as this enzyme only recognises and cleaves methylated DNA. Incubation at 20 °C for 5 minutes was sufficient to remove the methylated plasmid. During the removal of luciferase gene from pNL4.3Luc.R^-E^- to generate linearised pNL4.3R^-E^-, 3 µL of each NotI (20000 units/mL; NEB) and *XhoI* (ThermoFisher) were used in a 30 µL of reaction to digest a range of 4 - 5 µg of DNA. The digest mix was incubated at 37 °C overnight.

4.2.5 DNA gel purification

Post double digestion, in order to remove undesired DNA fragments, 30 µL of the digest mix was loaded into a 2 cm width well in a 1% w/v agarose gel. The gel was run at 60 V for 60 minutes until the small luciferase fragments were well separated from the pNL4.3R^-E^- vector. Under UV light, using a clean scalpel, the desired DNA

product was harvested by carefully and quickly removing the gel surrounding the correct band. The linearised DNA was harvested using a NcleoSpin® gel & PCR clean up kit (MACHEREY-NAGEL) according to manufacturer's manual.

4.2.6 Single round infection assay

HBV-PVs were generated by transfecting HEK 293 cells with a two-plasmid system comprising 2 µg of an HBV surface protein expressing plasmid (Table 4.2), and 2 µg of a lentiviral backbone and luciferase expressing plasmid. Plasmid mix was incubated with linear polyethylenimine (PEI, MW 25 kDa) at a ratio of 1:6 for 30 minutes prior to the transfection. Opti-MEM was used as the medium during the transfection. The density of HEK 293 cells was 1.2×10^6 in 10 cm^2 dish and seeded 24 hours prior the transfection. Seventy-two hours post transfection, PVs were harvested by passing the supernatant through a 0.45 µM syringe filter, 100 µL of which was added into target cells either HuH7 ± NTCP or HepG2 ± NTCP cells. These cells were seeded 24 hours prior to the infection in a white flat bottom 96 well plate (Costar) at cell density of 1.5×10^5 cells/mL (100 µL per well). Four hours after inoculation with the PVs, the supernatants were removed and 150 µL fresh complete DMEM was added into each well. Seventy-two hours post infection, cells were lysed using 50 µL/well fluorescence lysate 5x reagent (Promega) diluted with H_2O in 1:4 ratio. The resulting lysates were then read using FLUOstar Omega microplate reader (BMG LABTECH). Results were processed using GraphPad Prism 8.

4.2.7 Cell viability assay

The cell viability assay utilised CellTiter-blue (Promega). HuH7-NTCP cells were seeded into a Corning® 96-well flat clear bottom white plate with incremental density at 1.5×10^4, 2.0×10^4, 2.5×10^4 and 3.0×10^4 per well. Each cell density assessment

was performed in triplicates and repeated four times. Cells were maintained in complete DMEM and incubated at 37°C with 5% CO2. Seventy-two hours later, CellTiter-blue reagent (20 µL per well, Promega) was added and cells were incubated at 37°C for a further 1 hour before measuring fluorescence ($560(20)_{ex}$/$590(10)_{em}$) using FLUOstar Omega microplate reader.

4.2.8 Analysis of HBsAg expression

4.2.8.1 Western blot

On day 3 after transfection, the supernatants were collected and concentrated at 40,000 rpm (Beckman Coulter) at 4°C for 2.5 hours through a 20% sucrose cushion. The pellets were then resuspended in 100 µL PBS. The transfected cells were lysed as described in section 2.2.13 Lysis of mammalian cells. Before western blotting, total proteins were quantified using the Pierce™ BCA Protein Assay Kit (Thermo Fisher Scientific) as per the manufacturer's instruction. Ten to 20 µg of total proteins of each sample were then boiled at 95°C for 10 minutes in reducing buffer. Reduced proteins were separated by SDS-PAGE (12% Mini Protean® Precast gel, BioRad) before being transferred onto a PVDF membrane (Amersham Hybond P 0.45 µM, GE Healthcare) and probed with anti-HBsAg primary antibodies as indicated in the results session. Proteins were then visualised using a G:Box imaging system (SynGene).

4.2.8.2 Immunoprecipitation

Fifty µL Dyna-beads protein G (Invitrogen, ThermoFisher Scientific) was placed on a magnetic rack followed by removal of the supernatant (this step was required whenever the beads were resuspended in solution, for simplicity, this step is not described from this point onwards). The beads were mixed with antibody by the addition of one µL HBIG in 200 µL PBST (1% PBS with 0.02% Tween 20) and thoroughly mixed with rotation at room temperature. After 15 minutes, the beads / antibody mixture was washed twice with 200 µL PBST and then incubated with 1000

µL testing samples overnight at 4°C with agitation. On the following day, the sample was further washed three times with 200 µL PBST. Next, 20 µL of 50 mM glycine pH 2.8 and 20 µL of 2x Laemmli loading buffer were added into the mixture followed by heating at 70°C for 10mins. Finally, the sample was placed on the magnetic rack and 13 µL supernatant was directly loaded onto a 12% Mini-Protean TGX^{TM} precast gel (Bio-Rad) for polyacrylamide gel electrophoresis. The resulting gel was further processed via silver staining.

4.2.8.3 Silver staining

Silver stain was performed using SilverQuestTM Silver Staining Kit (ThermoFisher Scientific) as per the manufacturer's instruction. In brief, after protein separation, the polyacrylamide gel was placed in a clean staining tray and washed briefly with ultrapure water. The gel was then fixed with 20 mL of fixative solution containing 40% ethanol, 10% acetic acid and 50% ultrapure water for 20 minutes with gentle agitation. After that the gel was washed with 10 mL of 30% ethanol for 10 minutes followed by incubation with 10 mL of sensitising solution. After 10 minutes, the sensitising solution was removed and the gel was washed with 10 mL of 30% ethanol for 10 minutes and 10 mL of ultrapure water for 10 minutes. Afterwards, the gel was incubated in 10 mL of staining solution for 15 minutes before being again washed with 10 mL of ultrapure water for 30 seconds. Immediately after washing, the gel was incubated in 10 mL of developing solution for approximately 4 - 8 minutes until the desired bands intensity was achieved and then the action of developing solution was stopped by adding 10 mL of stopper with agitation until the colour turned from pink to colourless. Finally, the gel was washed with 10 mL of ultrapure water and was then ready for imaging.

4.2.8.4 Indirect immunofluorescence

Plasmid gtD4.LMS was transfected into HEK293T cells on coverslips in a 6-well plate. Forty-eight hours later, indirect immunofluorescence was performed and used the protocol as described in section 3.2.3 Indirect Immunofluorescence (IF) assay. The primary antibody used was 100 µL of human HBV immunoglobulin (HBIG) (a kind gift

from Dr S Ijaz, Public Health England) diluted at 1:100 in 1% BSA/PBS. The secondary antibody was 100 µL of mouse anti-human IgG (H+L) Alexa Fluor 488 (Southern Biotech) diluted 1:1000 in 1% BSA/PBS.

4.2.9 Statistical analysis

Statistical analyses were conducted as described in section 3.2.11.

4.3 Results

4.3.1 Site Directed Mutagenesis (SDM)

To obtain genome type B HBV. S (gtB.S, silencing start codons for L and M, i.e. only S expressed) and gtB.L (silencing start codon for M and S, i.e. only L expressed), the start codons of L, M or S were silenced by substituting ATG with ACG accordingly. The SDM used DNA PCR amplification with primers bearing the mutation, as specified in (Table 4.1). The PCR products were analysed on 2% agarose gels. PCR products with the expected size of >6000 bp were then transformed into NEB Q5-alpha Competent *E.Coli* cells and incubated in a LB agar plate containing 100 $\mu g.mL^{-1}$ ampicillin at 37°C overnight. Post overnight incubation, resulting colonies were screened using backbone specific primers T7F and bGHR. Colonies contained plasmids with expected bands at ~1300bp were assumed to be the desired colonies. Overnight minipreps were set up and the resultant plasmids were sent for Sanger sequencing. Figure 4.2 shows a typical example of successful start codon silencing where T was replaced by a C. Following successful single start codon knock out, mutants bearing L, M or S knock out were amplified with primers carrying second start codon knock out, thereby generating LM and MS start codon knock out to yield gtB.S and gtB.L.

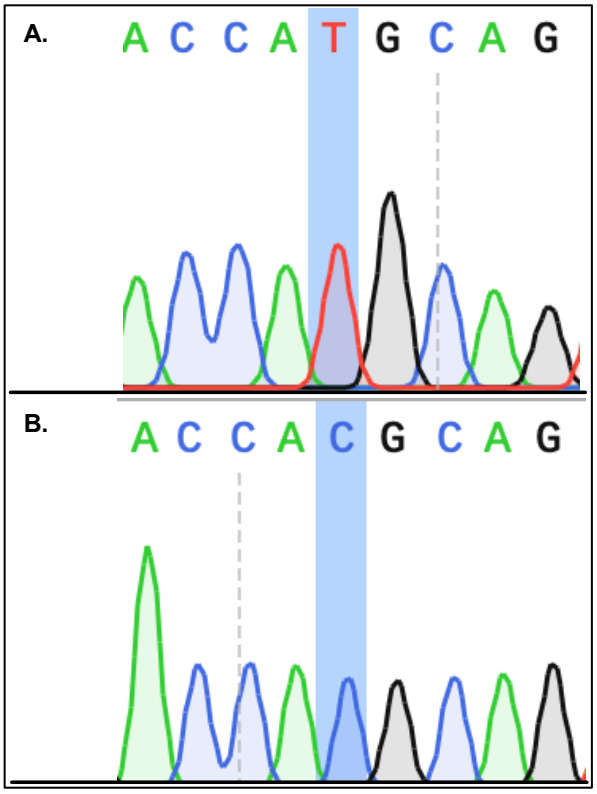

Figure 4.2 Site directed mutagenesis to silence start codons

Post exponential amplification, gel electrophoresis confirmed the right size of amplified DNA, Colony screening and gel electrophoresis was used to select colonies harbouring the right size of DNA fragments, i.e. ~1300kb; plasmid preps were sent for Sanger sequencing using T7F and bGHR, the reads were aligned against the reference sequence. ***A.*** *ATG in the reference sequence;* ***B.*** *ACG in the mutant.*

4.3.2 Protein detection

To determine whether the HBsAg expression plasmids developed in this experiment could produce correct size of HBV surface proteins, HEK293T cells were transfected with the HBsAg expression plasmids. The resulting supernatants and cells lysates were subject for western blot. However, choosing a suitable antibody for this task was proven challenging.

The first anti-HBsAg antibody tested was ab9193 (Abcam). This polyclonal antibody was raised in a horse. Very high background noise was observed in each attempt. Figure 4.3 shows a typical blotting image using this antibody. In this experiment, HEK293T cells were transiently transfected with 2 µg of gtB.LMS and increasing concentration of gtB.S (from 0.3 - 2 µg). Round leaf bat HBsAg (RLBS) expressing plasmid was used as a negative control (lane 11). A recombinant protein (ab91276, Abcam) in lane 1 was used as a positive control in these assays. However, as shown in Figure 4.3, the antibody ab9193 (Abcam) gave too much background staining and specific signal was hard to resolve. As such, this antibody was not used further.

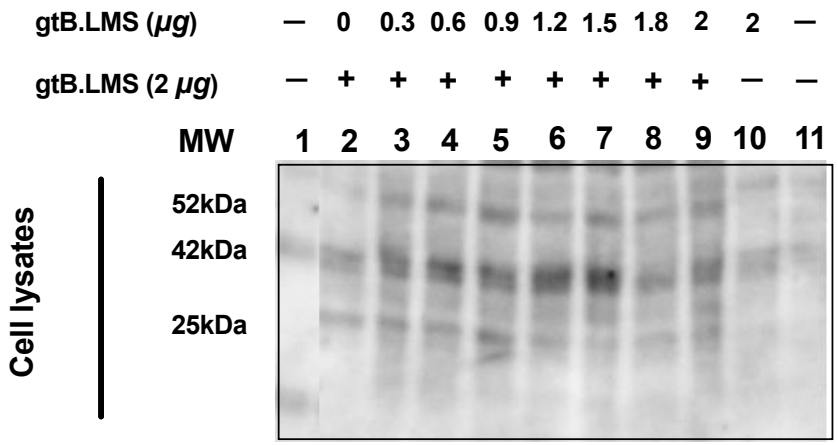

Figure 4.3 Detection of large (L), Middle (M) and Small (S) proteins in producer cells.

HEK293T cells were transfected with 2 μg of gtB.LMS and gtB.S at incremental concentration as indicated. Seventy-two hours later, cells were lysed in 1mL lysis buffer. Ten μL aliquot of each sample was used for western blotting. The primary antibody used in this assay was horse polyclonal anti-HBsAg antibody (ab9193, Abcam) at dilution 1:500. The secondary antibody was rabbit anti-horse IgG H+L (HRP) (ab6921, Abcam) at dilution 1:1000. Lane 1, HBsAg recombinant protein, ab91276 (Abcam); lane 11, RLBS (round leaf bat small surface antigen) was prepared as a negative control.

Due to unsuccessful detection of surface proteins, two more anti-HBsAg antibodies were tested. S26 (Ma1-7603, ThermoFisher) was a monoclonal antibody raised in mouse. Tested applications included IP, ELISA and immunocytochemistry (ICC). Monoclonal antibody 86C (ab20758, Abcam) was also raised in mouse and claimed suitable for ELISA and WB. However, application of both antibodies failed to detect the proteins of interest, although the positive control HBsAg protein ab9193 was detected by both antibodies. This raised a question whether there was any HBsAg presented in the samples or the concentration of HBsAg was just too low to be detectable.

Attempts were made to enhance the amounts of HBsAg recovery. Various methods were tried including RIPA (radioimmunoprecipitation assay) buffer instead of triton X-100 buffer and isolation of endoplasmic reticulum, but none yielded better protein detection (data not shown).

Finally, Mab1694, clone 1A10 (Abnova) was obtained. This is a mouse monoclonal antibody raised against the full-length recombinant HBV surface protein, which corresponds to the full length of a serotype HBsAg *ay*. Serotype *ay* is determined by an arginine (R) at amino acid-122 of the S protein. Whereas lysine (K) in this position is determined as serotype *ad* (Kramvis et al., 2005). Accordingly, the serotype of S expressed by gtB.S should be *ad,* while the S expressed by gtD4.S or gtD13.S should be *ay* (Figure 4.4A). After concentration optimisation, the optimal antibody working dilution was determined at 1:2000 (data not shown).

Using Mab1694, the first samples tested were the cellular lysates of HEK293T that had been transiently transfected with plasmids expressing S including gtB.S, gtD4.S, gtD13.S and L expressing plasmid gtB.L. Typical doublet bands representing the glycosylated and unglycosylated S protein at molecular weight (MW) 27 and 24kDa

respectively were observed in sample gtB.S, while bands at equivalent sizes but much lower intensity were observed in sample gtD13.S. Sample gtD4.S yielded visible S bands but the intensity was again much weaker than its counterparts. As expected, plasmid gtB.L yielded the glycosylated and unglycosylated bands at sizes of 42 and 39kDa. No bands were seen in the negative control where only cellular lysates of wild type HEK293T were used (Figure 4.4B).

To determine whether gtB.LMS could express L, M, S and any changes would occur if adding gtB.S, western blot of the cell lysates that transfected with gtB.LMS and gtB.S were performed. As demonstrated in Figure 4.4C, in sample gtB.LMS, the doublet of L appeared at highest intensity followed by the doublet of S, and a doublet of M was just fractionally visible. By adding gtB.S, the intensity of the S doublets was clearly enhanced (Figure 4.4C).

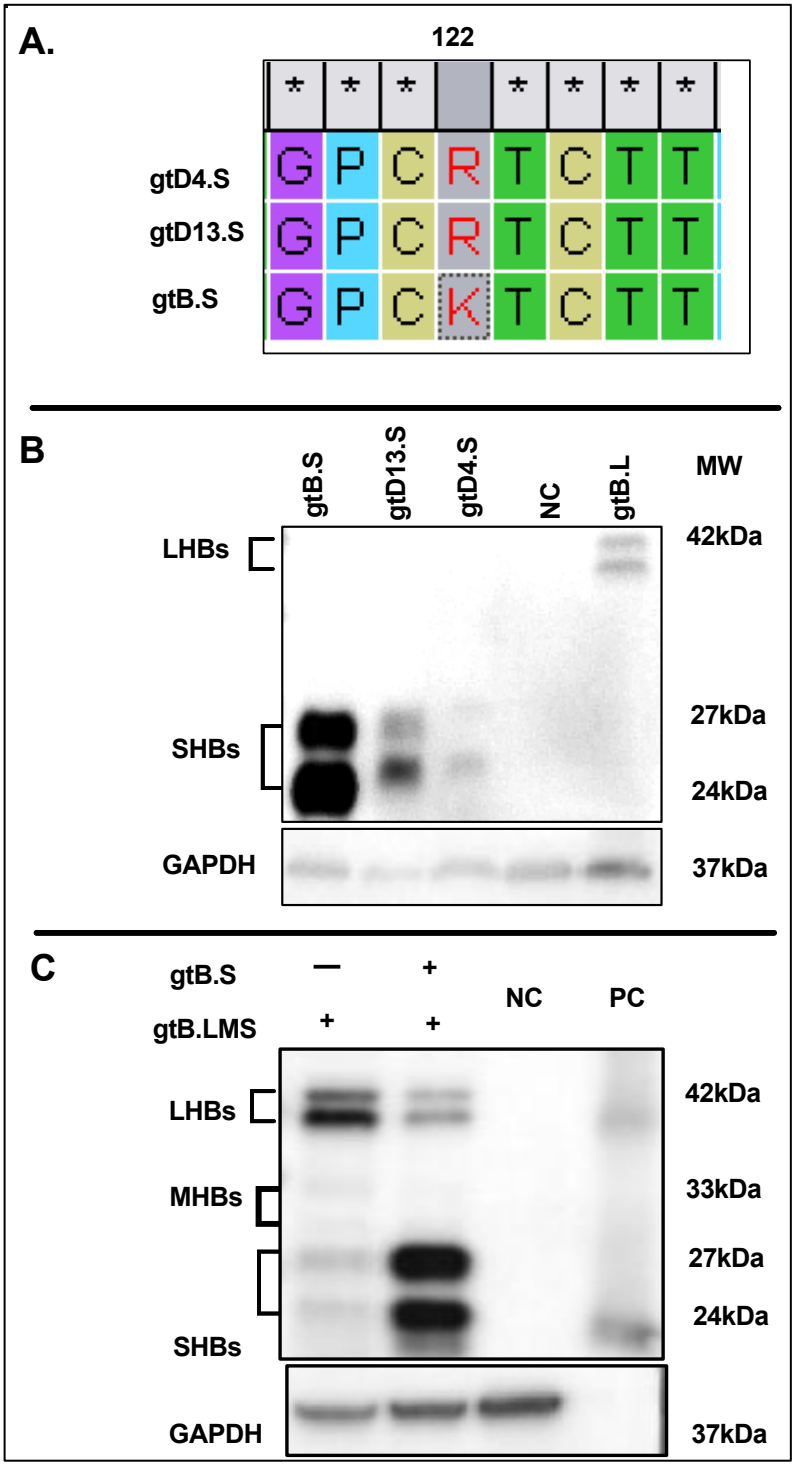

Figure 4.4 Western blot analysis of HBsAg production using Mab1694

***A**. S sequence of gtD4.S, gtD13.S and gtB.S were aligned in MEGA 7, the amino acid presented at position 122 is highlighted in red; **B**. Cellular lysates of transiently transfected HEK293T expressing the indicated various genotypes of HBsAg probed with anti-HBsAg monoclonal antibody Mab1694; **C**. Cellular lysates of transiently transfected HEK293T expressing gtB.LMS with or without adding gtB.S were probed*

with Mab1694. gtB.S: expressing the small (S) surface protein of genotype B HBV; gtD4.S and gtD13.S: expressing the S protein of genotype D (the numbering was determined in our lab for the purpose of differentiation); gtB.L: expressing L protein of genotype B; gtB.LMS: expressing L, M and S proteins of genotype B HBV. NC: negative control using cell lysates of HEK293T only; PC: positive control was a recombinant HBsAg (ab91276, Abcam). GAPDH: Glyceraldehyde 3-phosphate dehydrogenase, used as a loading control.

While experiencing difficulties in detecting HBsAg via WB at early stage, indirect IF was performed to observe the expression of HBsAg in producer cells. Using HBIG and mouse anti-human Fc Alexa Fluor 488 (Southern Biotech), HBsAg proteins were visible and they were mainly located in the endoplasmic reticulum (ER) region (Figure 4.5.1 & 2). However, less than 5% of transfected HEK293T cells were expressing HBsAg. To rule out whether underperformance was due to the secondary antibody or any technical errors, H77, a plasmid expressing E1E2 envelope proteins of hepatitis C virus, was studied in parallel. Anti HCV E2 Mab Ap33 was used to detect E1E2, while mouse anti-human Fc Alexa Fluor 488 was used as the secondary antibody. Greater than 15% HEK293T cells were observed expressing E1E2 (Figure 4.5.3). RLBS expression was also prepared in the same manner as that in gtB.LMS to rule out false positive signals. Figure 4.5.4 shows that no fluorescence was detected in this cell line staining with HBIG, which confirmed that the HBsAg observed in gtB.LMS transfected cells was genuine.

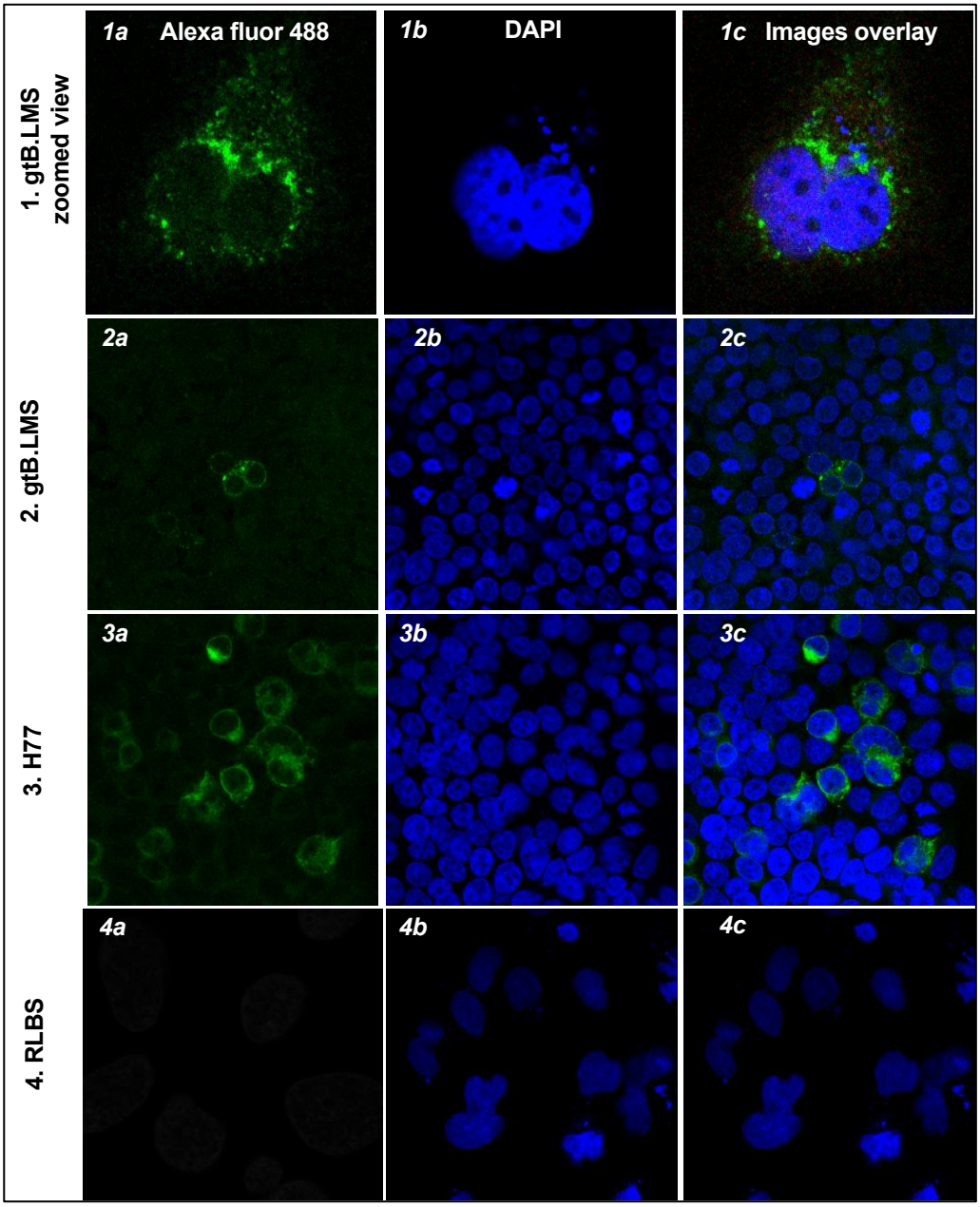

Figure 4.5 Assessment of HBsAg expression in producer cells.

Plasmid gtB.LMS was transfected into HEK293T cells on coverslips in a 6-well plate. Forty-eight hours later, cells were fixed with 4% formaldehyde and permeabilised with 0.1% Triton x-100 solution. After blocking with 1%BSA/PBS, cells were then incubated with Human HBV immunoglobulin (HBIG) diluted at 1:100 in 1%BSA/PBS at room temperature and followed by incubation with mouse anti-human IgG (H+L) Alexa Fluor 488 (Southern Biotech) diluted 1:1000 in 1%BSA/PBS. Cells were finally stained with 0.04 µg /mL DAPI (4',6-Diamidino-2-Phenylindole Dihydrochloride;

ThermoFisher). HBsAg expression was visualised using a Zeiss LSM710 Confocal microscope. ***1.*** *Zoomed view of HBsAg expression in HEK293T cells;* ***2.*** *HBsAg expression in HEK293T cells;* ***3.*** *H77 was a positive control to ensure the secondary antibody was functioning and the procedure was correct. Anti HCV E2 Mab Ap33 was used as the primary antibody;* ***4.*** *RLBS5 was prepared in parallel and used as a negative control.*

4.3.3 Single round infection assay

To limit the impact of technical errors and experimental bias, all infection assays in this study were performed in triplicate. To eliminate cell seeding technique as a source of error, cell viability assay using CellTiter-blue (Promega) was performed. HuH7-NTCP cells were seeded into a 96 well plate with increasing density at 1.5×10^4, 2×10^4, 2.5×10^4 and 3×10^4 cells per well. Each cell density was performed in triplicate and repeated four times. Each bar shown in Figure 4.6 was the mean value of the triplicate and the error bars showing the standard variation between the triplicates. The small errors bars indicated small variation among triplicates. The number of relative luminescence units (RLU) across groups also indicated even cell density among them (Figure 4.6).

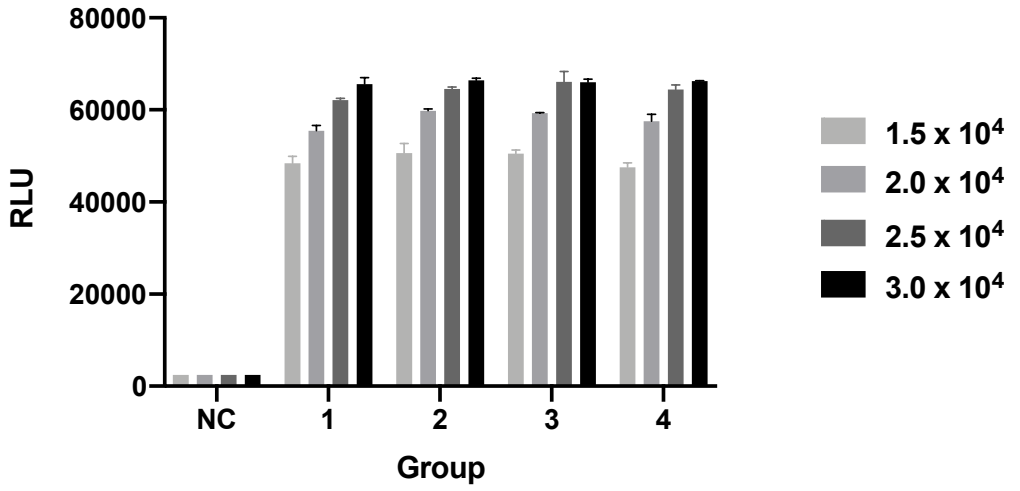

Figure 4.6 Cell viability was checked using Cell-Titer Blue (Promega).

Four experiments with increasing number of cells seeded in 96 well plate for 72 hours before cell viability was estimated. Blank wells were served as negative control (NC); RLU, relative luminescence unit.

4.3.4 Infectivity of HBV-PVs using HBsAg expression plasmid cloned in the pcDNA3.1 vector

HEK293T cells were used to produce HBV-PVs. A two-plasmid system, including an HIV-1 backbone/luciferase reporter plasmid (pNL4.3.luc.R-E-) and an HBsAg expression plasmid was used. All assays were performed in triplicate. As described in Chapter 3 (Application of NTCP-BSD overexpression cells), Delta E (ΔE) was used as a negative control. As a positive control, the E1E2 genes from widely used HCV strains either J6 or H77 were cloned into the pcDNA3.1 vector. These E1E2 expression plasmids were constructed previously in-house and was reported previously (Urbanowicz et al., 2015). For simplicity, these two E1E2 expression plasmids are referred to as J6 and H77. Initially, J6 was used, but it seemed to perform poorly across the lab after a period of use. While this issue was dealt with by senior researchers, the plasmid encoding the E1E2 genes of strain H77 was used instead.

Once the HuH7-NTCP cell line was established, HBV-PV infectivity assays were performed using this cell line. As shown in Figure 4.7a, the background signal from this cell line was as low as the wild type HuH7 cells. Though the signal from the gtB.LMS-possessing PV-infected HuH7-NTCP appeared to > 0.5 log greater compared to that from the ΔE infected HuH7-NTCP, the discrepancy in between was not statistically significant (Figure 4.7a). In search of a better negative control, plasmid RLBS was developed. As demonstrated in Figure 4.7b, the difference between these two negative controls did not reach statistical significance. To find out whether different producer cells could improve HBV-PV generation, HEK293T and HuH7-NTCP were co-transfected with pNL4.3.luc.R-E- and gtB.LMS. The resulting supernatants were collected and used to infect HuH7 and HuH7-NTCP cells. No significant higher infectivity (RLU) was observed in HBV-PVs generated in HuH7-NTCP cells (Figure 4.7c & d).

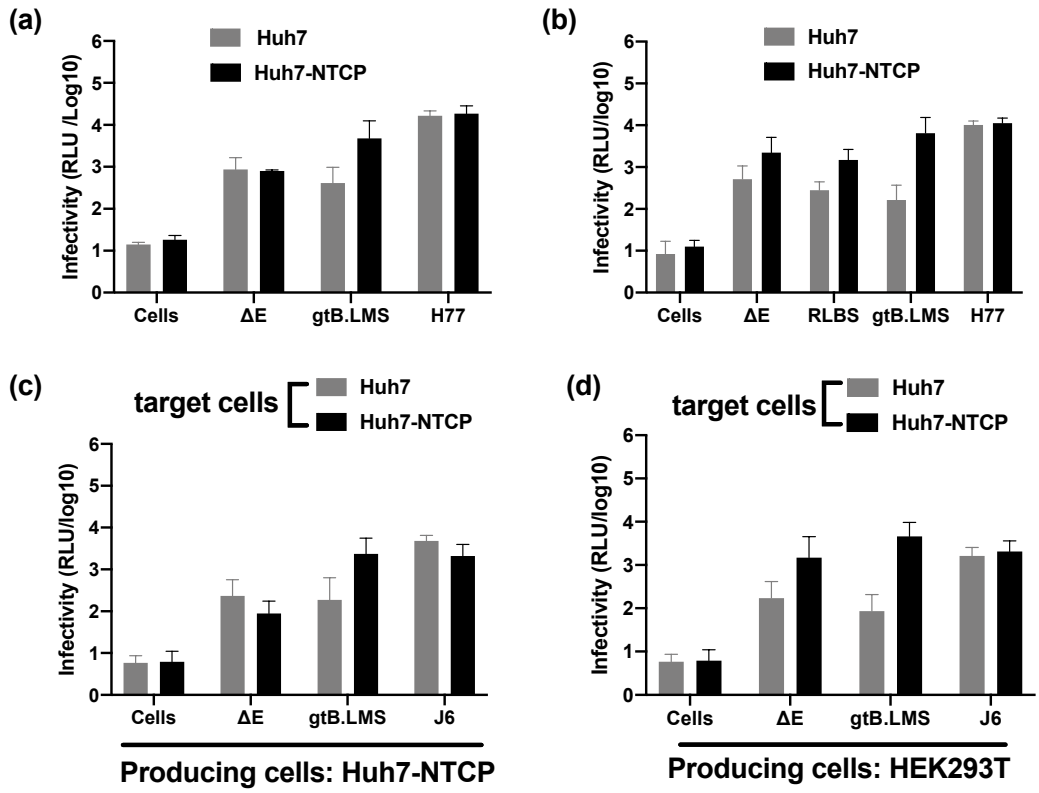

Figure 4.7 Optimisation of infection assay using transient HBsAg expression Luciferase reporter assay to measure HBV-PV infectivity

HEK293T cells were transfected with plasmid pNL4.3.Luc.R-E- (expressing lentiviral backbone and luciferase reporter gene) and plasmid expressing HBV surface antigen. Plasmids encoding the E1E2 genes of hepatitis C strains either J6 and H77 were prepared as a positive control; ΔE was lentiviral packaging without surface antigens prepared as a negative control; Cells were used as another form of negative control. Seventy-two hours later, PVs were harvested and used to infect cells as indicated in the graphs. After a further 72 hours incubation, HBV-PV infectivity was determined by reading luciferase activity shown as relative luminescence unit (RLU). ***(a).*** *HBV-PVs were prepared with pNL4.3.Luc.R-E- and used to infect HuH7 and HuH7-NTCP-eGFP, which is termed as HuH7-NTCP here onwards.* ***(b).*** *RLBS, a plasmid expressing the surface glycoprotein of roundleaf bat, was used as a negative control to compare with ΔE;* ***(c).*** *The infectivity of HBV-PVs produced in HuH7-NTCP cells was compared with that produced in* ***(d).*** *HEK293T cells.*

4.3.5 Development of pNL4.3.Nluc.RE- packaging vector

Higher luminescence intensity was consistently observed in HBV-PVs infected HuH7-NTCP cells compared to that in parental HuH7 cells and negative controls. However, statistical significance was not achieved on any occasion. Therefore, to improve sensitivity of this assay, a novel plasmid with HIV-1 backbone and nano luciferase (Nluc) reporter expressed *in cis* was developed. The Nluc gene (516 bp) from pNL1.1 was inserted into the *NotI/*XhoI site of pNL4.3.Luc.RE- (5026-6680 nt). These *NotI/XhoI* sites of pNL4.3.Luc.RE- was determined using SnapGene (Figure 4.1).

Post *NotI/XhoI* double digestion overnight at 37°C*,* the luciferase gene was removed from this vector. As shown in Figure 4.8A & B, colony 1 appeared having an unexpected band and colony 2 was free from this problem, as a result, 30 µL of colony 2 was further separated on 1% gel (Figure 4.8C). Post gel separation, the desired band was removed from the gel and purified using NucleoSpin® Gel and PCR Clean-up kit (Macherney-nagel). The DNA recovery was very low with a concentration at 39 ng/µL. Nevertheless, the linearised vector was used for In-Fusion cloning.

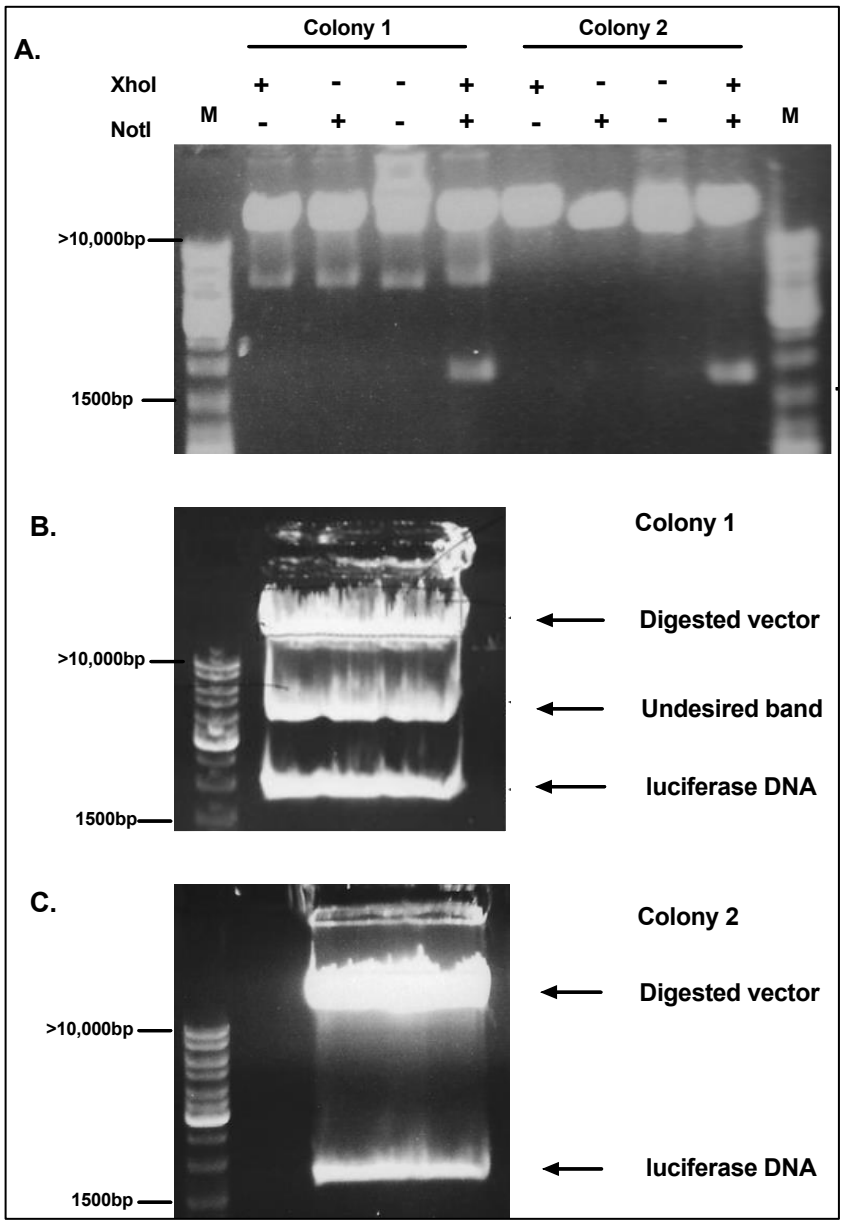

Figure 4.8 pNL4.3.luc.R-E- double digestion viewed in ethidium bromide stained 1% agarose gel

A. *Plasmids purified from two colonies were digested with either XhoI, NotI, or both. Incubation with buffer only were prepared in parallel as a negative control;* ***B.*** *Post double digested plasmid of the colony 1, three bands were observed. One band larger than 10,000bp was expected as the digested vector; the band size at ~1,600bp was assumed the luc gene; a band size at ~3,500bp was unexpected;* ***C.*** *Two bands with sizes indicating the digested vector and the luciferase fragment were obtained in post double digested plasmid of the colony 2.*

Meanwhile, the Nluc gene was amplified with primers that introduce *NotI* and *XhoI* restriction sticky ends which flank the sticky ends of digested pNL4.3.RE- using primer set Nluc1.f and Nluc1.r. These amplified Nluc fragments were incubated with linearized pNL4.3.RE- by In-Fusion cloning and subsequently transformed into *DH5a* cells. Small numbers of colonies were observed the following day. Using a pair of backbone primers, the colonies were screened and no colony carried the desired inserts (data not shown). This could be due to the poor preparation of the recipient vector.

To improve the quality of the recipient vector, an aliquot of pNL4.3.Luc.R-E- plasmid was digested and instead of visualising the DNA under ultraviolet (UV) light by binding to EtBr during loading, the DNA was stained with 1x SYBR Safe DNA gel stain (ThermoFisher) and viewed under a blue light transilluminator (Invitrogen). This blue light causes less damage to DNA than UV light during gel cutting, which in turn, could maximise the recovery of DNAs with improved quality and thus potentially improving cloning efficiency.

Attempts were made to purify recipient plasmid using SYBR Safe DNA gel stain. After double digestion, one half of the digested plasmid was subjected to gel purification and followed by gel clean up. The concentration of the recovered plasmid was 72 ng/µL. In an attempt to reduce the contamination from agarose gel, the other half of double digested plasmid was directly purified using the NucleoSpin gel & PCR clean up kit yielding an equal amount of linearised pNL4.3.RE- and the luciferase fragment. The concentration of the resultant plasmid was 91.3 ng/µl.

Meanwhile, an error was identified in the amplifying forward primer Nluc1.F where the “G” (underlined) was intended to be a “C” (underlined) as corrected in Nluc2.F (Table 4.4); furthermore, to improve annealing, a nucleotide ‘G’ (underlined) was added in

the 5’ end of reverse primer Nluc2.R (Table 4.4). Nluc was amplified using this set of updated primers.

Table 4.4 Primers used for amplifying Nano luciferase gene

NAME	PRIMER SEQUENCE 5’ → 3’
NLUC1.F	ATGGGTGGCGCGGC**G**GCCACCATGGTCTTCACA
NLUC1.R	TTTTCTAGGTCTCGATTACGCCAGAATGCGTTCG
NLUC2.F	ATGGGTGGCGCGGC**C**GCCACCATGGTCTTCACA
NLUC2.R	**G**TTTTTCTAGGTCTCGATTACGCCAGAATGCGTTCG
PNL4.3.SCR.F	CGCTGAGATAGGTGCCTCAC
PNL4.3.SCR.R	GAATCTCCTACAATATTGGAGTCAGG

In-Fusion cloning was performed again ligating the newly amplified Nluc into the recipient plasmid yielded from gel purification as well as the mixed plasmids purified from post double digestion. The recombinant plasmids were introduced into *DH5a* cells. Good numbers of colonies were formed after overnight incubation in the LB agar plate under ampicillin selection in both conditions. However, colony screening using the backbone primers yielded negative results. After aligning the primers against the backbone sequence, an error was identified that the forward primer (5’-CGCTGAGATAGGTGCCTCAC -3’) at position 1598-1617, was in the region of the ampicillin resistance gene and the reverse primer (5’-GAATCTCCTACAATATTGGAGTCAGG -3’) at position 6848-6873, was just 100bp outside of the gene of insert. The size of amplified sequence using this pair of primers

could had been greater than 5kb provided that the extension time was sufficient. In other words, given the extension time during the PCR (1 min), this 5kb long DNA could have never been amplified. To combat this problem, primer pair pNL4.3scr.r and Nluc2.r was used for the colony screening and results showed that Nluc was successfully cloned into linearised pNL4.3.R-E-.

Co-transfecting this plasmid with gtB.LMS into HEK293T cells generated HBV-PVs, which were then used to infect HuH7 ± NTCP cells. The infectivity was compared with HBV-PVs generated using the standard PVs produced using the pNL4.3.luc.R-E- backbone. One key difference between the two lentiviral packaging systems was their luciferase products. Firefly luciferase was produced in the cells that were infected with PVs generated from pNL4.3.luc.R-E-, whereas nano luciferase was produced in the cells infected with PVs generated from pNL4.3.Nluc.R-E-. As a result, different substrates were required for the bioluminescence reaction. D-luciferin was the substrate for firefly luciferase, while furimazine was the substrate for nano luciferase (Figure 4.9 A & B) (England et al., 2016). The preliminary results showed that the signal-noise ratio in HBV-PVs generated in pNL4.3.Nluc.R-E- was not superior than its counterparts (Figure 4.9 C & D). Interestingly, autoluminescence observed in HuH7-NTCP (including BSD and eGFP) cells diminished when using this Nluc packaging system. The substrate furimazine used in the Nano lucferase system might have contributed to this phenomenon. Moreover, standard deviation between triplicates was much smaller in pNL4.3.Nluc.R-E- generated HBV-PVs. Based on these results, pNL4.3.Nluc.R-E- was used as the packaging vector for the following experiments.

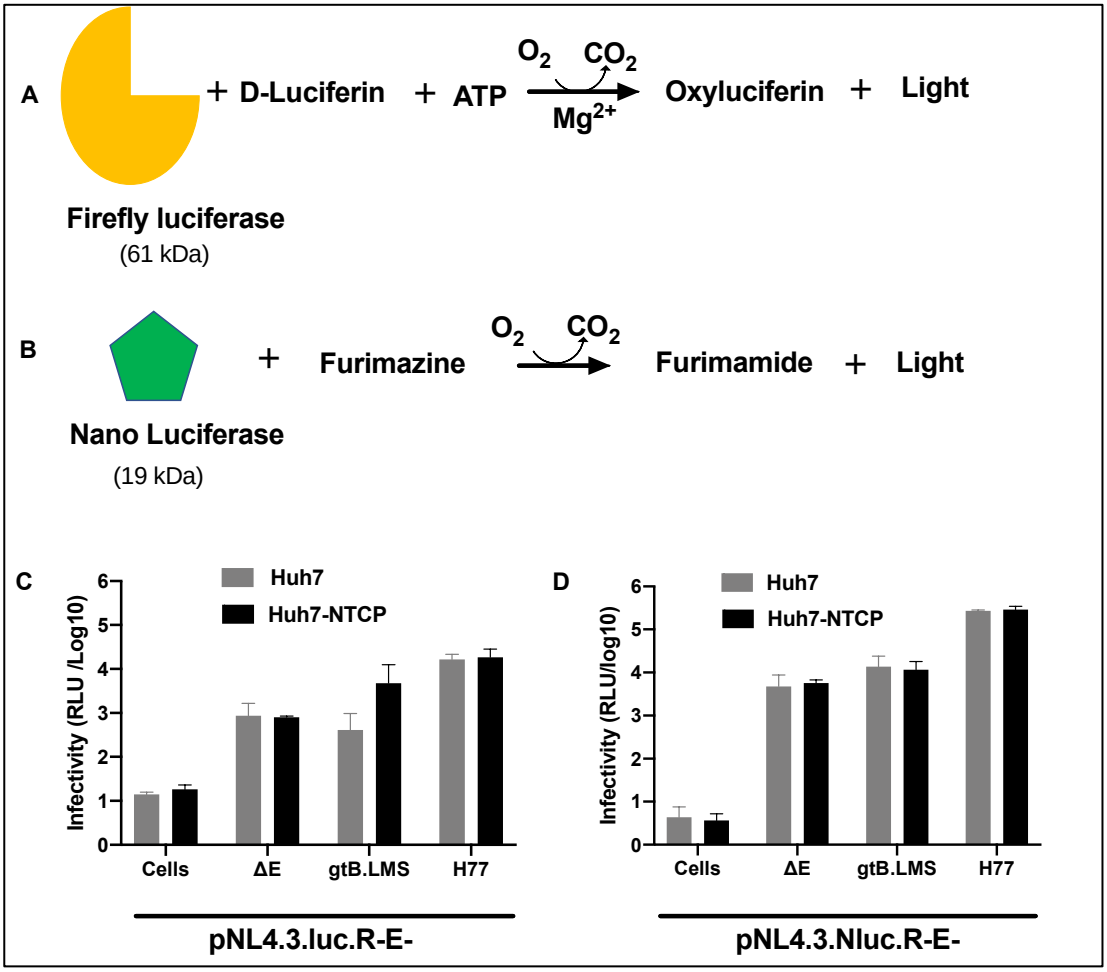

Figure 4.9 Comparison of pNL4.3.Nluc.R-E- to pNL4.3.luc.R-E-

HEK293T cells were transfected with gtB.LMS and lentiviral packaging plasmid; H77, a E1E2 of HCV expressing plasmid, was used as a positive control; ΔE was lentiviral packaging without surface antigens prepared as a negative control. Seventy-two hours later, pseudotyped viruses (PVs) were harvested and used to infect cells. After a further 72 hours incubation, HBV-PVs infectivity was determined by reading luciferase activity shown as relative luminescence unit (RLU). ***A.*** *firefly luciferase catalyses the oxidation of D-luciferin in the presence of oxygen, magnesium and adenosine triphosphate (ATP) to produces light;* ***B.*** *Nano luciferase catalyses the oxidation of furimazine to produces light;* ***C.*** *The infectivity of HBV-PVs packaged with pNL4.3.Luc.R-E-. The RLU was measured using (firefly) Luciferase Assay System (Promega);* ***D.*** *The infectivity of HBV-PVs packaged with pNL4.3.Nluc.R-E-. The RLU was measured using Nano-Glo® Luciferase Assay System (Promega).*

4.3.6 Infectivity of HBV-PVs using HBsAg expression plasmid cloned in the pI.18 vector

Due to the unsuccessful production of infectious HBV-PVs using pCDNA3.1 vector derived gtB.LMS or gtD4.LMS, the pI.18 vector was used for insertion of these HBsAg DNAs. The HBsAg gene from gtB.LMS was ligated into pI.18 vector via In-Fusion cloning. The resulting plasmid was designated as 'pI.18-gtB.LMS'. The PVs were generated using pI.18-gtB.LMS and pNL4.3.Nluc.R-E-. Their infectivity was compared with PVS generated using gtB.LMS (in the pcDNA3.1 vector). The results demonstrated that the infectivity of HBV-PVs produced by pI.18-gtB.LMS was not greater that that produced from gtB.LMS (Figure 4.10).

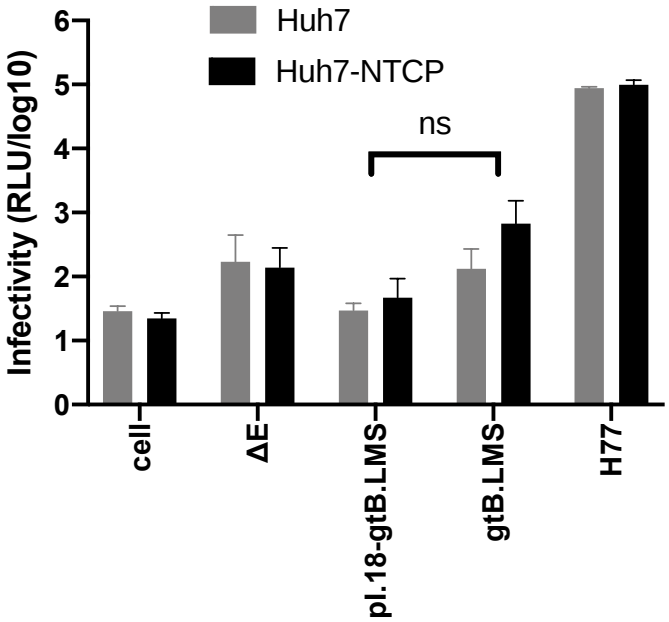

Figure 4.10 Comparison of HBV-PV infectivity using HBsAg expression vectors of pI.18 to pcDNA3.1

HBsAg ORF was ligated into pI.18 vector and the resulting plasmid was termed pI.18.gtB.LMS. Infectivity of HBV-PVs yielded using pI.18.gtB.LMS was compared with that yielded using gtB.LMS (ligated in pcDNA3.1 expression vector). H77, a E1E2 of HCV expressing plasmid, was used as a positive control; ΔE was lentiviral packaging without surface antigens prepared as a negative control. The RLU was measured using Nano-Glo® Luciferase assay system (Promega).

4.3.7 Infectivity of HBV-PVs from HBsAg expressed in 293T-gtD4.LMS cell lines

After a step-by-step optimisation procedure to improve the infectivity of HBV-PVs as described above, the target was not achieved, i.e. no statistically significant difference was observed between the entry of HBV-PVs and ΔE-PVs. According to the IF images shown in Figure 4.5.2, one of the possible reasons could be due to low production of HBsAg using transient transfection of HBsAg bearing plasmids. This prompted a change of approach, with generation of a HBsAg stably-expressing 293T-gtD4.LMS cell line.

The HBsAg gene from gtD4.LMS was ligated into pWPI-BSD vector via In-Fusion cloning, which obtained a plasmid assigned 'pWPI-gtD4.LMS-BSD'. Subsequently, PVs generated from pWPI-gtD4.LMS-BSD, HIV backbone (pcMVR.87.4), VSV glycoprotein (VSVG, pMD2) were transduced into HEK293T cells. HEK293T cells transduced with PVs generated from the HIV backbone and VSVG was served as negative control. Three days later, cells were challenged with blasticidin. The starting concentration was 5 $\mu g.mL^{-1}$, gradually increasing to 20 $\mu g.mL^{-1}$ where complete cell death was observed in the negative control. The blasticidin-resistant cells were further expanded under blasticidin. This cell line was designated '293T-gtD4.LMS'.

To confirm HBsAg proteins were expressed by this cell line, IF was performed. As illustrated in Figure 4.11A, the majority of 293-gtD4.LMS cells demonstrated HBsAg expression. This was a great improvement from the transiently transfected cells.

To generate HBV-PVs, pNL4.3.Nluc.R-E- was added into 293T-gtD4.LMS cells. The negative control ΔE was generated as above by adding pNL4.3.Nluc.R-E- into HEK293T cells. It was shown that less background noise was observed in HepG2-

NTCP compared to the HuH7-NTCP. Additionally, the error bars representing the standard deviation between the triplicates was shown to be much smaller when using HepG2 ± NTCP as target cells (Figure 4.11B). As such, for all future experiments, the target cells were changed from HuH7 ± NTCP to HepG2 ± NTCP. To further optimise this system, plasmid expressing S was added into this cell line as evidence shown in WB (Figure 4.4C) where transiently transfected HEK293T cells with gtB.LMS produced higher amounts of L, and by adding gtB.S, the intensity of S bands was enhanced. Using this system, the results demonstrated that production of infectious HBV-PVs required a fine balance between L and S proteins. Co-transfecting 1 µg gtD4.S expressing plasmid into this cell line yielded infectious HBV-PVs ($P<0.0001$) (Figure 4.11C). Interestingly, the infectivity of HBV-PVs produced by this cell line appeared to be independent of NTCP expression.

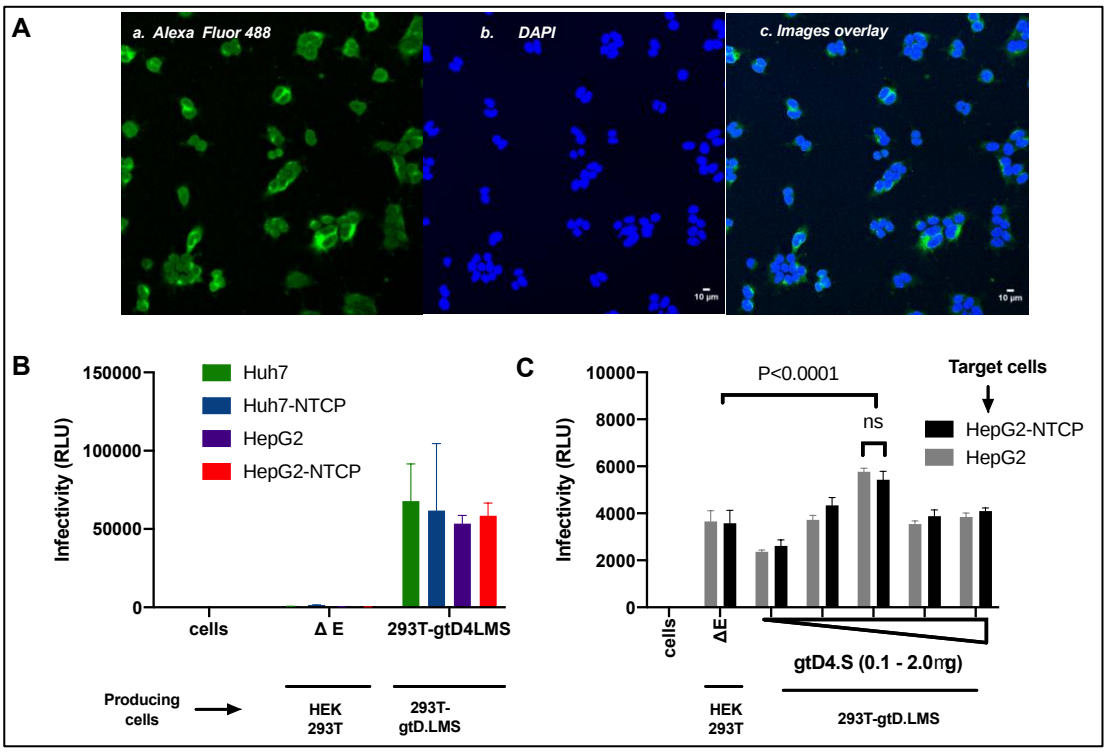

Figure 4.11 Infectivity of HBV-PVs produced from 293T-gtD4.LMS

293T-gtD4.LMS was a cell line expressing LMS proteins representative of a genotype D variant of HBV. ***A.*** *HBsAg expression in this cell line was analysed using indirect immunofluorescence (as detailed in (*Figure 4.5*). Cells were transfected with pNL4.3.Nluc.R-E-; Seventy-two later, supernatants were collected and used to infect target cells. Cells were incubated a further 72 hours and their bioluminescence emission was assessed as shown in RLU (relative luminescence unit). ΔE was produced by transfecting HEK293T cells with pNL4.3.Nluc.R-E-;* ***B.*** *the target cells were HuH7±NTCP and HepG2 ±NTCP;* ***C.*** *the 293T-gtD4.LMS cells were transfected with pNL4.3.Nluc.R-E- and incremental amounts of S expressing plasmid gtD4.S as indicated in the graph. The target cells were HepG2 ±NTCP. (*The infectivity of HBV-PVs generated from 293T-gtD4.LMS was inconsistent. Luciferase signal in HBV-PVs could be 40-50% more than the negative control in one setting, but could be below the negative control in another).

4.3.8 Additional approaches to optimising PV infectivity

Following the establishment of NTCP overexpressing cell lines and subsequent HBV-PV infectivity assays, low levels of infection had been observed throughout. To enhance infectivity of HBV-PVs, several experimental approaches were used.

It has been reported that Dimethyl sulfoxide (DMSO) enhances NTCP expression and promotes susceptibility to HBV infection (Li et al., 2016, Iwamoto et al., 2014). Therefore, enhancement of HBV-PV infectivity in the presence of DMSO was assessed. The target cells were pre-treated with DMSO (1-3%) overnight before incubation with HBV-PVs. Luciferase signal was not enhanced in DMSO treated NTCP cells (data not shown).

Polyethylene glycol (PEG) 8000 treatment of target cells has been reported to enhance HBV infectivity. It has been shown that PEG promotes HBV interaction with heparan sulphate proteoglycans resulting in enhanced infectivity (Michailidis et al., 2017). Accordingly, the action of PEG 8000 on HBV-PV infection efficiency was examined. HBV-PVs were mixed with 4% PEG 8000 during incubation period with the target cells. Four hours later cells were replaced with fresh media without 4% PEG 8000. Again, no enhancement on HBV-PV infectivity was detected after multiple attempts.

Sodium butyrate (NaB) is a butyric acid. It inhibits histone acetylase and has been used to promote production of recombinant proteins *in vitro* (Chai et al., 2007). Reported working concentrations vary from 10 µM to 10 mM (Carnell et al., 2015). To establish an optimal concentration, 293T-gtD4.LMS cells were treated with or without NaB for 24 hours. To simplify the experiment, the tested concentrations included 0.1, 1 and 10 mM. Cytotoxicity of NaB was determined by trypan blue staining assay.

Dead cells taking up this dye are stained blue. Increasing cell death was observed as the concentration of NaB increased.

To investigate whether sodium butyrate (NaB) would enhance HBsAg expression, cells were either untreated or treated with increasing concentration (0.1 to 10 mM) of NaB and the resulting supernatants were collected for immunoprecipitation followed by silver staining. As shown in Figure 4.12A, The heavy bands corresponding to the heavy (~50 kDa) and light chains (~25 kDa) of denatured antibodies indicated the precipitated HBIG from IP. bands at sizes of 42 and 39kDa indicating the glycosylated and unglycosylated L proteins were visible in samples treated with NaB but not in the untreated sample; band intensity was not enhanced as the concentration of NaB increased. The bands at sizes of 27 and 24kDa suggesting the glycosylated and unglycosylated S proteins were also visible and the band intensity was enhanced as the concentration of NaB increased (Figure 4.12A), suggesting 10 mM was an optimal concentration for HBsAg production.

To confirm whether NaB would enhance the production of PVs, at 24 hours post-transfection with pNL4.3.Nluc.R-E-, 293T-gtD.LMS were either left untreated or treated with NaB with concentrations ranging from 0.1 to 10 mM. The infectivity of the resulting PVs were determined as demonstrated in Figure 4.12B. The relative infectivity of the PVs produced from cells treated with 0.1 mM NaB was about >0.5 $\log_{10}$ higher compared to its counterparts, indicating NaB 0.1 mM was an optimal concentration.

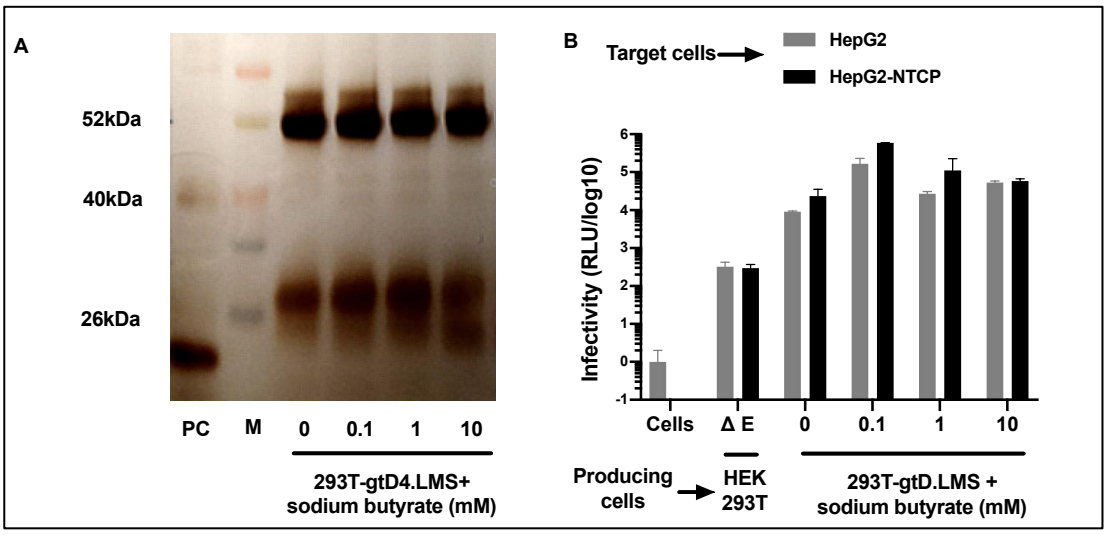

Figure 4.12 Sodium butyrate (NaB) enhanced the expression of HBsAg

A. *293T-gtD4.LMS cells were treated with (or without) increasing concentrations of NaB (0.1-10 mM) for 24 hours. One mL of the supernatants was immunoprecipitated by incubation with Dyna-beads protein G (Invitrogen, ThermoFisher Scientific) and Human Hepatitis B immunoglobulin (HBIG, kindly donated by the Public Health England) at a concentration of one µL per testing sample in 200 µL PBST (1% PBS with 0.02% Tween 20). Post immunoprecipitation, the proteins were separated by polyacrylamide gel electrophoresis. The separated proteins were detected by silver staining. PC (positive control), a recombinant HBsAg (ab91276, Abcam).* ***B****. 293T-gtD.LMS cells were transfected with pNL4.3.Nluc.R-E-. Twenty-four hours later, cells were incubated with (or without) NaB with concentrations ranging from 0.1 to 10 mM. Seventy-two later, supernatants were collected and used to infect HepG2 ±NTCP cells. Cells were incubated a further 72 hours and their bioluminescence emission was assessed as shown in RLU (relative luminescence unit). ΔE was produced by transfecting HEK293T cells with pNL4.3.Nluc.R-E-.*

4.3.9 Infectivity of HBV-PVs from HBsAg expressed in 293T-gtD4.S cell lines

A cell line expressing S only was developed. To achieve this, the S ORF of the gtD4.LMS was ligated into pWPI-BSD vector via In-Fusion cloning in the same manner as described in the generation of 293-gtD4.LMS. The resulting established cell line was designated '293T-gtD4.S'.

To generate HBV-PVs, L expressing plasmids were supplemented *in trans* in addition to pNL4.3.Nluc.R-E. Positive controls were accomplished by adding pMD2.G (VSVG) and pNL4.3.Nluc.R-E into this cell line, whereas negative controls were generated by addition of the pNL4.3.Nluc.R-E as well as an empty expression vector pCDNA3.1, the same vector as the L expressing plasmid. By doing so, the variable was restricted to the glycoprotein expressing plasmids.

To determine an optimal concentration for L expressing plasmids, five concentrations of L expressing plasmid were used (0.1, 0.5, 1.0, 1.5 and 2.0 µg per $1.2x10^6$ cells). At the time of this experiment, an L expressing plasmid for genotype D was not available, instead, plasmid gtD4.LM that expressed L and M was used. The results demonstrated a strong dose dependent effect. At 0.1 µg, the luciferase signal detected in HBV-PVs was 30% higher than the negative control ($P<0.001$). HBV-PV entry was reduced as the concentration of gtD4.LM increased (Figure 4.13A). To determine whether infectivity was correlated with better incorporation of HIV capsid or HBV glycoproteins into HBV-PV, western blots of PVs purified through 20% sucrose were conducted. The protein expression in corresponding producer cell lysates was also assessed. As expected, in the cell lysates, L expression was increased as the concentration of gtD4.LM increased, while the S decreased gradually (Figure 4.13A). However, in the pelleted PVs, no L proteins were observed,

while S was merely seen, but the expression of HIV capsid (p24) was reduced as the amounts of gtB.L increased. No correlation was seen between infectivity and HBV glycoprotein expression in HBV-PVs, but a strong association between infectivity and HIV capsid incorporation in HBV-PVs was seen (Figure 4.13A).

To determine whether S from genotype D could incorporate with L from genotype B into HBV-PVs, gtB.L was added into 293T-gtD4.S cells. The experimental design was similar to the addition of gtD4.LM. The results showed an opposite pattern to that observed when addition of gtD4.LM. No infectious HBV-PVs were generated when low concentrations of gtB.L was added. However, when the concentrations of gtB.L were increased to 1.5 and 2.0 µg, the luciferase signal was enhanced two-fold compared to the negative control ($P<0.006$ and $P<0.0001$ respectively) (Figure 4.13B). Western blot of the cell lysates detected increasing amounts of L as the concentration of gtB.L increased, while the S decreased gradually (Figure 4.13B). No HBsAg proteins were observed in HBV-PVs where infectivity was absent. However, small amounts of L and S proteins were observed in HBV-PVs when 1.5 and 2 µg of gtB.L were used; at which concentrations, the expression of HIV capsid (P24) was also the highest (Figure 4.13B).

In contrast to supplementing gtD4.LM where HBV-PV entry was independent of NTCP, the entry of HBV-PVs generated by the addition of gtB.L was in a NTCP dependent manner. As observed in Figure 4.13B, > 50% higher luciferase signal was detected in NTCP transduced HepG2 cells compared to parental cells when gtB.L concentration were at 1.5 and 2.0 µg ($P<0.02$ and $P<0.0002$ respectively).

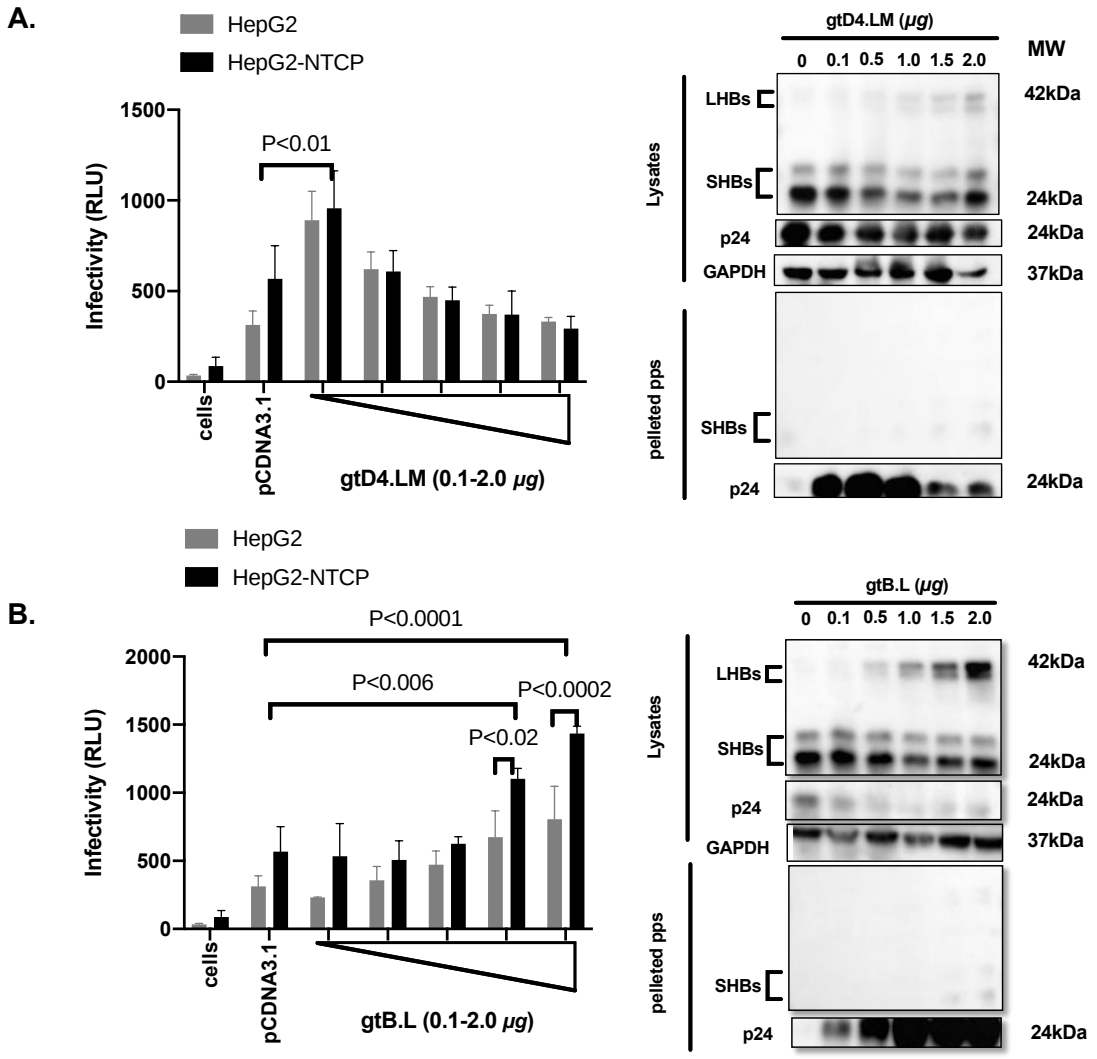

Figure 4.13 The infectivity of HBV-PVs produced from 293T-gtD4.S cell lines

*293T-gtD4.S cells were transfected with pNL4.3.Nluc.R-E- and increasing concentration (0.1-2.0 μg) of gtD4.LM **(A)** and gtB.L **(B)**. Seventy-two hours later, 100 μL of pseudotyped viruses (PVs) containing supernatant was used to infect HepG2 and HepG2-NTCP cells. Meanwhile, the remaining PVs were pelleted at 40,000 rpm for 2.5 hours at 4°C through a 20% sucrose cushion. The resulting pelleted PVs were resuspended in 100 μL PBS for western blotting. Cells were lysed in lysis buffer and supernatants were collected for western blotting. The primary antibody was mouse monoclonal anti-HBsAg antibody (Mab1964, Abnova) at dilution 1:2000 and secondary antibody was goat anti-mouse IgG/HRP (P0447, Dako) at dilution 1:1000. GAPDH: Glyceraldehyde 3-phosphate dehydrogenase, used as a loading control.*

4.4 Discussion

The original aims of this chapter were to explore the roles of L, M & S surface antigens in HBV entry using an *in vitro* pseudotyping model. The experimental evidence has demonstrated that HBsAg could not be easily pseudotyped. Following extensive optimisations, an in-house developed S protein expressing cell line 293T-gtD4.S was shown to produce infectious PVs when the cells were co-transfected with plasmids expressing L proteins.

4.4.1 The ratio of L to S plays a role in HBV-PV infectivity

The observation in Figure 4.4C demonstrated that there was expression of greater amounts of L than S when gtB.LMS alone was transfected into producer cells. It appeared that the expression of full-length L transcripts was more favoured under the cytomegalovirus (CMV) promoter compared with an internal promoter (SPII) residing in the PreS2 region that drives the production of S. This ratio of L and S expression was shifted when gtB.S was added where the S transcription was directly regulated under CMV promoter (Figure 4.4C).

Experiments have demonstrated that low levels or absence of S lead to intracellular retention of L, and consequently HBV retention (Melegari et al., 1997, Bruss and Ganem, 1991). The retention effect of L can be counteracted by co-expression of S (Bruss and Vieluf, 1995). In this current study, a distinct trend was observed that luciferase signals changed as the ratio of L and S changed in a dose dependent manner (Figure 4.13), which confirmed that the ratio between L and S was crucial for HBV-PV assembly and thereby its entry. It was proved that a too-high or too-low L/S ratio inhibited virion secretion, which is consistent with previous studies (Garcia et al., 2009). Interestingly, less gtD4.LM (0.1 μM) was required to produce the most

infectious HBV-PVs expressing gtD-derived HBsAg compared with supplementing gtB.L (2.0 µM). It is of interest to speculate whether the M played a role in particle formation, hence less gtD4.LM required. Another speculation could be that HBV-PVs enveloped by L and S from the same genotype when supplementing gtD4.LM made the envelopment more efficient compared with genotype D/B recombinant HBV-PV formation. The successful genotype D/B recombinant HBV-PVs also indicates that, *in vivo*, preS1 deletion of one genotype could be rescued by co-infection with a different genotype. The experiment also confirmed that S alone does not form infectious particles (Figure 4.13).

The experiment of supplementing gtB.L into 293T-gtD4.S cells *in trans* confirmed that viral particles can be assembled by surface proteins of different genotypes. This experiment has also demonstrated that a cell line expressing S offers flexibility in manipulating the relative amounts of L or M to be expressed, which can play an important role in investigating the role(s) of L, M and S in HBV assembly and entry. Interestingly, HBV-PVs containing gtB.L entered hepatocytes in a NTCP dependent manner, whereas, HBV-PVs containing gtD4.L gained entry independent of NTCP. A previous study showed that PreS1 5-20 aa peptide could effectively block HDV infection suggesting these residues were involved in viral-host receptor interaction (Barrera et al., 2005). Aligning the amino acid sequence of gtD4.L to gtB.L, residues 7 and 8 amino acids were the only ones to differ between the 2 sequences within 5-20 aa of PreS1. Polar uncharged side chains 7-threonine (T), serine (S)-8 were present in gtD4.L, whereas gtB.L is comprised of two hydrophobic amino acids valine (V) - proline (P) where P is a well-known amino acid that can introduce kinks in protein folding. Residues 7-VP-8 were also presented in the Myr-47 peptide which was used in the discovery of NTCP (Yan et al., 2012). This suggested that residues 7-TS-8 may promote PreS1 folding in a way that recognises a different receptor other than NTCP.

Future work focusing on mutational analysis of these two amino acids could be employed to identify their roles in viral-host interaction.

Western blots demonstrated that HBV-PV entry and subsequent reporter expression was not associated with glycoprotein incorporation, but with the levels of HIV-1 capsid p24 (Figure 4.13). This is in line with a previous analysis where undetectable (by western blots) levels of the HCV glycoproteins were nevertheless able to mediate infection (Urbanowicz et al., 2016). IP pelleted HBV-PVs followed by Western blotting also failed to detect the HBsAg incorporation, in agreement with previous findings (Chai et al., 2007). A more sensitive method is required. A small tetracysteine (TC) tag coding for C-C-P-G-C-C has been successfully incorporated into the HBV core gene and the resulting chimera produces infectious viruses that are comparable to the WT HBV (Sun et al., 2014). The TC tag can be bound with biarsenical dye and generate fluorescence that could be observed by confocal microscopy and transmission electron microscopy (TEM). By ligating this small tag with L, M or S, glycoprotein, incorporation with HIV capsid into HBV-PVs and their entry could be monitored in live cells.

4.4.2 Pseudotyping as a tool to investigate HBV entry pathway

In this study, extensive optimisations were explored to generate infectious HBV-PVs. Ultimately, supplementing L expressing plasmid *in trans* into the 293T-gtD4.S cell line appeared to achieve this goal. However, the efficiency of infectivity was very low compared to other glycoproteins such as VSVG and Ebola Env. Experiments demonstrated that VSV-Gpvs yielded luciferase signal >5 logs higher than HBV-PVs (data not shown). Previous studies have shown that VSVG and Ebola Env can be actively recruited to HIV-gag in the plasma membrane for viral assembly, and hence can be successfully pseudotyped. However, for some viruses, this incorporation might

be in a passive manner when any glycoproteins are present in the site of viral assembly (Jorgenson et al., 2009, Sandrin and Cosset, 2006). For instance, small amounts of HCV E1E2 glycoproteins escape the ER retention machinery and “leak” to the cell surface where they are pseudotyped (Op De Beeck et al., 2004). HBV capsid envelopment and virion formation occur in post-ER, pre-Golgi where the surface proteins are (Bruss, 2007). This suggests a low probability for HBsAg and HIV-gag incorporation into PVs as they are located in separate compartments. However, flow cytometry detected HBsAg protein expression in the plasma membrane, though it was 16-fold less compared with intracellular HBsAg expression (Chai et al., 2007). This may provide one explanation for the low HBV-PV infectivity observed in this study. Experimental evidence proved that this model was challenging to apply in a robust manner. HBsAg was expressed in producer cells, but no direct evidence showed whether HBV-PVs containing the HBsAg were generated.

The high background noise experienced in this study could be improved by using NanoBit® protein: protein interaction (PPI) system (Promega). A Nluc-large BiT gene (LgBiT; 17.6kDa) can be inserted into pNL4.3.R-E- vector, while a small BiT (SmBiT; 11 amino acids) can be tagged with HBsAg. As such, only incorporation of the HIV capsids and HBsAg could produce a functional enzyme capable of generating a luminescence signal and therefore eliminating the background noise. However, for a more sophisticated manner, it would be better incorporating the NanoBiT into the HBV trans-complementation system where part of the preCore in a 1.2-fold HBV genome is replaced with the LgBiT, and a second plasmid supplies the missing proteins with a silenced HBsAg encoding gene while a third plasmid provides SmBiT tagged HBsAg. By transfecting these three plasmids into producer cells, recombinant HBV virus containing the two subunits is generated. The strength of this system is that viral capsid assembly, HBsAg envelopment and virion budding are all in their native

manner. Moreover, it would maintain the strength of lentiviral HBV-PVs where only one single-round infection will occur, in turn, reducing biohazard risk.

4.4.3 Technical challenges

4.4.3.1 Site directed mutagenesis

The original aim of this study was to determine the roles of L, M and S in HBV entry. To achieve this, plasmids expressing L, M or S were developed via SDM strategies.

The SDM was not always successful. Quite often, colony PCR screening showed the correct size of genes, but Sanger sequencing confirmed that no mutation was introduced. It was difficult to distinguish wild type (WT) from mutants via colony PCR screening. One of the common reasons that colonies bearing the WT gene still presented was the failure of KLD enzyme treatment. This enzyme mix contains kinase, ligase and DpnI enzymes. Kinase mediates the transfer of a phosphate group onto the 5' end of polynucleotides, and ligase catalyses a phosphodiester bond and facilitates the joining of DNA strands together, which allows for circularisation of the amplified PCR products. DpnI restriction enzyme recognises methylated adenine (^{m}A) in a sequence G^{m}A| TC and cleaves the DNA. The methylated DNA is only presented in bacteria, while PCR derived DNA products are free of methylated adenine. Therefore, treatment with KLD enzyme should have removed the templates. Failure to do so could be due to suboptimal activity of the enzyme. One reason for failure could be due to the high amounts of input plasmid templates. Q5 SDM protocol recommends 1-25 ng/µl DNA templates, the better experimental concentration was below 2 ng/µl. Another possible reason was the incubation temperature. Though incubation was at room temperature for 5 minutes, room temperature may fluctuate. Therefore, subsequent incubation was conducted in a thermocycler where the

temperature was set at 22 °C. In general, after these optimisations, all required mutants were completed successfully.

4.4.3.2 Improving HBsAg detection in western blot

The first anti-HBsAg antibody used was ab9193 (Abcam) to detect HBsAg expression. Samples included 20% sucrose pelleted HBV-PVs and lysates of the producer cells. High background noise was observed in cell lysates (Figure 4.3). The multiple bands were observed in all samples including the negative control where RLBS was expressed. Therefore, it was not possible to analyse HBsAg expression in the cell lysates.

Subsequent attempts to detect HBsAg using other commercial antibodies including S26 and 86C did not achieve the goal. Finally, the use of Mab1694 successfully detected HBsAg in cell lysate, confirming the expression of HBsAg in cells. S or L were detected as shown in Figure 4.4. Interestingly, the interaction between HBsAg expressed by gtB.S with Mab1694 appeared to be more effective compared to its counterparts. The immunogen of this antibody is a recombinant protein corresponding to the full-length of HBsAg *ay*. As explained in 4.3.2 (Protein detection), the serotype of S expressed by gtB.S should be *ad,* while the S expressed by gtD4.S or gtD13.S should be *ay*. Accordingly, Mab1694 should have been more sensitive to detect S expressed by gtD4.S and gtD13.S. The opposite results suggest other epitopes could be involved in the binding between antigens and the antibody.

4.4.3.3 Factors involved in pseudotyping infection assay

A study of factors which might limit pseudotyping using patient derived HCV glycoproteins demonstrated that the ratio of packaging construct to the glycoprotein encoding plasmid was important for successful HCV-PV genesis. Furthermore, the selection of retroviral packaging construct also influences the function of HCV-PVs

(Urbanowicz et al., 2016). Similar experiments were adopted to enhance HBV-PV genesis and infectivity. However, no improvement was obtained (data not shown).

The observation of a high background signal generated by ΔE-PV led to the question of its suitability as a negative control. Sanger sequencing of pNL4.3.Luc.R-E revealed that there were residues of HIV-1 transmembrane glycoprotein GP41 in the vector forming a partial open reading frame. These residues of GP41 were able to incorporate into gag protein and were sufficient to assemble pseudotyped viruses and bud through a host-derived membrane to exit the cell (Johnson, 2011). This GP41 may mediate a fusion event and gain entry into target cells. Therefore, ΔE generated from HIV lentiviral vector may not be an ideal negative control. Based on this theory, a construct (RLBS) expressing envelope proteins of roundleaf bat hepatitis virus was developed. Roundleaf bat hepatitis virus was previously reported not to be able to infect human hepatocytes (Drexler et al., 2013). However, results showed no significance between both negative controls (Figure 4.7b) suggesting other factors might have contributed to the ineffective HBV-PV infection.

A recent study showed that infectious HBV-PV genesis was optimal in HuH7-NTCP cells compared to that generated in HEK293T cells, indicating that liver-specific factors could be required for assembly and/or infectivity (Meredith et al., 2016). Accordingly, a similar experiment was set up. Again, no significant difference was obtained between PVs generated by these two producer cells (Figure 4.7c & d). DMSO differentiated HepG2-NTCP (dHepG2-NTCP) cells were used as the target cells previously and showed a >4-fold infectivity in dHepG2-NTCP compared to NTCP negative HepG2 cells (Meredith et al., 2016). It would be interesting to obtain these dHepG2-NTCP cells for future work.

During the undertaking of these studies, Dr Emma Horncastle (a lab colleague) discovered that pI.18 plasmid bearing the DNA of rabies surface protein producing more infectious PVs than pCDNA3.1 vector derived plasmid (the results have not yet been published). Inspired by this discovery, pI.18.gtB.LMS was developed and following infection assay using this construct demonstrated that HBV-PVs generated from pI.18.gtB.LMS was not enhanced (Figure 4.10).

A recent literature reported that nano luciferase coupling with its substrate furimazine generated 150-fold greater signal than that of firefly luciferase (Hall et al., 2012). This led to the development of pNL4.3.Nluc.R-E-. During the process, an unexpected fragment at ~7kbp was identified in some of the pNL4.3.luc.R-E- plasmid preparations. After troubleshooting, the bands suggested that samples were contaminated with bacterial chromosomal DNA as a result of vigorous treatment of the lysates. This could be confirmed by digestion of the sample with *EcoRI* if a smear is observed (Qiagen, 2012). This observation highlights the importance of agarose gel analysis of purified plasmid.

To monitor HBsAg expression in producer cells, indirect IF was performed. Only ~5% of producer cells expressing HBsAg compare to ~15% producer cells expressing HCV E1E2 (H77) (Figure 4.5). Both gtB.LMS and E1E2 of H77 were cloned into pcDNA3.1 with their expression driven by a CMV promotor. Furthermore, both were introduced into HEK293T cells using linear polyethylenimine (PEI, MW 25 kDa) following the same protocol. Protein expression relies on functional amounts of mRNA transcribed from DNA. Assuming the same amounts of genetic materials were introduced into the host, then different levels of protein expression suggested there were different levels of DNA transcription and/or mRNA translation. One major difference between plasmids gtB.LMS and H77 is that hepatitis C is an RNA virus whereas hepatitis B is a DNA virus. In the literature, the majority of viruses that have been successfully

pseudotyped are RNA viruses, such as Rabies, VSV, Ebola, Lassa, Influenza and the most recent one, MERS-CoV viruses (Cronin et al., 2005, Grehan et al., 2015). There are many more questions to be answered why RNA viruses are more readily pseudotyped than DNA viruses. Many factors could be involved in this discrepancy, including foreign genetic material, host epi-genetic factors and the incorporation with the lentiviral vector. It has been reported that one of the limitations of NTCP expressing hepatoma cell lines is that the multiplicity of infection (MOI) required to achieve substantial infection is extremely high (Witt-Kehati et al., 2016). This combination of factors might contribute to the low infectivity observed in this study.

Having observed lower numbers of producing cells that expressed HBsAg, 293T-gtD4.LMS, a cell line constitutively expressing L, M and S was developed. Immunofluorescence showed that the majority 293T-geD4.LMS cells expressing HBsAg compared to only 5% of producer cells in the transient transfection (Figure 4.5 & Figure 4.11A). However, downstream pseudotyping using this cell line proved that there were drawbacks using this system. One of the drawbacks was too many variables in one setting. Two cell lines had to be employed to generate PVs in this system. 293T-gtD4.LMS cells were used for HBV-PVs, while HEK293T cells were used to generate PVs of the positive control (VSVG) and the negative control (ΔE). As the efficiency of generating PV is heavily dependent on the condition of the producer cells, it was difficult to normalise this between two cell lines. Experiments demonstrated that the infectivity of HBV-PVs generated from 293T-gtD4.LMS was inconsistent. It appeared that luciferase signal in HBV-PVs was 40-50% more than the negative control in one setting (Figure 4.11B), but it was below the negative control in another (data not shown). It is reasonable to speculate that the producer cells were playing a role in these inconsistent results.

Nevertheless, due to the drawbacks in the 293T-gtD4.LMS system, a different cell system was required. HBV entry is mediated by L, which implies that PVs enveloped by S are not able to enter target cells. This rationale prompted the generation of a stable cell line expressing S, the generation of 293T-gtD4.S.

Previous literature suggests that Dimethyl sulfoxide (DMSO) enhances NTCP expression and promotes susceptibility to HBV infection (Li et al., 2016, Iwamoto et al., 2014). Moreover, PEG has been shown to promote HBV interaction with heparan sulphate proteoglycans resulting in enhanced infectivity (Michailidis et al., 2017). However, repeated attempts using either DMSO or PEG did not achieve optimisation of HBV-PV infectivity. Nonetheless, the use of NaB showed promising results. As demonstrated in Figure 4.12A, HBsAg production was enhanced in NaB-treated producing cells and the bands corresponding to the glycosylated and unglycosylated S proteins were intensified as the concentration of NaB increased, suggesting 10 mM was the optimal concentration for HBsAg production. Interestingly, higher HBsAg production did not result in optimal HBV-PV infectivity (Figure 4.12B). The infection assay revealed that the optimal concentration of NaB to achieve a better HBV-PV infection was 0.1 mM. One explanation is that HBV-PV production was paradoxically reduced due to greater cell toxicity induced by NaB at concentrations above 0.1 mM.

4.5 Conclusions and future work

Building on the construction of NTCP overexpressing cell lines described in the previous chapter, the infectivity of lentiviral HBV-PVs was investigated. Particularly the importance of the ratio of L/S in HBV-PV assembly and infectivity. The study revealed that particles with S alone are not infectious which is consistent with previous findings. Furthermore, HBV-PVs.(gtD4.LM) generated by transfecting 293T-gtD4.S cells with 0.1 µg L expressing plasmid gtD4.LM were the most infectious; the infectivity of resulting HBV-PVs was decreased as the dose of gtD4.LM increased. Whereas, HBV-PVs.(gtB.L) that produced by transfecting 293T-gtD4.S cells with 2.0 µg gtB.L yielded the highest luciferase signal; the luciferase signal was decreased as the dose of gtB.L reduced. Interestingly, HBV-PVs.(gtD4.LM) entered cells in a NTCP independent manner while HBV-PVs.(gtB.L) required NTCP. Two amino acids 7-TS-8 in gtD4.L might have contributed to this NTCP independent behaviour. However, future mutational analysis is required to investigate this hypothesis.

Having established an optimal L/S ratio to generate infectious HBV-PVs, the S producing cell line 293T-gtD4.S can be utilised in a variety of future works. For example, it can be used to assess the neutralising antibody responses or can be used to validate diagnostic tests, such as ELISA, WB, or IF or PCR. Moreover, this approach can also be useful to interrogate other viruses that consist of multiple envelope glycoproteins, such as the influenza viruses including hemagglutinin (HA), neuraminidase (NA), and M2 ion channel. However, the system can be further optimised to tackle the issue of high background noise by incorporating NanoBiT® PPI with HBV trans-complementing system.

Chapter 5: Longitudinal Quasispecies Analysis of Hepatitis B virus using MinION Nanopore Sequencing

5.1 Introduction

5.1.1 HBV quasispecies

Although HBV is a partially double stranded DNA (dsDNA) virus, an unusual feature of its replication cycle, similar to retroviruses is that the genome replicates through an RNA intermediate by a low fidelity polymerase. As a result, mutations occur inevitably whenever the HBV replicates. However, it has a relative low mutation rate when compared with other retroviruses, such as human immunodeficiency virus type-1 (HIV-1). HBV and HIV-1 reverse transcriptase (RT) only share approximately 8% similarity in their amino acid sequences, but approximately 35% sequence identity has been identified in several moderately conserved motifs that form the deoxynucleotide-triphosphate (dNTP)-binding site. Hypothetically, HBV and HIV-1 share common catalytic mechanism when it comes to nucleotide addition to the 3'-end of the primer DNA (Yasutake et al., 2018). It is tempting to suggest that HBV and HIV-1 would share a similar mutation rate. Evidence demonstrates that the mutation rate in HIV-1 (mean substitution per site per cell infection (s/s/c) is 2.4×10^{-5}) is slightly higher than that in Duck HBV (DHBV) (mean s/s/c is 2.0×10^{-5}) (Sanjuan et al., 2010). However, this is a comparison between HIV-1 with DHBV, not with HBV directly. Moreover, the mutation rate is the rate at which errors arise from replication of viral genome (Peck and Lauring, 2018) and in some cases from genome segment reassortment and molecular recombination (Domingo et al., 2012), not taking into account of the fitness of the mutants within host overtime, which is a concept of the

rate of evolution (substitution per site per year). For example, all viral infections endure several selective constraints from the immune response, but there are added complications in the case of HIV-1 as subgroups of cells that participate in the immune response can also be infected resulting in infection-mediated immunosenescence overtime (Domingo et al., 2012). It has been reported that HBV has a rate of evolution (0.0005 substitutions per site per year) around 6-fold less than HIV (0.003 substitutions per site per year) (McNaughton et al., 2019a).

Comparing with other dsDNA viruses, HBV has a mutation rate of approximately 10-fold higher than viruses from other families. This high rate of mutation introduces a large coexisting pool of genetically distinct yet closely related viral variants within the same host (Du et al., 2017), creating a mutant spectrum that enables advantageous virus adaptation. This mutant spectrum will shift in response to selective constraints, or new mutant spectrum will be established as a results of a bottleneck or founder event arising from infection between cell to cell within a host (Domingo et al., 2012).

It is critical to investigate the dynamics of viral quasispecies and estimate an evolutionary rate as this can be utilised to predict disease progression and treatment outcome. For example, drug resistant mutants can be found in the viral population within the hosts who had never been exposed to the drug (Domingo et al., 2012). However, it is challenging to estimate the evolutionary rate for HBV due to its complex biology. There is considerable variation in substitution rate across the genome as a result of high constraints on this small sized genome that encompasses extensive numbers of overlapping reading frames and RNA secondary structure (McNaughton et al., 2019a). Moreover, there are a remarkable amount of interactions between the host and the HBV. For instance, in the early phase of the infection, HBeAg is expressed which stimulates regulatory T CD4 cells that suppress anti-viral T CD8 cells, in turn, assists in viral persistence. At the late chronic phase of the infection, as

a consequence of frequent host production of antibodies against HBeAg, mutants that can escape anti-HBeAg antibodies are selected, resulting in an HBeAg negative status. The most common mutation is an introduction of a stop codon at nucleotide position (nt.) 1896, where a G is replaced by A (G1896A), which prevents the production of HBeAg. This mutation also enhances viral replication rate. However, the mechanisms underlying this are still unclear. It is possible that the A1896 mutant strengthens the secondary structure of the encapsidation signal (ε) (Li et al., 1993). In addition, this HBeAg negative mutant allows the host to mount an unregulated immune response to against the virus. This increased selective pressure in combination with increased replication rate may contribute to the characterisation of lower viral load and high nucleotide diversity (Domingo et al., 2012, Harrison et al., 2011).

5.1.2 HBV structural variations

In general, genomic structural variations (SVs) contain canonical changes including insertions, deletions and duplications. Naturally occurring preS1/S2 mutations during the course of chronic HBV (CHB) infection have been widely reported. The preS1/S2 deletion was found significantly linked to the development of HBV related liver cirrhosis and HCC and most of the deletions involved B- and T-cell epitopes and other functional sites. In CHB, B- and T-cell epitope deletions may be as a consequence of selection for variants that escape host adaptive immune surveillance (Chen et al., 2007, Chen et al., 2006). Deletions ranging from 3 to 422 base pairs were frequently observed in preS1/S2 region in samples of 1 to 10 years prior to a diagnosis of HBV-related HCC. More than half of the deletions ending at aa140-142 of the preS1/S2 domain and more than 80% of them covered the region between aa100 and 140, indicating that this region could be a hot-spot for deletions (Zhang et al., 2015). It is possible that these deletions influence the pathogenesis of HBV infection. L and

truncated M have been reported to be involved in activating Ras/Raf-1/ERK signalling pathway, which acts as another oncogenic factor (Hildt et al., 2002).

In recent years, non-canonical forms of SVs have been reported, termed complex SVs. These complex SVs comprise two or more SVs at the same locus with a mixture of deletions, insertions and duplications (Fujiwara et al., 2018). An analysis of HBV strains discovered 70 HBV strains with complex SVs. These complex SVs were mostly located in the region between nt 1500 and 2000 which possesses the X ORF, PreCore/Core ORF as well as core upstream of regulatory sequence (CURS) and enhancer II regions involved in HBV transcription regulation (Fujiwara et al., 2018). Six classes were identified including class I (insertion ± deletion), class II (insertion ± deletion ± duplication), class III (insertion ± duplication), class IV (deletion ± duplication), class V (multiple duplications), and class VI (highly complicated) (Fujiwara et al., 2018). However, their roles in pathogenesis of HBV-related disease are still unclear.

5.1.3 Nanopore sequencing

From the success of first-generation genome sequencing by Sanger 40 years ago, sequencing technology has rapidly progressed to third-generation sequencing such as Pacific Biosciences (PacBio) and the Oxford Nanopore Technologies (ONT).

Both platforms offer the advantage of sequencing single molecule DNA without the need for PCR amplification and both sequencing ultralong DNA molecules up to 100kbp (Ameur et al., 2019). However, the technology utilised by both platforms is completely different. PacBio employs a sequencing-by-synthesis approach; nucleotides carrying fluorescent labels are incorporated into a single DNA molecule by polymerase, and the fluorescent signals are recorded in real-time by a camera.

Unique library preparation creates circular input molecules allows sequencing many times of one single DNA molecule resulting in highly accurate results (Ameur et al., 2019).

Nanopore sequencing is based on the use of transmembrane proteins that are able to form nanometer-sized pores across an electrically resistant polymer membrane. With a voltage applied across the polymer membrane, a current is passed through the nanopore and the change in current at any one pore can be detected by a sensor. Characteristic changes in electric current are measured as a strand of DNA passes through the pore. The software responsible for basecalling translates these electric currents into a DNA sequence (Figure 5.1) (Jain et al., 2016, Senol Cali et al., 2018). During sequencing, a motor protein unwinding the double-stranded DNA (dsDNA) and feeding a single-stranded DNA (ssDNA) through the pore, the resulting sequence is known as “1dementionaly (D) reads”; while a hairpin adapter ligated to the other end of the dsDNA enabling the sequencing of both strands, subsequently yielding “2D reads” (Rang et al., 2018).

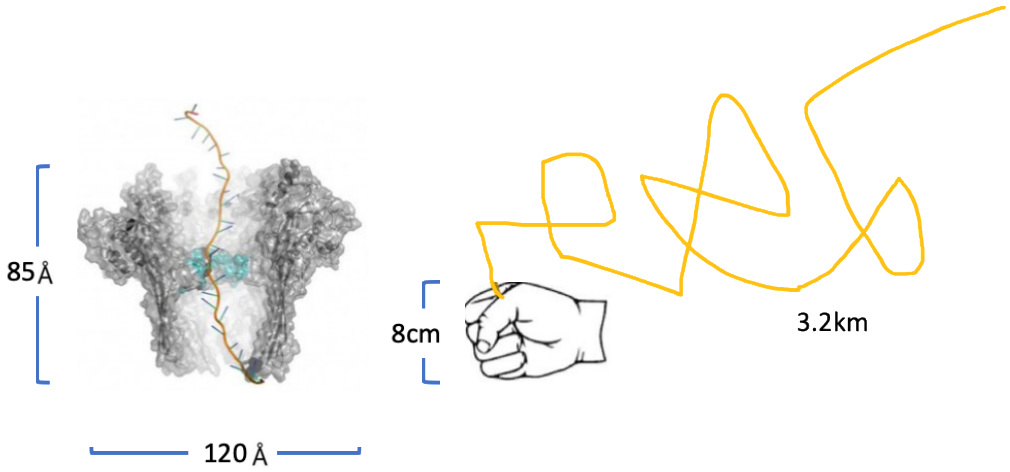

Figure 5.1 Schematic diagram of dimensions of the R9 nanopore and 1 Mbp strand of DNA

R9 nanopore is an engineered Escherichia coli Curlin sigma S-dependent growth (SsgG) membrane protein. Assuming the distance occupied by each base pair (bp) of DNA is 3.4Å, then the length of 1mega bp (Mbp) is 3,400,000Å, which is 40,000x height of the pore. (The diagram was originally designed by Adam Phillippy. Adapted from -https://twitter.com/aphillippy/status/941327354827345920)

5.1.3.1 Benefits of using MinION sequencing

MinION is the only portable, real-time sequencing device. Measuring at approximately 10 x 3 x 2 cm, the device consists of a flow cell where the samples are loaded. The other part is a USB port, through which the device is directly connected to a computer where data collection and analysis are performed. There is no additional lab infrastructure requirement and no complex installation required. Hence, it can be easily transported to remote and resource-limited areas (Senol Cali et al., 2018, Hoenen et al., 2016). In April 2015, a MinION sequencing system was transported in standard airline luggage to Guinea providing real-time genomic sequencing generating results in less than 24 hours (Quick et al., 2016).

Owing to its ability to sequence long-reads using direct RNA or DNA, MinION has been utilised in a wide range of scientific fields, including complete bacterial, fungal and viral genome sequencing, animal, plants, and basic human genome studies. Sequencing whole genomes in real-time allows full characterisation of structural variants (SVs), repetitive regions and RNA splice variants, offering a reliable platform for the analysis of clinical research samples, such as from patients with infectious disease or cancer. The sequence in real-time provides immediate access to results including species discovery, abundance, and antimicrobial resistance.

Traditional technologies sequence short fragments of DNA, for example, the typical read length is approximately 150-300bp by Illumina (Mantere et al., 2019). The short reads are then reassembled back into their original order. However, this creates significant challenges relating to the accuracy of assembly especially in regions where DNA is repetitive and having large structural variations that introduce bias and errors during genome assembly (Senol Cali et al., 2018). Nanopore sequencing, on the other hand, can process much longer reads (>50kb) (Quick et al., 2016). This offers several advantages. First, a longer sequencing read possesses more overlap with other

reads, which in turn, allows easier reassembling back into the correct order. Second, longer reads could potentially cover the full repetitive region, as such, gaps between fragments could be minimised during genome assembly (Senol Cali et al., 2018). This particularly beneficial in reconstruction haplotypes. Long reads obtained by PacBio or Nanopore allow directly linking variants, in particular, rare variants can be assigned properly to each haplotype (Ameur et al., 2019). Haplotype reconstruction enables detection of the population of viral variants within a given sample. Conversely, short reads obtained by Illumina may introduce bias into haplotype building as imputation is required to statistically assign variants to known alleles by using known haplotypes (Ameur et al., 2019) (Figure 5.2).

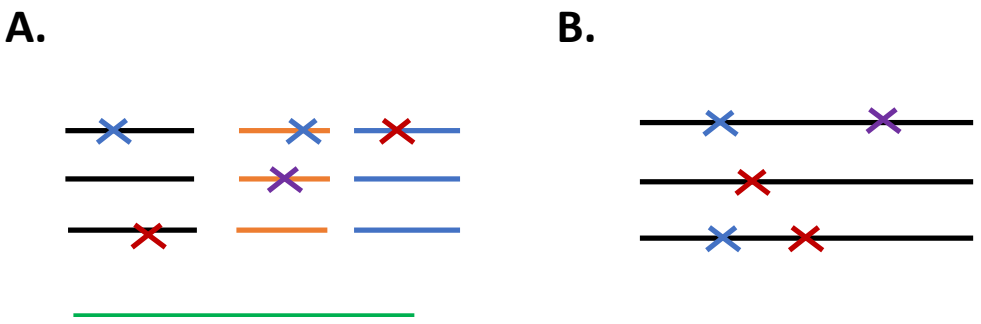

Figure 5.2 Next generation sequencing and haplotype building

A. *Millions of short reads are generated by Illumina. Imputation using a known haplotype is required to assign variants to known alleles;* ***B.*** *Without imputation, variants are directly linked in long reads generated by Oxford Nanopore sequencing or PacBio sequencing. Green line: reference haplotype; crosses: mutations; other colour lines: DNA sequences.*

5.1.3.2 Limitations of MinION sequencing

The lower read accuracy rate has been the major limitation of the MinION sequencing platform when compared with short-read technologies such as Illumina. The accuracy rate was recorded as low as 58.8% when the MinION was first introduced (Rang et al., 2018, Goodwin et al., 2015). As nanopore sequencing involves translating the electric current signal into DNA bases, errors of basecalling can be introduced during the step of recording the signal and the step of translating the signal. Therefore, the speed at which nucleotides pass through the pore, simultaneous multiple nucleotides within the pore, or signal output at which homopolymers pass through the pore could all influence the accuracy of sequencing (Rang et al., 2018).

Attempts have been made to tackle this low accuracy rate. For example, an algorithm known as Nanopolish has been implemented. Nanopolish conducts signal-level analysis using a hidden Markov model to calculate the probability of observing a sequence of electric current signals given an arbitrary nucleotide sequence (Loman et al., 2015). Post Nanopolish, the base-level accuracy improved to 99.5%, comprising 26 mismatches per 100kb and 371 indel errors per 100kb compared to 80 mismatches per 100kb and 921 indel errors per 100kb prior to Nanopolish (Loman et al., 2015).

ONT continues to optimise the sequencing chemistry of the MinION flow cells. Since the first generation of MinION flow cells, it has progressed from nanopore version R6 to the most current R10 that has a longer barrel and dual reader head, which in turn, improve resolution of homopolymeric regions and thereby improves the accuracy of nanopore sequencing data up to 99.995% (Nanopore, 2020).

5.1.3.3 Applications of MinION Sequencing and its use in HBV research

In recent years, single complete 3.2kb HBV molecules have been sequenced using MinION technology (Sauvage et al., 2018). Whilst sequencing libraries were prepared directly from the total products of HBV whole genome PCR reactions, the number of output reads was extremely low ranging from 72 to 4,190 that accounted for 1.9% to 21% of the total reads. In a larger experiment, a median of 60,000 reads ranging from 38,000 to 74,000, max. 139,000) yielded using the sequencing performance with the R7.3 flow cells (Sauvage et al., 2018, Ip et al., 2015).

To enhance the efficacy of sequencing output, Isothermal rolling-circle amplification (RCA) has been used to enrich the HBV genome (Margeridon et al., 2008, McNaughton et al., 2019b) . In this approach, the partially double-stranded HBV DNA is first converted to a complete double-stranded HBV molecule using a completion-ligation method, subsequently, using RCA methods to generate concatemeric amplicons that harbour multiple successive copies of the same genome. In a recent study, this RCA approach was used in MinION sequencing. To enhance accuracy, post sequencing, error correction was applied by comparing multiple genome copies combined into a single concatemer and by analysing reads generated from forward and reverse strands. A variant was considered if a non-consensus base appeared at >60% frequency within one or more concatemers. These potential variants were then further verified using Fisher's exact test and chi-squared contingency test. By applying these approaches, an accuracy rate of 99.7% was achieved (Rang et al., 2018, McNaughton et al., 2019b). While whole genome sequencing with 99.7% accuracy was achieved, the study only tested this approach in a small sample size including 3 sera and one HBV plasmid. In addition, the chosen samples were with very high viral load (>8.23 log IU/mL) and all were HBeAg positive where viral diversity was low.

Whether or not this approach can achieve this high accuracy rate in a larger study using samples from HBeAg negative patients plus low viral load is still unknown.

5.1.4 Aims of this study

In the past, resolving the individual members of quasispecies has proven laborious and time consuming. Previously, viral quasispecies was investigated by clonal analysis from samples and approximately 10 colonies were sequenced for each sample (Osiowy et al., 2006). Although sequencing purified plasmid can yield accurate reads, the amounts of colonies selected are limited using this method suggesting that the full dynamics of quasispecies could not be elucidated. Furthermore, sequence artefacts could be induced during extensive cloning and PCR. Bulk PCR amplification, an amplification reaction starting with multiple templates, has been widely used to interrogate HIV quasispecies genetic diversity. However, polymerase introduces recombination artifacts arising from PCRs starting with multiple genomes (Etemad et al., 2015). To avoid this drawback, Single Genome Amplification (SGA), amplification reaction presumed starting with a single genome, has emerged. This technique attempts to isolate single DNA molecule through serial dilutions. Additionally, a large number of independent SGAs are required to achieve the same level of detection of minority variants within a sample population compared to the bulk PCR amplification, which again is time and labour intensive as well as expensive (Butler et al., 2009).

In this study, MinION sequencing was explored to interrogate HBV quasispecies with the advantages of avoiding multiple PCRs, serial dilutions, cloning, plasmid purification. Ten longitudinal blood samples from 3 patients were used. The primary aims were to:

I. Establish workflow for next generation sequencing of HBV variants;

II. To characterise the intra-host quasispecies of HBV genome over a 2 to 4 year period;

III. To characterise naturally occurring variants in the preS1/S2/S region;

IV. Comparison of the viral diversity between HBeAg-positive and HBeAg-negative hepatitis.

5.2 Methods & Materials

5.2.1 Primers used

All primers used in this study are listed in Table 5.1. Figure 5.3 gives a visual guide on the coverage of the common amplicons and their corresponding primer pairs. Throughout this study, *EcoR1* numbering system was used, where (often hypothetical) the *EcoR1* restriction site is termed as the origin of the HBV genome (Hayer et al., 2013).

Table 5.1 Primers used in this experiment

PRIMER NAME	REFERENCE NAME	SEQUENCE (5'-3')	HBV GENOME POSITION
F1	1584f (Chook et al., 2015)	ACTTCGMBTCACCTCTGCACGT	1583-1604
R1	HBVM (PHE, 2007)	GACACATTTCCAATCAATNGG	989-970
F2	2819f (Chook et al., 2015)	ACCWTATWCYTGGGAACAA	2819-2837
R2	2835r(Chook et al., 2015)	GTTCCCAVGWATAWGGTGAYCC	2835-2814
F3	1584f-ONT	TTTCTGTTGGTGCTGATATTGC ACTTCGMBTCACCTCTGCACGT	1583-1604
R3	2835r-ONT	ACTTGCCTGTCGCTCTATCTTC GTTCCCAVGWATAWGGTGAYCC	2835-2814
F4	2819f-ONT	TTTCTGTTGGTGCTGATATTGC ACCWTATWCYTGGGAACAA	2819-2837
R4	HBVM-ONT	ACTTGCCTGTCGCTCTATCTTC GACACATTTCCAATCAATNGG	989-970
M13F	M13f (Source BioScience)	TGTAAAACGACGGCCAGT	N/A
M13R	M13r (Source BioScience)	CAGGAAACAGCTATGACC	N/A
F5	In-house	CACCATGGGAGGTTGGTCATCCAAA CC	2844-2870
R5	In-house	TTAAATGTATACCCAAAGACAAAAG AAAATTGG	803-836
QHBVF	Outer plus (Brechtbuehl et al., 2001)	GATGTGTCTGCGGCGTTTTA	376 - 395
QHBVR	Outer minus (Brechtbuehl et al., 2001)	CTGAGGCCCACTCCCATAGG	656-637

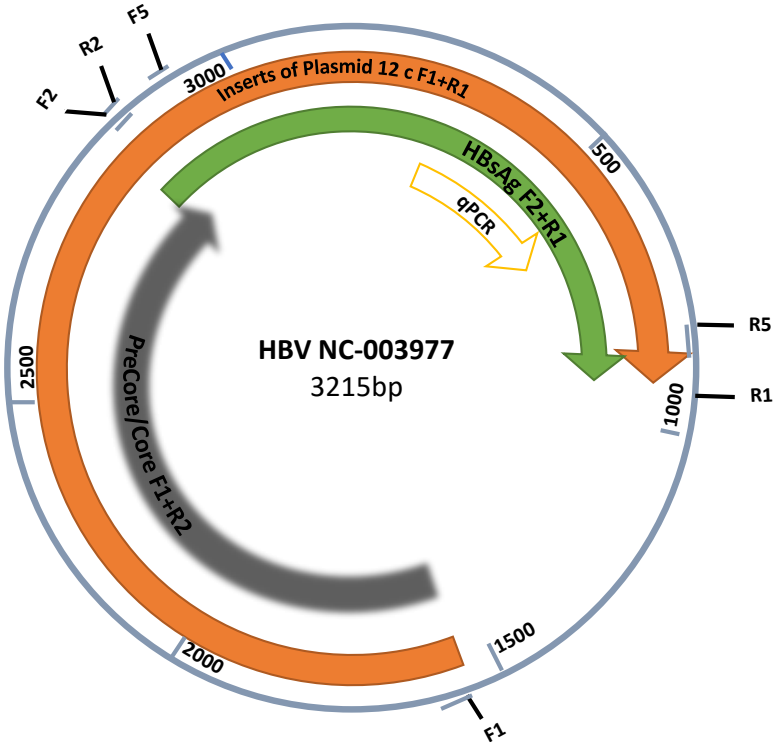

Figure 5.3 Schematic graph depicting the common amplicons in this study and their corresponding primer pairs (the graph is adopted from Chook et al).

In this study, four HBV DNA segments were amplified. The inserts of the Plasmid 12C, spanning from nucleotide (nt.) 1583-3215; 1-989 (a total of 2621 nucleotides), was the largest amplicon using primer pair F1 + R1. The segment preCore/Core was amplified using F1 + R2 spanning from nt.1583-2837 (a total of 1254 nucleotides). The full length of HBsAg was amplified using F2 + R1 spanning from nt. 2819-3215; 1-989, yielding a total of 1386 nucleotides. A short segment sequence, spanning from nt. 376-656 (total: 280 nucleotides), was amplified for the purpose of quantitative polymerase chain reaction (qPCR) using a primer pair qHBVf and qHBVr.

5.2.2 Plasmid 12C

Plasmid 12C was constructed by cloning a PCR amplicon of the HBV genome spanning from *nt. 1583-3215; 1-989* (from a genotype C HBV infected individual) (Figure 5.4). As such, this represents a single HBV template. Nucleotide variants would not be expected to be observed for this sample during downstream variant calling, allowing an assessment of the error introduced during the nanopore sequencing protocol.

A patient (12C) HBV DNA extract was amplified using F1 (nt.1583-1604) and R1 (nt. 989-970) (Figure 5.3). A 25 µL reaction contained 12.5 µL LongAmp Hot Start Taq 2x Master Mix (Biolab), 1 µL of each 10 µM primers, 1 µL 12C DNA and 9.5 µL of nuclease-free water. The cycle of PCR was initiated at 94°C for 30 seconds and followed by 55 cycles of 94°C for 30 seconds, 50°C for 30 seconds, 65°C for 2 and half minutes, finished with a final extension at 65°C for 10 minutes. The PCR products were analysed by 2% agarose gel electrophoresis stained with ethidium bromide (EtBr). The PCR products were then cloned into a pGEMT vector (Promega) following the manufacturer's instruction. In short, a ligation reaction mixture containing 5 µL 2x rapid ligation buffer, 1 µL pGEMT vector, 1 µL PCR product, 1 µL T4 DNA ligase and 2 µL nuclease-free water was incubated at 22°C for 2 hours and the resulting 2 µL of the ligation reaction product was added into 25 µL α-select gold efficiency chemically competent cells (Bioline). After incubation at 42°C for exactly 30 seconds, the cells were rescued from the heat shock by addition of 450 µL pre-warmed SOC media (containing 2% tryptone, 0.5% yeast extract, 0.4% glucose, 10 mM NaCl, 2.5 mM KCl, 10 mM $MgCl_2$ & 10 mM MgSO4). This mixture was then shaken with an agitation rate of 300 rpm at 37°C for one hour before spreading onto a LB agar plate containing 2% X-gal, 40 $\mu g.mL^{-1}$ Isopropyl β-D-1-thiogalactopyranoside (IPGT) and 100 $\mu g.mL^{-1}$ ampicillin. After incubation at 37°C overnight, any colonies with blue pigment were

excluded from the colony screening, whereas, white colonies were picked and added into a PCR mixture containing 0.6 µL dNTPs, 0.6 µL 10 µM M13F, 0.6 µL 10 µM M13R, 1.5 µL 10x Taq Buffer, 0.075 µL Hot Start Taq polymerase and 11.98 µL nuclease-free water. The PCR cycling condition was initiated at 95°C for 15 minutes, followed by 35 cycles of denaturation at 95°C for 20 seconds, annealing at 55°C for 20 seconds, extension at 72°C for 3 minutes and finished with a final extension at 72°C for 10 minutes. The PCR products were analysed by 2% agarose gel electrophoresis stained with EtBr. An overnight liquid bacterial culture containing a bacterial clone with the correct band confirmed by agarose gel electrophoresis was set up for miniprep. Plasmid purified from this bacterial culture was sent for Sanger sequencing. Two sets of primer pairs F1+R1 and F5 (nt. 2844-2870) +R5 (nt. 803-836) were utilised for Sanger sequencing to cover the full length.

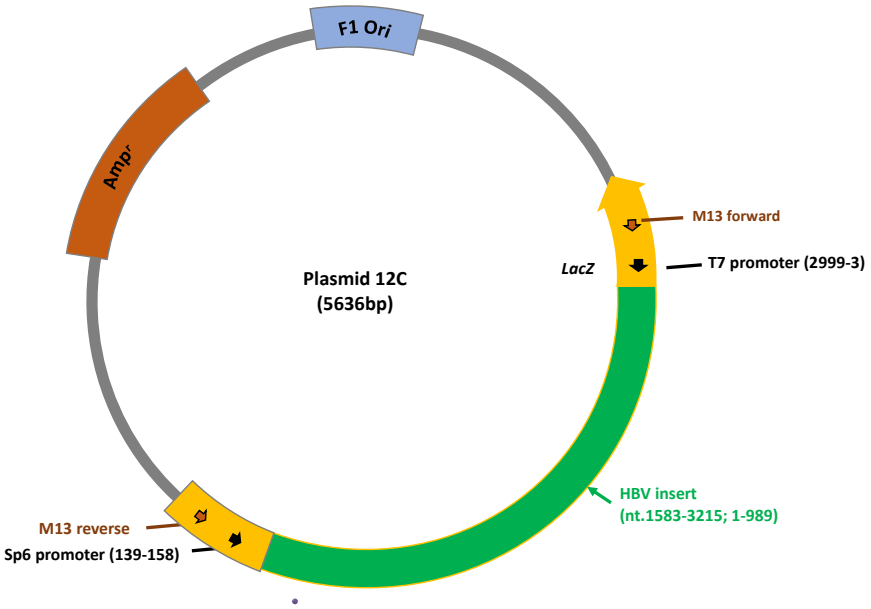

Figure 5.4 Schematic graph of plasmid 12C.

Plasmid 12C was created by inserting an HBV genotype C amplicon into the pGEM-T Easy vector. The 12C amplicon was amplified using primer pair F1 (nt.1583-1604) and R1 (nt. 989-970) resulting in a product size of ~2.5kb. This plasmid was used as a positive control. The graph is adapted from https://www.promega.com/-/media/files/resources/protocols/technical-manuals/0/pgem-t-and-pgem-t-easy-vector-systems-protocol.pdf

5.2.3 Clinical Samples

10 HBV DNA positive sera with known viral load were obtained from Nottingham University Hospital NHS Trust, Nottingham, UK and were held under ethical approval of the Nottingham Health Science Biobank. Patient 1 was tested negative for HBeAg; whereas, patients 2 and 3 were tested positive for HBeAg. The viral titres varied from 3.0×10^5 to $> 10^9$ IU/mL. Prior to viral DNA extraction, Samples were stored at -70°C. More details of the samples are listed in Table 5.2.

Table 5.2 Clinical samples used in this experiment

SUBJECTS	NAME OF SAMPLES	DATE OF SAMPLE COLLECTION	VIRAL LOAD (IU/ML)	ALT (0-45) IU/L	HBEAG	FIBROSCAN	THERAPY	GENOTYPE
PATIENT 1	P1-T1	29/01/2014	303,685	-	Negative	N/A	No	A1
	P1-T2	14/07/2016	328,414	29	Negative	normal	No	A1
	P1-T3	25/06/2018	648,900	331	Negative	N/A	No	A1
PATIENT 2	P2-T1	29/09/2014	15,285,400	92	Positive	N/A	Interferon	C1
	P2-T2	27/10/2014	6,969,317	422	Positive	N/A	Interferon	C1
	P2-T3	08/06/2015	16,941,931	63	Positive	N/A	Interferon	C1
	P2-T4	01/06/2016	$>10^9$	45	Positive	N/A	No	C1
PATIENT 3	P3-T1	12/03/2015	188,406,722	17	Positive	N/A	No	C1
	P3-T2	26/10/2017	54,265,835	15	Positive	N/A	No	C1
	P3-T3	09/11/2017	17,560,846	18	Positive	N/A	No	C1

5.2.4 HBV DNA extraction

Two different kits were evaluated for DNA extraction efficiency: QIAamp Viral RNA Mini Kit (VRMK) (Qiagen) and QIAamp MiniElute Virus Spin Kit (MEVK) (Qiagen). Both extractions were conducted in a Class I biological safety cabinet. Following the manufacturer's instructions, 140 µL serum was used for the VRMK. First, serum was incubated with 560 µL of lysis buffer for 10 minutes at room temperature followed by mixing with 560 µL of 100% ethanol thoroughly to obtain a homogeneous solution. The solution was then transferred to the QIAamp Mini column and centrifuged at 6000 x g for 1 minute. After discarding the flow though, the spin column was washed with 500 µL of buffer AW1 and further washed with 500 µL of buffer AW2. Finally, the viral DNA was eluted with 40 µL of buffer AVE. The viral DNA was stored at -20°C.

In order to minimise bias, 140 µL serum was also used in the MEVK extraction method. A 25 µL aliquot of Qiagen protease for each sample extraction was required in addition of 0.9% sodium chloride (0.9% normal saline) and sample to make up a total volume of 225 µL. Serum samples were lysed with 200 µL of Buffer AL for 15 minutes at 56°C and followed by mixing with 250 µL of 100% ethanol thoroughly and incubation for a further 5 minutes at room temperature. The total lysate was then transferred to a QiAamp MiniElute column and centrifuged at 6000 x g for 1 minute. After discarding the flow-through, the sample was washed with 500 µL of Buffer AW1 and further washed with 500 µL of Buffer AW2. Additionally, a final wash with 500 µL of 100% ethanol was required. After drying the column at 56°C for 3 minutes, the viral DNA was eluted with 40 µL of buffer AVE and the viral DNA was stored at -20°C.

The efficiency of DNA extraction was determined by performing qPCR obtaining cycle threshold (Ct) values, which is expressed as quantification cycle (Cq) value when BioRad CFX96 (Bio-Rad) is used. One primer pair qHBVf (nt. 376-395) and qHBVr

(nt. 656-637) to detect a short fragment of ~250 bp was used (Brechtbuehl et al., 2001). The qPCR was performed as described in Chapter 2 Materials and Methods.

5.2.5 PCR

Two fragments of HBV were amplified: PreCore/Core and PreS1/S2/S. PreCore/Core was amplified using F1 and R2 (nt. 2835-2814) while PreS1/S2/S was amplified using F2 (nt. 2819-2837) and R1. A 25 µL reaction containing 12.5 µL LongAmp HotStart *Taq* 2x Master Mix (Biolab), 1 µL of each 10 µM primers, 1 µL HBV DNA and 9.5 µL of nuclease-free water. The cycle of PCR was initiated at 94°C for 30 seconds and followed by 55 cycles of 94°C for 30 seconds, 50°C for 30 seconds, 65°C for 2 and half minutes, finished with a final extension at 65°C for 10 minutes. The PCR products were analysed by 2% agarose gel electrophoresis stained with EtBr.

5.2.6 Quantification of bands relative intensity

Where quantification was required, PCR gel images were analysed using Fiji (Schindelin et al., 2012). The output of the analysis was a histogram indicating intensity of each band in numerical value. Each value was divided by the value of a positive control (plasmid 12C), which gave a final relative intensity of the calculated band.

5.2.7 Sanger sequencing and genotyping

Based on agarose gel electrophoresis, PCR products with the brighter bands (relative intensity between 0.5-1.0) were diluted 1:10 in nuclease-free water, whereas PCR products with the weaker bands (relative intensity between 0.25-0.5) were diluted 1:5 in nuclease-free water and couriered at ambient temperature for sequencing by

Source Bioscience, Nottingham, UK. Each product was sequenced using F1 and R2 to cover PreCore/Core and F2 plus R1 to cover PreS1/S2/S. All Sanger reads were assembled to produce two separate complete contiguous sequences (contig) covering PreCore/Core and HBsAg including PreS1, PreS2 and S. These served as reference sequences for Nanopore reads. The genotypes were determined using a web-based tool HBV geno2pheno (https://www.geno2pheno.org/) (Beggel et al., 2012).

5.2.8 Library preparation

A total of 12 samples including ten HBV DNA extracts, one HBV plasmid (12C) and an aliquot of nuclease-free water were prepared for the MinION sequencing. A nested four-primer PCR approach using the SQK-LWB001 library preparation kit was employed to amplify the samples. The first round was carried out using 'inner' primers. These inner primers consisted of sequences specific for target DNA and a short fragment at 5' acting as a tag for the 'outer' primers. The 'outer' primers were universal primers provided by ONT composed of barcode and the linker sequences (Figure 5.5).

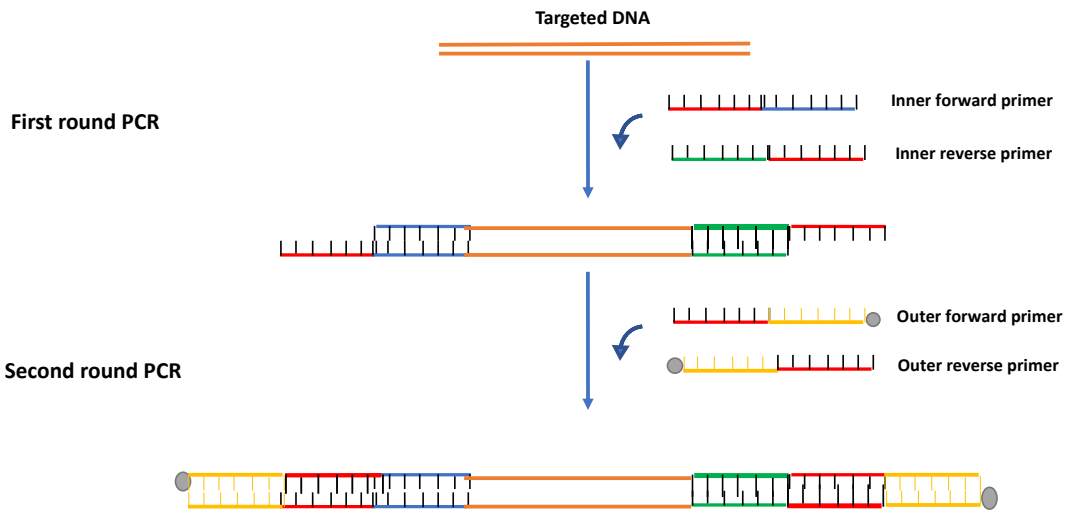

Figure 5.5 Schematic graph of the two-stage PCR

A nested four-primer PCR approach was employed to amplify the fragments. The first round was carried out using 'inner' primers. These primers consisted of target primers and tailed the 5' end with specific sequences allowing attachment by 'outer primer'. The outer primers, known as universal primers, were supplied by Oxford Nanopore technologies (ONT), consisting of barcode and linker sequences. Red represents the linker primers; blue represents forward inner primers; the green represents reverse inner primers; and the orange represents the barcodes.

Primers F3 (nt.1583-1604) and R3 (nt.2835-2814) were used to amplify PreCore/Core; while F4 (nt.2819-2837) and R4 (nt.989-970) were used to amplify PreS1/S2/S (Figure 5.3). These inner primers allow the enrichment of the target fragments. Post first round PCR, 2 µL of the PCR product including 1 µL from PreCore/Core reaction and 1 µL from PreS1/PreS2/S reaction was mixed together and used as the template for the second stage PCR. In this round, a 50 µL PCR reaction was set up containing 25 µL of 2x LongAmp HotStart *Taq* Master Mix (Biolabs), 1 µL of barcoded primer mix (including 'outer' primers, supplied by ONT) and 21.5 µL of nuclease-free water. Cycling conditions included 94°C for 1 minute, and 30 cycles of 94°C for 30 seconds, 62°C for 30 seconds, 65°C for 2 and half minutes, followed by a final extension at 65°C for 10 minutes. From this reaction, 3.5 µL of the product was analysed by 2% agarose gel electrophoresis, and once successful amplification was confirmed, the product was prepared following the SQK-LWB001 kit protocol. In brief, 40 µL of the PCR product was bound to equal amounts of AMPure XP beads (Beckman Coulter) and washed twice with 200 µL of freshly prepared nuclease-free 70% ethanol in the aid of a magnet. After briefly drying at room temperature, the PCR products were eluted in 10 µL 10 mM Tris-HCl (pH 8.0) with 50 mM NaCl. 1 µL of the cleaned products was then quantified by a Qubit fluorometer using the dsDNA high sensitivity kit (ThermoFisher). 0.5 µL of each product at a concentration of 25 ng/µL were pooled to achieve a total of 300 ng in 10 µL per MinION flow cell. This DNA library was then incubated with rapid sequencing adapters for 5 minutes at room temperature and ready to be loaded into the SpotON flow cell on a MinION Mk II device. The library was run for 48 hours through a computer running MinKNOW 1.10.16 in Microsoft Windows.

.

5.2.9 MinION basecalling, adapter trimming and quality control

MinION basecalling and sequencing analysis was conducted by Dr Stuart Astbury. The workflow is depicted in a flow chart (Figure 5.6). In brief, the work involved basecalling using Guppy v2.3.5 followed by adapter trimming in porechop 0.2.4-beta. Next, Nanofilt (https://github.com/wdecoster/nanofilt) was employed to remove ~200bp PCR artefacts and set a minimal read length at ~1000bp. The rationale behind this was that as the amplicons (either PreCore/Core or S) were both ~1200bp and potential ~200bp deletions were interested, hence the minimal read length threshold at ~1000bp.

5.2.10 Alignment and Variant calling

Filtered, adapter trimmed reads were aligned to their respective Sanger sequences using Minimap2 with the *map-ont* function optimised for long, high-noise reads. As both PreCore/Core and PreS1/S2/S sequences were pooled together all reads were first aligned to the PreS1/S2/S reference sequence and extracted, and non-aligned reads were aligned to the PreCore/Core reference sequence (Figure 5.6). This filtering was carried out on .bam filters produced in Minimap2 using Samtools. These alignments generated two .fastq files that were subsequently used for variant calling. Variants were called using LoFreq to generate a .vcf file of minor variants within each gene relative to the Sanger sequence.

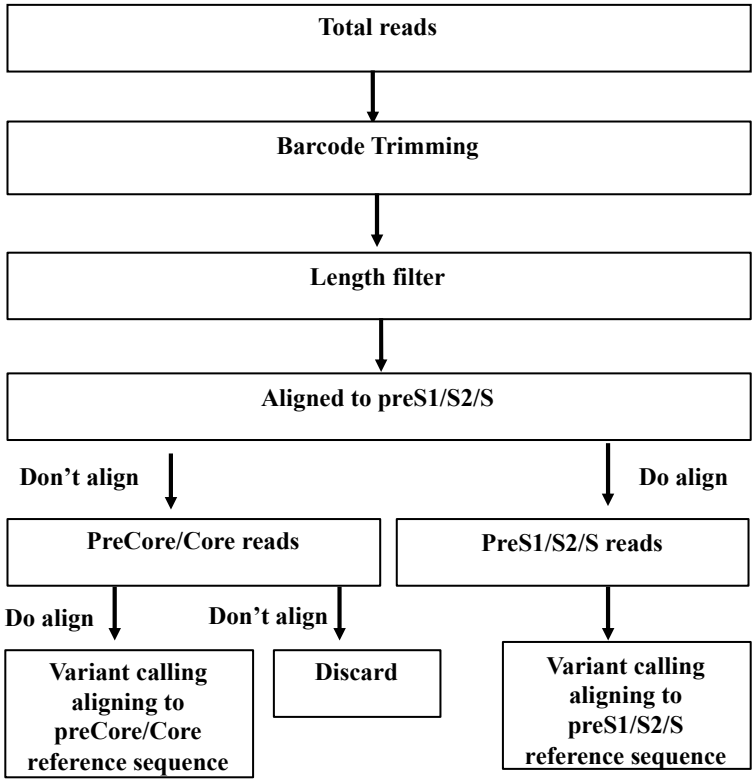

Figure 5.6 Flow chart of MinION sequencing data analysis

Raw data from the MinION sequencing was processed to remove the barcodes, only sequences with correct length were analysed further. The initial analysis was alignment to preS1/S2/S, the ones that did not align to this region were subsequently aligned to preCore/Core. Only reads matched to one of the two were used for further variant calling.

5.2.11 Intra-host evolution analysis

Variants of the overlapping (nt.2856-3210; nt.1-835) and the nonoverlapping preCore/Core (nt.1839-2306) regions from all time points of Patient 1 were processed for potential haplotypes (this analysis was performed by Dr Stuart Astbury). The resulting haplotypes including 137 from *pol* and 12 from preCore/Core were aligned in MEGA 7.0 based on amino acid sequence then exported as FASTA file. The evolutionary rate was estimated using BEAST v2.6.0 (Bouckaert et al., 2019). Following the protocol detailed previously to measure evolving populations (Drummond et al., 2015), the alignment was imported into BEAUti 2 to construct a .xmL file. Hasegawa-Kishino-Yano (HKY) model with empirical base frequencies was chosen for substitution model, and a strict molecular clock was applied. A Coalescent Constant Population model prior was used. The Markov Chain Monte Carlo (MCMC) analysis was run for 10×10^7 steps. After processing in BEAST v2.6.0, the output was analysed in Tracer v1.7.1, and the viral evolutionary rates were calculated (Rambaut et al., 2018). Time-scaled maximum clade credibility (MCC) tree was generated by TreeAnnotator v2.6.0 after discarding 10% as burn-ins, and the tree was viewed in FigTree v1.4.4 (Bambaut, 2018).

5.2.12 Identification of conserved and variable residues in the overlapping region

Amino acids alignments of the 137 variants from Patient 1 were imported into BioEdit Sequence Alignment Editor (Hall, 1999). The Shannon entropy of each residues in the overlapping region of *pol* and preS1/S2/S were calculated and displayed as entropy (Hx) plots.

5.3 Results

5.3.1 Plasmid 12C

Post overnight incubation, blue and white colonies appeared on Luria-Bertani (LB)/ampicillin selection agar plates consisting 40 µg.mL^{-1} IPTG and 2% X-gal. Approximately 400 blue colonies and 45 white colonies were observed. Out of 32 white colonies screened, 10 colonies harboured the DNA inserts with the desired sizes. Plasmid yielded from one of the colonies was sent for Sanger sequencing and confirmed that it was a genotype C.

Plasmid 12C was prepared in parallel with other HBV DNA extracts for MinION sequencing. The sequencing run yielded 66,466 reads with a modal read length of 1,250bp after applying Nanofilt. The reads were then aligned to Sanger sequences of either PreCore/Core or PreS1/S2/S. Variant calling results showed that 100% alignment was achieved. The only variant observed was a base “N’ in the primer site. The primers were included in the analysis intentionally to validate the basecalling ability of the nanopore pipeline.

5.3.2 The negative control in MinION sequencing

In order to monitor whether there were cross contaminations between samples, nuclease-free water was prepared in parallel with other samples. An aliquot of each library preparation was pooled together into one flow cell and performed multiplex sequencing.

After the first round of PCR amplification, the negative control template did not yield any products. As shown in Figure 5.11A & Figure 5.12A, no bands were visible in the

2% agarose gel electrophoresis. However, after the second round of PCR, electrophoresis revealed bright bands with a size of ~200bp (Figure 5.11B & Figure 5.12B). The size represented the barcode from each end that had been amplified. After MinION sequencing, this negative sample yielded a total of 722,456 reads and the modal read length was 56bp. None of the reads aligned to HBV genome when analysed using NCBI BLAST.

5.3.3 HBV DNA extraction

To eliminate bias introduced by sample volumes, 140µL of serum was used for both methods including VRMK and MEVK. A summary of similarity and differences between these two methods were listed in Table 5.3.

On completion of DNA extraction, real time-qPCR was performed. Primers qHBVf and qHBVr (Brechtbuehl et al., 2001) were used to cover approximately 250bp in the S region. Plasmid 12C was used as a positive control and nuclease-free water was used as a no template control (NTC). Both were performed to exclude false positives or negatives. As shown in Figure 5.7B, the amplicons of NTC sample melted at lower temperature indicating these amplicons were primer dimers. The minor peaks following the major ones observed in HBV DNA samples suggested secondary structures of DNA that required higher melting temperature. The quantification values (Cq) values for NTC sample was higher than the tested samples, which suggests that no false positive results were reported in all samples. The Cq values inversely correlated initial copy numbers of HBV DNA. Across the whole panel, Cq values of MEVK method were approximately one third lower compared to that of the VRMK (Figure 5.7).

Table 5.3 Comparison of QIAamp Viral RNA Mini Kit and QIAamp MinElute Viral Spin Kit.

KITS	STARTING VOLUME	ELUTE VOLUME	EXTRA REAGENT REQUIRED	COST PER 50 PREP	LABORATORY EQUIPMENT	TIME REQUIRED (MINUTES)
VRMK[1]	140µl serum	50µl	Carrier RNA	£205	Class I safety cabinet, Vortex, heating block; microcentrifuge	30-45
MEVK[2]	140µl serum + 60µl PBS[3]	50µl	Protease	£213	Class I safety cabinet, Vortex, heating block; microcentrifuge	45-60

1. VRMK- QiAamp Viral RNA Mini Kit; 2. MEVK – QIAamp MinElute Viral Spin Kit; 3. PBS – Phosphate-buffered Saline;

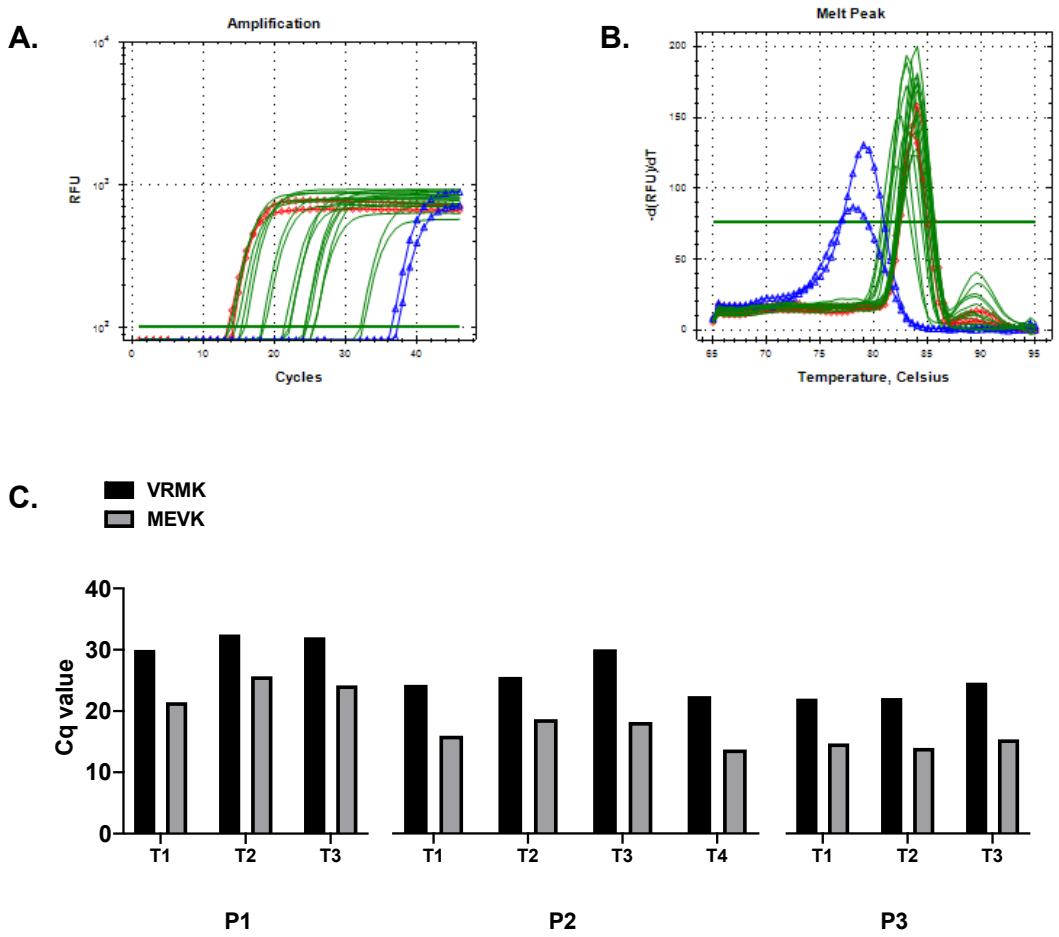

Figure 5.7 Comparison of HBV DNA extraction efficiency between QiaAmp viral RNA (VRMK) mini and QiaAmp MiniElute (MEVK) spin kits

HBV DNA was extracted from 10 HBV positive sera using QiaAmp viral RNA mini and QiaAmp MiniElute. The efficiency of DNA extraction of each method was assessed by comparing the quantification cycle (Cq) values of each sample obtained from real RT-qPCR using CFX96 (BioRad). ***A.*** *Amplification plot displays fluorescence signal (y-axis) against the number of cycles (x-axis) for each sample;* ***B.*** *Melting peak graph displays the dissociation characteristics of each amplicon during heating;* ***C.*** *The Cq values of all samples were displayed. Plasmid 12C was used as the positive control and nuclease-free water was used as the no template control (NTC). The lines with red diamond represent the positive controls; the lines with blue triangles represent the negative controls, and the greens lines represent samples of all HBV DNA extractions.*

5.3.4 Library preparation

Following generation of a consensus sequence for each sample, DNA isolates were prepared for MinION sequencing. A nested four-primer PCR approach was employed to amplify the fragments.

Initially, it was planned to amplify one large amplicon using primers F3 (nt. 1583-1604) and R4 (nt. 989-970), amplifying a ~2.5kb spanning from PreCore/Core to S. To confirm successful amplification, all PCR products were analysed by agarose gel electrophoresis. The first round PCR generated products of the desired size and satisfactory band intensity. As shown in Figure 5.8A, bright bands (relative intensity above 0.25) were seen for the majority of samples apart from P1-T2 (relative intensity ~0.05). P1-T2 had a low quantity of starting DNA templates, as a result, a bright band for this sample was not expected. Therefore, the second stage in an attempt to attach a barcode to each amplicon was proceeded. However, the second stage amplification was not successful as weak or even no bands were seen across the whole panel (Figure 5.8B). This could be due to sub-optimal annealing temperature during PCR, the concentration of input DNA templates or denatured “outer” primers.

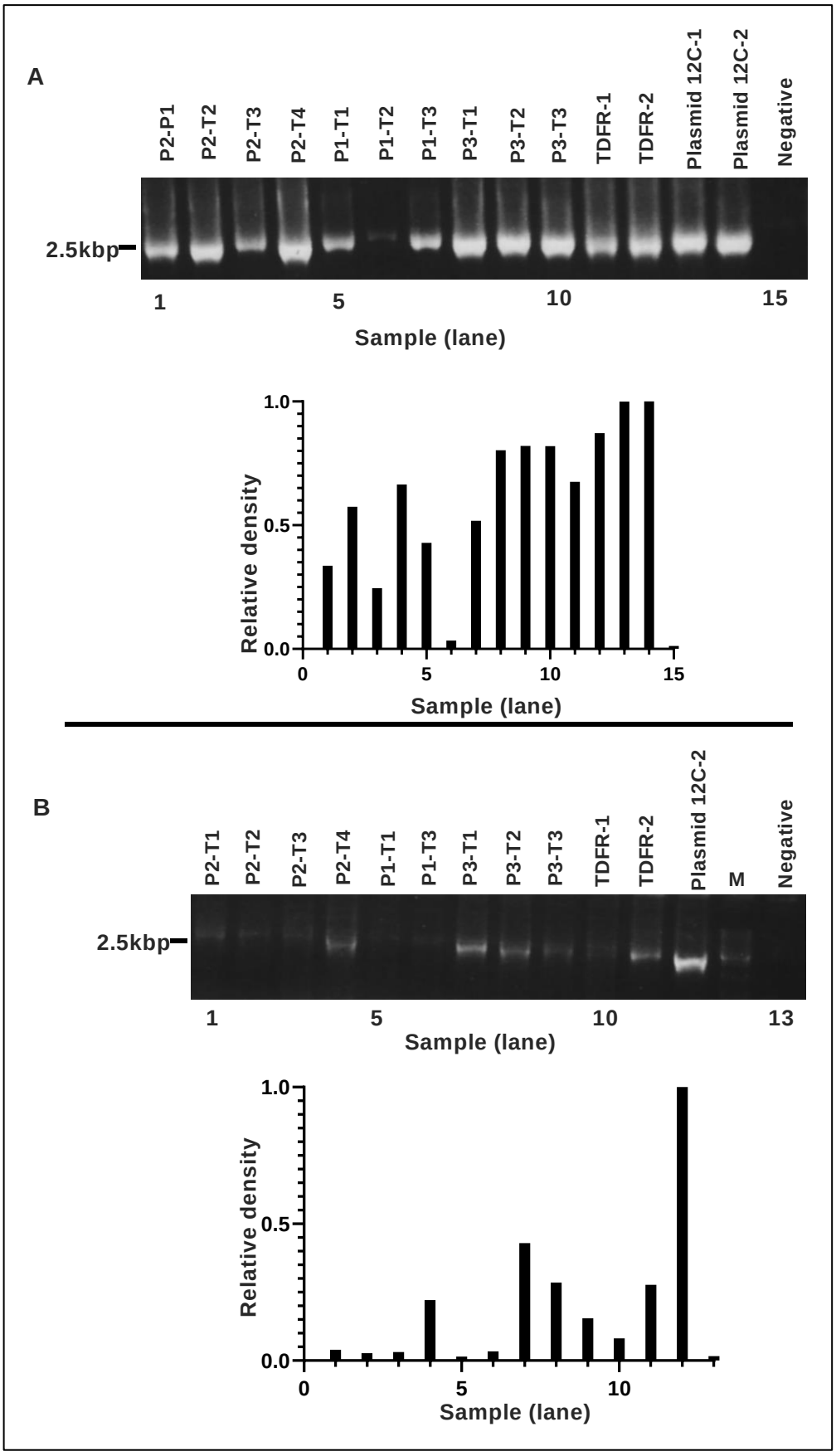

Figure 5.8 Failure of second stage PCR to amplify the 2.5kbp products

A. *One amplicon per sample using F3 (nt. 1583-1604) and R4 (nt. 989-970) amplifying ~2.5kb amplicons spanning from PreCore/Core (1584 to 3215 nt; 1 to 989 nt) to S;* ***B.*** *Barcoded PCR products of the second stage amplification. Samples TDFR-1 and TDFR-2 were both supplied by Dr Stuart Astbury, used as clinical positive controls. Plasmid 12C-1 (neat) and plasmid 12C-2 (1/10 dilution) were used as positive controls developed by ligation of HBV DNA into a plasmid vector (pGEM-T Easy). Nuclease free water was used as the Negative control. M, marker. Graphs of relative intensity are below their corresponding gel images. The relative intensity was calculated using Fiji.*

In order to find out the cause, a second round PCR varying annealing temperatures and DNA concentrations was performed. To simplify the experiment, only two samples were tested. P2-T3 represented the low DNA templates (band relative intensity ~0.25) and P2-T4 represented the high DNA templates (band relative intensity ~0.65). Both isolates were loaded for second round PCR at neat, 1 in10, 1 in 100 and 1in 1000 dilutions at annealing temperatures of 57°C, 62°C and 67°C. Results showed that no PCR products were amplified at any of the temperatures or dilutions in sample P2-T3. Sample P2-T4, on the other hand, was best amplified with neat concentration at either 57°C or 62°C (Figure 5.9).

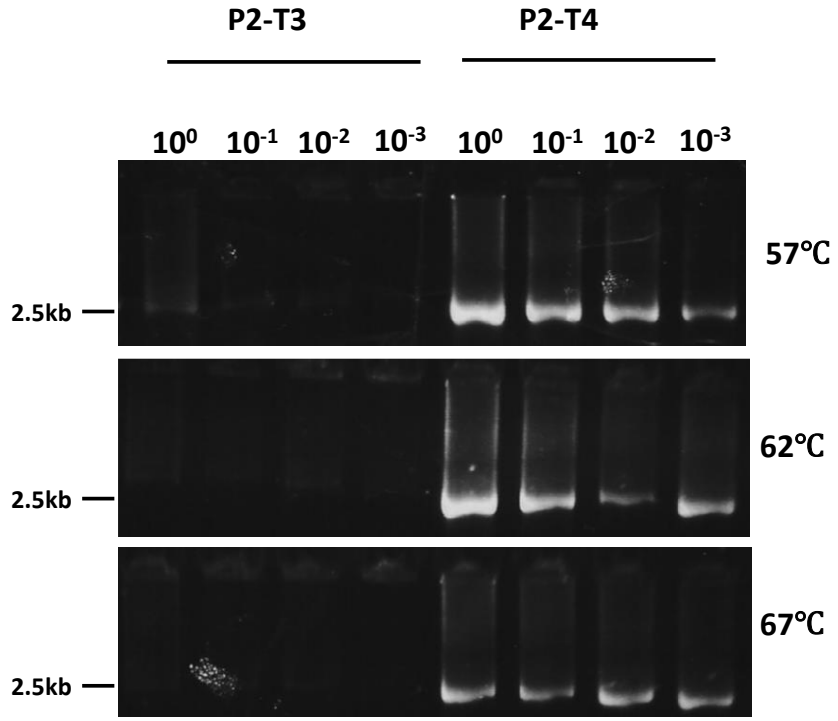

Figure 5.9 Optimisation of second round PCR

HBV DNA Isolates of samples P2-T3 and P2-T4 were loaded for second round PCR at neat, 1 in10, 1in 100 and 1in 1000 dilutions running at annealing temperatures of 57℃, 62℃ and 67℃. PCR products were analysed by 2% agarose gel electrophoresis.

Due to unsuccessful optimisation by varying annealing temperature and DNA concentrations, it was considered the possible cause could be due to denatured “outer” primers. Therefore, a new set of barcoded BCO1 was ordered. This barcode was the same as the outer primers described above but without ligation enzyme in order to reduce costs. This set of barcoded BC01 was used to set up a second run from the first round PreS1/S2/S PCR products (~1.5kb). Figure 5.10 demonstrates all samples amplified sufficiently except P1-T2. Continuing successful amplification using BC01, amplifying the 2.5kb amplicon was attempted. However, comparing with the 1.5kb products, the concentration of 2.5kb amplicons were potentially too low for downstream MinION sequencing (data not shown). Therefore, the library preparation with the 1.5kb amplicons was proceeded. As shown in Figure 5.11B & Figure 5.12B, all second round PCR yielded several sizes of amplicons including the desired ~1.5kb products as well as products with a size of 100bp-200bp.

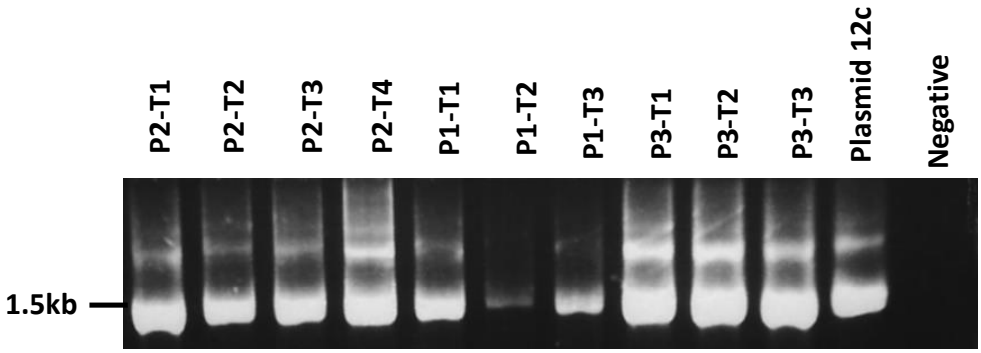

Figure 5.10 BC01 barcode PCR

Barcoded BCO1 was a set of outer primers without the ligation enzyme. The second round PCR using Barcode BCO1 forward and reverse primers using DNA templates from the first round PCR with PreS1/S2/S PCR products (~1.5kb).

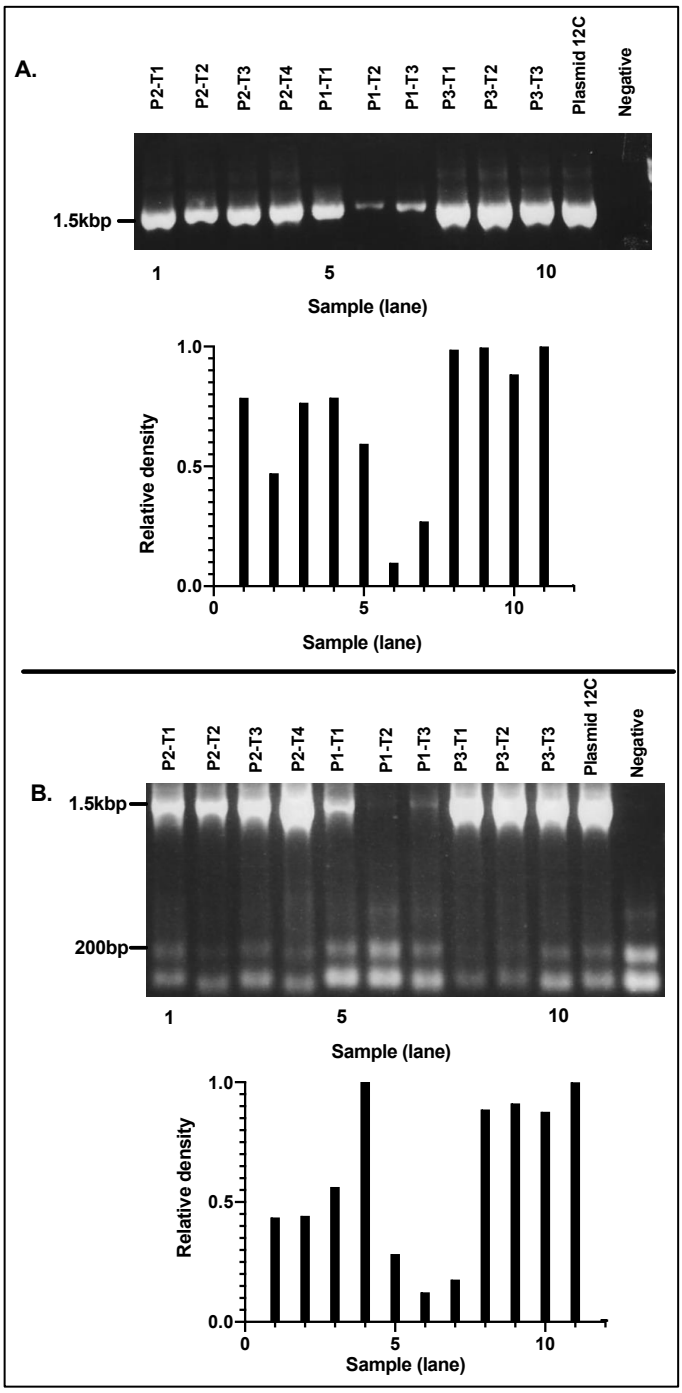

Figure 5.11 First and second stage PCR for preS1/S2/S

A. *First round PCR using primers F4 (nt. 2819-2837) and R4 (nt. 989-970) amplified preS1/S2/S (~1.5kb);* ***B****. Second round PCR attached barcodes to first round PCR products. preS1/S2/S and preCore/Core of each sample shared one barcode. Plasmid 12C was a positive control and nuclease-free water was the negative control. Graphs of relative intensity are below their corresponding gel images. The relative intensity was calculated using Fiji.*

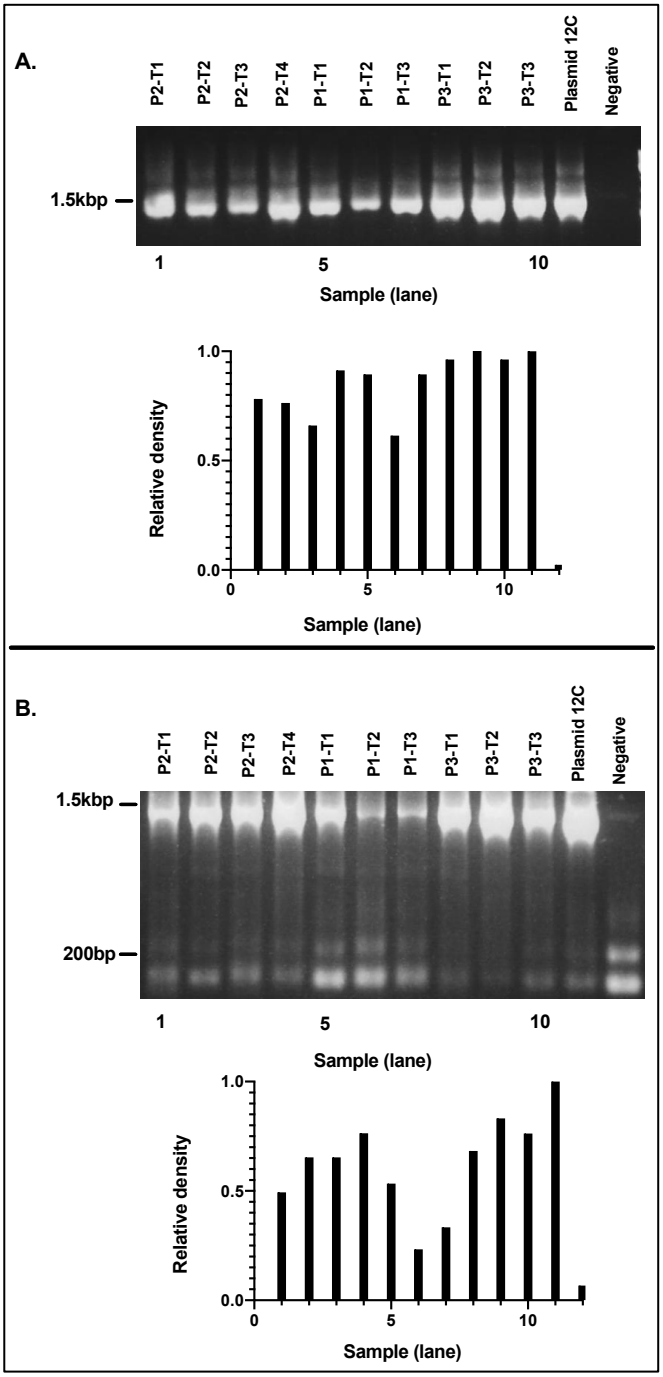

Figure 5.12 First and second stage PCR for preCore/Core

A. First round PCR using primers F3 (nt. 1583-1604) plus R3 (nt. 2835-2814) amplified preCore/Core (~1.3kb); B. Second round PCR attached barcodes to first round PCR products. preS1/S2/S and preCore/Core of each sample shared one barcode. Plasmid 12C was a positive control and nuclease-free water was the negative control. Graphs of relative intensity are below their corresponding gel images. The relative intensity was calculated using Fiji.

5.3.5 Nanopore sequencing

Upon first round PCR amplification, PCR products were processed for Sanger sequencing. Two pairs of primers were used. F1 (nt. 1583-1604) plus R2 (nt. 2835-2814) were used to obtain forward and reverse reads of preCore/Core, while F2 (nt. 2819-2837) plus R1 (nt. 989-970) were used to acquire forward and reverse reads of preS1/S2/S. Consensus sequences were generated by inputting both forward and reverse reads of each sample to generate contigs using SeqMan Pro (DNASTAR) (this analysis was performed by Dr Stuart Astbury).

Following library preparation, the DNA library was loaded into a SpotON flow cell and sequencing was run for 48 hours. The vast majority of reads were completed in the first 24 hours and rate of sequencing reduced sharply over time (Figure 5.13A). The total number of reads recovered from MinION sequencing was approximately 10 million. However, the modal read length was ~331 bp. After barcode trimming and Nanofilt to discard any reads with length below 1000 bp, the total number of remaining reads was 438482 with modal read length of 1257 bp. The least recovered sample was approximately 4000 reads for sample preS1/S2/S of P1-T2; and the highest recovered was approximately 26000 reads for sample preCore/Core of P3-T2 (Table 5.4). This difference in reads recovery correlated well with Cq values shown in Figure 5.7c.

These reads were aligned to either preS1/S2/S reference sequence or preCore/Core. This process generated two .fastq files, which was processed for variant calling. Patients 1, the HBeAg-negative patient, had the highest number of variants, while Patient 2 and 3 who were HBeAg-positive, had the least number of variants.

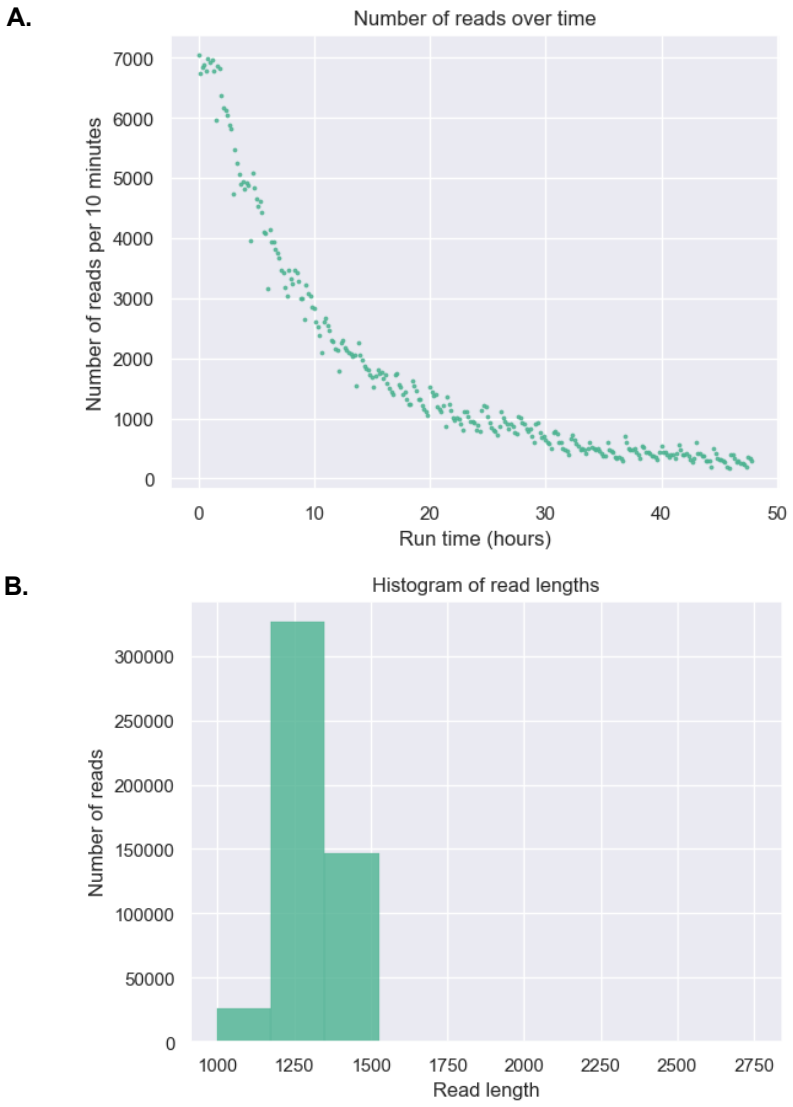

Figure 5.13 MinION sequencing data.

A. *Read acquisition as a factor of run time. The x-axis shows the run time in hours and y-axis shows the number of reads per 10 minutes;* ***B.*** *A histogram of read length post nanofilt. The highest bar represents the read length that occurred the most often in the data set.*

Table 5.4 Summary of the MinION sequencing runs

SAMPLE NAME	POST NANOFILT		NUMBER OF VARIANTS		ALIGNMENT TO REFERENCE SEQUENCE (%)
	Total reads	Modal read length (bp)	PreCore/Core	PreS1/S2/S	
P1-T1	38,109	1,250	1	7	100
P1-T2	11,876	1,252	2	8	100
P1-T2	26,730	1,250	3	7	100
P2-T1	71,139	1,306	5	0	100
P2-T2	66,028	1,260	3	0	100
P2-T3	39,867	1,287	2	0	100
P2-T4	51,657	1,251	2	0	100
P3-T1	45,004	1,256	1	1	100
P3-T2	39,775	1,255	1	1	100
P3-T3	42,751	1,255	1	1	100
POSITIVE CONTROL	66,466	1,250	0	0	100
NEGATIVE CONTROL	0	N/A	0	0	N/A
TOTAL	438,482	1,257	21	25	N/A

5.3.6 Validation of Nanopore sequencing analysis

To confirm whether the variants reported by MinION were legitimate, each variant was justified logically against Sanger trace files with a few criteria applied;

I. A variant reported by MinION was deemed legitimate if the mutation was also shown as a doublet in Sanger, for example in the region of preS1/S2/S of sample P1-T1, mutations at positions 60, 90, 246, 253, 414, 819, and 954 were classified as genuine (Appendix 1); if no corresponding doublet shown in Sanger;

II. A variant reported by MinION was deemed legitimate if the mutant was reported by Sanger in any time points of the same position of that patient. For example, in the region of preS1/S2/S of sample P1-T3, at position 90, 86% variants were G and 14% were A as reported by MinION, which was in line with the dominant base G as a single peak reported by Sanger. This A mutant appeared in P1-T3 (MinION) was deemed legitimate because the A was detected in sample P1-T1 and T2 (Sanger) (Appendix 1)

Sanger abi files were opened by SnapGene Viewer (SnapGene, 2019). The traces were examined to ensure the quality of raw signals were suitable for minor variant detection. Figure 5.14 shows three high-quality 4-colour chromatograms with evenly spaced peaks and no background noise. A variant was called where a single peak position within a trace presented two peaks of different colours. For example, a snapshot of trace file from sample P1-T1 shows double peaks including “A” and “C”, but the dominant peak “A” was reported. Similarly, a doublet was seen at the same position in sample P1-T2, but the dominant “A” was reported. However, in sample P1-T3, in the same position, the dominant peak was “C” while “A” became the minor variant (Figure 5.14). In MinION, these three events were reported along with variant frequency, 0.28, 0.40 and 0.25 respectively (Appendix 1).

The lowest variant frequency reported by the MinION platform was 0.11 (Appendix 1, Appendix 2 & Appendix 3). Based on these verification criteria, it was confident to confirm that all variants called were legitimate.

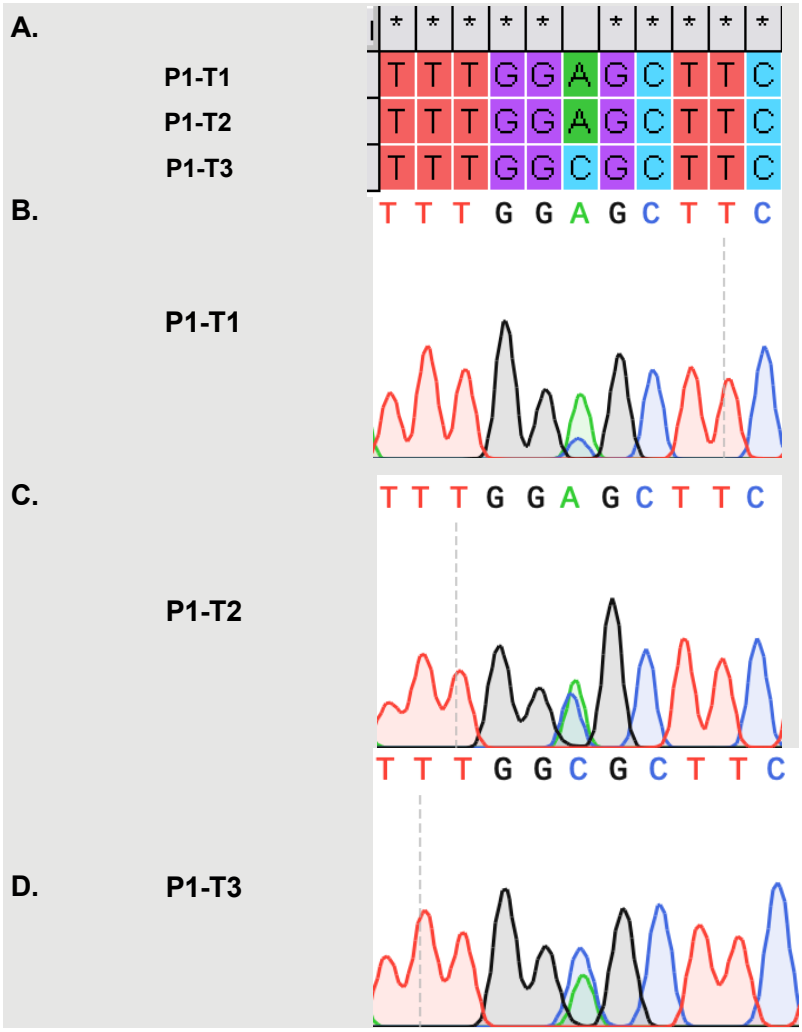

Figure 5.14 Sanger trace files highlighting polymorphic sites

Snapshots of Sanger trace files showing double peaks where variants were also detected in MinION sequencing. ***A.*** *MEGA alignments of the three-time points for Patient 1;* ***B.*** *Sanger trace file of sample P1-T1 where a single peak position harbours two peaks "A" and "C";* ***C.*** *Sanger trace file of sample P1-T2 showing double peaks "A" and "C";* ***D.*** *Sanger trace file of sample P1-T3 showing double peaks "A" and "C".*

5.3.7 Viral diversity

To appreciate baseline mutations among population diversity, the consensus sequences of Patient 1 was aligned to 10 consensus sequences of genotype A, and the consensus sequences of Patient 2 and 3 were aligned to 10 consensus sequences of genotype C. All sequences of genotype A and C were downloaded from GenBank at https://www.ncbi.nlm.nih.gov/nuccore/?term=HBV.

As shown in Table 5.5 & Table 5.6, across the analysed regions including preCore/Core, preS1/S2/S and partial polymerase, no insertions or deletions (indel) were observed. In this study, short randomly distributed indels (1 to 5bp) were ignored, as these commonly occur due to sequencing error (Sedlazeck et al., 2018).

In the preCore domain, three nonsynonymous mutations were observed (Table 5.5) in Patient 1 throughout the three time points, whilst in the core region, seven nonsynonymous mutations were observed in sequences generated from Patient 1 and four were reported in Patient 3, but only one was reported in Patient 3 (Table 5.5).

Table 5.5 Mutations in preCore/Core region

REGION	TYPE OF MUTATION	MUTATIONS		SAMPLE	REPORTED CHARACTERISTICS	PREVIOUS LITERATURE
		Amino acid	Nucleotide			
PRECORE	Sub[1]	I9V	A1838G	P1-T1/2/3	Highly associated with lower viral load, experiencing earlier viral suppression while on treatment	(Podlaha et al., 2019)
	Sub	W28*	G1896A	P1-T1/2/3	Defective HBeAg expression, but rare in gtA[2] HBV; association with ACLF[3]	(Kim et al., 2016, Hu et al., 2015)
	Sub	G29D	G1899A	P1-T1/2/3 P2-T1/2	Association to ACLF[3]	(Kim et al., 2016, Hu et al., 2015)
CORE	Sub	P5H	C1914A	P2-T1	May lead to hepatocarcinogenesis by inducing the ER[4] stress-ROS[5] axis	(Kim et al., 2016)
	Sub	T12S	A1934T	P1-T1/2/3	In T_h epitope, association with HCC[6]	(Kim et al., 2016, Datta et al., 2014)
	Sub	L31Q	T1992A	P1-T1/2/3	N/A	
	Sub	G63V	G2088T	P2-T1	Positively selected in lamivudine treated HBeAg seroconversion patients	(Cheng et al., 2013)
	Sub	E77Q	G2129C	P1-T1/2/3	Located in B-cell epitopes, correlate to disease progression to LC[7] & HCC	(Kim et al., 2016, Al-Qahtani et al., 2018)
	Sub	P130T	C2288A	P2-T3/4	Positively selected in lamivudine treated HBeAg seroconversion patients	(Cheng et al., 2013, Yuan and Shih, 2000)
	Sub	A131P	G2291C	P1-T3	Located in B-cell epitopes	(Jazayeri et al., 2004)
	Sub	P135L	C2304T	P1-T1/2/3	Under positive selection in seroconverters	(Lim et al., 2007)
	Sub	R150G	A2348G	P1-T1/2 P2-T1/2/3/4 P3-T1/2/3	Decreasing cell binding ability	(Bin Mohamed Suffian et al., 2015)
	Sub	S183P	T2447C	P1-T3	Increase the trend of progression from IC[8] to LC; this mutation is only reported in gtD[9]	(Sunbul, 2014, Ghosh et al., 2012)

1. Sub-substitution; 2. gtA – genotype A; 3. ACLF – acute-on-chronic liver failure; 4. ER – endoplasmic reticulum; 5. ROS - reactive oxygen species; 6. HCC – hepatocellular carcinoma; 7. LC – liver cirrhosis; 8. IC – inactive carrier; 9. gtD – genotype D HBV.

Table 5.6 Mutations in surface region and corresponding polymerase region

MUTATION			TYPE OF MUTATION	SAMPLES	REPORTED CHARACATERISTICS	REFERENCES
NUCLEOTIDE	**Amino acid**					
	HBsAg	Polymerase				
A2348G	NC[1]	NC	Sub[2]	P3-T1/2/3	N/A	
C2913T	NC	P203S	Sub	P1-T1/3	N/A	
A2943G	NC	R213G	Sub	P1-T1/2/3	N/A	
A2952G	NC	I216V	Sub	P1-T2	N/A	
C2988A	preS1.F45L	Q228K	Sub	P3-T1/2/3	N/A	
A3099G	NC	R265G	Sub	P1-T1/3	N/A	
C3106T	NC	T267I	Sub	P1-T1/2/3	N/A	
A3159G	NC	K285E	sub	P1-T1/3	N/A	
T46C	NC	S321P	Sub	P1-T1/2/3	N/A	
T451C	NC	NC	Sub	P1-T1	N/A	
T505C	NC	rt[3]Y126H	Sub	P1-T2	Anti-viral drug resistance	(Biswas et al., 2013)
T586C	NC	rtW153R	Sub	P1-T1/2/3	Anti-viral drug resistance; higher in LAM[4] +ADV[5] treated patients, but rare in treatment naïve patients	(Mirandola et al., 2012)
A589G	NC	rtK154E	Sub	P1-T3	N/A	

1. NC – no change; 2. Sub-substitution; 3. rt – reverse transcriptase; 4. LAM - Lamivudine; 5. ADV – adefovir dipivoxil.

On examination of the preS1/S2/S and its corresponding polymerase region, CHB patients who were classed as HBeAg-positive showed different profiles when compared with the ones classed as HBeAg-negative. No mutation was reported in samples from Patient 2, and one synonymous mutation in the spacer (SP) domain from Patient 3 was observed. Conversely, Patient 1 had the highest number of mutations detected and the majority of them were point mutations within preS1/S2/S region, all of which were synonymous. Nevertheless, these mutations introduced non-synonymous coding changes in the overlapping polymerase region (Table 5.6). Seven out 10 were in the spacer domain and three were in the reverse transcriptase (RT) domain (Figure 5.15). The entropy (Hx) plots demonstrated that (Figure 5.16), in preS1/S2/S region, no variable residue was observed; while in *pol*, the majority of variable residues were located in the Spacer domain.

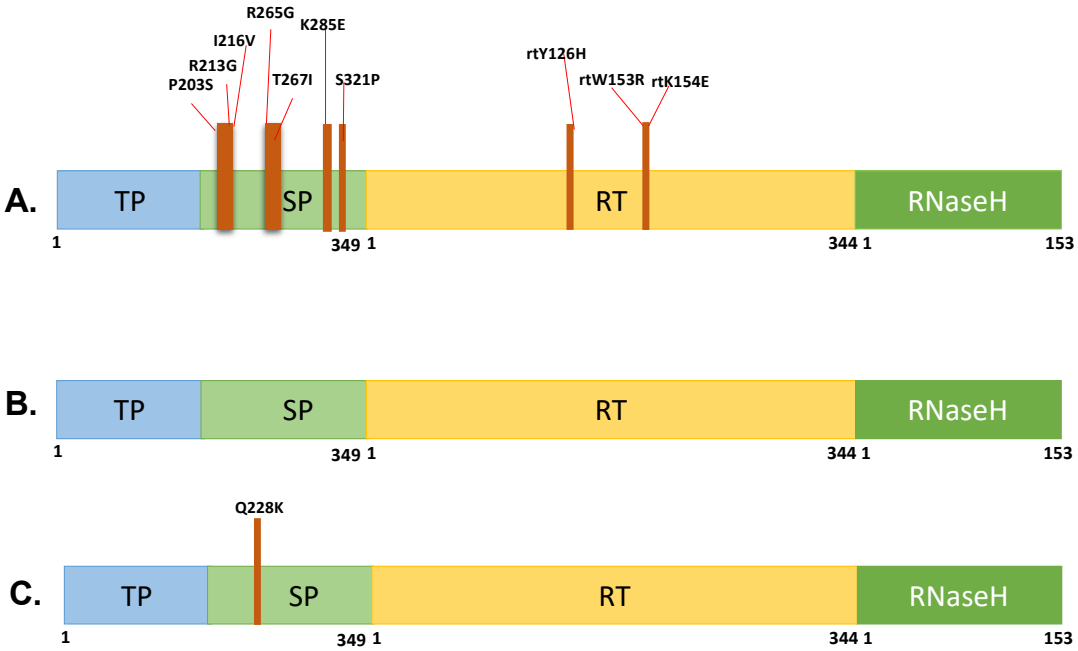

Figure 5.15 Schematic graph of observed polymerase mutations

A. *Point-mutations occurred in the polymerase region collected from all samples in Patient 1 (HBeAg negative);* ***B.*** *No mutation in this region was detected in Patient 2 (HBeAg-positive).* ***C.*** *One point-mutation occurred in Patient 3 (HBeAg-positive). (TP, terminal protein; SP, spacer protein; RT, reverse transcriptase; RNaseH, ribonuclease H).*

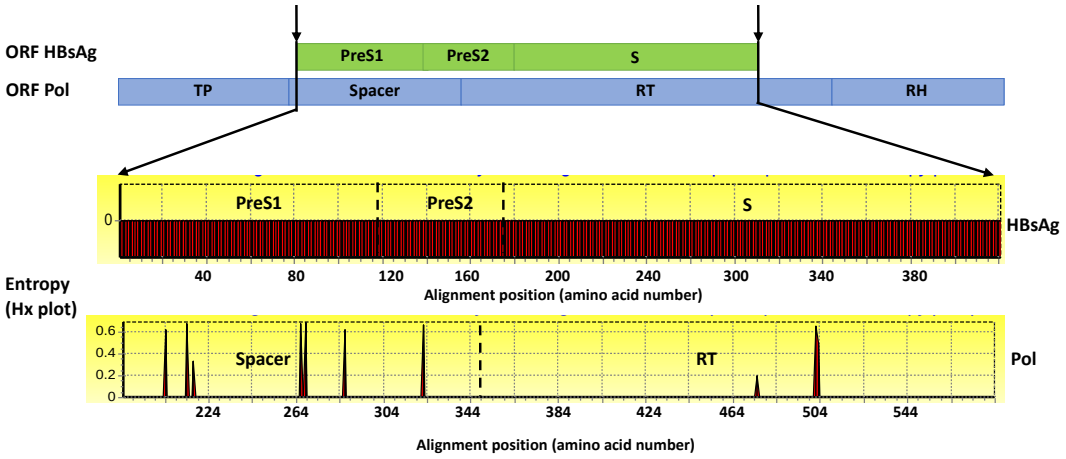

Figure 5.16 Entropy Hx plots of the overlapping region in samples from Patient 1

The major domains of the HBsAg and Polymerase (pol) are indicated in the schematic graphs at the top. The arrows show the overlapping regions of the two open reading frames (ORF). The Entropy (Hx) plots indicate the variations within the overlapping region of the HBsAg (the upper plot) and the pol (the lower plot). The X-axis denotes the position of the residues, while the Y-axis denotes the entropy of the individual positions, with a higher value indicating a more variable site. This map is adapted from (Chen et al., 2013).

5.3.8 Intra-host evolution

Of the three patients investigated, Patient 1 had the largest number of single nucleotide mutations, while Patient 2 and 3 had very few mutations recorded, therefore, Patient 1 was investigated further.

To compare the amino acid substitution rate in the overlapping region (nt.2856-3210; nt.1-835) and the nonoverlapping (nt.1839-2306) region, haplotypes with a frequency above 1% from all-time points of Patient 1 were selected for the intra-host evolution analysis using BEAST2. After applying Nanofilt, no variant was recalled in the positive control, achieving 100% accuracy. However, a 1% error rate was assumed to maximise the credibility of haplotypes used in the intra-host evolution analysis. The substitution rate in the overlapping region was 3.69 $x10^{-4}$ (95% higher posterior density (HPD) interval between 2.19 x 10^{-4}-5.42 x 10^{-4}) substitution/site/year (s/s/y) and 8.87 x 10^{-4} (95% HPD interval between 2.42 x 10^{-4}-1.68 x 10^{-3}) s/s/y in the nonoverlapping regions indicating an approximately 58% lower substitution rate in the overlapping region.

To further demonstrate the rapid change in viral diversity between the time points, Maximum Clad Credibility (MCC) trees were constructed for the *pol* region to address the relationship among variants across all time points.

The MCC tree for the overlapping (the *pol*) showed a distinct population of HBV variants, yet intermingled among the virus populations between the three time points (Figure 5.17). Patient 1 had never received therapy with anti-viral drugs. An approximately one log increase in ALT was reported between T2 (29 IU/L) and T3 (331 IU/L) while the viral load was increased approximately 50% (Table 5.2). The increase of ALT and viral load was linked to a concomitant increase in viral diversity.

The dominant haplotype at T2 with polymerase (P)I216V (20% frequency) mutation was phased out at T3; instead, one minor haplotype (4.1% frequency) at T1 with mutations including pP203S, pP213G, pR265G, P267I and rtW153R became a dominant one at T3 with a frequency of 11% after gaining an additional mutation rtK154E (data not shown).

Despite the large number of variants presented in *pol*, all variants were closely related. Due to this low input resolution, the posterior credibility of all branches was lower than 0.95.

An MCC tree for the nonoverlapping region (the Core) displayed a very different picture from the tree built from the overlapping *pol* gene. Far fewer haplotypes were observed in this region. However, similar to the MCC tree of the overlapping region of *pol*, the haplotypes were intermingled among the virus populations between the three time points (Figure 5.18).

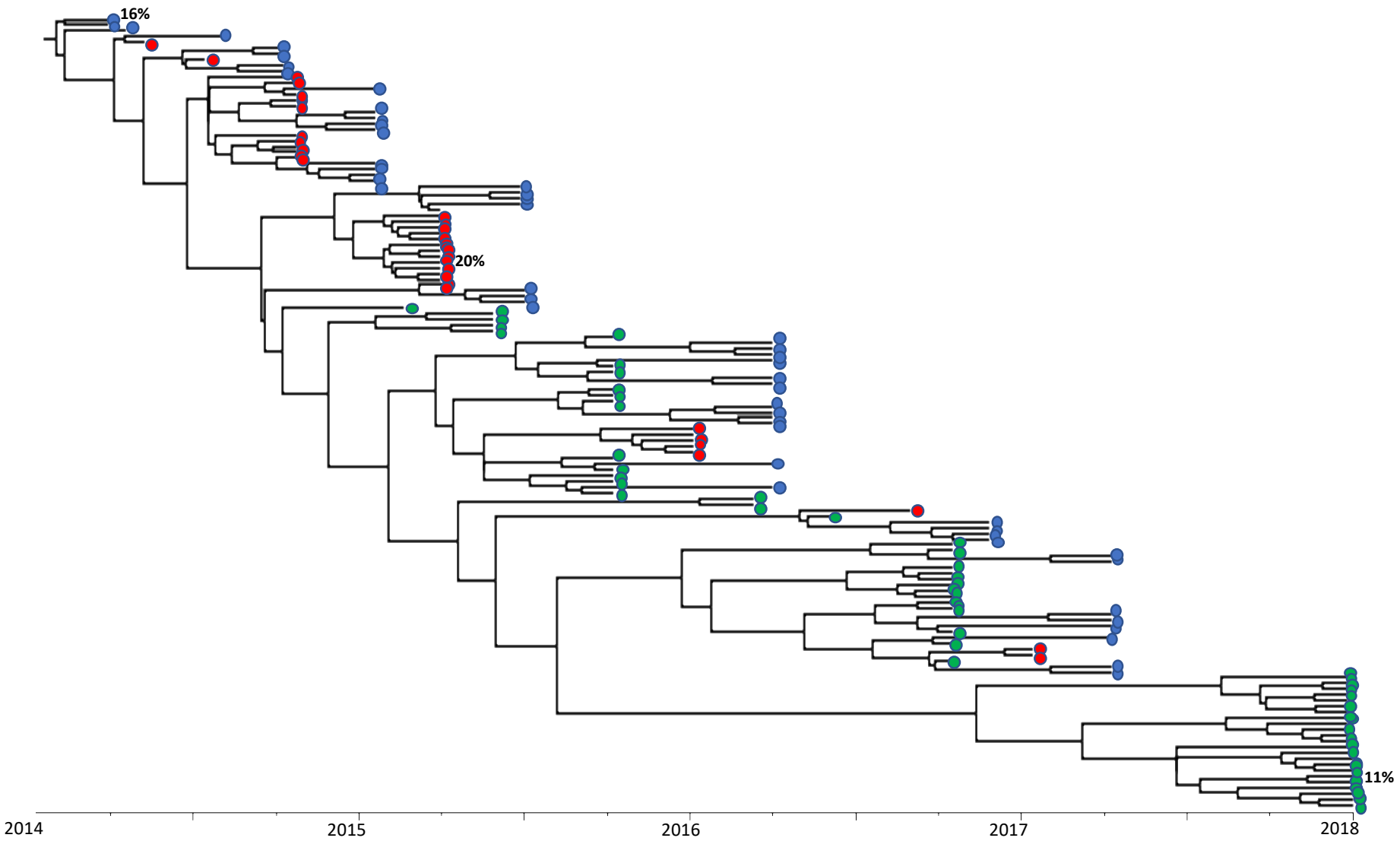
16%
20%
11%
2014
2015
2016
2017
2018

Figure 5.17 The time-scaled Maximum Clade Credibility (MCC) tree for *pol* (nt.2856-3210; nt.1-835) gene from Patient 1 based on amino acid sequence

Haplotypes (total 137) with frequency above 1% were selected to build an evolutionary tree. All sequences were aligned in MEGA 7.0 based on amino acid coding frame and then analysed in BEASTv2.6.0. The MCC tree was generated by TreeAnnotator v2.6.0 after removing 10% as burn-ins and viewed in FigTree v1.4.4. Blue, red and green circles represent samples from 2014, 2016 and 2018 respectively. Consensus haplotypes were labelled with frequency in percentage against corresponding circles. The phylogenetic tree was rooted with consensus sequence from time point 1. The x-axis displays the timescale in years. The divergence time was inferred by BEASTv2.9.0 using molecular clock model. Posterior credibility is not shown as all values were below 0.95.

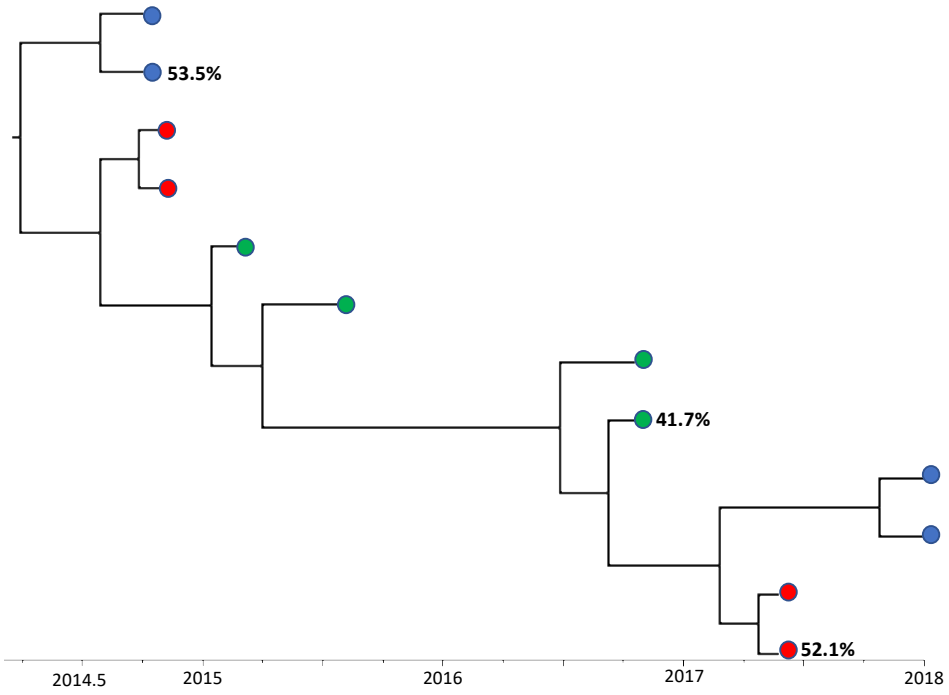

Figure 5.18 The time-scaled Maximum Clade Credibility (MCC) tree for the nonoverlapping region (nt.1839-2306) from Patient 1 based on amino acid sequence

Haplotypes (total 12) were used to build an evolutionary tree. All sequences were aligned in MEGA 7.0 based on amino acid coding frame and then analysed in BEASTv2.6.0. The MCC tree was generated by TreeAnnotator v2.6.0 after removing 10% as burn-ins and viewed in FigTree v1.4.4. Blue, red and green circles represent samples from 2014, 2016 and 2018 respectively. Consensus haplotypes were labelled with frequency in percentage against corresponding circles. The phylogenetic tree was rooted with consensus sequence from time point 1. The x-axis displays the timescale in years. The divergence time was inferred by BEASTv2.9.0 using molecular clock model. Posterior credibility is not shown as all values were below 0.95.

5.4 Discussion

In this study, the MinION sequencing platform was employed to investigate the dynamics of HBV variance in both HBeAg positive and negative patients. The application of deep sequencing and haplotype reconstruction allowed the detection of a variant population down to a frequency of 1%. This novel approach revealed a much higher level of genetic complexity in the viral population in a CHB patient with HBeAg negative compared to the ones who were HBeAg positive.

As a cell model (chapter 3) and HBV pseudotyping system (chapter 4) were established to investigate the contribution of L, M and S in virus entry, it would be interesting to phenotype naturally occurring variants in the preS1/S2/S gene. Hence, one of the primary aims of this study was to interrogate the viral diversity in this region. Surprisingly, only one non-synonymous variant was observed in the preS1/S2/S in the samples examined; instead, non-synonymous variant residues were mainly located in the Spacer domain of polymerase (Figure 5.16). Therefore, to further understand HBV molecular biology, the substitution rate in overlapping vs nonoverlapping regions was investigated.

5.4.1 Substitution rate in overlapping vs nonoverlapping regions

An approximately 58% lower substitution rate in the overlapping region compared to its counterparts was observed. It has been explained that this phenomenon is due to constraint to nucleotide substitution in the overlapping region arising from regulatory elements as well as encoded RNA secondary structures required for replication (McNaughton et al., 2019a). As the HBsAg ORF is entirely overlapped by the *pol* ORF with the HBsAg ORF shifted downstream by one nucleotide (Cento et al., 2013). Following the rules of the degeneracy of codons, a point mutation at s1/p2 would lead

to non-synonymous mutation in both HBsAg and the *pol*. This situation would be detrimental for the viral evolution; therefore, the s1/p2 position showed the highest conservation (Zaaijer et al., 2007). As a point mutation at the third codon generally leads to synonymous mutation, nucleotide substitution at s2/p3 and s3/p1 would only affect either HBsAg or *pol* (Zaaijer et al., 2007). Indeed, this was observed in this study where >90% of nucleotide substitutions resulted in synonymous mutation in the HBsAg ORF but non-synonymous codon changes in the *pol*, particularly in the Spacer domain. The results were in consistent with previous study that in the HBsAg and *pol* overlapping region, 87.7% of the positive selection were located in the spacer domain (Chen et al., 2013). While it has previously been suggested that the Spacer domain, acting as a tether between the terminal protein (TP) and the reverse transcriptase (RT), is dispensable, the recent study has proposed that this domain acts as a harbour for the accumulation of mutations, offering flexibility for conformational changes and responding to immune pressure (Chen et al., 2013). However, the dN/dS ratio was not calculated to determine whether these mutations were under positive, negative selection or random drift as dN/dS ratio may not be an appropriate indicator in this situation. The dN/dS was originally developed to determine evolutionary forces in distantly related genomes, not in closely related quasispecies. In a quasispecies swarm, only one or a few genomic sites may be subjected to a positive selection event, while many unrelated sites undergo synonymous substitution, which produces a low dN/dS ratio. Nonetheless, this low dN/dS ratio still drives virus evolution as viral quasispecies act as a unit of selection (Domingo et al., 2012).

A previous finding obtaining a rate of substitution in overlapping regions was 40% less than the nonoverlapping regions (Paraskevis et al., 2015), lower than the 58% reported in this study. The substitution rate in this study was derived from samples over a short time period (~4 years) from a single patient, while Paraskevis et al. analysed archived 105 samples from the HBVdb, including samples from the major

HBV genotypes A-H and J and from four African and Asian primate species. It is unclear whether the discrepancy between these two studies was due to differences between genotypes or due to the coverage of sampling period investigated. There is evidence showing that, during the course of chronic infection, many substitutions that occur in the viral genome might not lead to variation, but reverse back to the consensus sequence (McNaughton et al., 2019a). As shown in this study, mutant I216V appeared at time point 2 was reverted back to wildtype at time point 3 (Table 5.6). Therefore, it is justifiable to observe a higher substitution rate over a short term. Moreover, unlike this study that only a short region of nonoverlapping region of Core gene and one section of *pol* gene were included in the evolution analysis, Paraskevis and colleagues had combined all overlapping regions against all nonoverlapping regions of X, C and P for the analysis. Some nonoverlapping regions overlapping with other regulatory elements may provide further constraint to nucleotide substitution. For instance, X region overlaps with the basal core promoter, a regulatory element that controls transcription of preCore mRNA and pregenomic RNA, which may contribute to the lower viral diversity within nonoverlapping regions of X in relation to the other nonoverlapping regions (McNaughton et al., 2019a).

5.4.2 Viral diversity is higher in HBeAg negative patients

Single nucleotide polymorphism (SNP) point mutations were largely discovered in sera from Patient 1, who was HBeAg-negative. A total of 21 mutations were identified compared to a total of 7 mutations in sera from the HBeAg-positive patients (patients 2 and 3). This finding is consistent with a previous study that the viral diversity in e antigen positive patients was 2.4-fold significantly lower than that in e antigen negative patients (Lim et al., 2007). It has been hypothesised that, during the HBeAg positive phase, the HBeAg stimulates regulatory T cells that suppress anti-viral T cell responses; therefore, the selective pressure is low and fewer adaptive mutations are

observed. During the HBeAg negative phase, with an unclear mechanism, viral replication is enhanced, meanwhile, the regulatory immune response is lifted, resulting in stronger selection immune response. This combination results in greater selection for escape mutants, hence generating viral diversity (Lim et al., 2007, Harrison et al., 2011).

5.4.3 Variant evolution and clinical significance

Following European Association for the Study of the Liver (EASL) guideline, Patient 1 did not require treatment as the laboratory results indicated normal ALT and normal fibroscan, even though HBV DNA was > 2000 IU/mL (European Association for the Study of the Liver. Electronic address and European Association for the Study of the, 2017). However, deep sequencing revealed that many variants circulating in Patient 1 were linked to poor disease outcome and therefore warranted early intervention. For example, W28*, G29D, T12S, E77Q, S183P have been reported to associate with acute-on chronic-liver failure (ACLF), liver cirrhosis (LC) and hepatocellular carcinoma (HCC) (Podlaha et al., 2019, Kim et al., 2016, Hu et al., 2015, Sunbul, 2014, Ghosh et al., 2012). The data also showed Patient 1 carrying a mutant rtW153R that was linked to anti-viral drug resistance, particularly higher in lamivudine (LMV) and adefovir dipivoxil (ADV) treated individuals (Mirandola et al., 2012), suggesting LMV and ADV should not be chosen for this patient should anti-viral treatment is required. Having identified the dominant haplotype in samples P1-T3, when the patient had ALT flare, harbouring pP203S, pP213G, pI216V, pR265G, pP267I, rtW153R and rtK154E, it would be interesting to phenotype these variants *in-vivo* in future research, especially the phenotypic impact of pR265G where a bulky positively charged side group was replaced by a small non-polar side chain; rtW153R where a non-polar side chain was replaced by a bulky positively charged side group; and rtK154E, where a positively charged side group was replaced by a negatively charged

side chain. SNP rtK154E is a novel mutation discovered in this study, only appeared in sample P1-T3; it is unclear what impact this mutation would have on disease progression.

P5H, a mutant associated with hepatocarcinogenesis (Podlaha et al., 2019, Kim et al., 2016, Hu et al., 2015), along with G29D were discovered in samples P2-T1 and P2-T1/2 respectively. Patient 2 commenced interferon treatment since time point 1 and discontinued after time point 3. Although no HBeAg seroconversion achieved after treatment, P5H and G29D that were associated with poor disease outcome were reverted back to wildtype. However, it is unclear whether the disappearance of these two mutants was as a result of the treatment or genetic drift.

5.4.4 Nanopore sequencing and its accuracy in variant calling

Several controls were in place to ensure the credibility of the variants discovered in this study. First, plasmid 12C was created as a positive control. There was one single HBV template in this sample, and the results confirmed that the positive control achieved 100% accuracy rate with no variants was observed. This gives confidence in interpreting the sequencing results of the clinical samples. Second, the negative control did not yield reads that were aligned to the HBV genome, which confirmed there were no contaminations between samples. Third, the accuracy of the variants reported by MinION sequencing was verified by comparing each variant against base Sanger chromatogram files.

It can be stated with high confidence that in this experiment, the variants recorded using filtered reads by MinION were reliable. However, all variants might not have been fully recovered in all samples by MinION. In this study, during the variant calling, the threshold was set at 10%, i.e. only the variant frequency above 10% would be reported, which means a potentially lower population of true variants could be missed.

Furthermore, variants recovery might have been influenced by the quality of library preparation. This was evident in sample P1-T2 where the amounts of input DNA templates was low, which in turn, resulted in ~4000 reads for preS1/S2/S region. At the same region, there were ~13000 reads recovered in sample P1-T1. The variant at position 60 reported in sample P1-T1 was not recorded in P1-T2; however, this variant reappeared in sample P1-T3. This example showed that all variants in a given sample might not have been fully identified even though the recovered ones were legitimate.

This highlights the limitation of nanopore sequencing. PacBio sequencing yields a much higher accuracy rate and allows detection of variant frequency down to 1% (Betz-Stablein et al., 2016). One can be attributed to its unique library preparation creating circular DNA molecules that can be subsequently sequenced many times to maximise accuracy. This approach has been adopted in nanopore sequencing by McNaughton et al. who exploited isothermal rolling-circle amplification (RCA) to enrich the HBV genome and achieved 99.7% accuracy rate (McNaughton et al., 2019b). However, the PacBio sequencing is much more costly compared to MinION sequencing, and it does not provide flexibility and portability as what MinION can provide.

5.4.5 Technical challenges

5.4.5.1 HBV DNA extraction

VRMK and MEVK are two common QiaAmp kits used in our laboratory. Both kits utilise silica membrane columns to adsorb DNA/RNA material. VRMK requires a supplement of carrier RNA to enhance genome recovery, especially the starting genomic material is low. The carrier RNA was not used in this experiment as the viral load of each sample were sufficiently high. MEVK, on the other hand, requires the

addition of protease to denature nuclease in the specimen to keep genomic material intact (Table 5.3).

Viral genomic extraction is the first important step in NGS. Poor extraction might artificially lower the genome diversity in a sample. Furthermore, Viral RNA/DNA extraction is a laborious and time-consuming process. If sample availability is limited, then the choice of DNA/RNA extraction method becomes crucial, hence the rationale for comparing the extraction efficiency of these two methods.

Following viral DNA extraction, the extraction efficiency was analysed by qPCR. As shown in Figure 5.7, MEVK yielded lower Ct (equivalent to Cq value) values compared to that of VRMK. This result disagrees the finding reported by Klenner and his colleagues, who reported that MEVK had the highest Ct values in all virus tested compared to other extraction methods, although MEVK recovered the highest reads in RNA viral extraction in the NGS analysis (Klenner et al., 2017). It was reported that the ratio of viral reads and coverage yielded by NGS analysis was not well correlated with Ct values of the same sample (Zhang et al., 2018). This suggests that although the Ct value yielded by VRMK was higher than that by MEVK in this study, the reads recovery in the NGS analysis could be the opposite, i.e. VRMK extracted HBV genomes could be recovered higher reads in MinION sequence.

According to the 2.5kb PCR products using primer pair of F1 (nt. 1583-1604) and R1(nt. 989-970) shown in Figure 5.8A, the band intensity was well correlated to the Ct values of each sample (Figure 5.7). The dimmest band with relative density at ~0.05 was sample P1-T2 (Ct value 32.5) and the brightest band (relative density of ~0.85 was sample P3-T2 (Ct value 22.1). This correlation continued to read recovery in MinION sequencing, where P1-T2 had approximately 11000 reads whereas P3-T2 had approximately 44000 reads. This implies that the Ct values in qPCR are

consistent with the read recovery in the NGS analysis when comparing viral genomes extracted from one type of extraction method.

5.4.5.2 Library preparation

In this study, a nested four-primer approach was used to target and enrich HBV DNA for analysis. This approach required two stages of PCR to amplify targeted fragments at the first stage and then tag barcodes at the second stage. As shown in Figure 5.8A, the first stage successfully amplified the 2.5kb amplicons. However, the second stage amplification was not achieved (Figure 5.8B). Troubleshooting was attempted by optimising annealing temperatures and varying concentration of input templates, but none of these optimisations achieved better amplification. Although a new library preparation kit made some improvement, the amplification was still not ideal. As this method did not allow for amplification of the larger amplicon, two short amplicons of the preCore/Core and preS1/S2/S region were amplified instead.

Successful amplification relies heavily on optimal ratios between PCR reagents, including concentrations of primers, DNA templates, dNTPs, DNA polymerase, enzymes and buffers. Few measures were adopted to eliminate bias arising from PCR procedures. For example, the longAmp Hot Start *Taq* 2x Master Mix, which offers two-fold higher fidelity than *Taq* was used to limit bias. To minimise bias arising from the process of PCR as preferential binding between primers and DNA sequences, the inner primers used in this study were universal primers designed by Chook et al. (Chook et al., 2015) and PHE (PHE, 2007), aiming to cover the complete HBV genomes.

5.4.5.3 Short DNA fragments

The 100bp and 200bp bands presented in the second round PCR indicated the presence of higher amounts of primers dimers or self-amplified hairpins (Figure 5.11 & Figure 5.12). These smaller size products would consume valuable space on the flow cell in the downstream sequencing run and consequently reduce sequencing

efficiency. Ideally, these undesired products should be removed. Although PCR purification using AMPure XP beads was performed prior to sequencing runs, the AMPure XP beads are not able to remove DNA fragment greater than 100bp. Ideally, these should be removed by gel extraction after second round PCR and only taking the desired bands for the downstream sequencing. Gel purification method was attempted to recover PCR products (detailed in Chapter 4); however, the DNA recovery was extremely low. Thus, the sequencing run using AMPure XP beads purified DNA products was proceeded because the volumes of the second round PCR products were limited to undergo any unpredictable purification procedures.

5.5 Conclusions and future work

Deep sequencing of the HBV genome extracted from 10 clinical samples was performed using the MinION nanopore platform. The samples were collected from 3 patients over a 2 to 4 year period. MinION sequencing yielded ~ 438,482 qualified reads in total for the quasispecies analysis. The viral diversity between HBeAg-positive and HBeAg-negative as well as between overlapping (nt.2856-3210; nt.1-835) and nonoverlapping (nt.1839-2306) genome regions were compared. Viral diversity was higher in the patient in an HBeAg-negative phase in contrast to 2 patients in an HBeAg-positive phase. When comparing the viral diversity between nonoverlapping and the overlapping regions, the results showed a 58% lower substitution rate in the overlapping region. This was in agreement with previous studies that lower substitution rate has been observed in the overlapping region compared with the nonoverlapping region. However, the substitution rate (58%) observed in this study was higher than the reported rate (40%). This could be due to smaller sample size as well as a shorter coverage of sampling period investigated in the current study. In addition, this study also discovered a rare mutation in the reverse transcriptase gene. The mutation rtK154E occurred at the time point where the patient

had a disease flare. Future work in an animal model would be required to phenotype this mutation.

HBV haplotype building was a novel tool employed in this study to investigate the intra-host evolution of a viral population. Multiple approaches were performed to ensure the accuracy of haplotypes. Firstly, a plasmid of a single HBV template was sequenced to rule out false variants recall. Library preparation use nuclease free water acting as a negative control was another quality control. After applied Nanofilt, 100% accuracy in variant recall was achieved. Secondly, all single nucleotide polymorphisms were validated against Sanger sequence trace files. However, this validation could be improved upon comparison with PacBio sequencing. The longitudinal quasispecies analysis could be further improved by sequencing the complete genome and in a larger sample size.

Worldwide, HBV genome sequencing is not a common practice. In the UK, this has only been introduced recently and is usually by standard Sanger sequencing. This study provides evidence that the standard Sanger sequencing does not allow interrogation of possible quasispecies in the host. In addition, this study also provides evidence that the MinION sequencing platform is potentially a reliable tool for identifying sequence variability in addition to its advantages of being affordable and small portable size. This provides detailed molecular information on the virus populations for better treatment management.

Chapter 6: Final conclusions and future work

In this study, in-house developed Huh7-NTCP and HepG2-NTCP cell lines were used as the target cells to measure the infectivity of HBV-PVs. Following extensive optimisations, an in-house developed PV producing cell line 293T-gtD4.S was shown to produce infectious HBV-PVs when the cells were supplemented with plasmids expressing L proteins. The results confirm previous knowledge that S alone does not form infectious particles. Moreover, the L/S ratio is crucial in viral assembly, secretion and infectivity. Having established an optimal L/S ratio to generate infectious HBV-PVs, the S producing cell line 293T-gtD4.S can be utilised in a variety of future works. For example, it can be used to assess the potency of novel entry inhibitors for HBV; it also can be used to assess the entry fitness of many naturally occurring variants, such as preS1 deletion mutants, of which some strains are missing up to one-half of the preS1 domain.

Having gained this expertise in pseudotyping, it would be interesting to look at how this approach might be useful for other viruses. For example, the entry of other viruses that contain multiple envelope glycoproteins, such as highly pathogenic avian H5N1 influenza could be investigated using this system. Influenza pseudotyped viruses have been extensively used to avoid the production of highly pathogenic viruses. Up to five plasmids have been used for the production of influenza PVs, including plasmids expressing hemagglutinin (HA), neuraminidase (NA), M2 ion channel the HIV core and a reporter respectively. Alternatively, exogenous NA treatment is used to replace the NA expression plasmid. M2 ion channel is often omitted as it is not required for the production of PVs despite some reports that showed M2 incorporation enhances the yields and infectivity (Carnell et al., 2015). To reduce plasmid load and burden, NA and M2 can be both transduced and constitutively expressed in HEK293T cells similar to the S producing cell line 293T-gtD4.S described in the current study.

The use of in-house developed NTCP overexpressing hepatocellular carcinoma cell lines to evaluate HBV-PVs infectivity proved to be challenging. It was difficult to determine whether the low infectivity observed throughout the study arose through some deficit in the NTCP cell lines or in the production of HBV-PVs. For future reference, the study could be improved by using a cell line that is known to be susceptible to HBV infection as the target cells to determine the infectivity of HBV-PVs. Examples of other choices include primary hepatocytes or terminally differentiated HepaRG cells.

The data generated here demonstrate that the current pseudotyping system using a lentiviral backbone to package HBV glycoproteins can be achieved, but the efficiency was low. The number of HBsAg molecules presented in the plasma membrane is approximately 16-fold less than that in intracellular compartments (Chai et al., 2007). As HIV assembly occurs in the plasma membrane, HBV-PVs using a lentiviral backbone requires HBsAg "leaking" from the ER to the plasma membrane, which may partly explain the low efficiency in HBV-PVs production. Building on HBV-PVs production using a lentiviral pseudotyping system, a better approach may be to use an HBV trans-complementation system where one plasmid expresses a reporter gene with HBV packaging signal, one plasmid expresses the necessary proteins for HBV assembly and viral replication including the core and reverse transcriptase, and one plasmid expresses the HBsAg. As such, HBV-PVs assembly would occur in the same manner as for native HBV; therefore, not only will it ensure the production of HBV-PVs but would also truly reflect the natural behaviour of the virus. Moreover, phenotyping analysis of HBsAg mutations would remain as easily manipulated as within the lentiviral pseudotyping system.

The second part of the study using the Nanopore sequencing platform for longitudinal quasispecies analysis revealed that mutations were more common in HBeAg-

negative patients compared to HBeAg-positive individuals. Furthermore, mutations occurred more frequently (~58%) in nonoverlapping regions than in overlapping regions. A caveat is that only a very small number of individuals was studied in the current work, larger sample size would be required to give higher confidence level. The longitudinal quasispecies analysis on clinical samples demonstrated that variants co-exist in a host and the frequency of individual variants evolve, which may link to the disease progression. For example, a novel mutation rtK154E discovered in this study evolved from a minor variant to a dominant one at the time point when there was an ALT flare, indicating that this rtK154E might have contributed to disease progression. Obviously, the phenotypic behaviour of this variant needs to be further evaluated.

Combining with all the techniques and knowledge obtained from this study, future efforts will concentrate on monitoring mutations in clinical samples using Nanopore sequencing, followed by phenotyping the genetic variants exploiting the HBV trans-complementation system in a safer CL2 environment. By doing so, a greater understanding of the viral life cycle will provide valuable information in the development of novel antiviral compounds and therefore towards the goal of HBV elimination by 2030 set by the WHO.

Chapter 7: References

ABOU-JAOUDE, G. & SUREAU, C. 2007. Entry of hepatitis delta virus requires the conserved cysteine residues of the hepatitis B virus envelope protein antigenic loop and is blocked by inhibitors of thiol-disulfide exchange. *J Virol,* 81**,** 13057-66.

AHN, S. H., PARK, Y. K., PARK, E. S., KIM, J. H., KIM, D. H., LIM, K. H., JANG, M. S., CHOE, W. H., KO, S. Y., SUNG, I. K., KWON, S. Y. & KIM, K. H. 2014. The impact of the hepatitis B virus polymerase rtA181T mutation on replication and drug resistance is potentially affected by overlapping changes in surface gene. *J Virol,* 88**,** 6805-18.

AL-ALLAF, F. A., TOLMACHOV, O. E., ZAMBETTI, L. P., TCHETCHELNITSKI, V. & MEHMET, H. 2013. Remarkable stability of an instability-prone lentiviral vector plasmid in Escherichia coli Stbl3. *3 Biotech,* 3**,** 61-70.

AL-QAHTANI, A. A., AL-ANAZI, M. R., NAZIR, N., ABDO, A. A., SANAI, F. M., AL-HAMOUDI, W. K., ALSWAT, K. A., AL-ASHGAR, H. I., KHAN, M. Q., ALBENMOUSA, A., EL-SHAMY, A., ALANAZI, S. K., DELA CRUZ, D., BOHOL, M. F. F. & AL-AHDAL, M. N. 2018. The Correlation Between Hepatitis B Virus Precore/Core Mutations and the Progression of Severe Liver Disease. *Front Cell Infect Microbiol,* 8**,** 355.

ALLWEISS, L. & DANDRI, M. 2016. Experimental in vitro and in vivo models for the study of human hepatitis B virus infection. *J Hepatol,* 64**,** S17-S31.

ALTER, M. J. & MARGOLIS, H. S. 1990. The emergence of hepatitis B as a sexually transmitted disease. *Med Clin North Am,* 74**,** 1529-41.

ALTER, M. J. & MAST, E. E. 1994. The epidemiology of viral hepatitis in the United States. *Gastroenterol Clin North Am,* 23**,** 437-55.

AMEUR, A., KLOOSTERMAN, W. P. & HESTAND, M. S. 2019. Single-Molecule Sequencing: Towards Clinical Applications. *Trends Biotechnol,* 37**,** 72-85.

ANDERSON, R. T., LIM, S. G., MISHRA, P., JOSEPHSON, F., DONALDSON, E., GIVEN, B. & MILLER, V. 2019. Challenges, Considerations, and Principles to Guide Trials of Combination Therapies for Chronic Hepatitis B Virus. *Gastroenterology,* 156**,** 529-533 e4.

APPELMAN, M. D., CHAKRABORTY, A., PROTZER, U., MCKEATING, J. A. & VAN DE GRAAF, S. F. 2017. N-Glycosylation of the Na+-Taurocholate Cotransporting Polypeptide (NTCP) Determines Its Trafficking and Stability and Is Required for Hepatitis B Virus Infection. *PLoS One,* 12**,** e0170419.

BAMBAUT, A. 2018. *FigTree 1.4.4 software* [Online]. Available: http://tree.bio.ed.ac.uk/software/figtree/ [Accessed 20 Nov 2019].

BARRERA, A., GUERRA, B., NOTVALL, L. & LANFORD, R. E. 2005. Mapping of the hepatitis B virus pre-S1 domain involved in receptor recognition. *J Virol,* 79**,** 9786-98.

BAYER, M. E., BLUMBERG, B. S. & WERNER, B. 1968. Particles associated with Australia antigen in the sera of patients with leukaemia, Down's Syndrome and hepatitis. *Nature,* 218**,** 1057-9.

BEASLEY, R. P., HWANG, L. Y., LIN, C. C., STEVENS, C. E., WANG, K. Y., SUN, T. S., HSIEH, F. J. & SZMUNESS, W. 1981. Hepatitis B immune globulin (HBIG) efficacy in the interruption of perinatal transmission of hepatitis B virus carrier state. Initial report of a randomised double-blind placebo-controlled trial. *Lancet,* 2**,** 388-93.

BECK, J. & NASSAL, M. 2007. Hepatitis B virus replication. *World J Gastroenterol,* 13**,** 48-64.

BEGGEL, B., NEUMANN-FRAUNE, M., DORING, M., LAWYER, G., KAISER, R., VERHEYEN, J. & LENGAUER, T. 2012. Genotyping hepatitis B virus dual infections using population-based sequence data. *J Gen Virol,* 93**,** 1899-907.

BERK, P. D. & POPPER, H. 1978. Fulminant hepatic failure. *Am J Gastroenterol,* 69**,** 349-400.

BERTING, A., HAHNEN, J., KROGER, M. & GERLICH, W. H. 1995. Computer-aided studies on the spatial structure of the small hepatitis B surface protein. *Intervirology,* 38**,** 8-15.

BETZ-STABLEIN, B. D., TOPFER, A., LITTLEJOHN, M., YUEN, L., COLLEDGE, D., SOZZI, V., ANGUS, P., THOMPSON, A., REVILL, P., BEERENWINKEL, N., WARNER, N. & LUCIANI, F. 2016. Single-Molecule Sequencing Reveals Complex Genome Variation of Hepatitis B Virus during 15 Years of Chronic Infection following Liver Transplantation. *J Virol,* 90**,** 7171-7183.

BILL, C. A. & SUMMERS, J. 2004. Genomic DNA double-strand breaks are targets for hepadnaviral DNA integration. *Proc Natl Acad Sci U S A,* 101**,** 11135-40.

BIN MOHAMED SUFFIAN, I. F., NISHIMURA, Y., MORITA, K., NAKAMURA-TSURUTA, S., AL-JAMAL, K. T., ISHII, J., OGINO, C. & KONDO, A. 2015. Mutation of arginine residues to avoid non-specific cellular uptakes for hepatitis B virus core particles. *J Nanobiotechnology,* 13**,** 15.

BIOLABS, N. E. 2020. *NEB® Stable Competent E. coli (High Efficiency)* [Online]. Available: https://international.neb.com/products/c3040-neb-stable-competent-e-coli-high-efficiency#Product%20Information [Accessed 4th April 2020].

BISWAS, A., PANIGRAHI, R., CHANDRA, P. K., BANERJEE, A., DATTA, S., PAL, M., CHAKRABORTY, S., BHATTACHARYA, P., CHAKRABARTI, S. & CHAKRAVARTY, R. 2013. Characterization of the occult hepatitis B virus variants circulating among the blood donors from eastern India. *ScientificWorldJournal,* 2013**,** 212704.

BLESCH, A. 2004. Lentiviral and MLV based retroviral vectors for ex vivo and in vivo gene transfer. *Methods,* 33**,** 164-72.

BLOCK, T. M., ALTER, H. J., LONDON, W. T. & BRAY, M. 2016. A historical perspective on the discovery and elucidation of the hepatitis B virus. *Antiviral Res,* 131**,** 109-23.

BLUMBERG, B. S., ALTER, H. J. & VISNICH, S. 1965. A "New" Antigen in Leukemia Sera. *JAMA,* 191**,** 541-6.

BOND, W. W., FAVERO, M. S., PETERSEN, N. J., GRAVELLE, C. R., EBERT, J. W. & MAYNARD, J. E. 1981. Survival of hepatitis B virus after drying and storage for one week. *Lancet,* 1**,** 550-1.

BORDEAUX, J., WELSH, A., AGARWAL, S., KILLIAM, E., BAQUERO, M., HANNA, J., ANAGNOSTOU, V. & RIMM, D. 2010. Antibody validation. *Biotechniques,* 48**,** 197-209.

BORGIA, G., CARLEO, M. A., GAETA, G. B. & GENTILE, I. 2012. Hepatitis B in pregnancy. *World J Gastroenterol,* 18**,** 4677-83.

BOUCHARD, M. J. & SCHNEIDER, R. J. 2004. The enigmatic X gene of hepatitis B virus. *J Virol,* 78**,** 12725-34.

BOUCKAERT, R., VAUGHAN, T. G., BARIDO-SOTTANI, J., DUCHENE, S., FOURMENT, M., GAVRYUSHKINA, A., HELED, J., JONES, G., KUHNERT, D., DE MAIO, N., MATSCHINER, M., MENDES, F. K., MULLER, N. F., OGILVIE, H. A., DU PLESSIS, L., POPINGA, A., RAMBAUT, A., RASMUSSEN, D., SIVERONI, I., SUCHARD, M. A., WU, C. H., XIE, D., ZHANG, C., STADLER, T. & DRUMMOND, A. J. 2019. BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. *PLoS Comput Biol,* 15**,** e1006650.

BRECHTBUEHL, K., WHALLEY, S. A., DUSHEIKO, G. M. & SAUNDERS, N. A. 2001. A rapid real-time quantitative polymerase chain reaction for hepatitis B virus. *J Virol Methods,* 93**,** 105-13.

BRUSS, V. 2007. Hepatitis B virus morphogenesis. *World J Gastroenterol,* 13**,** 65-73.

BRUSS, V. & GANEM, D. 1991. The role of envelope proteins in hepatitis B virus assembly. *Proc Natl Acad Sci U S A,* 88**,** 1059-63.

BRUSS, V. & THOMSSEN, R. 1994. Mapping a region of the large envelope protein required for hepatitis B virion maturation. *J Virol,* 68**,** 1643-50.

BRUSS, V. & VIELUF, K. 1995. Functions of the internal pre-S domain of the large surface protein in hepatitis B virus particle morphogenesis. *J Virol,* 69**,** 6652-7.

BURTON, A. R., PALLETT, L. J., MCCOY, L. E., SUVEIZDYTE, K., AMIN, O. E., SWADLING, L., ALBERTS, E., DAVIDSON, B. R., KENNEDY, P. T., GILL, U. S.,

MAURI, C., BLAIR, P. A., PELLETIER, N. & MAINI, M. K. 2018. Circulating and intrahepatic antiviral B cells are defective in hepatitis B. *J Clin Invest,* 128**,** 4588-4603.

BUTLER, D. M., PACOLD, M. E., JORDAN, P. S., RICHMAN, D. D. & SMITH, D. M. 2009. The efficiency of single genome amplification and sequencing is improved by quantitation and use of a bioinformatics tool. *J Virol Methods,* 162**,** 280-3.

CARNELL, G. W., FERRARA, F., GREHAN, K., THOMPSON, C. P. & TEMPERTON, N. J. 2015. Pseudotype-based neutralization assays for influenza: a systematic analysis. *Front Immunol,* 6**,** 161.

CENTO, V., MIRABELLI, C., DIMONTE, S., SALPINI, R., HAN, Y., TRIMOULET, P., BERTOLI, A., MICHELI, V., GUBERTINI, G., CAPPIELLO, G., SPANO, A., LONGO, R., BERNASSOLA, M., MAZZOTTA, F., DE SANCTIS, G. M., ZHANG, X. X., VERHEYEN, J., D'ARMINIO MONFORTE, A., CECCHERINI-SILBERSTEIN, F., PERNO, C. F. & SVICHER, V. 2013. Overlapping structure of hepatitis B virus (HBV) genome and immune selection pressure are critical forces modulating HBV evolution. *J Gen Virol,* 94**,** 143-9.

CHAI, N., CHANG, H. E., NICOLAS, E., GUDIMA, S., CHANG, J. & TAYLOR, J. 2007. Assembly of hepatitis B virus envelope proteins onto a lentivirus pseudotype that infects primary human hepatocytes. *J Virol,* 81**,** 10897-904.

CHAKIATH, C. S. & ESPOSITO, D. 2007. Improved recombinational stability of lentiviral expression vectors using reduced-genome Escherichia coli. *Biotechniques,* 43**,** 466, 468, 470.

CHANG, H. K. & TING, L. P. 1989. The surface gene promoter of the human hepatitis B virus displays a preference for differentiated hepatocytes. *Virology,* 170**,** 176-83.

CHANG, H. K., WANG, B. Y., YUH, C. H., WEI, C. L. & TING, L. P. 1989. A liver-specific nuclear factor interacts with the promoter region of the large surface protein gene of human hepatitis B virus. *Mol Cell Biol,* 9**,** 5189-97.

CHAO, M., HSIEH, S. Y. & TAYLOR, J. 1990. Role of two forms of hepatitis delta virus antigen: evidence for a mechanism of self-limiting genome replication. *J Virol,* 64**,** 5066-9.

CHEN, B. F., LIU, C. J., JOW, G. M., CHEN, P. J., KAO, J. H. & CHEN, D. S. 2006. High prevalence and mapping of pre-S deletion in hepatitis B virus carriers with progressive liver diseases. *Gastroenterology,* 130**,** 1153-68.

CHEN, B. K., SAKSELA, K., ANDINO, R. & BALTIMORE, D. 1994. Distinct modes of human immunodeficiency virus type 1 proviral latency revealed by superinfection of nonproductively infected cell lines with recombinant luciferase-encoding viruses. *J Virol,* 68**,** 654-60.

CHEN, C. H., HUNG, C. H., LEE, C. M., HU, T. H., WANG, J. H., WANG, J. C., LU, S. N. & CHANGCHIEN, C. S. 2007. Pre-S deletion and complex mutations of hepatitis

B virus related to advanced liver disease in HBeAg-negative patients. *Gastroenterology,* 133**,** 1466-74.

CHEN, P., GAN, Y., HAN, N., FANG, W., LI, J., ZHAO, F., HU, K. & RAYNER, S. 2013. Computational evolutionary analysis of the overlapped surface (S) and polymerase (P) region in hepatitis B virus indicates the spacer domain in P is crucial for survival. *PLoS One,* 8**,** e60098.

CHENG, X., GUAN, W., SUN, S., LI, B., LI, H., KANG, F., KANG, J., YANG, D., NASSAL, M. & SUN, D. 2015. Stable Human Hepatoma Cell Lines for Efficient Regulated Expression of Nucleoside/Nucleotide Analog Resistant and Vaccine Escape Hepatitis B Virus Variants and Woolly Monkey Hepatitis B Virus. *PLoS One,* 10**,** e0145746.

CHENG, Y., GUINDON, S., RODRIGO, A. & LIM, S. G. 2013. Increased viral quasispecies evolution in HBeAg seroconverter patients treated with oral nucleoside therapy. *J Hepatol,* 58**,** 217-24.

CHERESIZ, S. V., KONONOVA, A. A., RAZUMOVA, Y. V., DUBICH, T. S., CHEPURNOV, A. A., KUSHCH, A. A., DAVEY, R. & POKROVSKY, A. G. 2014. A vesicular stomatitis pseudovirus expressing the surface glycoproteins of influenza A virus. *Arch Virol,* 159**,** 2651-8.

CHOOK, J. B., TEO, W. L., NGEOW, Y. F., TEE, K. K., NG, K. P. & MOHAMED, R. 2015. Universal Primers for Detection and Sequencing of Hepatitis B Virus Genomes across Genotypes A to G. *J Clin Microbiol,* 53**,** 1831-5.

CHOTIYAPUTTA, W. & LOK, A. S. 2009. Hepatitis B virus variants. *Nat Rev Gastroenterol Hepatol,* 6**,** 453-62.

COORAY, S., HOWE, S. J. & THRASHER, A. J. 2012. Retrovirus and lentivirus vector design and methods of cell conditioning. *Methods Enzymol,* 507**,** 29-57.

CORCHERO, J. L. & VILLAVERDE, A. 1998. Plasmid maintenance in Escherichia coli recombinant cultures is dramatically, steadily, and specifically influenced by features of the encoded proteins. *Biotechnol Bioeng,* 58**,** 625-32.

CRONIN, J., ZHANG, X. Y. & REISER, J. 2005. Altering the tropism of lentiviral vectors through pseudotyping. *Curr Gene Ther,* 5**,** 387-98.

DALBA, C., BELLIER, B., KASAHARA, N. & KLATZMANN, D. 2007. Replication-competent vectors and empty virus-like particles: new retroviral vector designs for cancer gene therapy or vaccines. *Mol Ther,* 15**,** 457-66.

DANE, D. S., CAMERON, C. H. & BRIGGS, M. 1970. Virus-like particles in serum of patients with Australia-antigen-associated hepatitis. *Lancet,* 1**,** 695-8.

DASGUPTA, U., WESTON-HAFER, K. & BERG, D. E. 1987. Local DNA sequence control of deletion formation in Escherichia coli plasmid pBR322. *Genetics,* 115**,** 41-9.

DATTA, S., GHOSH, A., DASGUPTA, D., GHOSH, A., ROYCHOUDHURY, S., ROY, G., DAS, S., DAS, K., GUPTA, S., BASU, K., BASU, A., DATTA, S., CHOWDHURY, A. & BANERJEE, S. 2014. Novel point and combo-mutations in the genome of hepatitis B virus-genotype D: characterization and impact on liver disease progression to hepatocellular carcinoma. *PLoS One,* 9**,** e110012.

DELIUS, H., GOUGH, N. M., CAMERON, C. H. & MURRAY, K. 1983. Structure of the hepatitis B virus genome. *J Virol,* 47**,** 337-43.

DILL, J. A., CAMUS, A. C., LEARY, J. H., DI GIALLONARDO, F., HOLMES, E. C. & NG, T. F. 2016. Distinct Viral Lineages from Fish and Amphibians Reveal the Complex Evolutionary History of Hepadnaviruses. *J Virol,* 90**,** 7920-33.

DOITSH, G. & SHAUL, Y. 2004. Enhancer I predominance in hepatitis B virus gene expression. *Mol Cell Biol,* 24**,** 1799-808.

DOMINGO, E. & PERALES, C. 2019. Viral quasispecies. *PLoS Genet,* 15**,** e1008271.

DOMINGO, E., SHELDON, J. & PERALES, C. 2012. Viral quasispecies evolution. *Microbiol Mol Biol Rev,* 76**,** 159-216.

DORING, B., LUTTEKE, T., GEYER, J. & PETZINGER, E. 2012. The SLC10 carrier family: transport functions and molecular structure. *Curr Top Membr,* 70**,** 105-68.

DREXLER, J. F., GEIPEL, A., KONIG, A., CORMAN, V. M., VAN RIEL, D., LEIJTEN, L. M., BREMER, C. M., RASCHE, A., COTTONTAIL, V. M., MAGANGA, G. D., SCHLEGEL, M., MULLER, M. A., ADAM, A., KLOSE, S. M., CARNEIRO, A. J., STOCKER, A., FRANKE, C. R., GLOZA-RAUSCH, F., GEYER, J., ANNAN, A., ADU-SARKODIE, Y., OPPONG, S., BINGER, T., VALLO, P., TSCHAPKA, M., ULRICH, R. G., GERLICH, W. H., LEROY, E., KUIKEN, T., GLEBE, D. & DROSTEN, C. 2013. Bats carry pathogenic hepadnaviruses antigenically related to hepatitis B virus and capable of infecting human hepatocytes. *Proc Natl Acad Sci U S A,* 110**,** 16151-6.

DRUMMOND, A. J., BOUCKAERT, R. & XIE, W. 2015. Measurably evolving populations.

DU, Y., CHI, X., WANG, C., JIANG, J., KONG, F., YAN, H., WANG, X., LI, J., WU, N. C., DAI, L., ZHANG, T. H., SHU, S., ZHOU, J., YOSHIZAWA, J. M., LI, X., BHATTACHARYA, D., WU, T. T., NIU, J. & SUN, R. 2017. Quantifying perinatal transmission of Hepatitis B viral quasispecies by tag linkage deep sequencing. *Sci Rep,* 7**,** 10168.

DULL, T., ZUFFEREY, R., KELLY, M., MANDEL, R. J., NGUYEN, M., TRONO, D. & NALDINI, L. 1998. A third-generation lentivirus vector with a conditional packaging system. *J Virol,* 72**,** 8463-71.

EBLE, B. E., LINGAPPA, V. R. & GANEM, D. 1990. The N-terminal (pre-S2) domain of a hepatitis B virus surface glycoprotein is translocated across membranes by downstream signal sequences. *J Virol,* 64**,** 1414-9.

EBLE, B. E., MACRAE, D. R., LINGAPPA, V. R. & GANEM, D. 1987. Multiple topogenic sequences determine the transmembrane orientation of the hepatitis B surface antigen. *Mol Cell Biol,* 7**,** 3591-601.

ENGLAND, C. G., EHLERDING, E. B. & CAI, W. 2016. NanoLuc: A Small Luciferase Is Brightening Up the Field of Bioluminescence. *Bioconjug Chem,* 27**,** 1175-1187.

ETEMAD, B., GHULAM-SMITH, M., GONZALEZ, O., WHITE, L. F. & SAGAR, M. 2015. Single genome amplification and standard bulk PCR yield HIV-1 envelope products with similar genotypic and phenotypic characteristics. *J Virol Methods,* 214**,** 46-53.

EUROPEAN ASSOCIATION FOR THE STUDY OF THE LIVER. ELECTRONIC ADDRESS, E. E. E. & EUROPEAN ASSOCIATION FOR THE STUDY OF THE, L. 2017. EASL 2017 Clinical Practice Guidelines on the management of hepatitis B virus infection. *J Hepatol,* 67**,** 370-398.

EVRIPIOTI, A. A., ORTEGA-PRIETO, A. M., SKELTON, J. K., BAZOT, Q. & DORNER, M. 2019. Phosphodiesterase-induced cAMP degradation restricts hepatitis B virus infection. *Philos Trans R Soc Lond B Biol Sci,* 374**,** 20180292.

FOURATI, S. & PAWLOTSKY, J. M. 2016. Recent advances in understanding and diagnosing hepatitis B virus infection. *F1000Res,* 5.

FUJIWARA, K., MATSUURA, K., MATSUNAMI, K., IIO, E. & NOJIRI, S. 2018. Characterization of hepatitis B virus with complex structural variations. *BMC Microbiol,* 18**,** 202.

GALLE, P. R., HAGELSTEIN, J., KOMMERELL, B., VOLKMANN, M., SCHRANZ, P. & ZENTGRAF, H. 1994. In vitro experimental infection of primary human hepatocytes with hepatitis B virus. *Gastroenterology,* 106**,** 664-73.

GALLINA, A., DE KONING, A., ROSSI, F., CALOGERO, R., MANSERVIGI, R. & MILANESI, G. 1992. Translational modulation in hepatitis B virus preS-S open reading frame expression. *J Gen Virol,* 73 (Pt 1)**,** 139-48.

GANEM, D. & PRINCE, A. M. 2004. Hepatitis B virus infection--natural history and clinical consequences. *N Engl J Med,* 350**,** 1118-29.

GARCIA, T., LI, J., SUREAU, C., ITO, K., QIN, Y., WANDS, J. & TONG, S. 2009. Drastic reduction in the production of subviral particles does not impair hepatitis B virus virion secretion. *J Virol,* 83**,** 11152-65.

GEHRING, A. J. & PROTZER, U. 2019. Targeting Innate and Adaptive Immune Responses to Cure Chronic HBV Infection. *Gastroenterology,* 156**,** 325-337.

GENCAY, M., HUBNER, K., GOHL, P., SEFFNER, A., WEIZENEGGER, M., NEOFYTOS, D., BATRLA, R., WOESTE, A., KIM, H. S., WESTERGAARD, G., REINSCH, C., BRILL, E., THU THUY, P. T., HOANG, B. H., SONDERUP, M., SPEARMAN, C. W., PABINGER, S., GAUTIER, J., BRANCACCIO, G., FASANO, M.,

SANTANTONIO, T., GAETA, G. B., NAUCK, M. & KAMINSKI, W. E. 2017. Ultra-deep sequencing reveals high prevalence and broad structural diversity of hepatitis B surface antigen mutations in a global population. *PLoS One,* 12**,** e0172101.

GERLICH, W. H. 2013. Medical virology of hepatitis B: how it began and where we are now. *Virol J,* 10**,** 239.

GHOSH, S., MONDAL, R. K., BANERJEE, P., NANDI, M., SARKAR, S., DAS, K., SANTRA, A., BANERJEE, S., CHOWDHURY, A. & DATTA, S. 2012. Tracking the naturally occurring mutations across the full-length genome of hepatitis B virus of genotype D in different phases of chronic e-antigen-negative infection. *Clin Microbiol Infect,* 18**,** E412-8.

GIROGLOU, T., CINATL, J., JR., RABENAU, H., DROSTEN, C., SCHWALBE, H., DOERR, H. W. & VON LAER, D. 2004. Retroviral vectors pseudotyped with severe acute respiratory syndrome coronavirus S protein. *J Virol,* 78**,** 9007-15.

GLEBE, D., ALIAKBARI, M., KRASS, P., KNOOP, E. V., VALERIUS, K. P. & GERLICH, W. H. 2003. Pre-s1 antigen-dependent infection of Tupaia hepatocyte cultures with human hepatitis B virus. *J Virol,* 77**,** 9511-21.

GLEBE, D. & URBAN, S. 2007. Viral and cellular determinants involved in hepadnaviral entry. *World J Gastroenterol,* 13**,** 22-38.

GLEBE, D., URBAN, S., KNOOP, E. V., CAG, N., KRASS, P., GRUN, S., BULAVAITE, A., SASNAUSKAS, K. & GERLICH, W. H. 2005. Mapping of the hepatitis B virus attachment site by use of infection-inhibiting preS1 lipopeptides and tupaia hepatocytes. *Gastroenterology,* 129**,** 234-45.

GOODWIN, S., GURTOWSKI, J., ETHE-SAYERS, S., DESHPANDE, P., SCHATZ, M. C. & MCCOMBIE, W. R. 2015. Oxford Nanopore sequencing, hybrid error correction, and de novo assembly of a eukaryotic genome. *Genome Res,* 25**,** 1750-6.

GREGORY, D. A., OLINGER, G. Y., LUCAS, T. M. & JOHNSON, M. C. 2014. Diverse viral glycoproteins as well as CD4 co-package into the same human immunodeficiency virus (HIV-1) particles. *Retrovirology,* 11**,** 28.

GREHAN, K., FERRARA, F. & TEMPERTON, N. 2015. An optimised method for the production of MERS-CoV spike expressing viral pseudotypes. *MethodsX,* 2**,** 379-84.

GRIPON, P., CANNIE, I. & URBAN, S. 2005. Efficient inhibition of hepatitis B virus infection by acylated peptides derived from the large viral surface protein. *J Virol,* 79**,** 1613-22.

GRIPON, P., DIOT, C., THEZE, N., FOUREL, I., LOREAL, O., BRECHOT, C. & GUGUEN-GUILLOUZO, C. 1988. Hepatitis B virus infection of adult human hepatocytes cultured in the presence of dimethyl sulfoxide. *J Virol,* 62**,** 4136-43.

GRIPON, P., RUMIN, S., URBAN, S., LE SEYEC, J., GLAISE, D., CANNIE, I., GUYOMARD, C., LUCAS, J., TREPO, C. & GUGUEN-GUILLOUZO, C. 2002. Infection of a human hepatoma cell line by hepatitis B virus. *Proc Natl Acad Sci U S A,* 99**,** 15655-60.

GUO, H., JIANG, D., ZHOU, T., CUCONATI, A., BLOCK, T. M. & GUO, J. T. 2007. Characterization of the intracellular deproteinized relaxed circular DNA of hepatitis B virus: an intermediate of covalently closed circular DNA formation. *J Virol,* 81**,** 12472-84.

HAGENBUCH, B. & MEIER, P. J. 1994. Molecular cloning, chromosomal localization, and functional characterization of a human liver Na+/bile acid cotransporter. *J Clin Invest,* 93**,** 1326-31.

HALL, M. P., UNCH, J., BINKOWSKI, B. F., VALLEY, M. P., BUTLER, B. L., WOOD, M. G., OTTO, P., ZIMMERMAN, K., VIDUGIRIS, G., MACHLEIDT, T., ROBERS, M. B., BENINK, H. A., EGGERS, C. T., SLATER, M. R., MEISENHEIMER, P. L., KLAUBERT, D. H., FAN, F., ENCELL, L. P. & WOOD, K. V. 2012. Engineered luciferase reporter from a deep sea shrimp utilizing a novel imidazopyrazinone substrate. *ACS Chem Biol,* 7**,** 1848-57.

HALL, T. A. 1999. BioEdit: A user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Scientific Research,* Nucleic Acids Symposium series, 41, 95-98.

HALLEN, S., MARENINOVA, O., BRANDEN, M. & SACHS, G. 2002. Organization of the membrane domain of the human liver sodium/bile acid cotransporter. *Biochemistry,* 41**,** 7253-66.

HARRISON, A., LEMEY, P., HURLES, M., MOYES, C., HORN, S., PRYOR, J., MALANI, J., SUPURI, M., MASTA, A., TERIBORIKI, B., TOATU, T., PENNY, D., RAMBAUT, A. & SHAPIRO, B. 2011. Genomic analysis of hepatitis B virus reveals antigen state and genotype as sources of evolutionary rate variation. *Viruses,* 3**,** 83-101.

HAYER, J., JADEAU, F., DELEAGE, G., KAY, A., ZOULIM, F. & COMBET, C. 2013. HBVdb: a knowledge database for Hepatitis B Virus. *Nucleic Acids Res,* 41**,** D566-70.

HBVDB. 2020. *Genome organisation* [Online]. Available: https://hbvdb.lyon.inserm.fr/HBVdb/HBVdbGenome [Accessed 13th Jan 2020].

HEDEGAARD, D. L., TULLY, D. C., ROWE, I. A., REYNOLDS, G. M., BEAN, D. J., HU, K., DAVIS, C., WILHELM, A., OGILVIE, C. B., POWER, K. A., TARR, A. W., KELLY, D., ALLEN, T. M., BALFE, P. & MCKEATING, J. A. 2017. High resolution sequencing of hepatitis C virus reveals limited intra-hepatic compartmentalization in end-stage liver disease. *J Hepatol,* 66**,** 28-38.

HILDT, E., MUNZ, B., SAHER, G., REIFENBERG, K. & HOFSCHNEIDER, P. H. 2002. The PreS2 activator MHBs(t) of hepatitis B virus activates c-raf-1/Erk2 signaling in transgenic mice. *EMBO J,* 21**,** 525-35.

HILL, J. B., SHEFFIELD, J. S., KIM, M. J., ALEXANDER, J. M., SERCELY, B. & WENDEL, G. D. 2002. Risk of hepatitis B transmission in breast-fed infants of chronic hepatitis B carriers. *Obstet Gynecol,* 99**,** 1049-52.

HO, R. H., LEAKE, B. F., ROBERTS, R. L., LEE, W. & KIM, R. B. 2004. Ethnicity-dependent polymorphism in Na+-taurocholate cotransporting polypeptide (SLC10A1) reveals a domain critical for bile acid substrate recognition. *J Biol Chem,* 279**,** 7213-22.

HOBRO, A. J. & SMITH, N. I. 2016. An evaluation of fixation methods: Spatial and compositional cellular changes observed by Raman imaging. *Elsevier Vibrational Spectroscopy,* 91 (2017) 31-45.

HOENEN, T., GROSETH, A., ROSENKE, K., FISCHER, R. J., HOENEN, A., JUDSON, S. D., MARTELLARO, C., FALZARANO, D., MARZI, A., SQUIRES, R. B., WOLLENBERG, K. R., DE WIT, E., PRESCOTT, J., SAFRONETZ, D., VAN DOREMALEN, N., BUSHMAKER, T., FELDMANN, F., MCNALLY, K., BOLAY, F. K., FIELDS, B., SEALY, T., RAYFIELD, M., NICHOL, S. T., ZOON, K. C., MASSAQUOI, M., MUNSTER, V. J. & FELDMANN, H. 2016. Nanopore Sequencing as a Rapidly Deployable Ebola Outbreak Tool. *Emerg Infect Dis,* 22**,** 331-4.

HOLMES, K., SHEPHERD, D. A., ASHCROFT, A. E., WHELAN, M., ROWLANDS, D. J. & STONEHOUSE, N. J. 2015. Assembly Pathway of Hepatitis B Core Virus-like Particles from Genetically Fused Dimers. *J Biol Chem,* 290**,** 16238-45.

HOOFNAGLE, J. H., SHAFRITZ, D. A. & POPPER, H. 1987. Chronic type B hepatitis and the "healthy" HBsAg carrier state. *Hepatology,* 7**,** 758-63.

HSU, M., ZHANG, J., FLINT, M., LOGVINOFF, C., CHENG-MAYER, C., RICE, C. M. & MCKEATING, J. A. 2003. Hepatitis C virus glycoproteins mediate pH-dependent cell entry of pseudotyped retroviral particles. *Proc Natl Acad Sci U S A,* 100**,** 7271-6.

HU, F., BI, S., YAN, H., SHI, Y. & SHENG, J. 2015. Associations between hepatitis B virus basal core promoter/pre-core region mutations and the risk of acute-on-chronic liver failure: a meta-analysis. *Virol J,* 12**,** 87.

HU, H. H., LIU, J., LIN, Y. L., LUO, W. S., CHU, Y. J., CHANG, C. L., JEN, C. L., LEE, M. H., LU, S. N., WANG, L. Y., YOU, S. L., YANG, H. I., CHEN, C. J. & GROUP, R.-H. S. 2016. The rs2296651 (S267F) variant on NTCP (SLC10A1) is inversely associated with chronic hepatitis B and progression to cirrhosis and hepatocellular carcinoma in patients with chronic hepatitis B. *Gut,* 65**,** 1514-21.

HWANG, S. B. & LAI, M. M. 1993. Isoprenylation mediates direct protein-protein interactions between hepatitis large delta antigen and hepatitis B virus surface antigen. *J Virol,* 67**,** 7659-62.

IP, C. L. C., LOOSE, M., TYSON, J. R., DE CESARE, M., BROWN, B. L., JAIN, M., LEGGETT, R. M., ECCLES, D. A., ZALUNIN, V., URBAN, J. M., PIAZZA, P., BOWDEN, R. J., PATEN, B., MWAIGWISYA, S., BATTY, E. M., SIMPSON, J. T., SNUTCH, T. P., BIRNEY, E., BUCK, D., GOODWIN, S., JANSEN, H. J., O'GRADY,

J., OLSEN, H. E., MIN, I. O. N. A. & REFERENCE, C. 2015. MinION Analysis and Reference Consortium: Phase 1 data release and analysis. *F1000Res,* 4**,** 1075.

IWAMOTO, M., SASO, W., SUGIYAMA, R., ISHII, K., OHKI, M., NAGAMORI, S., SUZUKI, R., AIZAKI, H., RYO, A., YUN, J. H., PARK, S. Y., OHTANI, N., MURAMATSU, M., IWAMI, S., TANAKA, Y., SUREAU, C., WAKITA, T. & WATASHI, K. 2019. Epidermal growth factor receptor is a host-entry cofactor triggering hepatitis B virus internalization. *Proc Natl Acad Sci U S A,* 116**,** 8487-8492.

IWAMOTO, M., WATASHI, K., TSUKUDA, S., ALY, H. H., FUKASAWA, M., FUJIMOTO, A., SUZUKI, R., AIZAKI, H., ITO, T., KOIWAI, O., KUSUHARA, H. & WAKITA, T. 2014. Evaluation and identification of hepatitis B virus entry inhibitors using HepG2 cells overexpressing a membrane transporter NTCP. *Biochem Biophys Res Commun,* 443**,** 808-13.

JACQUET, S., PONS, J. B., DE BERNARDO, A., NGOUBANGOYE, B., COSSET, F. L., REGIS, C., ETIENNE, L. & PONTIER, D. 2019. Evolution of Hepatitis B Virus Receptor NTCP Reveals Differential Pathogenicities and Species Specificities of Hepadnaviruses in Primates, Rodents, and Bats. *J Virol,* 93.

JAIN, M., OLSEN, H. E., PATEN, B. & AKESON, M. 2016. The Oxford Nanopore MinION: delivery of nanopore sequencing to the genomics community. *Genome Biol,* 17**,** 239.

JANUSZKIEWICZ-LEWANDOWSKA, D., RUCKA, A., KOWALA-PIASKOWSKA, A., BERESZYNSKA, I., MOZER-LISEWSKA, I., ZAJAC-SPYCHALA, O., WYSOCKI, J. & NOWAK, J. 2014. Mutations in Pol gene of hepatitis B virus in patients with chronic hepatitis B before and after therapy with nucleoside/nucleotide analogues. *Acta Virol,* 58**,** 185-9.

JAOUDE, G. A. & SUREAU, C. 2005. Role of the antigenic loop of the hepatitis B virus envelope proteins in infectivity of hepatitis delta virus. *J Virol,* 79**,** 10460-6.

JAZAYERI, S. M., DORNAN, E. S., BONER, W., FATTOVICH, G., HADZIYANNIS, S. & CARMAN, W. F. 2004. Intracellular distribution of hepatitis B virus core protein expressed in vitro depends on the sequence of the isolate and the serologic pattern. *J Infect Dis,* 189**,** 1634-45.

JOHNSON, M. C. 2011. Mechanisms for Env glycoprotein acquisition by retroviruses. *AIDS Res Hum Retroviruses,* 27**,** 239-47.

JORGENSON, R. L., VOGT, V. M. & JOHNSON, M. C. 2009. Foreign glycoproteins can be actively recruited to virus assembly sites during pseudotyping. *J Virol,* 83**,** 4060-7.

KAHILA BAR-GAL, G., KIM, M. J., KLEIN, A., SHIN, D. H., OH, C. S., KIM, J. W., KIM, T. H., KIM, S. B., GRANT, P. R., PAPPO, O., SPIGELMAN, M. & SHOUVAL, D. 2012. Tracing hepatitis B virus to the 16th century in a Korean mummy. *Hepatology,* 56**,** 1671-80.

KANEKO, S., KAKINUMA, S., ASAHINA, Y., KAMIYA, A., MIYOSHI, M., TSUNODA, T., NITTA, S., ASANO, Y., NAGATA, H., OTANI, S., KAWAI-KITAHATA, F., MURAKAWA, M., ITSUI, Y., NAKAGAWA, M., AZUMA, S., NAKAUCHI, H., NISHITSUJI, H., UJINO, S., SHIMOTOHNO, K., IWAMOTO, M., WATASHI, K., WAKITA, T. & WATANABE, M. 2016. Human induced pluripotent stem cell-derived hepatic cell lines as a new model for host interaction with hepatitis B virus. *Sci Rep,* 6**,** 29358.

KIM, H., LEE, S. A., DO, S. Y. & KIM, B. J. 2016. Precore/core region mutations of hepatitis B virus related to clinical severity. *World J Gastroenterol,* 22**,** 4287-96.

KING, B. & TARR, A. 2017. How have retrovirus pseudotypes contributed to our understanding of viral entry? *Future Medicine*.

KING, B., TEMPERTON, N. J., GREHAN, K., SCOTT, S. D., WRIGHT, E., TARR, A. W. & DALY, J. M. 2016. Technical considerations for the generation of novel pseudotyped viruses. *Future Medicine,* 11(1), 47–59.

KLAVER, B. & BERKHOUT, B. 1994. Comparison of 5' and 3' long terminal repeat promoter function in human immunodeficiency virus. *J Virol,* 68**,** 3830-40.

KLENNER, J., KOHL, C., DABROWSKI, P. W. & NITSCHE, A. 2017. Comparing Viral Metagenomic Extraction Methods. *Curr Issues Mol Biol,* 24**,** 59-70.

KNOWLES, B. B., HOWE, C. C. & ADEN, D. P. 1980. Human hepatocellular carcinoma cell lines secrete the major plasma proteins and hepatitis B surface antigen. *Science,* 209**,** 497-9.

KO, C., LEE, S., WINDISCH, M. P. & RYU, W. S. 2014. DDX3 DEAD-box RNA helicase is a host factor that restricts hepatitis B virus replication at the transcriptional level. *J Virol,* 88**,** 13689-98.

KOCK, J., NASSAL, M., MACNELLY, S., BAUMERT, T. F., BLUM, H. E. & VON WEIZSACKER, F. 2001. Efficient infection of primary tupaia hepatocytes with purified human and woolly monkey hepatitis B virus. *J Virol,* 75**,** 5084-9.

KONIG, A., DORING, B., MOHR, C., GEIPEL, A., GEYER, J. & GLEBE, D. 2014. Kinetics of the bile acid transporter and hepatitis B virus receptor Na+/taurocholate cotransporting polypeptide (NTCP) in hepatocytes. *J Hepatol,* 61**,** 867-75.

KRAM, K. E. & FINKEL, S. E. 2015. Rich Medium Composition Affects Escherichia coli Survival, Glycation, and Mutation Frequency during Long-Term Batch Culture. *Appl Environ Microbiol,* 81**,** 4442-50.

KRAMVIS, A., ARAKAWA, K., YU, M. C., NOGUEIRA, R., STRAM, D. O. & KEW, M. C. 2008. Relationship of serological subtype, basic core promoter and precore mutations to genotypes/subgenotypes of hepatitis B virus. *J Med Virol,* 80**,** 27-46.

KRAMVIS, A., KEW, M. & FRANCOIS, G. 2005. Hepatitis B virus genotypes. *Vaccine,* 23**,** 2409-23.

LAMBERT, C. & PRANGE, R. 2001. Dual topology of the hepatitis B virus large envelope protein: determinants influencing post-translational pre-S translocation. *J Biol Chem,* 276**,** 22265-72.

LE DUFF, Y., BLANCHET, M. & SUREAU, C. 2009. The pre-S1 and antigenic loop infectivity determinants of the hepatitis B virus envelope proteins are functionally independent. *J Virol,* 83**,** 12443-51.

LE SEYEC, J., CHOUTEAU, P., CANNIE, I., GUGUEN-GUILLOUZO, C. & GRIPON, P. 1998. Role of the pre-S2 domain of the large envelope protein in hepatitis B virus assembly and infectivity. *J Virol,* 72**,** 5573-8.

LE SEYEC, J., CHOUTEAU, P., CANNIE, I., GUGUEN-GUILLOUZO, C. & GRIPON, P. 1999. Infection process of the hepatitis B virus depends on the presence of a defined sequence in the pre-S1 domain. *J Virol,* 73**,** 2052-7.

LEE, J., ZONG, L., KROTOW, A., QIN, Y., JIA, L., ZHANG, J., TONG, S. & LI, J. 2018. N-Linked Glycosylation Is Not Essential for Sodium Taurocholate Cotransporting Polypeptide To Mediate Hepatitis B Virus Infection In Vitro. *J Virol,* 92.

LEISTNER, C. M., GRUEN-BERNHARD, S. & GLEBE, D. 2008. Role of glycosaminoglycans for binding and infection of hepatitis B virus. *Cell Microbiol,* 10**,** 122-33.

LEMPP, F. A., SCHLUND, F., RIEBLE, L., NUSSBAUM, L., LINK, C., ZHANG, Z., NI, Y. & URBAN, S. 2019. Recapitulation of HDV infection in a fully permissive hepatoma cell line allows efficient drug evaluation. *Nat Commun,* 10**,** 2265.

LEPERE-DOUARD, C., TROTARD, M., LE SEYEC, J. & GRIPON, P. 2009. The first transmembrane domain of the hepatitis B virus large envelope protein is crucial for infectivity. *J Virol,* 83**,** 11819-29.

LEUNG, K., KIM, J. O., GANESH, L., KABAT, J., SCHWARTZ, O. & NABEL, G. J. 2008. HIV-1 assembly: viral glycoproteins segregate quantally to lipid rafts that associate individually with HIV-1 capsids and virions. *Cell Host Microbe,* 3**,** 285-92.

LEWELLYN, E. B. & LOEB, D. D. 2011. The arginine clusters of the carboxy-terminal domain of the core protein of hepatitis B virus make pleiotropic contributions to genome replication. *J Virol,* 85**,** 1298-309.

LI, J., ZONG, L., SUREAU, C., BARKER, L., WANDS, J. R. & TONG, S. 2016. Unusual Features of Sodium Taurocholate Cotransporting Polypeptide as a Hepatitis B Virus Receptor. *J Virol,* 90**,** 8302-13.

LI, J. S., TONG, S. P., WEN, Y. M., VITVITSKI, L., ZHANG, Q. & TREPO, C. 1993. Hepatitis B virus genotype A rarely circulates as an HBe-minus mutant: possible contribution of a single nucleotide in the precore region. *J Virol,* 67**,** 5402-10.

LIANG, T. J. 2009. Hepatitis B: the virus and disease. *Hepatology,* 49**,** S13-21.

LIM, S. G., CHENG, Y., GUINDON, S., SEET, B. L., LEE, L. Y., HU, P., WASSER, S., PETER, F. J., TAN, T., GOODE, M. & RODRIGO, A. G. 2007. Viral quasi-species evolution during hepatitis Be antigen seroconversion. *Gastroenterology,* 133**,** 951-8.

LIN, C. L. & KAO, J. H. 2015. Hepatitis B virus genotypes and variants. *Cold Spring Harb Perspect Med,* 5**,** a021436.

LIU, J., LIANG, W., JING, W. & LIU, M. 2019. Countdown to 2030: eliminating hepatitis B disease, China. *Bull World Health Organ,* 97**,** 230-238.

LIU, N., JI, L., MAGUIRE, M. L. & LOEB, D. D. 2004. cis-Acting sequences that contribute to the synthesis of relaxed-circular DNA of human hepatitis B virus. *J Virol,* 78**,** 642-9.

LOK, A. S. 2002. Chronic hepatitis B. *N Engl J Med,* 346**,** 1682-3.

LOMAN, N. J., QUICK, J. & SIMPSON, J. T. 2015. A complete bacterial genome assembled de novo using only nanopore sequencing data. *Nat Methods,* 12**,** 733-5.

LOPEZ-TERRADA, D., CHEUNG, S. W., FINEGOLD, M. J. & KNOWLES, B. B. 2009. Hep G2 is a hepatoblastoma-derived cell line. *Hum Pathol,* 40**,** 1512-5.

MAINI, M. K. & PALLETT, L. J. 2018. Defective T-cell immunity in hepatitis B virus infection: why therapeutic vaccination needs a helping hand. *Lancet Gastroenterol Hepatol,* 3**,** 192-202.

MANTERE, T., KERSTEN, S. & HOISCHEN, A. 2019. Long-Read Sequencing Emerging in Medical Genetics. *Front Genet,* 10**,** 426.

MARCELLIN, P., BONINO, F., YURDAYDIN, C., HADZIYANNIS, S., MOUCARI, R., KAPPRELL, H. P., ROTHE, V., POPESCU, M. & BRUNETTO, M. R. 2013. Hepatitis B surface antigen levels: association with 5-year response to peginterferon alfa-2a in hepatitis B e-antigen-negative patients. *Hepatol Int,* 7**,** 88-97.

MARENINOVA, O., SHIN, J. M., VAGIN, O., TURDIKULOVA, S., HALLEN, S. & SACHS, G. 2005. Topography of the membrane domain of the liver Na+-dependent bile acid transporter. *Biochemistry,* 44**,** 13702-12.

MARGERIDON, S., CARROUEE-DURANTEL, S., CHEMIN, I., BARRAUD, L., ZOULIM, F., TREPO, C. & KAY, A. 2008. Rolling circle amplification, a powerful tool for genetic and functional studies of complete hepatitis B virus genomes from low-level infections and for directly probing covalently closed circular DNA. *Antimicrob Agents Chemother,* 52**,** 3068-73.

MATHER, S., SCOTT, S., TEMPERTON, N., WRIGHT, E., KING, B. & DALY, J. 2013. Current progress with serological assays for exotic emerging/re-emerging viruses. *Future Medicine,* 8(8), 745-755.

MCMAHON, B. J., ALWARD, W. L., HALL, D. B., HEYWARD, W. L., BENDER, T. R., FRANCIS, D. P. & MAYNARD, J. E. 1985. Acute hepatitis B virus infection: relation of age to the clinical expression of disease and subsequent development of the carrier state. *J Infect Dis,* 151**,** 599-603.

MCNAUGHTON, A. L., D'ARIENZO, V., ANSARI, M. A., LUMLEY, S. F., LITTLEJOHN, M., REVILL, P., MCKEATING, J. A. & MATTHEWS, P. C. 2019a. Insights From Deep Sequencing of the HBV Genome-Unique, Tiny, and Misunderstood. *Gastroenterology,* 156**,** 384-399.

MCNAUGHTON, A. L., ROBERTS, H. E., BONSALL, D., DE CESARE, M., MOKAYA, J., LUMLEY, S. F., GOLUBCHIK, T., PIAZZA, P., MARTIN, J. B., DE LARA, C., BROWN, A., ANSARI, M. A., BOWDEN, R., BARNES, E. & MATTHEWS, P. C. 2019b. Illumina and Nanopore methods for whole genome sequencing of hepatitis B virus (HBV). *Sci Rep,* 9**,** 7081.

MEIRELES, L. C., MARINHO, R. T. & VAN DAMME, P. 2015. Three decades of hepatitis B control with vaccination. *World J Hepatol,* 7**,** 2127-32.

MELEGARI, M., SCAGLIONI, P. P. & WANDS, J. R. 1997. The small envelope protein is required for secretion of a naturally occurring hepatitis B virus mutant with pre-S1 deleted. *J Virol,* 71**,** 5449-54.

MEREDITH, L. W., HU, K., CHENG, X., HOWARD, C. R., BAUMERT, T. F., BALFE, P., VAN DE GRAAF, K. F., PROTZER, U. & MCKEATING, J. A. 2016. Lentiviral hepatitis B pseudotype entry requires sodium taurocholate co-transporting polypeptide and additional hepatocyte-specific factors. *J Gen Virol,* 97**,** 121-127.

MICHAILIDIS, E., PABON, J., XIANG, K., PARK, P., RAMANAN, V., HOFFMANN, H. H., SCHNEIDER, W. M., BHATIA, S. N., DE JONG, Y. P., SHLOMAI, A. & RICE, C. M. 2017. A robust cell culture system supporting the complete life cycle of hepatitis B virus. *Sci Rep,* 7**,** 16616.

MINA, T., AMINI BAVIL OLYAEE, S., TACKE, F., MAES, P., VAN RANST, M. & POURKARIM, M. R. 2015. Genomic Diversity of Hepatitis B Virus Infection Associated With Fulminant Hepatitis B Development. *Hepat Mon,* 15**,** e29477.

MINOR, M. M. & SLAGLE, B. L. 2014. Hepatitis B virus HBx protein interactions with the ubiquitin proteasome system. *Viruses,* 6**,** 4683-702.

MIRANDOLA, S., SEBASTIANI, G., ROSSI, C., VELO, E., ERNE, E. M., VARIO, A., TEMPESTA, D., ROMUALDI, C., CAMPAGNOLO, D. & ALBERTI, A. 2012. Genotype-specific mutations in the polymerase gene of hepatitis B virus potentially associated with resistance to oral antiviral therapy. *Antiviral Res,* 96**,** 422-9.

NAKABAYASHI, H., TAKETA, K., YAMANE, T., MIYAZAKI, M., MIYANO, K. & SATO, J. 1984. Phenotypical stability of a human hepatoma cell line, HuH-7, in long-term culture with chemically defined medium. *Gan,* 75**,** 151-8.

NANOPORE, O. 2020. *R10.3: the newest nanopore for high accuracy nanopore sequencing* [Online]. Available: https://nanoporetech.com/about-us/news/r103-newest-nanopore-high-accuracy-nanopore-sequencing-now-available-store [Accessed 20th May 2020].

NASSAL, M. 1992. The arginine-rich domain of the hepatitis B virus core protein is required for pregenome encapsidation and productive viral positive-strand DNA synthesis but not for virus assembly. *J Virol,* 66**,** 4107-16.

NEURATH, A. R., KENT, S. B., STRICK, N. & PARKER, K. 1986. Identification and chemical synthesis of a host cell receptor binding site on hepatitis B virus. *Cell,* 46**,** 429-36.

NI, Y., LEMPP, F. A., MEHRLE, S., NKONGOLO, S., KAUFMAN, C., FALTH, M., STINDT, J., KONIGER, C., NASSAL, M., KUBITZ, R., SULTMANN, H. & URBAN, S. 2014. Hepatitis B and D viruses exploit sodium taurocholate co-transporting polypeptide for species-specific entry into hepatocytes. *Gastroenterology,* 146**,** 1070-83.

NI, Y., SONNABEND, J., SEITZ, S. & URBAN, S. 2010. The pre-s2 domain of the hepatitis B virus is dispensable for infectivity but serves a spacer function for L-protein-connected virus assembly. *J Virol,* 84**,** 3879-88.

NICE. 2019. *Hepatitis B - diagnosis* [Online]. Available: https://cks.nice.org.uk/hepatitis-b#!diagnosisSub:1 [Accessed 13 Feb 2020].

NIH. 2017. *NIH AIDs Reagent Programe: HIV-1 NL4-3 ΔEnv Vpr Luciferase Reporter Vector (pNL4-3.Luc.R-E-)* [Online]. [Accessed 20th June 2020].

NISHITSUJI, H., UJINO, S., SHIMIZU, Y., HARADA, K., ZHANG, J., SUGIYAMA, M., MIZOKAMI, M. & SHIMOTOHNO, K. 2015. Novel reporter system to monitor early stages of the hepatitis B virus life cycle. *Cancer Sci,* 106**,** 1616-24.

NKONGOLO, S., NI, Y., LEMPP, F. A., KAUFMAN, C., LINDNER, T., ESSER-NOBIS, K., LOHMANN, V., MIER, W., MEHRLE, S. & URBAN, S. 2014. Cyclosporin A inhibits hepatitis B and hepatitis D virus entry by cyclophilin-independent interference with the NTCP receptor. *J Hepatol,* 60**,** 723-31.

NOLAN, T., HUGGETT, J. & SANCHEZ, E. 2013. Good practice guide for the application of quantittive PCR (qPCR). *LGC*.

OKAMOTO, H., TSUDA, F., SAKUGAWA, H., SASTROSOEWIGNJO, R. I., IMAI, M., MIYAKAWA, Y. & MAYUMI, M. 1988. Typing hepatitis B virus by homology in nucleotide sequence: comparison of surface antigen subtypes. *J Gen Virol,* 69 (Pt 10)**,** 2575-83.

OKOCHI, K. & MURAKAMI, S. 1968. Observations on Australia antigen in Japanese. *Vox Sang,* 15**,** 374-85.

OP DE BEECK, A., VOISSET, C., BARTOSCH, B., CICZORA, Y., COCQUEREL, L., KECK, Z., FOUNG, S., COSSET, F. L. & DUBUISSON, J. 2004. Characterization of functional hepatitis C virus envelope glycoproteins. *J Virol,* 78**,** 2994-3002.

OSIOWY, C., GILES, E., TANAKA, Y., MIZOKAMI, M. & MINUK, G. Y. 2006. Molecular evolution of hepatitis B virus over 25 years. *J Virol,* 80**,** 10307-14.

PAN, W., SONG, I. S., SHIN, H. J., KIM, M. H., CHOI, Y. L., LIM, S. J., KIM, W. Y., LEE, S. S. & SHIN, J. G. 2011. Genetic polymorphisms in Na+-taurocholate co-transporting polypeptide (NTCP) and ileal apical sodium-dependent bile acid transporter (ASBT) and ethnic comparisons of functional variants of NTCP among Asian populations. *Xenobiotica,* 41**,** 501-10.

PARASKEVIS, D., ANGELIS, K., MAGIORKINIS, G., KOSTAKI, E., HO, S. Y. & HATZAKIS, A. 2015. Dating the origin of hepatitis B virus reveals higher substitution rate and adaptation on the branch leading to F/H genotypes. *Mol Phylogenet Evol,* 93**,** 44-54.

PASTOR, F., HERRSCHER, C., PATIENT, R., EYMIEUX, S., MOREAU, A., BURLAUD-GAILLARD, J., SEIGNEURET, F., DE ROCQUIGNY, H., ROINGEARD, P. & HOURIOUX, C. 2019. Direct interaction between the hepatitis B virus core and envelope proteins analyzed in a cellular context. *Sci Rep,* 9**,** 16178.

PASTORINO, B., DE LAMBALLERIE, X. & CHARREL, R. 2017. Biosafety and Biosecurity in European Containment Level 3 Laboratories: Focus on French Recent Progress and Essential Requirements. *Front Public Health,* 5**,** 121.

PATTERSON ROSS, Z., KLUNK, J., FORNACIARI, G., GIUFFRA, V., DUCHENE, S., DUGGAN, A. T., POINAR, D., DOUGLAS, M. W., EDEN, J. S., HOLMES, E. C. & POINAR, H. N. 2018. The paradox of HBV evolution as revealed from a 16th century mummy. *PLoS Pathog,* 14**,** e1006750.

PECK, K. M. & LAURING, A. S. 2018. Complexities of Viral Mutation Rates. *J Virol,* 92.

PENG, L., ZHAO, Q., LI, Q., LI, M., LI, C., XU, T., JING, X., ZHU, X., WANG, Y., LI, F., LIU, R., ZHONG, C., PAN, Q., ZENG, B., LIAO, Q., HU, B., HU, Z. X., HUANG, Y. S., SHAM, P., LIU, J., XU, S., WANG, J., GAO, Z. L. & WANG, Y. 2015. The p.Ser267Phe variant in SLC10A1 is associated with resistance to chronic hepatitis B. *Hepatology,* 61**,** 1251-60.

PEREZ-VARGAS, J., AMIRACHE, F., BOSON, B., MIALON, C., FREITAS, N., SUREAU, C., FUSIL, F. & COSSET, F. L. 2019. Enveloped viruses distinct from HBV induce dissemination of hepatitis D virus in vivo. *Nat Commun,* 10**,** 2098.

PERRILLO, R. P., GISH, R. & FALCK-YTTER, Y. T. 2015. American Gastroenterological Association Institute technical review on prevention and treatment of hepatitis B virus reactivation during immunosuppressive drug therapy. *Gastroenterology,* 148**,** 221-244 e3.

PHE. 2007. *Recommended Protocol for HBV genotyping and for antiviral resistance analysis* [Online]. Available: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/342414/cfi_vrd_PolHBsAgRecommendationsFINAL_HBV_protocol.pdf [Accessed].

PIRATVISUTH, T. 2013. Optimal management of HBV infection during pregnancy. *Liver Int,* 33 Suppl 1**,** 188-94.

PODLAHA, O., GANE, E., BRUNETTO, M., FUNG, S., CHUANG, W. L., PAN, C. Q., JIANG, Z., LIU, Y., BHARDWAJ, N., MUKHERJEE, P., FLAHERTY, J., GAGGAR, A., SUBRAMANIAN, M., IZUMI, N., SHALIMAR, LIM, Y. S., MARCELLIN, P., BUTI, M., CHAN, H. L. Y. & AGARWAL, K. 2019. Large-scale viral genome analysis identifies novel clinical associations between hepatitis B virus and chronically infected patients. *Sci Rep,* 9**,** 10529.

POLLICE, A. A., MCCOY, J. P., JR., SHACKNEY, S. E., SMITH, C. A., AGARWAL, J., BURHOLT, D. R., JANOCKO, L. E., HORNICEK, F. J., SINGH, S. G. & HARTSOCK, R. J. 1992. Sequential paraformaldehyde and methanol fixation for simultaneous flow cytometric analysis of DNA, cell surface proteins, and intracellular proteins. *Cytometry,* 13**,** 432-44.

QIAGEN 2012. QIAGEN® Plasmid Purification Handbook. *In:* QIAGEN (ed.).

QUICK, J., LOMAN, N. J., DURAFFOUR, S., SIMPSON, J. T., SEVERI, E., COWLEY, L., BORE, J. A., KOUNDOUNO, R., DUDAS, G., MIKHAIL, A., OUEDRAOGO, N., AFROUGH, B., BAH, A., BAUM, J. H., BECKER-ZIAJA, B., BOETTCHER, J. P., CABEZA-CABRERIZO, M., CAMINO-SANCHEZ, A., CARTER, L. L., DOERRBECKER, J., ENKIRCH, T., DORIVAL, I. G. G., HETZELT, N., HINZMANN, J., HOLM, T., KAFETZOPOULOU, L. E., KOROPOGUI, M., KOSGEY, A., KUISMA, E., LOGUE, C. H., MAZZARELLI, A., MEISEL, S., MERTENS, M., MICHEL, J., NGABO, D., NITZSCHE, K., PALLASH, E., PATRONO, L. V., PORTMANN, J., REPITS, J. G., RICKETT, N. Y., SACHSE, A., SINGETHAN, K., VITORIANO, I., YEMANABERHAN, R. L., ZEKENG, E. G., TRINA, R., BELLO, A., SALL, A. A., FAYE, O., FAYE, O., MAGASSOUBA, N., WILLIAMS, C. V., AMBURGEY, V., WINONA, L., DAVIS, E., GERLACH, J., WASHINGTON, F., MONTEIL, V., JOURDAIN, M., BERERD, M., CAMARA, A., SOMLARE, H., CAMARA, A., GERARD, M., BADO, G., BAILLET, B., DELAUNE, D., NEBIE, K. Y., DIARRA, A., SAVANE, Y., PALLAWO, R. B., GUTIERREZ, G. J., MILHANO, N., ROGER, I., WILLIAMS, C. J., YATTARA, F., LEWANDOWSKI, K., TAYLOR, J., RACHWAL, P., TURNER, D., POLLAKIS, G., HISCOX, J. A., MATTHEWS, D. A., O'SHEA, M. K., JOHNSTON, A. M., WILSON, D., HUTLEY, E., SMIT, E., DI CARO, A., WOELFEL, R., STOECKER, K., FLEISCHMANN, E., GABRIEL, M., WELLER, S. A., KOIVOGUI, L., DIALLO, B., KEITA, S., RAMBAUT, A., FORMENTY, P., et al. 2016. Real-time, portable genome sequencing for Ebola surveillance. *Nature,* 530**,** 228-232.

RAMBAUT, A., DRUMMOND, A. J., XIE, D., BAELE, G. & SUCHARD, M. A. 2018. Posterior Summarization in Bayesian Phylogenetics Using Tracer 1.7. *Syst Biol,* 67**,** 901-904.

RANG, F. J., KLOOSTERMAN, W. P. & DE RIDDER, J. 2018. From squiggle to basepair: computational approaches for improving nanopore sequencing read accuracy. *Genome Biol,* 19**,** 90.

RENZETTE, N., PFEIFER, S. P., MATUSZEWSKI, S., KOWALIK, T. F. & JENSEN, J. D. 2017. On the Analysis of Intrahost and Interhost Viral Populations: Human Cytomegalovirus as a Case Study of Pitfalls and Expectations. *J Virol,* 91.

REVILL, P., TESTONI, B., LOCARNINI, S. & ZOULIM, F. 2016. Global strategies are required to cure and eliminate HBV infection. *Nat Rev Gastroenterol Hepatol,* 13**,** 239-48.

ROE, T., REYNOLDS, T. C., YU, G. & BROWN, P. O. 1993. Integration of murine leukemia virus DNA depends on mitosis. *EMBO J,* 12**,** 2099-108.

RUDD, P. M. & DWEK, R. A. 1997. Glycosylation: heterogeneity and the 3D structure of proteins. *Crit Rev Biochem Mol Biol,* 32**,** 1-100.

SAI, L. T., YAO, Y. Y., GUAN, Y. Y., SHAO, L. H., MA, R. P. & MA, L. X. 2016. Hepatitis B virus infection and replication in a new cell culture system established by fusing HepG2 cells with primary human hepatocytes. *J Microbiol Immunol Infect,* 49**,** 471-6.

SAIGO, K., YOSHIDA, K., IKEDA, R., SAKAMOTO, Y., MURAKAMI, Y., URASHIMA, T., ASANO, T., KENMOCHI, T. & INOUE, I. 2008. Integration of hepatitis B virus DNA into the myeloid/lymphoid or mixed-lineage leukemia (MLL4) gene and rearrangements of MLL4 in human hepatocellular carcinoma. *Hum Mutat,* 29**,** 703-8.

SAKURAI, F., MITANI, S., YAMAMOTO, T., TAKAYAMA, K., TACHIBANA, M., WATASHI, K., WAKITA, T., IIJIMA, S., TANAKA, Y. & MIZUGUCHI, H. 2017. Human induced-pluripotent stem cell-derived hepatocyte-like cells as an in vitro model of human hepatitis B virus infection. *Sci Rep,* 7**,** 45698.

SALISSE, J. & SUREAU, C. 2009. A function essential to viral entry underlies the hepatitis B virus "a" determinant. *J Virol,* 83**,** 9321-8.

SANDRIN, V. & COSSET, F. L. 2006. Intracellular versus cell surface assembly of retroviral pseudotypes is determined by the cellular localization of the viral glycoprotein, its capacity to interact with Gag, and the expression of the Nef protein. *J Biol Chem,* 281**,** 528-42.

SANJUAN, R., NEBOT, M. R., CHIRICO, N., MANSKY, L. M. & BELSHAW, R. 2010. Viral mutation rates. *J Virol,* 84**,** 9733-48.

SARGIACOMO, C., EL-KEHDY, H., POURCHER, G., STIEGER, B., NAJIMI, M. & SOKAL, E. 2018. Age-dependent glycosylation of the sodium taurocholate cotransporter polypeptide: From fetal to adult human livers. *Hepatol Commun,* 2**,** 693-702.

SARKAR, S., BANANIS, E., NATH, S., ANWER, M. S., WOLKOFF, A. W. & MURRAY, J. W. 2006. PKCzeta is required for microtubule-based motility of vesicles containing the ntcp transporter. *Traffic,* 7**,** 1078-91.

SAUVAGE, V., BOIZEAU, L., CANDOTTI, D., VANDENBOGAERT, M., SERVANT-DELMAS, A., CARO, V. & LAPERCHE, S. 2018. Early MinION nanopore single-molecule sequencing technology enables the characterization of hepatitis B virus genetic complexity in clinical samples. *PLoS One,* 13**,** e0194366.

SCHAEFER, S. 2007. Hepatitis B virus taxonomy and hepatitis B virus genotypes. *World J Gastroenterol,* 13**,** 14-21.

SCHINDELIN, J., ARGANDA-CARRERAS, I., FRISE, E., KAYNIG, V., LONGAIR, M., PIETZSCH, T., PREIBISCH, S., RUEDEN, C., SAALFELD, S., SCHMID, B., TINEVEZ, J. Y., WHITE, D. J., HARTENSTEIN, V., ELICEIRI, K., TOMANCAK, P. & CARDONA, A. 2012. Fiji: an open-source platform for biological-image analysis. *Nat Methods,* 9**,** 676-82.

SCHULZE, A., GRIPON, P. & URBAN, S. 2007. Hepatitis B virus infection initiates with a large surface protein-dependent binding to heparan sulfate proteoglycans. *Hepatology,* 46**,** 1759-68.

SCHULZE, A., SCHIECK, A., NI, Y., MIER, W. & URBAN, S. 2010. Fine mapping of pre-S sequence requirements for hepatitis B virus large envelope protein-mediated receptor interaction. *J Virol,* 84**,** 1989-2000.

SEDLAZECK, F. J., RESCHENEDER, P., SMOLKA, M., FANG, H., NATTESTAD, M., VON HAESELER, A. & SCHATZ, M. C. 2018. Accurate detection of complex structural variations using single-molecule sequencing. *Nat Methods,* 15**,** 461-468.

SELZER, L. & ZLOTNICK, A. 2015. Assembly and Release of Hepatitis B Virus. *Cold Spring Harb Perspect Med,* 5.

SENOL CALI, D., KIM, J. S., GHOSE, S., ALKAN, C. & MUTLU, O. 2018. Nanopore sequencing technology and tools for genome assembly: computational analysis of the current state, bottlenecks and future directions. *Brief Bioinform*.

SHIN, M. K., LEE, J. & RYU, W. S. 2004. A novel cis-acting element facilitates minus-strand DNA synthesis during reverse transcription of the hepatitis B virus genome. *J Virol,* 78**,** 6252-62.

SIMMONDS, P. & MIDGLEY, S. 2005. Recombination in the genesis and evolution of hepatitis B virus genotypes. *J Virol,* 79**,** 15467-76.

SIMONETTI, J., BULKOW, L., MCMAHON, B. J., HOMAN, C., SNOWBALL, M., NEGUS, S., WILLIAMS, J. & LIVINGSTON, S. E. 2010. Clearance of hepatitis B surface antigen and risk of hepatocellular carcinoma in a cohort chronically infected with hepatitis B virus. *Hepatology,* 51**,** 1531-7.

SNAPGENE. 2019. *SnapGene software-clone smarter and faster* [Online]. Available: https://www.snapgene.com/ [Accessed 10th Oct 2019].

SOMIYA, M., LIU, Q., YOSHIMOTO, N., IIJIMA, M., TATEMATSU, K., NAKAI, T., OKAJIMA, T., KUROKI, K., UEDA, K. & KURODA, S. 2016. Cellular uptake of hepatitis B virus envelope L particles is independent of sodium taurocholate cotransporting polypeptide, but dependent on heparan sulfate proteoglycan. *Virology,* 497**,** 23-32.

STEEDS, K., HALL, Y., SLACK, G. S., LONGET, S., STRECKER, T., FEHLING, S. K., WRIGHT, E., BORE, J. A., KOUNDOUNO, F. R., KONDE, M. K., HEWSON, R., HISCOX, J. A., POLLAKIS, G. & CARROLL, M. W. 2020. Pseudotyping of VSV with Ebola virus glycoprotein is superior to HIV-1 for the assessment of neutralising antibodies. *Sci Rep,* 10**,** 14289.

STELMA, F., WILLEMSE, S. B., ERKEN, R., DE NIET, A., SINNIGE, M. J., VAN DORT, K., ZAAIJER, H. L., VAN LEEUWEN, E. M. M., KOOTSTRA, N. A. & REESINK, H. W. 2017. Dynamics of the Immune Response in Acute Hepatitis B Infection. *Open Forum Infect Dis,* 4**,** ofx231.

SU, I. J., WANG, L. H., HSIEH, W. C., WU, H. C., TENG, C. F., TSAI, H. W. & HUANG, W. 2014. The emerging role of hepatitis B virus pre-S2 deletion mutant proteins in HBV tumorigenesis. *J Biomed Sci,* 21**,** 98.

SUN, S., YAN, J., XIA, C., LIN, Y., JIANG, X., LIU, H., REN, H., YAN, J., LIN, J. & HE, X. 2014. Visualizing hepatitis B virus with biarsenical labelling in living cells. *Liver Int,* 34**,** 1532-42.

SUN, Y., QI, Y., PENG, B. & LI, W. 2017. NTCP-Reconstituted In Vitro HBV Infection System. *Methods Mol Biol,* 1540**,** 1-14.

SUN, Y., WANG, S., YI, Y., ZHANG, J., DUAN, Z., YUAN, K., LIU, W., LI, J. & ZHU, Y. 2018. The Hepatitis B Surface Antigen Binding Protein: An Immunoglobulin G Constant Region-Like Protein That Interacts With HBV Envelop Proteins and Mediates HBV Entry. *Front Cell Infect Microbiol,* 8**,** 338.

SUNBUL, M. 2014. Hepatitis B virus genotypes: global distribution and clinical importance. *World J Gastroenterol,* 20**,** 5427-34.

SUREAU, C. & SALISSE, J. 2013. A conformational heparan sulfate binding site essential to infectivity overlaps with the conserved hepatitis B virus a-determinant. *Hepatology,* 57**,** 985-94.

SUSLOV, A., BOLDANOVA, T., WANG, X., WIELAND, S. & HEIM, M. H. 2018. Hepatitis B Virus Does Not Interfere With Innate Immune Responses in the Human Liver. *Gastroenterology,* 154**,** 1778-1790.

TAN, A., KOH, S. & BERTOLETTI, A. 2015. Immune Response in Hepatitis B Virus Infection. *Cold Spring Harb Perspect Med,* 5**,** a021428.

TANI, H., MORIKAWA, S. & MATSUURA, Y. 2011. Development and Applications of VSV Vectors Based on Cell Tropism. *Front Microbiol,* 2**,** 272.

TATEMATSU, K., TANAKA, Y., KURBANOV, F., SUGAUCHI, F., MANO, S., MAESHIRO, T., NAKAYOSHI, T., WAKUTA, M., MIYAKAWA, Y. & MIZOKAMI, M. 2009. A genetic variant of hepatitis B virus divergent from known human and ape genotypes isolated from a Japanese patient and provisionally assigned to new genotype J. *J Virol,* 83**,** 10538-47.

THOMPSON, S. R. 2012. So you want to know if your message has an IRES? *Wiley Interdiscip Rev RNA,* 3**,** 697-705.

TONG, M. J., BLATT, L. M., TYSON, K. B. & KAO, V. W. 2006. Death From Liver Disease and Development of Hepatocellular Carcinoma in Patients With Chronic Hepatitis B Virus Infection: A Prospective Study. *Gastroenterol Hepatol (N Y),* 2**,** 41-47.

TONG, S. & REVILL, P. 2016. Overview of hepatitis B viral replication and genetic variability. *J Hepatol,* 64**,** S4-S16.

TREPO, C., CHAN, H. L. & LOK, A. 2014. Hepatitis B virus infection. *Lancet,* 384**,** 2053-63.

TREY, C. & DAVIDSON, C. S. 1970. The management of fulminant hepatic failure. *Prog Liver Dis,* 3**,** 282-98.

TU, T., BUDZINSKA, M. A., SHACKEL, N. A. & URBAN, S. 2017. HBV DNA Integration: Molecular Mechanisms and Clinical Implications. *Viruses,* 9.

URBANOWICZ, R. A., MCCLURE, C. P., BROWN, R. J., TSOLERIDIS, T., PERSSON, M. A., KREY, T., IRVING, W. L., BALL, J. K. & TARR, A. W. 2015. A Diverse Panel of Hepatitis C Virus Glycoproteins for Use in Vaccine Research Reveals Extremes of Monoclonal Antibody Neutralization Resistance. *J Virol,* 90**,** 3288-301.

URBANOWICZ, R. A., MCCLURE, C. P., KING, B., MASON, C. P., BALL, J. K. & TARR, A. W. 2016. Novel functional hepatitis C virus glycoprotein isolates identified using an optimized viral pseudotype entry assay. *J Gen Virol,* 97**,** 2265-2279.

VAN DE WIEL, S., MERKX, M. & VAN DE GRAAF, S. 2016. Real Time Monitoring of Intracellular Bile Acid Dynamics Using a Genetically Encoded FRET-based Bile Acid Sensor. *J Vis Exp*.

VAN DER VELDEN, L. M., GOLYNSKIY, M. V., BIJSMANS, I. T., VAN MIL, S. W., KLOMP, L. W., MERKX, M. & VAN DE GRAAF, S. F. 2013. Monitoring bile acid transport in single living cells using a genetically encoded Forster resonance energy transfer sensor. *Hepatology,* 57**,** 740-52.

VERRIER, E. R., COLPITTS, C. C., BACH, C., HEYDMANN, L., WEISS, A., RENAUD, M., DURAND, S. C., HABERSETZER, F., DURANTEL, D., ABOU-

JAOUDE, G., LOPEZ LEDESMA, M. M., FELMLEE, D. J., SOUMILLON, M., CROONENBORGHS, T., POCHET, N., NASSAL, M., SCHUSTER, C., BRINO, L., SUREAU, C., ZEISEL, M. B. & BAUMERT, T. F. 2016. A targeted functional RNA interference screen uncovers glypican 5 as an entry factor for hepatitis B and D viruses. *Hepatology,* 63**,** 35-48.

WALLIN, M., EKSTROM, M. & GAROFF, H. 2005. The fusion-controlling disulfide bond isomerase in retrovirus Env is triggered by protein destabilization. *J Virol,* 79**,** 1678-85.

WALTER, E., KEIST, R., NIEDEROST, B., PULT, I. & BLUM, H. E. 1996. Hepatitis B virus infection of tupaia hepatocytes in vitro and in vivo. *Hepatology,* 24**,** 1-5.

WARNER, N. & LOCARNINI, S. 2008. The antiviral drug selected hepatitis B virus rtA181T/sW172* mutant has a dominant negative secretion defect and alters the typical profile of viral rebound. *Hepatology,* 48**,** 88-98.

WATASHI, K., URBAN, S., LI, W. & WAKITA, T. 2014. NTCP and beyond: opening the door to unveil hepatitis B virus entry. *Int J Mol Sci,* 15**,** 2892-905.

WEDEMEYER, H., BOGOMOLOV, P., BLANK, A., ALLWEISS, L., DANDRI-PETERSEN, M., BREMER, B., VORONKOVA, N., SCHÖNEWEIS, K., PATHIL, A., BURHENNE, J., HAAG, M., SCHWAB, M., HAEFELI, W.-E., WIESCH, J. S. Z., ALEXANDROV, A. & URBAN, S. 2018. Final results of a multicenter, open-label phase 2b clinical trial to assess safety and efficacy of Myrcludex B in combination with Tenofovir in patients with chronic HBV/HDV co-infection. *Journal of Hepatology,* vol. 68.

WELM, B. E., DIJKGRAAF, G. J., BLEDAU, A. S., WELM, A. L. & WERB, Z. 2008. Lentiviral transduction of mammary stem cells for analysis of gene function during development and cancer. *Cell Stem Cell,* 2**,** 90-102.

WHITT, M. A. 2010. Generation of VSV pseudotypes using recombinant DeltaG-VSV for studies on virus entry, identification of entry inhibitors, and immune responses to vaccines. *J Virol Methods,* 169**,** 365-74.

WHO. 2003. *WHO biosafety guidelines for handling of SARS specimens* [Online]. World Health Organization Available: https://www.who.int/csr/sars/biosafety2003_04_25/en/ [Accessed 18 Dec 2019].

WHO 2015. GUIDELINES FOR THE PREVENTION, CARE AND TREATMENT OF PERSONS WITH CHRONIC HEPATITIS B INFECTION.

WHO 2016. COMBATING HEPATITIS B AND C TO REACH ELIMINATION BY 2030.

WHO. 2019a. *Hepatitis B* [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/hepatitis-b [Accessed 14 Feb 2020].

WHO. 2019b. *Hepatitis B* [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/hepatitis-b [Accessed 13 Feb 2020].

WHO. 2019c. *Hepatitis D* [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/hepatitis-d [Accessed 6th Jan 2020].

WILL, H., REISER, W., WEIMER, T., PFAFF, E., BUSCHER, M., SPRENGEL, R., CATTANEO, R. & SCHALLER, H. 1987. Replication strategy of human hepatitis B virus. *J Virol,* 61**,** 904-11.

WITT-KEHATI, D., BITTON ALALUF, M. & SHLOMAI, A. 2016. Advances and Challenges in Studying Hepatitis B Virus In Vitro. *Viruses,* 8.

WOOL-LEWIS, R. J. & BATES, P. 1998. Characterization of Ebola virus entry by using pseudotyped viruses: identification of receptor-deficient cell lines. *J Virol,* 72**,** 3155-60.

WYNNE, S. A., CROWTHER, R. A. & LESLIE, A. G. 1999. The crystal structure of the human hepatitis B virus capsid. *Mol Cell,* 3**,** 771-80.

XIA, Y., CARPENTIER, A., CHENG, X., BLOCK, P. D., ZHAO, Y., ZHANG, Z., PROTZER, U. & LIANG, T. J. 2017. Human stem cell-derived hepatocytes as a model for hepatitis B virus infection, spreading and virus-host interactions. *J Hepatol,* 66**,** 494-503.

XU, D. Z., YAN, Y. P., CHOI, B. C., XU, J. Q., MEN, K., ZHANG, J. X., LIU, Z. H. & WANG, F. S. 2002. Risk factors and mechanism of transplacental transmission of hepatitis B virus: a case-control study. *J Med Virol,* 67**,** 20-6.

XU, Z. Y., LIU, C. B., FRANCIS, D. P., PURCELL, R. H., GUN, Z. L., DUAN, S. C., CHEN, R. J., MARGOLIS, H. S., HUANG, C. H. & MAYNARD, J. E. 1985. Prevention of perinatal acquisition of hepatitis B virus carriage using vaccine: preliminary report of a randomized, double-blind placebo-controlled and comparative trial. *Pediatrics,* 76**,** 713-8.

YAN, H., PENG, B., HE, W., ZHONG, G., QI, Y., REN, B., GAO, Z., JING, Z., SONG, M., XU, G., SUI, J. & LI, W. 2013. Molecular determinants of hepatitis B and D virus entry restriction in mouse sodium taurocholate cotransporting polypeptide. *J Virol,* 87**,** 7977-91.

YAN, H., PENG, B., LIU, Y., XU, G., HE, W., REN, B., JING, Z., SUI, J. & LI, W. 2014. Viral entry of hepatitis B and D viruses and bile salts transportation share common molecular determinants on sodium taurocholate cotransporting polypeptide. *J Virol,* 88**,** 3273-84.

YAN, H., ZHONG, G., XU, G., HE, W., JING, Z., GAO, Z., HUANG, Y., QI, Y., PENG, B., WANG, H., FU, L., SONG, M., CHEN, P., GAO, W., REN, B., SUN, Y., CAI, T., FENG, X., SUI, J. & LI, W. 2012. Sodium taurocholate cotransporting polypeptide is a functional receptor for human hepatitis B and D virus. *Elife,* 1**,** e00049.

YANG, J., YANG, Y., XIA, M., WANG, L., ZHOU, W., YANG, Y., JIANG, Y., WANG, H., QIAN, J., JIN, L. & WANG, X. 2016. A genetic variant of the NTCP gene is associated with HBV infection status in a Chinese population. *BMC Cancer,* 16**,** 211.

YANG, W. & SUMMERS, J. 1995. Illegitimate replication of linear hepadnavirus DNA through nonhomologous recombination. *J Virol,* 69**,** 4029-36.

YASUTAKE, Y., HATTORI, S. I., HAYASHI, H., MATSUDA, K., TAMURA, N., KOHGO, S., MAEDA, K. & MITSUYA, H. 2018. HIV-1 with HBV-associated Q151M substitution in RT becomes highly susceptible to entecavir: structural insights into HBV-RT inhibition by entecavir. *Sci Rep,* 8**,** 1624.

YATSUJI, H., NOGUCHI, C., HIRAGA, N., MORI, N., TSUGE, M., IMAMURA, M., TAKAHASHI, S., IWAO, E., FUJIMOTO, Y., OCHI, H., ABE, H., MAEKAWA, T., TATENO, C., YOSHIZATO, K., SUZUKI, F., KUMADA, H. & CHAYAMA, K. 2006. Emergence of a novel lamivudine-resistant hepatitis B virus variant with a substitution outside the YMDD motif. *Antimicrob Agents Chemother,* 50**,** 3867-74.

YOO, B. C., KIM, J. H., CHUNG, Y. H., LEE, K. S., PAIK, S. W., RYU, S. H., HAN, B. H., HAN, J. Y., BYUN, K. S., CHO, M., LEE, H. J., KIM, T. H., CHO, S. H., PARK, J. W., UM, S. H., HWANG, S. G., KIM, Y. S., LEE, Y. J., CHON, C. Y., KIM, B. I., LEE, Y. S., YANG, J. M., KIM, H. C., HWANG, J. S., CHOI, S. K., KWEON, Y. O., JEONG, S. H., LEE, M. S., CHOI, J. Y., KIM, D. G., KIM, Y. S., LEE, H. Y., YOO, K., YOO, H. W. & LEE, H. S. 2007. Twenty-four-week clevudine therapy showed potent and sustained antiviral activity in HBeAg-positive chronic hepatitis B. *Hepatology,* 45**,** 1172-8.

YUAN, T. T. & SHIH, C. 2000. A frequent, naturally occurring mutation (P130T) of human hepatitis B virus core antigen is compensatory for immature secretion phenotype of another frequent variant (I97L). *J Virol,* 74**,** 4929-32.

YUH, C. H. & TING, L. P. 1990. The genome of hepatitis B virus contains a second enhancer: cooperation of two elements within this enhancer is required for its function. *J Virol,* 64**,** 4281-7.

ZAAIJER, H. L., VAN HEMERT, F. J., KOPPELMAN, M. H. & LUKASHOV, V. V. 2007. Independent evolution of overlapping polymerase and surface protein genes of hepatitis B virus. *J Gen Virol,* 88**,** 2137-43.

ZHANG, A. Y., LAI, C. L., HUANG, F. Y., SETO, W. K., FUNG, J., WONG, D. K. & YUEN, M. F. 2015. Evolutionary Changes of Hepatitis B Virus Pre-S Mutations Prior to Development of Hepatocellular Carcinoma. *PLoS One,* 10**,** e0139478.

ZHANG, D., LOU, X., YAN, H., PAN, J., MAO, H., TANG, H., SHU, Y., ZHAO, Y., LIU, L., LI, J., CHEN, J., ZHANG, Y. & MA, X. 2018. Metagenomic analysis of viral nucleic acid extraction methods in respiratory clinical samples. *BMC Genomics,* 19**,** 773.

ZHANG, Q. & CAO, G. 2011. Genotypes, mutations, and viral load of hepatitis B virus and the risk of hepatocellular carcinoma: HBV properties and hepatocarcinogenesis. *Hepat Mon,* 11**,** 86-91.

ZHANG, Z., ZEHNDER, B., DAMRAU, C. & URBAN, S. 2016. Visualization of hepatitis B virus entry - novel tools and approaches to directly follow virus entry into hepatocytes. *FEBS Lett,* 590**,** 1915-26.

ZHAO, L. H., LIU, X., YAN, H. X., LI, W. Y., ZENG, X., YANG, Y., ZHAO, J., LIU, S. P., ZHUANG, X. H., LIN, C., QIN, C. J., ZHAO, Y., PAN, Z. Y., HUANG, G., LIU, H., ZHANG, J., WANG, R. Y., YANG, Y., WEN, W., LV, G. S., ZHANG, H. L., WU, H., HUANG, S., WANG, M. D., TANG, L., CAO, H. Z., WANG, L., LEE, T. L., JIANG, H., TAN, Y. X., YUAN, S. X., HOU, G. J., TAO, Q. F., XU, Q. G., ZHANG, X. Q., WU, M. C., XU, X., WANG, J., YANG, H. M., ZHOU, W. P. & WANG, H. Y. 2016. Genomic and oncogenic preference of HBV integration in hepatocellular carcinoma. *Nat Commun,* 7**,** 12992.

ZUFFEREY, R., DULL, T., MANDEL, R. J., BUKOVSKY, A., QUIROZ, D., NALDINI, L. & TRONO, D. 1998. Self-inactivating lentivirus vector for safe and efficient in vivo gene delivery. *J Virol,* 72**,** 9873-80.

ZUFFEREY, R., NAGY, D., MANDEL, R. J., NALDINI, L. & TRONO, D. 1997. Multiply attenuated lentiviral vector achieves efficient gene delivery in vivo. *Nat Biotechnol,* 15**,** 871-5.

Chapter 8: Appendices

Appendix 1. Sanger vs. MinION in variant calling in Patient 1

REGION	POSITION	P1-T1			P1-T2			P1-T3		
		sanger	ABI alt*	MinION alt (freq.)	sanger	ABI alt	MinION alt (freq.)	sanger	ABI alt	MinION alt (freq.)
PRES1/S2/S	**60**	C	T	T (0.21)	C	T	-	T	C	C (0.26)
	90	A	G	G(0.37)	A	G	G (0.30)	G	-	A (0.14)
	99	A	G	-	A	G	G (0.22)	A	-	-
	109	A	G (fwd)	-	A	-	-	A	-	-
	246	A	G	G (0.25)	A	-	-	G	-	-
	253	C	T	T (0.23)	C	T	T (0.17)	T	C	C 0.17)
	306	A	G	-	A	-	G (0.14)	G	A (fwd.)	-
	414	T	C	C (0.32)	T	-	C (0.18)	C	T	T (0.21)
	819	T	C	C (0.31)	T	-	C (0.14)	C	T (fwd.)	T (0.35)
	873	T	-	-	T	C (rev.)	C (0.14)	T	-	-
	954	T	C	C (0.27)	T		C (0.12)	C	T	T (0.34)
	957	A	-	-	A	-	-	A	G (rev.)	G (0.18)
PRECORE/CORE	**117**	A	C (fwd)	C (0.28)	A	C	C (0.40)	C	A (fwd)	A (0.25)
	478	G	-	-	G	-	-	G	-	C (0.24)
	535	A	-	G (0.12)	A	-	G (0.13)	A	-	-
	634	T	-	-	T	-	-	T	C	C (0.29)

*alt: alternative

Appendix 2. Sanger vs. MinION in variant calling in Patient 2

REGION	POSITION	P2-T1			P2-T2			P2-T3			P2-T4		
		sanger	ABI alt	MinION alt (freq.)	sanger	ABI alt	MinION alt (freq.)	sanger	ABI alt	MinION alt (freq.)	sanger	ABI alt	MinION alt (freq.)
PRECORE/CORE	33	A	-	T (0.13)	A	-	T (0.11)	A	-	-	-	-	-
	86	G	-	A (0.18)	G	-	A (0.15)	G	-	-	-	-	-
	101	C	-	A (0.17)	C	-	-	C	-	-	-	-	-
	275	G	-	T (0.19)	G	-	-	G	-	-	-	-	-
	475	C	-	-	C	-	-	C	-	A (0.23)	A	-	C (0.40)
	535	A	-	G (0.17)	A	-	G (0.15)	A	-	G (0.15)	A	-	G (0.16)

1. PreCore/Core_1 & 2 & 3 &4 forward reads (1584f) are masked by high background noise, unreadable. No minor peaks seen in 2835R reads for all time points.

2. NO variant detected in preS1/S2/S region by MinION sequencing for all time points

Appendix 3. Sanger vs. MinION in variant calling in Patient 3

REGION	POSITION	P3-T1			P3-T2			P3-T3		
		sanger	ABI alt	MinION alt (freq.)	sanger	ABI alt	MinION alt (freq.)	sanger	ABI alt	MinION alt (freq.)
PRES1/S2/S	141	C	-	A (0.20)	C	-	A (0.21)	C	-	A (0.20)
PRECORE/CORE	535	A	-	G (0.16)	A	-	G (0.16)	A	-	G (0.16)